

DÉSAISONNALISER UNE SÉRIE TEMPORELLE



3 - Exploration des séries et décomposition

ALAIN QUARTIER-LA-TENTE

Objectifs de cette séquence

Présenter les concepts élémentaires relatifs aux séries temporelles, ainsi que les outils de base pour les appréhender.

Après cette séquence vous saurez :

- la définition d'une série temporelle ainsi que ses principales composantes
- utiliser les outils graphiques pour explorer une série temporelle
- identifier le schéma de décomposition d'une série

Sommaire

1. Quelques définitions

1.1 Définitions d'une série temporelle

1.2 Descriptions des différentes composantes d'une série

2. L'exploration d'une série

3. Schémas de décomposition

4. Conclusion

Questions de positionnement

Quelle définition d'une série temporelle ?

Qu'est-ce qu'une tendance ?

Qu'est-ce qu'un cycle ?

Que sont les effets de calendrier ?

Qu'est-ce que l'irrégulier ?

Définition d'une série temporelle

Définition : Suite de valeurs numériques ordonnées et indicées par le temps, généralement mesurées à des intervalles réguliers :

- valeurs continues (cours d'une action boursière)
- ou discrètes (la plupart des séries économiques)

Se rencontrent dans tous les domaines : économie, démographie, météorologie, astronomie, etc.

Définition statistique d'une série temporelle

Statistiquement on peut définir une série temporelle comme une suite de *variables aléatoires* $(X_t)_t$ dont on observe une réalisation $(X_t(\omega))_t$

La suite $(X_t)_t$ est appelée *processus stochastique*

➡ Il n'y a pas d'hypothèse sur la relation entre les X_t

Les composantes d'une série

La *série brute* se décompose en plusieurs éléments de base inobservables :

- la composante tendance-cycle
 - tendance
 - cycle
- la composante saisonnière
- les autres effets de calendrier
 - la composante « jours ouvrables »
 - l'effet de graduel de Pâques (fêtes mobiles)
- la composante irrégulière (« l'irrégulier »)

La série désaisonnalisée est constituée des composantes tendance-cycle **et** irrégulière

La composante tendance-cycle

- La tendance est l'évolution de longue durée de la série
Cette composante traduit les variations de fond de la série observées sur une longue période de temps. C'est une composante « lisse » ne traduisant « pas d'évolution brusque »
- Le cycle est le mouvement lisse et presque périodique autour de la tendance.
Il se caractérise par une alternance de périodes d'expansion et de contraction, dont la longueur varie entre une année et demie et dix ans (pour les « cycle de croissance »).

Difficile de les estimer séparément en pratique car les séries sont courtes (et ce n'est pas l'objectif !).

Les méthodes de décomposition estiment ensemble tendance et cycle.

La composante tendance-cycle n'est publiée que par ABS et Statistique Canada.

Les effets de calendrier

- Effet des jours ouvrables
Lié à la composition journalière du mois ou du trimestre : un dimanche de plus ou de moins peut affecter la production.
Effets des jours fériés, le mois de mai en particulier.
- Effet de Pâques (fêtes mobiles)
Pâques est en mars ou en avril, au 1er ou au 2eme trimestre
Or Pâques affecte les ventes de chocolat, fleurs, agneau etc.
Les effets sont *ponctuel* et/ou *graduels*

La composante irrégulière

La composante irrégulière est faite des fluctuations résiduelles et erratiques qui ne peuvent être attribuées aux autres composantes : tendance-cycle, saisonnalité, effets de calendrier.

Exemples :

- Événements inhabituels : grèves, inondations, etc.
- Erreurs de mesures sur la variable elle-même dues à la collecte et au traitement des données.
- Fin d'une mesure fiscale ponctuelle, d'un programme d'aide ponctuel, etc.

