

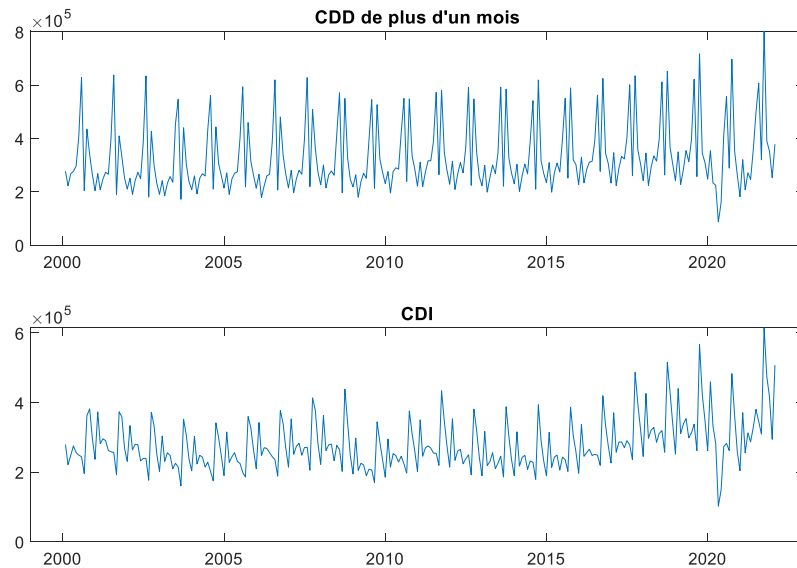
Projet séries temporelles

V2

2IA - 2^e année - 2022

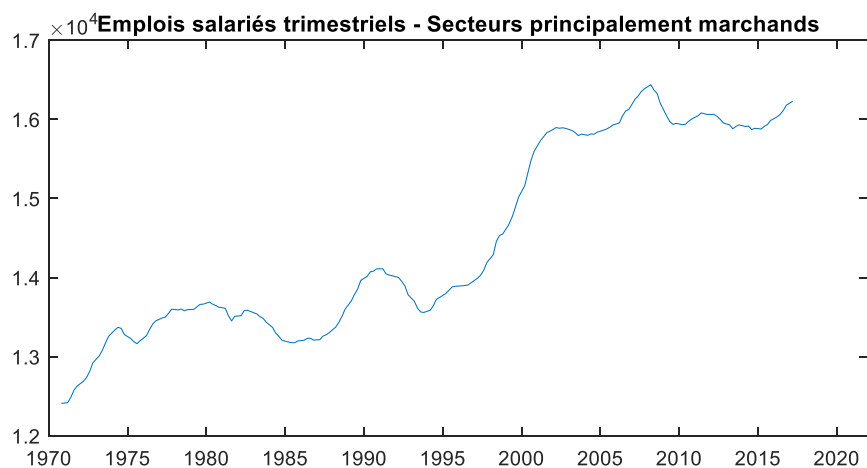
(B. Xu, P. Guyot)

Déclarations préalables à l'embauche (DAPE) mensuelles de plus d'un mois, France entière (2000 – 2022)



- *Année*: annee[date] Année
- *Trimestre*: trimestre[int] Trimestre (1, 2, 3 ou 4)
- *Dernier jour du mois*: dernier_jour_du_mois[date] Dernier jour du mois
- *Durée de contrat*: duree_de_contrat[text] Durée de contrat
- *Nature de contrat*: nature_de_contrat[text] Nature de contrat (CDD et CDI de plus d'un mois)
- *DPAE (brut)*: dpae_brut[int] Nombre de déclarations d'embauche (données brutes)
- *DPAE (cvs)*: dpae_cvs[int] Nombre de déclarations d'embauche (données correction des variations saisonnières)

Estimations d'emploi salarié par secteur d'activité (EESSA), 80 secteurs, 1970 – 2022



Question 1 : visualisation des deux bases de données

- Dans les données DPAE, on dispose de deux séries temporelles en fonction du contrat (CDD ou CDI). Créer deux figures pour visualiser ces deux séries temporelles. Effectuez des statistiques sur chaque type de contrat, de manière trimestrielle et annuelle. *Par exemple, vous pourrez afficher par année et par trimestre les statistiques classiques : moyennes, médianes, écart type, variance. Vous pourrez également afficher des boîtes à moustaches (par année et/ou par trimestre).* Quelles sont les tendances pour l'évolution des nombres de CDD et de CDI ?
- Dans les données EESSA, 80 séries temporelles existent.
 - Nettoyer ces données EESSA (de manière à ce qu'elles soient utilisables en python). *Vous pourrez par exemple remplacer les valeurs manquantes « (O) » par des « pandas.NA », utiliser les temps en ligne et les secteurs en colonnes, etc.*
 - Créer une seule série temporelle qui somme l'ensemble de secteurs. Quelles sont vos observations ?
 - Sélectionner 2 à 5 secteurs qui vous intéressent. Comment évoluent ces secteurs ?
 - Affichez des statistiques classiques sur ces séries temporelles

Question 2 : Saisonnalité et corrélations

- Évaluer la saisonnalité des CDD et CDI sur DPAE
- Évaluer la saisonnalité des secteurs sélectionnés sur EESSA

Vous pourrez par exemple afficher les graphiques superposés sur une année (voir TP « chickenpox »).

- Existe-il une corrélation entre les séries de DPAE et les séries sélectionnées d'EESSA ?

Remarque : *attention à sélectionner la même période. De plus, les deux bases de données n'ont pas la même résolution temporelle. Vous pourrez effectuer une interpolation ou un sous-échantillonnage du signal.*

Question 3 : prédiction par lissage exponentiel

Séparer vos données en ensemble d'apprentissage et de test pour effectuer des prédictions par lissage temporel sur les données DPAE (CDD et CDI) et EESSA (secteurs sélectionnés).

Question 4 : évaluation de la prédiction

Évaluer votre prédiction par le calcul d'erreur.

Question 5 : application et évaluation d'autres modèles de prédiction

Utiliser d'autres modèles (par exemple ARMA) pour effectuer la prédiction. Comparer les résultats.

Question 6 : bonus cadeau

- DPAE : prédire les emplois (CDD et CDI) de février à mai 2022
- EESSA : prédire les emplois (de secteurs sélectionnés) du 1^{er} trimestre de 2022