



2 - Exploration des séries et décomposition

ANNA SMYK ET TANGUY BARTHÉLÉMY
Division Recueil et Traitement de l'Information
Département des Méthodes Statistiques

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 connaîtrez la définition d'une série temporelle ainsi que ses principales composantes
- vous saurez utiliser les outils graphiques pour l'explorer
- vous pourrez identifier son schéma de décomposition

Sommaire

1. Quelques définitions

1.1 Définition d'une série temporelle

1.2 Description des différentes composantes d'une série

2. L'exploration d'une série

3. Schémas de décomposition

4. Conclusion

Définition d'une série temporelle

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

- valeurs discrètes (mensuelles, trimestrielles)
- valeurs “continues” (cours d'une action boursière)

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

Les composantes d'une série

Afin d'enlever l'effet saisonnier, on doit décomposer la série.

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

- la composante saisonnière
- la composante tendance-cycle
 - tendance
 - cycle
- les effets de calendrier
- la composante irrégulière (« l'irrégulier »)

La composante saisonnière (1/2)

C'est celle que l'on cherche à retirer.

Vu précédemment :

- elle suit un profil annuel périodique
- elle est « localement » stable : saisonnalité supposée constante sur plusieurs années consécutives
- on la modélise par des coefficients associés à chaque période (mois ou trimestre)

La composante saisonnière (2/2)

Saisonnalité constante :

Pour chaque date t (par exemple un mois) le coefficient saisonnier reste constant sur deux années consécutives

$$S_t = S_{t+12}$$

Comme cela est vrai pour t :

$$\sum_{i=0}^{11} S_{t+i} = \sum_{i=1}^{12} S_{t+i}$$

$$\sum_{i=0}^{11} S_{t+i} = \sum_{i=k}^{11+k} S_{t+i} = cst$$

Finalement

$$\sum_{i=0}^{11} S_{t+i} = 0$$

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 (par exemple pour les « cycles de croissance »).

Les méthodes de désaisonnalisation ne distinguent pas, en général, le cycle de la tendance. On parle alors d'une seule composante : la tendance. La composante tendance n'est pas publiée à l'Insee, qui diffuse des séries brutes et des séries cvs-cjo.

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 1^{er} ou au 2^e trimestre
Or Pâques affecte les ventes de chocolat, fleurs, agneau etc.
Les effets sont *ponctuels* 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, saisonnalité, effets de calendrier. Elle est déduite par différence entre la série brute et les autres composantes, dont les caractéristiques sont plus formalisées.

Exemples :

- Événements inhabituels : grèves, inondations, etc.
- Erreurs de mesure sur la variable elle-même dues à la collecte et au traitement des données.

La série CVS

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

Attention : l'utilisateur s'attend parfois à un résultat nécessairement lisse, car il pense avoir affaire à une tendance.

Sommaire

1. Quelques définitions

2. L'exploration d'une série

2.1 L'analyse d'une série temporelle

2.2 L'analyse graphique

2.3 Autres outils disponibles

3. Schémas de décomposition

4. Conclusion

Les objectifs usuels de l'analyse d'une série

- Décrire son évolution
- En extraire ce qui est interprétable : décomposition puis désaisonnalisation
- La modéliser (formalisation de la dynamique)
- Prévoir les valeurs futures (notamment pour améliorer la désaisonnalisation)

Il est utile de se renseigner sur la série (que mesure-t-on, objectifs, mode de collecte, etc.) : y a-t-il une saisonnalité et des effets de calendrier "attendus" ?

L'analyse graphique

Objectifs

Prédéterminer certaines caractéristiques de la série :
composantes (tendance, saisonnalité...), points atypiques, ruptures de série
ou de profil saisonnier...

Les guidelines d'Eurostat recommandent explicitement une analyse
graphique, au moins une fois par an, sur les séries "les plus importantes
pour l'utilisateur".