Sommaire

1. Quelques définitions

2. L'exploration d'une série

2.1 L'exploration : une phase indispensable

2.2 L'analyse graphique

2.3 Autres outils disponibles

3. Schémas de décomposition

4. Conclusion

Questions de positionnement

Comment et par quoi démarrer l'étude d'une série temporelle ?

Quels sont les outils simples que l'on peut utiliser ?

Quand et pourquoi étudier les données dans le domaine des fréquences ?

Qu'est ce qu'un spectre ?

Que faire avec une série ?

- Se renseigner sur la série (mode de collecte, etc.) : y a-t-il une saisonnalité ?
- Décrire : déterminer des composantes
- Filtrer : transformer la série par moyenne mobile par exemple
- Modéliser : rechercher les causalités, les corrélations
- Prévoir : anticiper sur les informations futures

L'exploration : une phase essentielle !

Objectifs :

- Prédéterminer certaines caractéristiques de la série : Composantes (tendance, saisonnalité...), points atypiques, ruptures de série ou de profil saisonnier...
- Deviner le type de schéma de décomposition pour la série Schéma additif ou multiplicatif ?

Défi à l'empan mnésique !

Que pouvez-vous dire sur cette série ? (IPI France)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1990	92,1	92,3	102,1	93,0	93,3	100,8	92,9	66,7	95,8	105,0	96,7	89,2
1991	92,5	89,2	97,4	93,8	87,5	100,3	93,4	64,3	96,9	103,5	94,0	92,1
1992	90,7	89,0	99,4	93,7	86,1	101,3	90,4	62,9	96,6	98,4	91,9	92,6
1993	82,3	84,0	95,6	88,3	82,2	97,9	85,5	61,3	93,7	93,0	88,3	92,1
1994	83,6	83,7	97,0	88,3	88,3	102,9	87,3	65,9	98,2	98,0	96,8	98,0
1995	91,8	90,1	102,9	90,4	91,6	103,7	90,6	66,8	98,7	101,4	97,2	94,8
1996	92,0	91,1	98,1	94,3	90,5	101,8	96,1	66,3	98,9	105,0	95,0	96,0
1997	91,9	91,3	99,1	102,8	93,2	108,2	100,4	70,5	107,3	114,1	99,6	106,7
1998	98,2	98,7	109,3	103,7	97,6	114,7	106,1	72,1	111,5	112,6	105,6	107,4
1999	97,2	98,3	114,5	104,8	99,9	120,2	105,7	76,1	115,2	115,1	111,1	114,0
2000	103,4	107,5	121,7	105,7	113,1	119,4	108,1	82,0	116,4	121,3	117,2	111,9
2001	110,7	108,9	124,0	109,3	109,8	121,9	112,4	85,5	114,1	123,4	114,2	104,9
2002	108,4	106,7	118,5	113,4	105,6	119,2	113,9	81,4	115,6	121,7	111,0	105,2
2003	106,9	105,4	117,1	112,0	101,5	115,2	111,2	75,7	117,5	122,4	107,8	109,3
2004	104,7	106,7	122,8	112,7	104,5	126,5	111,1	79,7	121,9	118,8	112,2	112,6
2005	107,6	106,3	118,8	113,7	109,7	125,0	106,4	81,7	123,0	115,1	115,5	111,6
2006	108,8	105,9	124,8	108,0	113,1	126,7	108,7	84,1	121,0	121,5	116,6	108,2
2007	111,5	109,6	124,0	111,7	111,7	126,6	116,6	87,1	117,3	127,2	118,0	106,5
2008	113,2	114,4	117,5	120,2	107,6	121,4	115,1	78,8	118,9	118,8	99,5	99,2
2009	88,7	88,1	100,0	88,0	88,0	105,5	87,0	70,0	100,5	104,7	99,0	100,0
2010	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Défi à l'empan mnésique ! (code)

