Sébastien **Saurin**

Ph.D. in Economics | Advanced Credit Risk Analytics: Fairness, Interpretability, Homogeneity

Link +33 6 34 36 64 79 sebastien.saurin@univ-orleans.fr

Professional Experience

Ongoing — Postdoctoral Researcher, University of Orléans, France

2024 – 2021 — Ph.D. in Economics, University of Orléans, France
- Thesis title: "Advanced Credit Risk Analytics: Fairness, Interpretability, Homogeneity"
- PhD supervisors: Christophe Hurlin (University of Orléans) and Christophe Pérignon (HEC Paris)
- Jury: Bart Baesens (KU Leuven), Raffaella Calabrese (University of Edinburgh), Emmanuel Flachaire (Aix-Marseille School of Economics), Thierry Foucault (HEC Paris), Olivier Scaillet (University of Geneva)

August 2023 — Visiting PhD student, Macquarie University, Australia
- Supervisor: Shuping Shi

May 2023 – February 2023 — Visiting PhD student, HEC Paris, France
- Supervisor: Christophe Pérignon
- Courses attended:
  - Blockchains and Cryptocurrencies (HEC, 12h) by Bruno Biais
  - Fintech and Decentralized Finance (HEC, 12h) by Simon Mayer
  - Machine Learning for Econometrics (ENSAE, 24h): by Anthony Strittmatter and Bruno Crépon
  - Advanced Econometrics: Semi-parametric and Simulations (ENSAE, 24h) by Thierry Kamionka and Xavier D'Haultfoeuille

July 2022 – May 2022 — Visiting PhD student, HEC Paris, France
- Supervisor: Christophe Pérignon

2021 – 2020 — Research Assistant, HEC Paris, France
- Supervisor: Christophe Pérignon
- Research Reproductibilty and Fairness in Finance

July 2021 – April 2021 — Research Internship, University of Orléans, France
- Supervisor: Christophe Hurlin
- Machine learning interpretability: Shapley value adaptation to decompose model performance metrics, study of causality.

July 2020 – March 2020 — Research Internship, University of Orléans, France
- Supervisor: Christophe Hurlin
- Study of discrimination biases in Machine Learning
- Empirical analysis on python

Education

2019-2021 Master in Econometrics and Applied Statistics at the University of Orléans (highest honour)

2016-2019 Bachelor in Economics, specialized in Econometrics at the University of Orléans (highest honour)

Research Interests

Machine Learning : Supervised and unsupervised learning, Algorithmic fairness, eXplainable Artificial Intelligence (XAI), Model performance, Feature importance, Sustainable artificial intelligence, Conformal prediction, Model overfitting

Econometrics : Credit scoring, Logistic regression, Inference, Model heterogeneity, Homogeneity measure

Finance : Credit risk, Operational risk, Financial Regulation, Regulatory capital, Governance and regulation of artificial intelligence, Internal Ratings-Based approach

Publications

2024 **Hurlin C., Pérignon C. and Saurin S.** (2024), The Fairness of Credit Scoring Models. HEC Paris Research Paper No. FIN-2021-1411.

- **Forthcoming** in *Management Science*.
- Abstract: In credit markets, screening algorithms aim to discriminate between good-type and bad-type borrowers. However, when doing so, they can also discriminate between individuals sharing a protected attribute (e.g. gender, age, racial origin) and the rest of the population. This can be unintentional and originate from the training dataset or from the model itself. We show how to formally test the *algorithmic fairness* of scoring models and how to identify the variables responsible for any lack of fairness. We then use these variables to optimize the fairness-performance trade-off. Our framework provides guidance on how algorithmic fairness can be monitored by lenders, controlled by their regulators, improved for the benefit of protected groups, while still maintaining a high level of forecasting accuracy.

Working papers

2024 **Saurin S.** (2024), Homogeneity Test for Credit Scoring Models: A Conformal-Prediction Approach