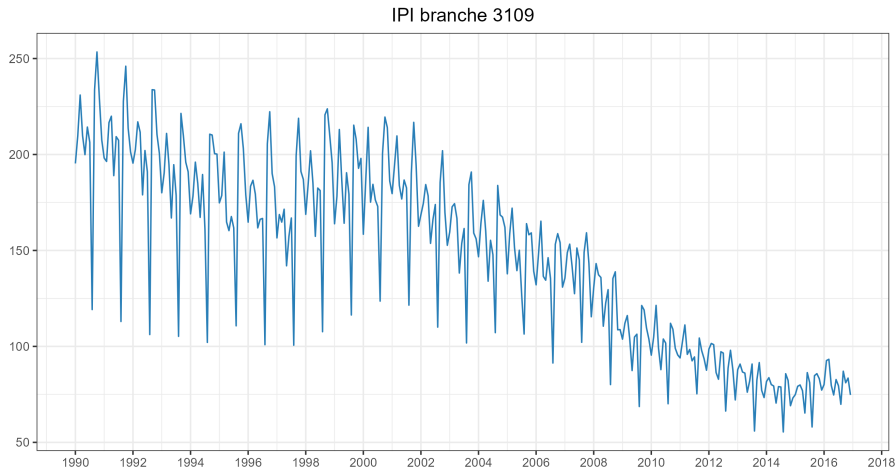
Présentation en tableau

Que peut-on dire sur cette série ? (IPI branche 3109)

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1990 | 195,33 | 209,79 | 230,98 | 209,86 | 199,90 | 214,29 | 206,33 | 119,19 | 233,69 | 253,42 | 229,36 | 207,51 |
| 1991 | 198,17 | 196,37 | 216,60 | 219,95 | 188,98 | 209,30 | 207,38 | 112,98 | 227,73 | 246,00 | 213,70 | 201,32 |
| 1992 | 195,50 | 202,55 | 217,00 | 211,60 | 178,97 | 202,07 | 190,92 | 106,17 | 233,78 | 233,58 | 210,52 | 200,69 |
| 1993 | 180,05 | 190,32 | 210,98 | 194,65 | 166,92 | 194,67 | 178,67 | 105,24 | 221,41 | 209,85 | 195,99 | 190,98 |
| 1994 | 169,08 | 178,21 | 196,02 | 184,90 | 167,24 | 189,55 | 161,07 | 102,08 | 210,58 | 210,12 | 200,42 | 200,21 |
| 1995 | 174,80 | 178,62 | 201,24 | 164,83 | 160,32 | 167,66 | 161,75 | 110,72 | 211,01 | 216,00 | 203,02 | 179,07 |
| 1996 | 164,74 | 183,47 | 186,58 | 179,33 | 161,75 | 166,23 | 166,67 | 100,88 | 205,55 | 222,28 | 190,07 | 183,21 |
| 1997 | 156,54 | 168,71 | 164,77 | 171,48 | 142,00 | 157,40 | 166,91 | 100,62 | 199,26 | 218,87 | 191,15 | 186,99 |
| 1998 | 168,81 | 184,43 | 201,92 | 185,14 | 157,35 | 182,51 | 181,22 | 107,65 | 220,65 | 223,78 | 210,02 | 195,71 |
| 1999 | 163,90 | 178,28 | 213,01 | 186,94 | 164,17 | 190,49 | 179,33 | 116,36 | 215,28 | 207,98 | 192,81 | 197,97 |
| 2000 | 158,44 | 186,47 | 214,15 | 175,17 | 184,45 | 176,31 | 172,95 | 123,59 | 200,67 | 219,50 | 214,03 | 186,17 |
| 2001 | 179,60 | 194,49 | 209,67 | 183,87 | 176,82 | 186,69 | 182,59 | 121,49 | 190,21 | 216,75 | 194,34 | 162,60 |
| 2002 | 168,62 | 174,46 | 184,33 | 178,46 | 153,70 | 166,04 | 173,93 | 110,05 | 185,39 | 201,96 | 169,80 | 152,68 |
| 2003 | 159,59 | 172,72 | 174,39 | 166,76 | 138,23 | 153,66 | 161,42 | 101,79 | 184,30 | 190,87 | 159,08 | 156,06 |
| 2004 | 146,71 | 164,16 | 176,09 | 160,05 | 133,98 | 155,29 | 147,78 | 107,16 | 183,86 | 168,51 | 167,37 | 162,04 |
| 2005 | 137,83 | 158,73 | 172,04 | 151,52 | 139,49 | 150,11 | 126,54 | 106,44 | 163,96 | 158,17 | 159,16 | 139,20 |
| 2006 | 132,09 | 148,08 | 165,26 | 136,51 | 134,49 | 146,21 | 135,37 | 91,34 | 153,22 | 158,74 | 154,08 | 130,89 |
| 2007 | 135,54 | 148,87 | 153,28 | 141,61 | 127,48 | 151,28 | 145,04 | 102,13 | 149,04 | 159,20 | 143,35 | 115,45 |
| 2008 | 130,57 | 143,16 | 137,32 | 135,92 | 110,53 | 122,60 | 129,57 | 80,10 | 135,40 | 138,87 | 108,66 | 108,77 |
| 2009 | 103,77 | 112,04 | 116,08 | 104,01 | 87,41 | 104,90 | 106,37 | 68,67 | 121,31 | 118,99 | 109,61 | 103,77 |
| 2010 | 95,45 | 104,94 | 121,37 | 99,15 | 87,89 | 103,86 | 101,73 | 70,11 | 112,04 | 108,92 | 98,96 | 95,58 |
| 2011 | 93,99 | 102,55 | 111,18 | 95,86 | 98,42 | 92,47 | 94,52 | 75,29 | 104,37 | 97,79 | 93,57 | 87,59 |
| 2012 | 98,56 | 101,54 | 100,93 | 86,39 | 82,90 | 97,23 | 96,51 | 66,32 | 88,64 | 97,99 | 87,79 | 72,11 |
| 2013 | 87,90 | 90,80 | 86,68 | 86,10 | 76,14 | 81,98 | 90,82 | 55,91 | 82,54 | 91,57 | 77,09 | 73,35 |
| 2014 | 81,69 | 83,71 | 80,12 | 79,38 | 70,47 | 79,06 | 78,86 | 55,43 | 85,77 | 82,29 | 69,10 | 73,12 |
| 2015 | 74,85 | 79,21 | 79,91 | 76,97 | 65,28 | 86,35 | 81,16 | 58,02 | 84,75 | 85,81 | 83,17 | 77,23 |
| 2016 | 80,12 | 92,62 | 93,30 | 79,68 | 74,70 | 82,74 | 79,51 | 69,83 | 87,08 | 81,05 | 83,50 | 74,71 |