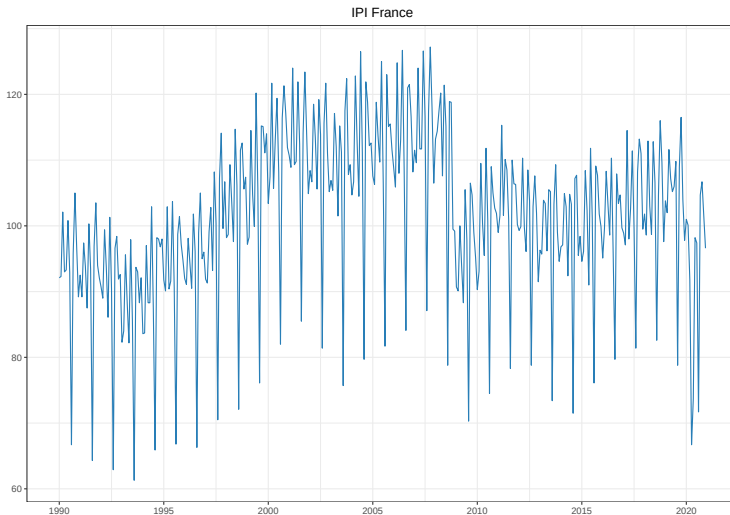
```
y <- ipi_c_eu[, "FR"]
mysa <- x13(y, spec = "RSA5c")

ipi_fr <- mysa$final$series
ipi_fr_y <- mysa$final$series[, "y"]
# Extraction effets JO
tde <- mysa$regarima$model$effects[, "tde"]
ipi_fr <- ts.union(ipi_fr, tde)
colnames(ipi_fr) <- gsub("^ipi_fr\\.\\.\"", "", colnames(ipi_fr))

tableau_ipifr <- gsub("\\\\.\"", "",
                    .preformat.ts(round(ipi_fr_y, 2)))
# knitr::kable(tableau_ipifr, align = "c")
library(kableExtra)
kbl(tableau_ipifr) %>%
  kable_styling(bootstrap_options = c("striped", "condensed", "responsive"),
               latex_options = "scale_down")
```


Rôle fondamental du graphique !

Et maintenant ?



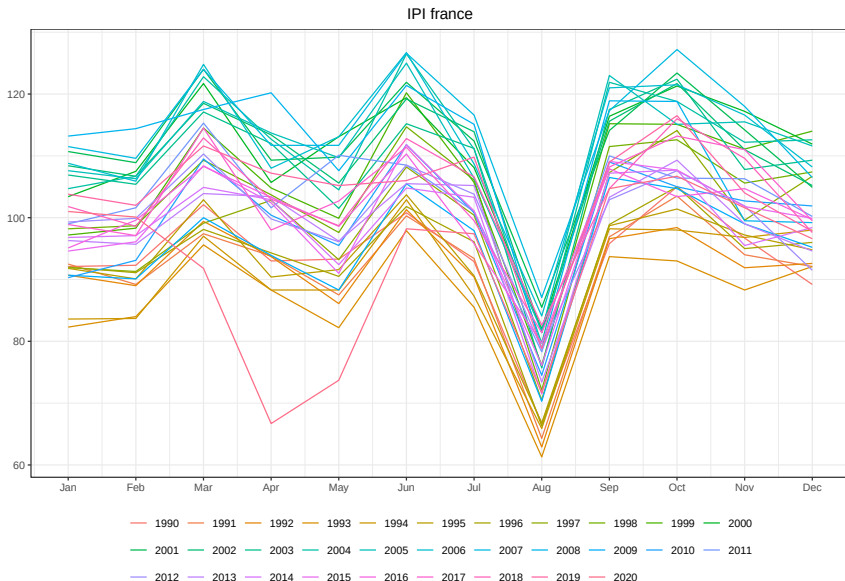
Rôle fondamental du graphique ! (code)

```
graph_comp <- data.frame(date = time(ipi_fr), ipi_fr)
graph_ipi_fr <- ggplot(graph_comp, aes(x = date, y = y)) +
  geom_line(colour="#287EB7") + theme_bw() +
  scale_x_continuous(breaks = scales::pretty_breaks(n = 12)) +
  labs(x=NULL, y=NULL, title = "IPI France") +
  theme(plot.title = element_text(hjust = 0.5))
graph_ipi_fr
```

