Note d’étape

# Présentation du projet

L’objectif à terme de ce projet de statistiques appliquées est de créer une prévision à court terme (un « nowcast ») de la consommation en France à l’aide autant de données officielles qu’alternatives (Sources officielles de vente de détail, sources officielles liées à l’économie, sources de données en temps réel des douanes, autres données alternatives (Google Trends, etc.)

Pour faire cela, il nous faut étudier la littérature sur les modèles de « nowcast » utilisés par les instituts économiques pour identifier les modèles les mieux adaptés à notre situation et les implémenter. Il nous faut aussi diviser la consommation des ménages en différents items et identifier les composantes qui expliquent le mieux les fluctuations macroéconomiques pour les intégrer à nos modèles. Puis il faut chercher à reproduire les évolutions des variables officielles et comparer les différentes approches pour enfin obtenir un indicateur répliquant en temps réel la dépense de consommation finale des ménages en France.

# Travail effectué

Pour commencer nous avons dû faire un tour d’horizon de la littérature sur le « nowcast » de la consommation afin de trouver un article de recherche intéressant sur la base duquel nous pourrions travailler. Nous avons donc consacré novembre à la recherche d’un article à répliquer.

Après avoir lu de nombreux articles nous avons décidés que chacun devrait présenter 1 article lors d’une réunion entre nous. A la suite de cette réunion nous avons décidé de présenter à QuantCube un seul article. Cet article de la banque d’Espagne de 2018 nous a semblé particulièrement pertinent puisqu’il cherchait à établir un « nowcast » de la consommation en Espagne et ce à l’aide autant de données officielles qu’alternatives.

Voici cet article : [nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data](https://www.semanticscholar.org/paper/Nowcasting-private-consumption%3A-traditional-credit-Gil-P%C3%A9rez/f1a508491cc16f9d3872d70931b7029bf89b86b8?fbclid=IwAR0PLEaJI1E0NIlM_CnmIsS5GeR0fmYOnHT-fKWzllt8ROynwLYHUiTkpJo).

Nous l’avons donc présenté à QuantCube. Ils l’ont trouvé fort intéressant et nous avons donc décidé de nous baser sur celui-ci pour la suite. Nous avons ensuite déterminé la marche à suivre pour la suite du projet. Il a été décidé que nous devrions faire l’étude de la variance de la dépense de consommation finale des ménages (étudier comment elle se décompose entre les différentes composantes de la consommation) et étudier le calcul des dépenses de consommation pour la France. Puis que nous devrions constituer un dataset avec les variables officielles et faire une étude statistique entre les variables officielles et la dépense de consommation finale des ménages. Ceci pour pouvoir ensuite faire un premier modèle de prédiction de la consommation à l’aide de la consommation passée. Ensuite il nous faudra faire un second modèle incorporant en plus les données officielles et enfin un troisième en ajoutant les données google trend. Finalement il nous faudra comparer les résultats des 3 modèles.

Nous nous sommes donc mis à la tâche. Nous avons tout d’abord cherché des données fiables. Nous avons commencé par trouver un ensemble de données contenant la dépense de consommation finale des ménages, 20 fonctions de consommation et 577 catégories de biens et services et ce depuis 1959 à un rythme annuel. Nous avons donc analysé cet ensemble de données en décomposant la variance et en faisant une analyse en composantes principales et ce à la fois sur les fonctions de consommations et les différents produits.

Puis nous avons cherché d’autres bases de données afin d’obtenir une meilleure granularité.

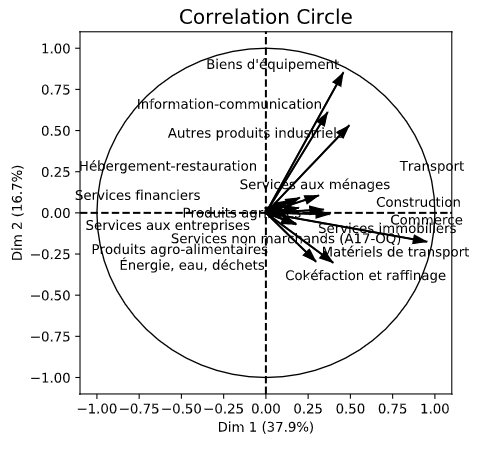
## Difficultés rencontrées

Nous avons été confrontés à des problèmes de compréhension lors de la recherche des articles car ils utilisaient des concepts ou des techniques que nous ne connaissions pas. Notamment pour bien les comprendre il eût fallu maîtriser les différentes techniques que nous voyons en cours de séries temporelles. De même nous avons eu quelques difficultés à comprendre exactement ce que recouvrait la consommation et l’évolution des différentes techniques utilisées pour la calculer, ainsi que les différentes manières de l’exprimer (en valeur ou en volume, en prix chaînés ou en prix courant). Néanmoins en approfondissant la question et demandant au besoin à QuantCube de nous éclairer nous avons été en mesure d’avancer.

Mais la plus grande difficulté à laquelle nous avons été confrontée a été de trouver des bases de données avec la bonne granularité. En effet pour faire un « nowcast » efficace il nous faut la granularité la plus fine possible or en fonction des sources et des variables la granularité varie fortement.

## Quelques résultats préliminaires

Nous avons réalisé une analyse en composante principales et une analyse de la variance à la fois sur les fonctions de consommation (voir les captures d’écran ci-dessous) et sur toutes les variables officielles à notre disposition.





Néanmoins nous pouvons observer de très fortes corrélations entre les variables ce qui traduit de forts effets externes qui influencent toutes les variables (comme par exemple un effet revenu).

# Prochaines étapes

Il nous reste à améliorer et affiner notre analyse en composantes principales et notre analyse de la variance de la dépense de consommation finale des ménages à l’aide des variables officielles. Et ce notamment en faisant varier les périodes considérées afin de s’assurer qu’il n’y a pas eu de changements à même de biaiser notre prédiction.

Il nous faudra aussi faire des statistiques descriptives pour étudier l’évolution des variables officielles et leur corrélation avec la dépense de consommation des ménages

## Difficultés anticipées

Nous risquons de nous retrouver face à un problème de granularité dans notre recherche des données sur les variables officielles. Il y aura aussi sûrement des problèmes de saisonnalité qu’il faudra corriger.

Mais un des points les plus compliqués à effectuer sera la sélection (en utilisant l’acp et l’analyse de la variance) des variables officielles à prendre en compte.

Enfin la sélection des modèles et leur application sera normalement la dernière étape difficile à effectuer.