- Abstract: Since the signature of the Basel II Accords in 2004, most international banks have been implementing the Internal Ratings-Based (IRB) approach to determine their capital requirements for credit risk. This approach involves estimating the default risk of each loan lying in the balance-sheet of the bank through a credit scoring model and then, allocate these loans into *homogenous* risk grades (or risk classes) gathering credits with similar default risk. Yet, effective solutions for testing this homogeneity remain lacking. In response, we introduce the Risk Homogeneity Coefficient (RHC), a novel measure that quantifies the degree of homogeneity within risk grades. RHC is derived from confidence intervals for the difference between each credit's estimated probability of default and the risk grade's default probability. The key insight is that a wider confidence interval suggests lower homogeneity within the risk grade. To build these confidence intervals, we adapt to our context the conformal prediction approach—a modern framework widely used in machine learning to construct confidence intervals for predictions without relying on distributional assumptions or asymptotic convergence results. Through numerical illustrations, we demonstrate that the RHC effectively measures homogeneity within risk grades. Applying our methodology to data simulated under the IRB framework, we observe significant variation in homogeneity across risk grades, with the overall level of homogeneity remaining moderate, even with a seemingly optimal credit segmentation. This finding raises important questions about the feasibility of achieving perfectly homogeneous risk grades in practice.

2023 **Hué S., Hurlin C., Pérignon C. and Saurin S.** (2023), Measuring the Driving Forces of Predictive Performance: Application to Credit Scoring. HEC Paris Research Paper No. FIN-2022-1463

- **Revised & Resubmit** at *Management Science*.
- Python package: github.com/hi-paris/XPER
- Blog article on Towards Data Science
- Abstract: In credit scoring, machine learning models are known to outperform standard parametric models. As they condition access to credit, banking supervisors and internal model validation teams need to monitor their predictive performance and to identify the features with the highest impact on performance. To facilitate this, we introduce the XPER methodology to decompose a performance metric (e.g., AUC, R^2) into specific contributions associated with the various features of a classification or regression model. XPER is theoretically grounded on Shapley values and is both model-agnostic and performance metric-agnostic. Furthermore, it can be implemented either at the model level or at the individual level. Using a novel dataset of car loans, we decompose the AUC of a machine-learning model trained to forecast the default probability of loan applicants. We show that a small number of features can explain a surprisingly large part of the model performance. Furthermore, we find that the features that contribute the most to the predictive performance of the model may not be the ones that contribute the most to individual forecasts (SHAP). We also show how XPER can be used to deal with heterogeneity issues and significantly boost out-of-sample performance.

2024 **De Stefano F., Hurlin C., Pérignon C. and Saurin S.** (2024), Artificial Intelligence in hiring and gender-based differences in job applicants' outcomes: Evidence from a large retail organization.

- Abstract: We test whether Artificial Intelligence (AI) can mitigate discrimination in the workplace. Using more than 40,000 individual job applications submitted to a multinational company, we show that, on average, women have a lower probability to obtain a job or to move to the next screening phase, while controlling for education, skills, and work experience. We then compare (1) job openings for which all applicants are assessed by an AI recruiting tool and (2) similar job openings for which applicants are not assessed by AI. While men and women end up with similar AI score distributions, we find that AI does not mitigate gender effect in hiring. Our conclusion is robust to methodological variation: treatment effects, fairness tests, and inclusion of applicants refusing to disclose their gender.

2024 **Hurlin C., Pérignon C. and Saurin S.** (2024), Fairness Equivalence in Credit Scoring: A Framework to Treat Algorithms like Prescription Drugs.

- Abstract: We introduce the concept of *fairness equivalence* by building on both the regulatory paradigm in drug development and on the fairness literature in machine learning. Equivalence tests allow banks and fintech to prove the fairness of the predictions of their credit scoring models, similarly as drug manufacturers proving the effectiveness of their drug products. In the equivalence approach, an algorithm is said to be fair if all individuals have an *equivalent* probability of positive outcome, regardless of their group membership. It means that the difference between group probabilities remains below a given *tolerance level*. The equivalence approach brings four main advantages. First, it allows to formally test any fairness definition by relying on an inference test while fully controlling for the risk of wrongly validating an unfair model (similar to the FDA tests on new drugs). Second, by introducing a tolerance level, the equivalence approach allows to control for any residual heterogeneity among individuals. Third, it can accommodate any high-stakes algorithms by adjusting the tolerance level according to the societal importance of the AI application. Fourth, it allows to identify the features at the origin of the fairness problem and to mitigate their effects.

2024 **Digalakis V., Sentenac F., Pérignon C., and Saurin S.** (2024), When Should the Challenger Model Replace the Status Quo?

