

# Sujet Post-Doc: Machine learning algorithms to forecast inflation in a data-rich environment

*Randal Douc et Sylvain Le Corff.*

## Résumé

Dans de nombreux secteurs d'activité, il est crucial de connaître un indicateur fiable des prix des biens du marché en temps réel. Cet indicateur est un outil utile à tous les acteurs économiques dans le but de suivre l'évolution des prix afin d'optimiser les décisions d'investissement, de prédire les prix futurs et d'adapter les politiques fiscales et monétaires. L'objectif de ce projet est de développer un modèle statistique permettant de décrire la dynamique de l'inflation et notamment d'anticiper en temps réel ses points d'inflexion et de rupture. Il s'agit plus particulièrement d'identifier les covariables économiques, environnementales ou sociales expliquant la distribution de l'inflation au cours du temps, tels que les prix des fruits et légumes, de l'énergie ou de l'habillement par exemple. Nous proposons l'utilisation de nombreuses covariables dont les mesures sont obtenues de façon automatique : marché de l'hôtellerie, du transport aérien, du carburant, de l'électricité, du gaz naturel, de l'habillement et des légumes/fruits frais. Les observations de ces variables au cours du temps permettent de définir un modèle statistique dans le but d'obtenir une prédiction consistante de la distribution de l'inflation plus performante que les modèles utilisés à ce jour : marche aléatoire, modèles autorégressifs, modèles à volatilité stochastique, etc. Très récemment, de premiers résultats ont permis de justifier les approches basées sur des techniques d'apprentissage modernes, telles que les forêts aléatoires, afin d'expliquer les non linéarités présentes dans la dynamique de l'inflation au cours du temps. Dans ce projet, nous mettrons l'accent sur différents challenges théoriques et algorithmiques : assimilation de données hétérogènes, sélection de modèles pour choisir si une covariable doit être ajoutée dans le modèle ou non, quantification de l'incertitude de prédiction.

## Abstract

Forecasting inflation in real time based on a data-rich environment is a crucial challenge in economics. Developing an accurate indicator to explain derivatives prices is pivotal to predict economic growth, goods and services production and consumption and to design monetary and fiscal policies. In this project, we propose to develop a new statistical model to forecast inflation based on numerous covariates identified as key factors of the inflation dynamics: hotels and restaurants industry, air transport, fuel consumptions, apparel market, fresh fruits and vegetables sales and production, etc. The aim is then to design an algorithm to estimate the posterior distribution of the inflation given the observations of all these covariates and to detect regime shifts in its dynamics based on machine learning methods prone to be more efficient than state of the art methodology in macroeconomy such as autoregressive models or stochastic volatility models for instance. This research project sets the focus on several theoretical and statistical challenges: heterogeneous data assimilation, model selection to decide whether a covariate should be included in the model or not, model calibration, estimation of the prediction uncertainty.

**Contact:** randal.douc@telecom-sudparis.eu, or sylvain.lecorff@gmail.com