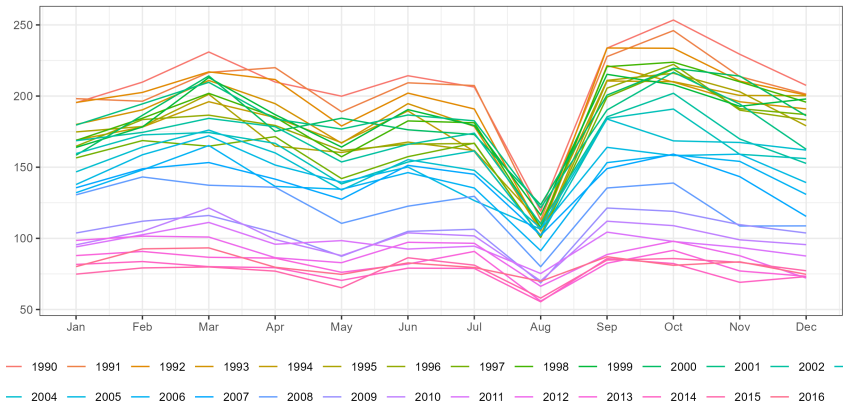
Rôle fondamental du graphique !

Et maintenant ?



Représentation par année

IPI branche 3109



Informations fournies par les graphiques

Forte saisonnalité en août (point bas) ; octobre est plutôt un point haut

Évolution tendancielle à la baisse

Pas de rupture visible (en tendance ou en saisonnalité)

Variance « proportionnelle » au niveau de la série

La représentation en tableau est très utile dans un deuxième temps (précision). Elle est aussi utilisée pour des tests par les algorithmes de désaisonnalisation. (Analyse de la variance)

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)

