Prédiction de la Production de Bière en Australie

Par :

Philippine RENAUDIN
Elvina EURY

Master 2 Mathématiques et Applications
Parcours Data Science

Sous la direction de M. Frédéric Proia

2020-2021

Introduction

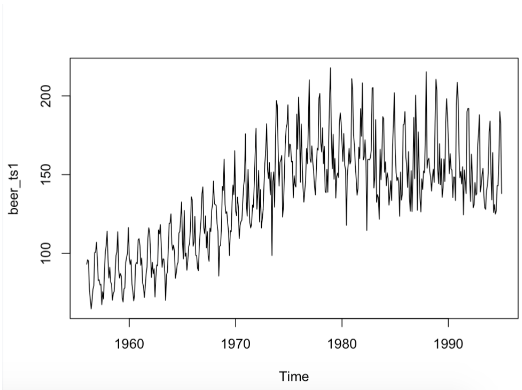
L'objectif de cet étude est de prédire la production de bière en Australie pour les trois prochaines années à partir de séries chronologiques. La production de bière est une variable endogène car elle est dépendante de d'autres facteurs comme le climat. Nous nous basons sur des données mensuelles, de janvier 1956 à août 1995. Comme il est difficile de visuellement discerner le modèle multiplicatif de l'additif nous analysons à la fois le modèle initial et le modèle transformée. Notre approche de vise en premier lieu à transformer les données, à éliminer la périodicité et la tendance avant de rendre le processus stationnaire. Puis des tests statistiques sont effectués sur les résidus afin de cibler les modèles les plus robustes. Finalement, nous utilisons les modèles choisis afin de prédire la production de bière sur les trois prochaines années. A COMPLETER/ MODIFIER...

Méthologie

.1 Analyse initiale

Les séries chronologiques ont 3 composantes principales : La tendance (T), qui décrit les mouvement à long terme, la périodicité (S), qui est cyclique, et les fluctuations (ϵ).

Nous débutons notre approche en faisant une analyse visuelle des séries chronologiques.



À première vue nous voyons que nos séries sont non-stationnaire avec une tendance et une périodicité annuelle. Cette dernière est d'ailleurs confirmé sur les sorties ACF et PACF des séries initiales. Il est dur à dire si la tendance est quadratique ou linéaire. Nous nous baserons sur des tests supplémentaires pour être sûr. De plus, nous remarquons un accroissement subtile de la saisonnalité, nous laissant un doute quant à la nature multiplicative ou additive du modèle. Nous avons ainsi décidé de comparer les résultats obtenus des séries initiales, avec les résultats du logarithme des séries.

.2 Élimination de la tendance et de la saisonnalité –MEILLEUR TITRE ?

À partir de ces deux modèles, nous commencons par faire une différenciation $(I - B)$, soit d=1. Les sorties ACF et PACF nous affiche des pics à 12, 24, et plus, nous poussant à faire une prochaine différenciation, $(I - B^{12})$, soit D=1. Nous prenons ainsi les cas de différenciation suivants :

1. d=0, D=1
2. d=1, D=1

Pour le premier cas, la série est non-stationnaire, tandis que lorsque nous prenons d=1 et D=1, nous obtenons une série stationnaire, un résultat qui est vu à la fois via les sorties ACF, PACF et à la fois grâce aux tests ADF et KPSS. Nous établirons nos prochains modèles à partir de Y_t (cas non-transformé) et YL_t (cas transformé)
Soit $(X_t)_{t\in\mathbb{Z}}$ le processus initial

$$Y_t = (I - B) * (1 - B^{12}) * X_t$$
$$YL_t = (I - B) * (1 - B^{12}) * log(X_t)$$

.3 Création de modèles

Processus Y_t