Ou avec le package forecast :

```
library(forecast)
ipi_fr %>% autoplot() +
  theme_bw() +
  labs(x=NULL, y=NULL, title = "IPI France") +
  theme(plot.title = element_text(hjust = 0.5))
```

Représentation par année



Représentation par année (code)

```
graphique_par_annee <- .preformat.ts(ipi_fr_y)
noms_mois <- colnames(graphique_par_annee)
# on arrondi pour éviter des problèmes sur trunc si chiffres non entiers
annees <- as.character(trunc(round(time(ipi_fr_y),3)))
data_par_annee <- data.frame(annee = annees,
                             mois = noms_mois[cycle(ipi_fr_y)],
                             mois_num = cycle(ipi_fr_y),
                             value = ipi_fr_y)
ggplot(data_par_annee, mapping = aes(x = mois_num,y = value, color = annee)) +
  geom_line() +
  scale_x_continuous(breaks=1:12,
                     labels=function(x)noms_mois[x])+ theme_bw()+
  theme(legend.position="bottom",plot.title = element_text(hjust = 0.5))+
  guides(colour=guide_legend(nrow=3,byrow=TRUE,title = NULL))+
  labs(x=NULL,y=NULL,title = "IPI france")
```

Avec forecast :

```
ipi_fr_y %>% ggseasonplot() + theme_bw()
```

Sous JDemetra+, dans un graphique clic droit + split into yearly components

Quelles premières informations ?

Forte saisonnalité en août, point bas – octobre plutôt un point haut

Pas de tendance nette

Pas de rupture (en tendance ou en saisonnalité) visible

Série non stationnaire

Schéma plutôt additif

Autocorrélogramme

On regarde les corrélations de la série avec elle-même, retardée de 1, 2, 3... mois (ou trimestres).

Fonction d'autocorrélation d'ordre k (empirique)

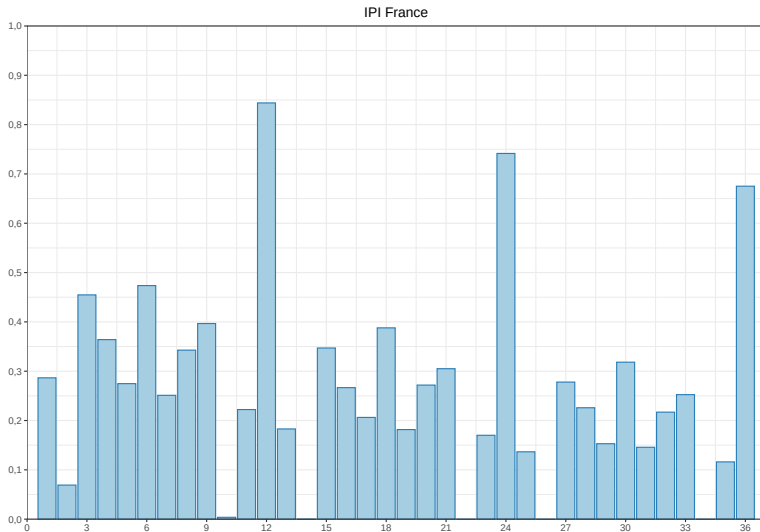
Coefficient de corrélation linéaire entre la série observée et elle-même retardée de k périodes, i.e. : rapport entre

- la covariance empirique entre la série t et la série retardée $t - k$
- la variance empirique de la série

$$\varrho_k = \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

Autocorrélogramme : le graphique qui représente ces autocorrélations

Exemple d'autocorrélogramme



Que voit-on ?

