

Axe 5: New Challenges for New Data

Anna SIMONI (CREST-ENSAE-Ecole
Polytechnique)

Reunion du Board of Trustees
-- 3 septembre 2019 --

PI et Co-PI

➤ Anna SIMONI (PI, ENSAE)



➤ Ludovic STOURM (Co-PI, HEC)



➤ Benoit SCHMUTZ (Co-PI, X)



Membres:

17 chercheurs/ chercheuses:

- HEC (5) : K. De Valck, P. Ebbes, J. Hombert, C. Yang, L. Stourm
- ENSAE (10) : P. Alquier, V-E. Brunel, C. Butucea, O. Catoni, N. Chopin, A. Dalalyan, G. Lecué, E. Matzan-Lober, A. Simoni, A. Tsybakov
- X (2) : C. Belzil, B. Schmutz

Thèmes de Recherche:

Interdisciplinarité: statistique, économétrie et marketing.

Nouvelles sources de données

- Réseaux sociaux (Facebook)
- Mesures de comportements de consommateurs (Internet companies: Google, Amazon; retailers, cartes de crédit,...)
- Données administratives (santé, emploi, impôts, education,...)

Nouvelles opportunités

Nouveaux défis

- Masse des données ('Big data')
- Parfois plus de variables que d'observations
- Données non-structurées

L'aide du Labex ECODEC nous a permis (1/4)

Fonds utilisés pour:

- Postes: doctorants, post-doctorants
- Invitations de professeurs étrangers (ou provinciaux)
- Conférences
- Séries de workshops spécialisés
- Financement de projets de recherche individuels de 5 k€ (assistants de recherche, achat de données, conduite d'expériences, mais aussi missions, ordinateurs)
- Co-financement de outils informatiques (serveurs) : un serveur GPU (10K€, Ecole Polytechnique), un serveur CPU (13K€, ENSAE), en 2019

L'aide du Labex ECODEC nous a permis (2/4)

Doctorants et Post-doctorants (sur les 3 institutions):

- 18 doctorants (qui ont commencé \geq 2016)
- 4 doctorants en fin de thèse (2018-2019)
- 1 Post-doc financé Labex ECODEC (\geq 2016)
- 1 visite à une université au Japon pour un doctorant financé par le Labex ECODEC (2019)

Conférences et Workshops (\geq 2017):

1. “2017 Labex workshop series on marketing analytics” (HEC, organisé par P. EBBES)
2. “2018 LABEX workshop series on marketing analytics” (HEC, organisé par P. EBBES)
3. “Workshop on Modeling Variance Risk” (2019) (ENSAE, organisé par A. VIOLANTE)
4. “2019 LABEX workshop series on marketing analytics” (HEC, organisé par P. EBBES)
5. “Meeting in Econometrics” (2017) (TSE, co-organisé par A. SIMONI)
6. “Journées non-stationnaires” (2018) (ENSAE, co-organisé par P. ALQUIER)

L'aide du Labex ECODEC nous a permis (3/4)

Conférences et Workshops (>=2017):

7. “Big Data Day @HEC” (2018) (HEC)
8. “Improving marketing performance” (2018), (HEC)
9. “Small Area Estimation Conference 2017”, (ENSAE)
10. “Bayesian Non Parametrics conference” (2017), (Univ. Dauphine)
11. “Structured Regularization for High-Dimensional Data Analysis” (2017)

Achat d'équipements :

- Achat d'un serveur de calcul pur CPU
- Achat d'un serveur GPU et d'ordinateurs
- Serveur pour le laboratoire d'expérimentation (participation)

L'aide du Labex ECODEC nous a permis (4/4)

Financement de projets de recherche:

- En 2018: 13 projets, dont 4 projets de chercheurs externes à l'Axe 5 mais sur les sujets de l'Axe 5.
- En 2019: 16 projets, dont 5 projets de chercheurs externes à l'Axe 5 mais sur les sujets de l'Axe 5.

Résultats (1/2)

Production scientifique:

- Publications: 44 entre 2015 et 2018 ! (dont 12 publications dans « top 3 journals » en statistique).
- Travaux en cours: plusieurs articles dans des séries de Working Paper (ArXiv, Crest, Banque de France, ...) et articles en révisions.

Collaborations:

- Rapprochements entre les institutions.
- Nouvelles collaborations avec chercheurs d'autres universités.

Formations:

- Forte implication des membres de l'Axe 5 dans la voie Data Science de l'**ENSAE**: Data Science, Statistique & Apprentissage (responsable: Arnak DALAYAN), Data Science & Sciences Sociales (co-responsable: Anna SIMONI).
- Graduate degrees et Bachelor degrees de l'**Ecole Polytechnique**.

Résultats (2/2)

Projets individuels de recherche réalisés en 2018:

- 1) Robust community detection via Median-of-means estimators, G. Lecué (ENSAE)
- 2) Optimal and adaptive variable selection in *high-dimensional* regression, A. Tsybakov (ENSAE)
- 3) Les sentiers de la gloire : Les déterminants du succès des films de cinéma, R. Nicolas (ENSAE)
- 4) *HD*-Bayes, N. Chopin (ENSAE)
- 5) Causal inference and nowcasting in presence of *large data sets*, A. Simoni (ENSAE)
- 6) How do organizations learn from *online customer* feedback? G. Di Stefano (HEC)
- 7) The Evolution of Educational Inequality and Selectivity, C. Belzil (École polytechnique)
- 8) Gone with the Wind: on the Evolution of Spatial Inequality in French Agglomerations, M. Eric (HEC)
- 9) Fostering Innovation through *Open Data*, S.C. Srivastava (HEC)
- 10) Identification of social interactions with *genetic data*: new insights into an old problem, Goffette (ENSAE)
- 11) Gremlins in the *data*, P. EBBES (HEC)
- 12) Blank project, L. STOURM (HEC)
- 13) Linear Functional Estimation under Column *Sparsity* and Robustness, A. Dalalyan (ENSAE)

Visibilité: prix et récompenses

- Cristina **BUTUCEA**: nommée *fellow de l'Institute of Mathematical Statistics* (IMS) – 2019
- Anna **SIMONI**: Médaille de bronze du CNRS 2019
- Alexandre **TSYBAKOV**: Invited Lecture, International Congress of Mathematicians 2014

Merci de votre attention!