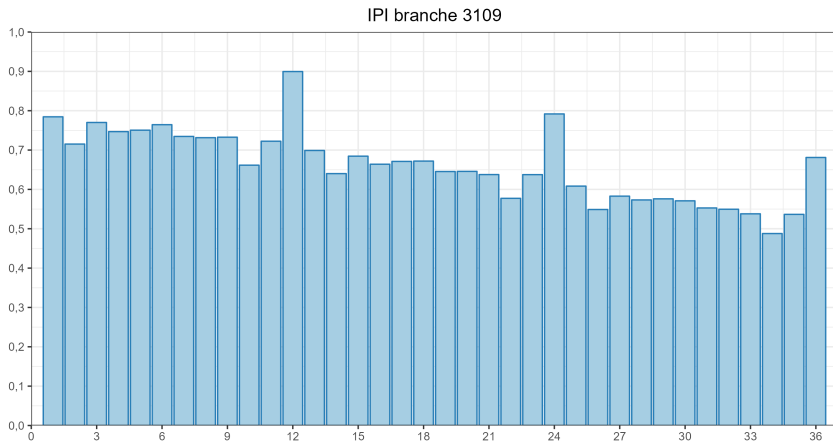
Coefficient de corrélation linéaire entre la série observée et elle-même retardée de k périodes, i.e. rapport entre :

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

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

Autocorrélogramme : le diagramme en bâtons qui représente ces autocorrélations

Exemple d'autocorrélogramme



Que voit-on ?

Le spectre d'une série

Représentation usuelle d'une série : variations de la série en fonction du temps

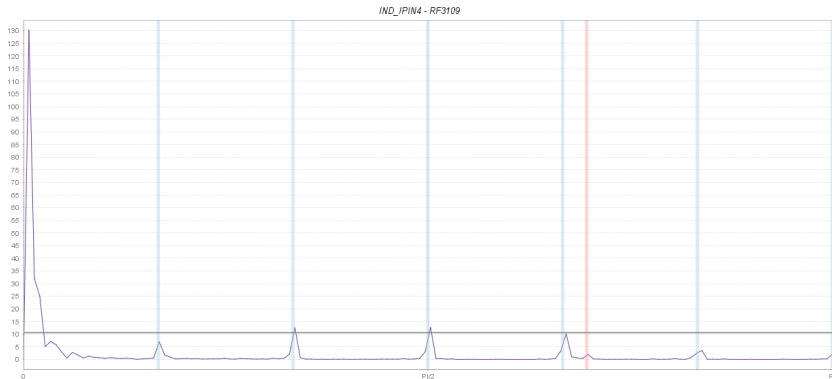
Représentation spectrale : une série peut être représentée dans le domaine des fréquences, en utilisant la transformée de Fourier. Un spectre montre les fluctuations de la série par fréquence.

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

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

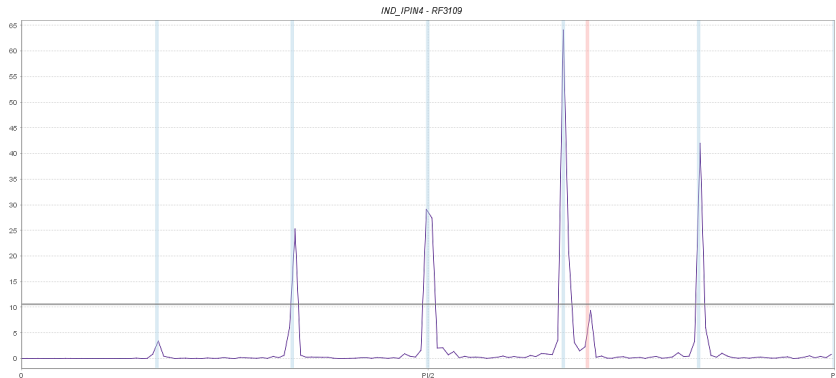
Intérêt : les différentes composantes (tendance, saisonnalité et effets de calendrier) ont des fréquences caractéristiques. Le spectre permet de détecter leur présence.

Exemple d'un spectre (1/2)



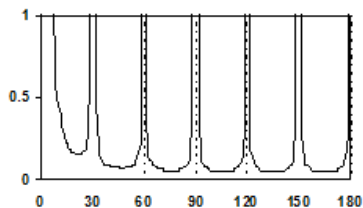
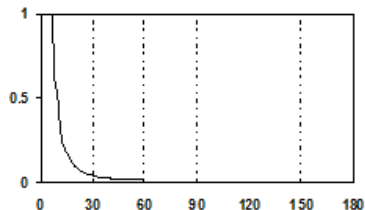
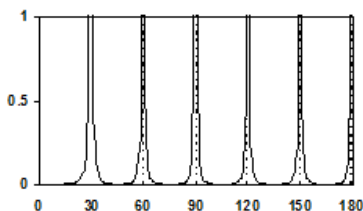
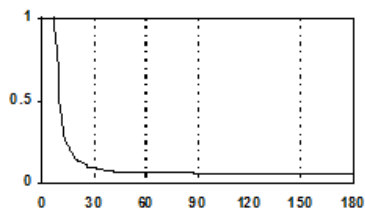
Pic spectral aux basses fréquences (Les basses fréquences correspondent à une "période infinie" donc à une tendance)