Le Spectre d'une série

Représentation classique d'une série : fluctuations de la série en fonction du temps. . .

Représentation spectrale : une série peut être représentée dans le domaine des fréquences. Un spectre montre les fluctuations de la série par fréquence.

$$\text{fréquence} = \frac{2\pi}{\text{période}}$$

Interprétation : **décomposition de la variance de la série** selon les différentes fréquences

Intérêt : Il y a des fréquences qui correspondent aux composantes : tendance-cycle, saisonnalité et effets de calendrier

! plusieurs façons de calculer le spectre ! Deux méthodes principales : périodogramme et spectre autorégressif

Sous **R** : `spec.pgram()` ou `spectrum(method = "pgram")`

Exemple d'un Spectre (1/2)

Que voit-on ?

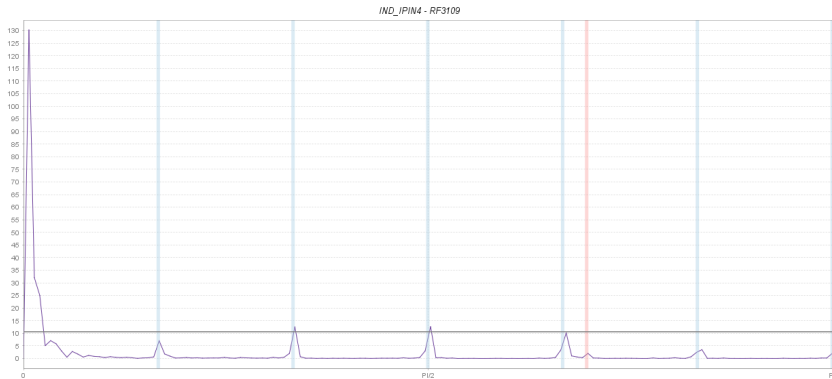


Figure 1 – Spectre

Pic spectral aux basses fréquences

Exemple d'un Spectre (2/2)

Que voit-on ?

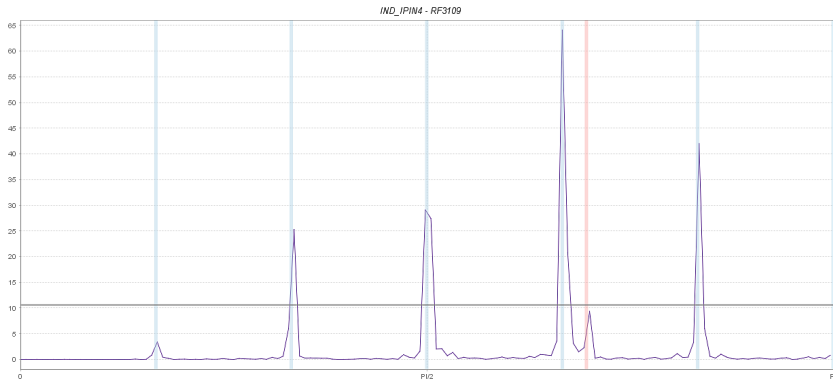


Figure 2 – Spectre

Pics spectraux aux fréquences $k \frac{\pi}{6} = k \frac{2\pi}{12}$

Périodogramme (1)

Sous **R** : `spec.pgram()` ou `spectrum(method = "pgram")`

Soit $x_1, \dots, x_n$ des observations d'une série temporelle. On suppose $x_1, \dots, x_n$ les valeurs d'une fonction en $1, \dots, n$:

$$x_t = \frac{1}{\sqrt{n}} \sum_{\pi < \omega_j \leq \pi} a_j e^{it\omega_j}, \quad \omega_j = \frac{2\pi j}{n}$$

Rmq : c'est bien possible d'écrire sous cette forme car $e_j = n^{-1/2}(e^{i\omega_j}, \dots, e^{in\omega_j})$ base orthonormée de $\mathbb{C}^n$

