



Université Paris 1  
Panthéon Sorbonne  
12, place du Panthéon  
75005 Paris

CES

Centre d'Économie de la Sorbonne  
UMR 8174



Année universitaire 2023-2024

M1 Économétrie-Statistiques

Mémoires du semestre 1

## Présentation générale

Ci-dessous une liste (non exhaustive) de mémoires. Je vous rappelle que le mémoire sera forcément écrit en TeX (Overleaf, SWP 6.0,...), en anglais, avec tous les calculs sous SAS. Le nombre maximal de pages avec bibliographie est d'environ 30 pages. La présentation est de 10 minutes et est en anglais.

Un mémoire doit contenir :

1. Un titre,
2. Les noms des auteurs,
3. Un résumé,
4. Une introduction qui présente votre sujet de façon générale, et qui introduit ensuite votre problématique (e.g. « The goal of this report is to investigate the informational content of SOC in order to forecast a financial crisis »). L'introduction contient obligatoirement un survey, i.e. une revue de la littérature, et l'annonce du plan. A la fin de l'introduction, on peut annoncer les résultats clefs,
5. Une partie présentant la théorie, les techniques économétriques employées ainsi que les données,
6. Une partie présentant les résultats,
7. Une conclusion.

La conclusion résume les résultats obtenus et ouvre sur de nouvelles perspectives.

8. Une bibliographie. Celle-ci doit être parfaitement formatée. Elle sera faite avec BibTeX en utilisant par exemple Jabref.

[1] Bill, J.E. & H.R. Rowling (2017): Harry Potter and Bounded Rationality, *Journal of Non-Sense Economics* 15, 110-113.

Toutes les références apparaissant dans le papier (Bill and Rowling, 2017) doivent être dans la bibliographie. La bibliographie est classée par ordre alphabétique, pas par ordre d'apparition dans le texte. Pas de citation du style : le cours de M.Xxxx ou l'interview de Mme Yyyy

Prenez un sujet qui vous plaise, pas forcément le plus technique ou le plus difficile. Si vous choisissez un sujet autre que ceux-ci-dessous, commencez impérativement par vérifier la disponibilité des données.



Université Paris 1  
Panthéon Sorbonne  
12, place du Panthéon  
75005 Paris

**CES**  
Centre d'Économie de la Sorbonne  
UMR 8174



## Sujets

### Sujet 1. Statistical learning vs machine learning

En économétrie / machine learning, se pose la question de la sélection automatique des variables, particulièrement lorsque la base de données contient beaucoup de prédicteurs potentiels. On peut (artificiellement) différencier les procédures avec inférence (statistical learning) des procédures sans inférence (machine learning). Par exemple, les procédures stepwise (forward ou backward) utilisent un critère statistique (T-stat, F-test, AIC,...), à l'inverse du LASSO qui utilise une procédure pour contracter les coefficients ayant peu d'influence économique. Après avoir présenté les différentes méthodes et les validations associées (tests statistiques, critères d'information, K-fold, leave-one-out...), ce mémoire se propose à partir de données simulées, d'estimer la performance de ces dernières, en considérant différents critères d'arrêt. En particulier, seront considérées les méthodes stepwise forward et backward, Forward Stagewise, LARS et LASSO. Pour cette dernière on s'intéressera typiquement au k-fold, leave-one-out (PRESS), CP, etc. On s'intéressera aussi à l'impact de la modalité du choix du paramètre  $\lambda$  qui diffère selon que nous considérons un critère explicatif ou prédictif. Les données générées le seront sous l'hypothèse nulle d'indépendance, et sous l'alternative de dépendance. Sous cette dernière, seront pris en compte des dépendances linéaires, des dépendances linéaires avec rupture(s) et de la multi-colinéarité. Des données avec valeurs extrêmes seront aussi considérées. Une application empirique conclura le mémoire.