Exemple d'un spectre (2/2)



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

Spectres théoriques des composantes

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

Sommaire

1. Quelques définitions

2. L'exploration d'une série

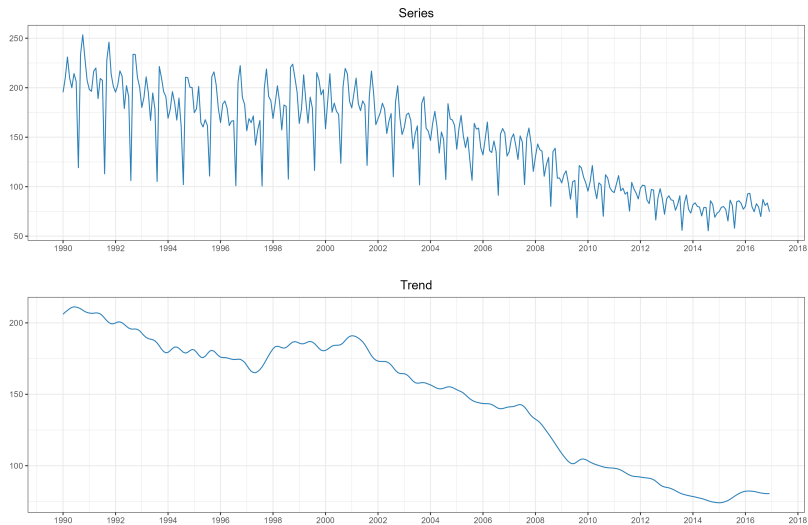
3. Schémas de décomposition

3.1 Exemple des différentes composantes

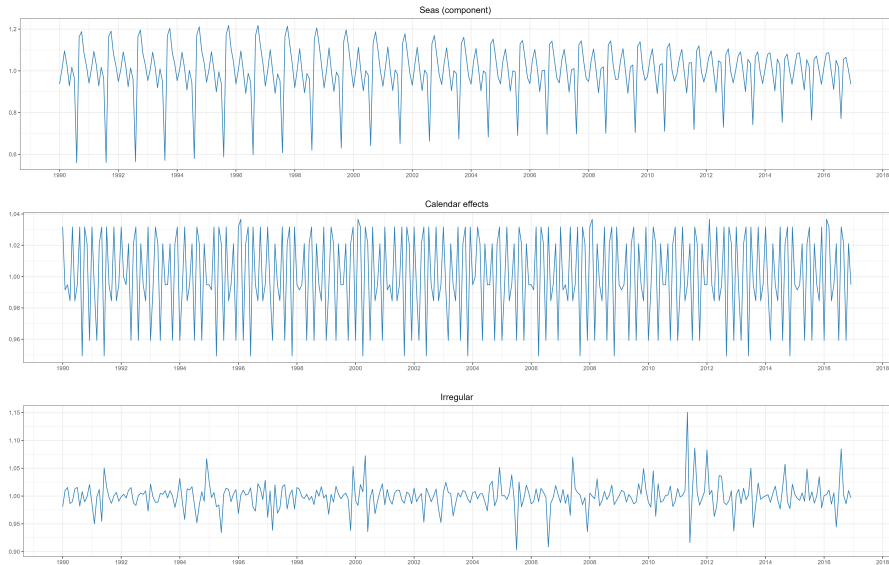
3.2 Les modèles de décomposition

4. Conclusion

Exemple (1/2)



Exemple (2/2)



Modèles de décomposition (1/2)

Schéma additif :

$$X_t = T_t + S_t + I_t$$

Schéma multiplicatif :

$$X_t = T_t \times S_t \times I_t$$

Série CVS = série brute d'où l'on a éliminé S_t :

$$CVS_t = X_t - S_t \text{ ou } CVS_t = X_t / S_t$$

Si l'on tient compte des effets de calendrier, on se réfère au schéma suivant (idem en multiplicatif) :

$$X_t = T_t + S_t + I_t + C_t$$

où C_t représente les effets de calendrier.