Il vient :

$$x = \sum_{j|\pi < \omega_j \leq \pi} a_j e_j \quad \text{avec} \quad a_j = \langle x, e_j \rangle = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-it\omega_j}$$

Périodogramme (2)

La valeur du périodogramme $I(\omega_j)$ est :

$$I(\omega_j) = |a_j|^2 = \frac{1}{n} \left| \sum_{t=1}^n x_t e^{-it\omega_j} \right|^2$$

On a donc une décomposition de la variance :

$$\|x\|^2 = \sum I(\omega_j)$$

Rmq : dans le cas réel, $I(\omega_j) = I(-\omega_j)$

Rmq : généralement x est normalisé (en enlevant également la tendance)

Donne les mêmes informations que les autocorrélogrammes :

$$I(\omega_j) = \sum_{|k| < n} \hat{\gamma}(k) e^{-ik\omega_j}$$

Spectre autorégressif

Sous **R** : `spec.ar()` ou `spectrum(method = "ar")`

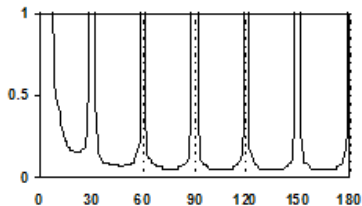
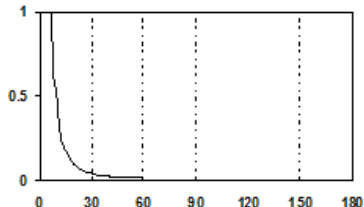
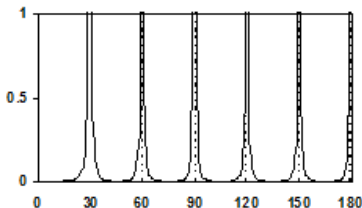
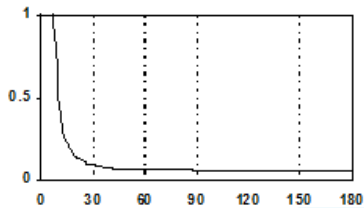
Autre méthode : calculer le spectre d'un processus AR :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

$$f(\omega) = \sigma^2 \frac{1}{2\pi \left| 1 - \sum_{j=1}^p \phi_j e^{ij\omega} \right|^2}$$

Rmq : sous R $f(\omega) = \sigma^2 \frac{1}{\text{freq} \left| 1 - \sum_{j=1}^p \phi_j e^{ij\omega} \right|^2}$