Seminar and Conferences

2024 Conference on New Financial Actors, New Technologies and Risks, ACPR, Paris, December 4

2024 23th Workshop in Econometrics for Finance, Nanterre University, Paris, November 13

2024 European meeting of the Econometric Society, Rotterdam, Netherlands, August 26-30

2024 Economics of Financial Technology Conference, Edinburgh, Scotland, June 19-21

2024 Workshop ANR Machine Learning and Econometrics for Risk Measurement in Finance (MLEforRisk), Orléans, France, May 17

2023 Workshop in Econometrics for Finance, Nanterre University, Paris, November 15

2023 Econometric Society Australasian Meeting, Sydney, August 7-10

2023 Asian Meeting of the Econometric Society (AMES), Singapore, July 28-30

2023 Big Data and Econometrics seminar, Aix-Marseille School of Economics (AMSE), June 20

2023 Quantitative Finance and Financial Econometrics, Aix-Marseille School of Economics (AMSE), June 8-9

2023 Doctoral Days, University of Orléans, June 1-2

2023 16th Financial Risks International Forum, March 20-21

2023 ACSS-PSL Institute 3rd Research Seminar, Paris Dauphine University, February 17

2022 16th International Conference on Computational and Financial Econometrics (CFE), December 17-19

2022 22th Workshop in Econometrics for Finance, Nanterre University, Paris, November 16

2022 International Conference on Statistics and Econometrics (CISEM), University of Monastir, May 13-15

2022 Quantitative Finance and Financial Econometrics, Aix-Marseille School of Economics (AMSE), June 14-15

2022 Online PhD seminar, University of Orléans, January 20

2021 21th Workshop in Econometrics for Finance, Nanterre University, Paris, November 10

2021 Online Workshop in Applied Econometrics, University of Nantes, September 24

2021 Online Congress of the French Economic Association (AFSE), June 8-10

Teaching

2024 | **HEC Paris - Ecole Polytechnique | Master of Data Science & AI for Business, 2nd year**
Lecture: Interpretability and Algorithmic Fairness (English, 24h)
- Joint with Christophe Pérignon, Finance professor, HEC Paris
- Evaluation: 4.34/5

2021-2023 | **HEC Paris - Ecole Polytechnique | Master of Data Science for Business, 2nd year**
Lecture: Interpretability and Algorithmic Fairness (English, 24h)
- Joint with Christophe Pérignon, Finance professor, HEC Paris
- Evaluation: 4.33/5 (2021) and 4.62/5 (2022)

2023 | University PSL | PSL week: AI for Economics and Finance - Dauphine, ENS, Mines Paris, ENSAD, ESPCI
Lecture: Interpretable Machine Learning (English, 6h)

2022-2024 | University of Orléans | Master in Econometrics and Applied Statistics, 2nd year
Lecture: Interpretable Machine Learning (French, 12h)
- Evaluation: 8.5/10 (2022) and 8.8/10 (2023)

2022-2024 | University of Orléans | Master in Econometrics and Applied Statistics, 2nd year
Lecture: Support Vector Machine (SVM) (French, 12h)
- Evaluation: 8.5/10 (2022) and 8.9/10 (2023)

2021 | HEC Paris - Ecole Polytechnique | Master of Data Science for Business, 2nd year
Lecture: Algorithmic Fairness (English, 8h)
- Joint with Christophe Pérignon, Finance professor, HEC Paris
- Evaluation: 4.5/5

2021 | University of Orléans | Master in Econometrics and Applied Statistics, 1st year
Tutorial: Time Series (French, 2x15h)

2021-2023 | University of Orléans | Master in Econometrics and Applied Statistics, 1st year
Tutorial: Mathematical statistics (French, 2x15h)

2021 | University of Orléans | Bachelor in Economics, 2nd year
Tutorial: Personal and Professional Project (French, 4h)

Competitions

2021-2022-2023 Econometric Games

- Econometric Modeling competition with 30 teams from the best international universities (Harvard, Oxford, Cambridge, etc.) (website)
- Captain of the University of Orléans's team (2022-2023)
- Subjects: Preventing Malnutrition (2023), Inflation (2022), Airbnb effect on house prices in Amsterdam (2021)
- Results: Top 10 (2023)

Languages and Computer Skills

- **Languages:** French (native), English (C1, TOEIC 970/990)
- **Computer:** Python, SAS, R, Stata, LaTeX

Responsabilities

2024 Referee: Economic Modelling

2023-2024 PhD seminar organizer

2022-2024 PhD students' representative

References

Christophe Hurlin
Professor
University of Orléans
christophe.hurlin@univ-orleans.fr

Christophe Pérignon
Professor
HEC Paris
perignon@hec.fr

Sullivan Hué
Assistant Professor
Aix-Marseille School of Economics
sullivan.hue@univ-amu.fr