Série CVS-CJO = série brute d'où l'on a éliminé S_t et C_t :

$$CVSCJO_t = X_t - S_t - C_t \text{ ou } CVSCJO_t = X_t / (S_t \times C_t)$$

Modèles de décomposition (2/2)

Dans le schéma global de décomposition d'une série, le plus souvent, on n'écrit pas explicitement l'effet de calendrier C_t

On suppose que S_t contient C_t

S_t est en fait S_t (coefficient saisonnier "pur") $+ / \times C_t$ (effet de calendrier)

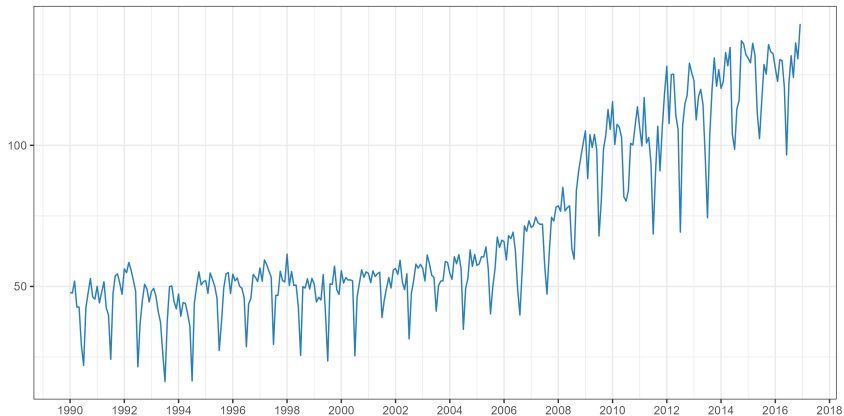
Ainsi on dit couramment "série cvs" pour désigner une "série cvs-cjo".

Dans les logiciels, le coefficient S_t final représente bien les effets saisonniers et de calendrier cumulés.

Exemples (1/2)

Schéma additif

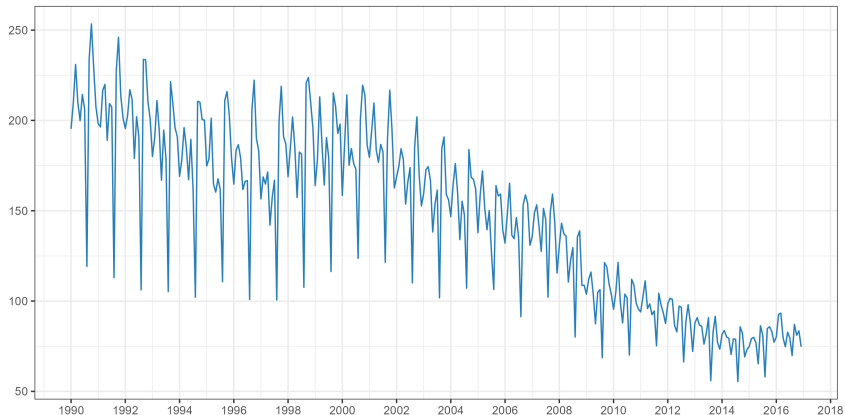
IPI branche 1041



Exemples (2/2)

Schéma multiplicatif

IPI branche 3109



Sommaire

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

Difficultés de résolution

On doit résoudre une équation à plusieurs inconnues car les composantes sont toutes inobservables.

→ Il n'existe pas de solution unique ! Mais une classe de solutions plus ou moins restreinte selon les contraintes que l'on impose sur les différentes composantes et la manière choisie pour les calculer.

Les hypothèses privilégiées sur la forme des composantes sont *in fine* à la base des différences entre les méthodes, telles que X-13 Arima vs Tramo-Seats.

Les essentiels

- Une série CVS donne le mouvement de moyen et long terme de la série (tendance-cycle : tendance), ainsi que des informations ponctuelles et le bruit (irrégulier).
- Il faut démarrer par une exploration graphique de ses séries.
- 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 des hypothèses de travail (contraintes imposées sur les différentes composantes).