Spectres théoriques des composantes

Series**Trend****Seasonality****Seasonally adjusted series**

Sommaire

1. Quelques définitions

2. L'exploration d'une série

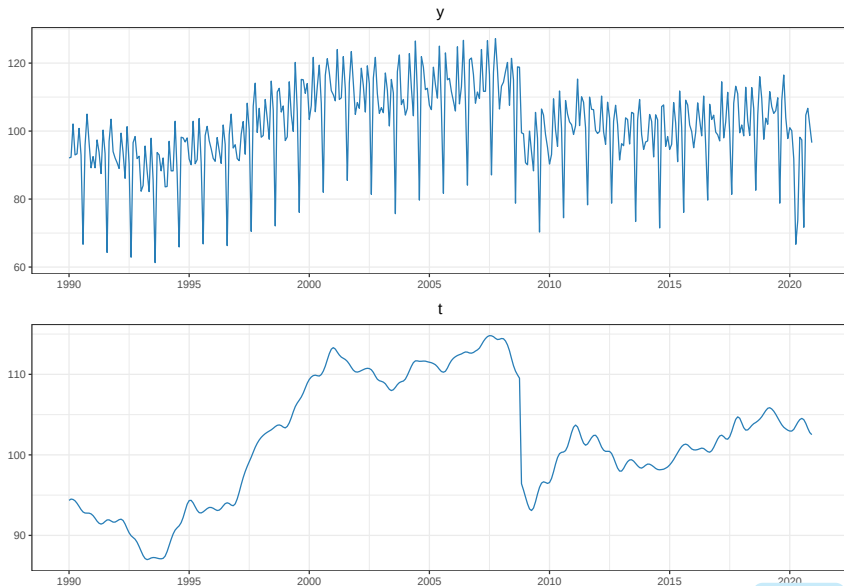
3. Schémas de décomposition

3.1 Exemple des différentes composantes

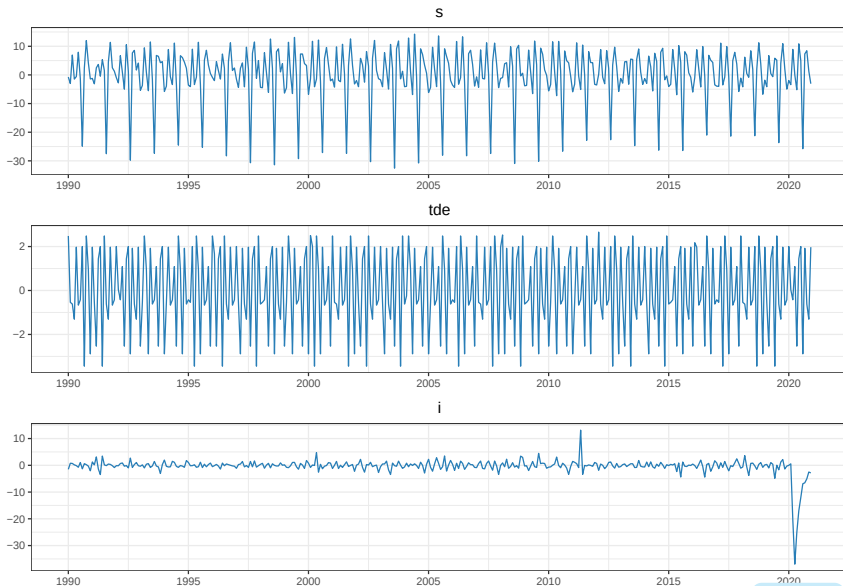
3.2 Les modèles de décomposition

4. Conclusion

Exemple (1/2)



Exemple (2/2)



Exemple (code)

```
library(patchwork)
plot_comp <- function(comp){
  ggplot(graph_comp,aes_string(x = "date",y = comp))+
    geom_line(colour="#287EB7")+theme_bw() +
    labs(x=NULL,y=NULL,title = comp)+
    theme(plot.title = element_text(hjust = 0.5)) +
    scale_x_continuous(breaks = scales::pretty_breaks(n = 8)) +
    scale_y_continuous(breaks = scales::pretty_breaks(n = 4),
                      labels = function(x) format(x, decimal.mark = ","))
}
plot_comp("y") / plot_comp("t")
plot_comp("s") / plot_comp("tde") / plot_comp("i")
```

Modèles de décomposition

Schéma additif (amplitude de S_t et I_t indépendants du niveau) :

$$X_t = TC_t + S_t + I_t$$

Schéma multiplicatif (amplitudes de S_t et I_t varient en fonction de la tendance) :

$$X_t = TC_t \times S_t \times I_t \text{ ou } X_t = TC_t \times (1 + \tilde{S}_t) \times (1 + \tilde{I}_t)$$

Schéma pseudo-additif (permet de traiter des cas où X_t proche de 0) :

$$X_t = TC_t \times (S_t + I_t)$$

Schéma log-additif :