Mots clefs : Forward / backward / Stepwise selection, Statistical learning, Machine learning, Forward Stagewise / LARS / LASSO, k-fold, cross-validation

SAS : proc glmselect, proc iml

### Sujet 2. Tests de ruptures structurelles : taille et puissance du test d'Andrews et Ploberger

Il est peu probable que la relation liant variable expliquée et variables explicatives dans une régression linéaire reste stable dans le temps. A cette fin, plusieurs tests ont été introduits, comme le test de Chow, le Cusum, Cusum carré, les résidus récurrents, et plus récemment le test d'Andrews et Ploberger (1993, 1994). Ce mémoire se propose d'estimer la taille et la puissance du test d'Andrews et Ploberger (1993, 1994) et au moins d'un autre test (Cusum,...) par simulations de Monte Carlo. Sous l'hypothèse nulle vous considérerez un modèle sans rupture et calculerez l'erreur de type 1 pour différentes tailles d'échantillons et de facteurs. Les valeurs critiques du test seront obtenues par bootstrap. Sous l'alternative vous calculerez l'erreur de type 2 pour un modèle avec une ou plusieurs cassures, de différentes tailles, en faisant varier la ou les dates de cassures. Vous considérerez aussi des modèles avec changement



Université Paris 1  
Panthéon Sorbonne  
12, place du Panthéon  
75005 Paris

CES

Centre d'Économie de la Sorbonne  
UMR 8174

continus, time-varying parameters, ou le stop-break model d'Engle et Smith (1999). Il sera aussi judicieux de jouer sur le paramètre Lambda (bord) et sur la taille de la cassure structurelle. Idéalement, le test d'Andrews et Ploberger sera aussi appliqué sur la variance des résidus, et comparé au test ICSS. On ne se focalisera ici que sur une rupture.

Mots clefs : Simulations de Monte Carlo, Bootstrap, changements structurels, Andrews et Ploberger.

SAS : proc IML, GRAPH.

### Sujet 3. Allocation de portefeuille et théorie des graphes



Ce mémoire se propose de construire un portefeuille d'actifs en liant résultats économétriques, théorie des graphes et allocation de portefeuille. En théorie, les actifs entrant dans un portefeuille doivent être non-corrélés. Cette vision exclue de facto les effets de feedback et de spillover et donc d'interactions dynamiques. En se focalisant sur un indice de votre choix (CAC40, DJIA, MSCI) vous construirez de façon dynamique, ie sur fenêtre roulante un réseau, les nœuds étant les actifs et les arcs la corrélation croisée à plus ou moins une période. Une fois la topologie retrouvée, vous construirez la matrice d'adjacence et trouverez une mesure de clustering pertinente. Les actifs les moins systémiques seront alors sélectionnés et alloués (equally weighted) sur une période temps donnée. En rebalançant votre portefeuille toutes les x semaines, vous comparerez vos résultats avec l'indice de référence, l'objectif étant de battre le marché pour un risque équivalent ou moindre

Techniques SAS : Proc IML, proc timeseries, théorie des graphes

### Sujet 4. Synchronisation des cycles

Beaucoup d'articles traitent de la relation entre le cycle d'activité, le cycle du crédit et le cycle monétaire, le but étant de savoir si les phases d'expansion et de contraction du crédit (et de la monnaie) sont pro-cycliques ou contra cycliques, c'est-à-dire favorisant l'activité économique ou non. Ce mémoire se propose d'étudier la synchronisation entre les trois différents cycles. La première étape consistera à extraire les cycles par des méthodes de lissage (filtre HP), de datation (Bry et Boschan, 1971) ou à rechercher les cycles sur données de l'OCDE, ou du NBER. Ensuite, plusieurs méthodes pourront être utilisées comme les corrélations croisées, ou des mesures de synchronisation (Harding et Pagan, 2002, 2006).