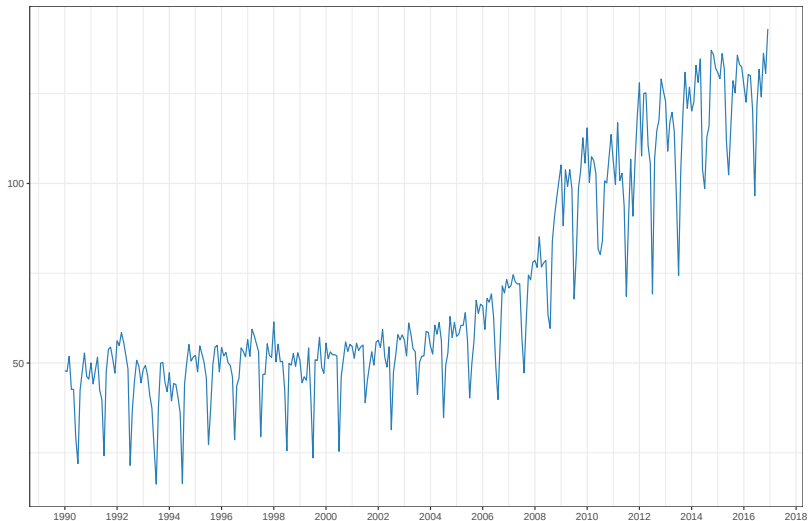
$$\log(X_t) = \log(TC_t) + \log(S_t) + \log(I_t)$$

Voir *cette page* pour des exemples

En pratique, également effets de calendrier, points aberrants, etc.

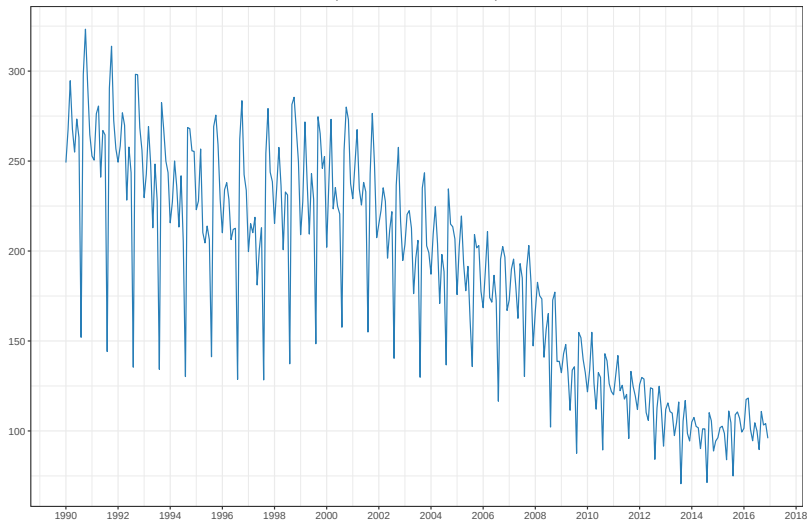
Exemples (1/2) : Schéma additif

IPI branche 1041
(Fabrication d'huiles et graisses)



Exemples (2/2) : Schéma multiplicatif

IPI branche 3109
(Fabrication d'autres meubles)



Sommaire

1. Quelques définitions
2. L'exploration d'une série
3. Schémas de décomposition
- 4. Conclusion**

Tout est relatif !

On doit résoudre une équation à plusieurs inconnues, composantes inobservables.

Il n'existe pas de solution unique !

En particulier, une série ne possède pas UNE seule tendance : cela dépend du problème et des hypothèses que l'on prend.

⇒ Une solution est donc relative au choix du point de vue et des hypothèses

La statistique est l'art de résumer des informations, il n'y a pas qu'un “bon” résumé.

Les essentiels

- Une CVS donne le mouvement de moyen et long terme de la série (tendance-cycle), ainsi que des informations ponctuelles (irrégulier)
- Démarrer par une exploration graphique de ses séries
- Une série peut être étudiée dans sa dimension temporelle ou sa dimension fréquentielle (le spectre)
- Le schéma de décomposition peut être additif ou multiplicatif, ce dernier est plus fréquent
- Le résultat de la décomposition n'est pas unique et dépend du point de vue et des hypothèses de travail