Techniques: proc IML, proc timeseries, proc expand, filtrage HP

### Sujet 5. Prévisions de crises financières (écono-physique)

En écologie, mais aussi plus généralement dans les systèmes complexes, se pose la question de la détection des points de ruptures, bifurcations ou transitions critiques qui correspondent au passage d'un état à un autre. Les régulateurs sont



Université Paris 1  
Panthéon Sorbonne  
12, place du Panthéon  
75005 Paris

CES

Centre d'Économie de la Sorbonne  
UMR 8174

donc intéressés par la mise en place d'indicateurs avancés « early warnings signals». Un ensemble d'indicateurs est basé sur le manque de résilience d'un système à l'approche d'une bifurcation et se rattachent à la théorie du Critical Slowing Down (CSD). Ce mémoire se propose de calculer ces indicateurs sur données financières (détrendées) et d'évaluer leur pertinence à l'approche d'une crise financière, comme celle de 2007. On s'intéressera aussi à un indice bitcoins, et à des données météo. On s'intéressera à la robustesse des résultats en fonction des techniques de detrending.

Techniques : proc IML.

### Sujet 6. Prévisions de crises financières (écono-physique)



Il a été noté que certains systèmes dits complexes, pouvaient avoir une tendance à s'auto-organiser, présenter un bris de symétrie, et ainsi une transition de phase. Typiquement, sur données financières, ce qui est observé est une soudaine augmentation des corrélations entre actifs financiers. L'augmentation des corrélations semblant précéder une crise financière. Ce mémoire se propose, sur fenêtres roulantes, de calculer des matrices de corrélations. Les résultats seront présentés de façon graphique sous forme de heatmaps, et on analysera l'évolution des heatmaps à l'approche de crash. On pourra aussi calculer la première valeur propre de la matrice de corrélations, ou se référer à la théorie de matrices aléatoires (RMT). On utilisera des données journalières sectorielles, typiquement DJIA, Euronext. Le mémoire sera étendu à des matrices de corrélations croisées, en prenant plus ou moins un retard.

Techniques : proc IML

### Sujet 7. Particle Swarm Algorithm (PSO) et optimisation

En économétrie et Machine Learning, l'optimisation d'une fonction objectif est requise. Cependant, la majorité des algorithmes renvoient des optima locaux, comme la descente de gradient, ou Newton-Raphson. Ont émergé depuis quelques années de nouvelles classes d'algorithmes dont l'algorithme Particle Swarm Algorithm (PSO). Ce mémoire se propose après avoir présenté de façon rigoureuse l'algorithme PSO et ses évolutions, de présenter un codage sous IML, puis d'appliquer l'algorithme à divers fonctions, linéaires et non hautement non-linéaires pour le valider. Une mise en parallèle avec une descente de gradient est un plus.

Techniques : SAS IML, optimisation

### Sujet 8. Sphère financière, économie réelle

De plus en plus de travaux s'intéressent à la relation entre économie réelle et sphère financière. En choisissant judicieusement un ensemble de variables le mémoire consiste à construire sur fenêtres roulantes un réseau temporel dynamique. La relation entre sphère économique s'analysera en utilisant des indicateurs de centralité dynamique (typiquement Katz) permettant de savoir qui

broadcast et qui reçoit l'information. Enfin, sur un réseau unique, on construira des mesures de centralité en tenant compte des communautés. Il n'y a à ma connaissance pas de travaux traitant le sujet de cette façon. L'analyse permettra de savoir quelles variables sont centrales dans le système complexe.

Technique : Proc IML, théorie des graphes

### Sujet 9. Prévion d'inflation

Dans un article de 1989, Hallman, Porter et Small ont présenté un nouvel indicateur, le  $p^*$  (p-star). Leur but était de réconcilier l'évolution de la masse monétaire avec l'évolution des prix. Ils proposent une extension de la théorie quantitative de la monnaie basée sur des valeurs de long terme servant à calculer un prix d'équilibre. La dynamique entre le prix d'équilibre et le prix observé est alors utilisé pour prévoir l'inflation. Ce mémoire se propose de vérifier empiriquement la validité du concept, sur plusieurs pays, en utilisant plusieurs définitions de la masse monétaire. Une approche graphique sera dans un premier temps utilisée, suivie d'une approche économétrique. L'ordre d'intégration sera testé, ainsi que la cointégration, et un modèle à correction d'erreurs estimé. Dans ce dernier, des simulations dynamiques seront effectuées.

Technique : Intégration / co-intégration, proc model.

\*

\* \*