

Lucas Girard

Post-doctoral researcher – CREST, ENSAE, IP Paris

Job Market Candidate 2025-26

+33 (0)6 86 93 75 32

lucas.girard@ensae.fr

<https://faculty.crest.fr/lgirard>

5 avenue Henry Le Chatelier, 91120 Palaiseau,
France. French citizen, born in 1993

Fields and research interests

Primary: econometric theory (non-asymptotic inference, causal inference, nonparametric methods)

Secondary: applied microeconometrics (urban economics, segregation measurements)

References

- Xavier D'Haultfœuille – CREST, ENSAE, IP Paris – xavier.dhaultfoeuille@ensae.fr, [website](#)
- Guillaume Lécué – ESSEC Business School – lecue@essec.edu, [website](#)
- Frédéric Loss – Cergy Paris University, THEMA – frederic.loss@cyu.fr, [website](#)
- Elia Lapenta – CREST, ENSAE, IP Paris – elia.lapenta@ensae.fr, [website](#)

Experience

2024-26	Post-doctoral researcher – CREST, ENSAE, IP Paris, Economics department
2022-24	Teaching coordinator in mathematics – ENSAE Paris, Head of Studies division
2020-22	Teaching fellow (192 hours of teaching/year) – ENSAE Paris
2017-21	Ph.D. candidate in economics – CREST, ENSAE, IP Paris Supervisor: Xavier D'Haultfœuille Committee: Bruno Crépon and Alessandro Riboni Jury: Russell Davidson (McGill University), Maximilian Kasy (University of Oxford), Roland Rathelot (CREST, ENSAE, IP Paris), Daniel Wilhelm (Ludwig Maximilian University of Munich)

Education

2016-17	M.Phil. in Quantitative Economics (with specialization in Econometrics) – IP Paris
2015-17	M.Sc. in Engineering (with specialization in Data Science) – ENSAE Paris
2013-15	M.Sc. in Management – HEC Paris
2013-14	B.Sc. in Applied Mathematics – Paris-Saclay University
2011-13	Preparatory class B/L (humanities and social sciences track) – Lycée Henri IV, Paris
2008-11	High school diploma in Sciences – Lycée Victor Duruy, Paris

Research

Job Market Paper

Can we have it all? Non-asymptotically valid and asymptotically exact confidence intervals for expectations and linear regressions

joint with Alexis Derumigny and Yannick Guyonvarch

[Link to latest version](#)

Takeaway: We construct confidence intervals for expectations and linear regressions' coefficients with non-asymptotic validity guarantees while maintaining the precision of standard intervals as sample size grows to infinity.

Abstract: We contribute to bridging the gap between large- and finite-sample inference by studying confidence sets (CSs) that are both non-asymptotically valid and asymptotically exact uniformly (NAVAE) over semi-parametric statistical models. NAVAE CSs are not easily obtained; for instance, we show they do not exist over the set of Bernoulli distributions.

We first derive a generic sufficient condition: NAVAE CSs are available as soon as uniform asymptotically exact CSs are. Second, building on that connection, we construct closed-form NAVAE confidence intervals (CIs) in two standard settings – scalar expectations and linear combinations of OLS coefficients – under moment conditions only. For expectations, our sole requirement is a bounded kurtosis. In the OLS case, our moment constraints accommodate heteroskedasticity and weak exogeneity of the regressors. Under those conditions, we enlarge the Central Limit Theorem-based CIs, which are asymptotically exact, to ensure non-asymptotic guarantees. Those modifications vanish asymptotically so that our CIs coincide with the classical ones in the limit. We illustrate the potential and limitations of our approach through a simulation study.

Publications

Explicit non-asymptotic bounds for the distance to the first-order Edgeworth expansion
joint with Alexis Derumigny and Yannick Guyonvarch – *Sankhya A*, 86(1), pp. 261-336, 2024

[Link to journal article](#)

Takeaway: We derive explicit, improved finite-sample bounds for approximating the distribution of standardized sums of independent random variables, providing tools for more reliable statistical inference with small samples.

Abstract: We obtain explicit bounds on the uniform distance between the cumulative distribution function of a standardized sum S_n of n independent centered random variables with moments of order four and its first-order Edgeworth expansion. Those bounds are valid for any sample size with $n^{-1/2}$ rate under moment conditions only and n^{-1} rate under additional regularity constraints on the tail behavior of the characteristic function of S_n . In both cases, the bounds are further sharpened if the variables involved in S_n are unskewed. We also derive new Berry-Esseen-type bounds from our results and discuss their links with existing ones. Following these theoretical results, we discuss the practical use of our bounds, which depend on possibly unknown moments of the distribution of S_n . Finally, we apply our bounds to investigate several aspects of the non-asymptotic behavior of one-sided tests: informativeness, sufficient sample size in experimental design, and distortions in terms of levels and p-values.

Bridging methodologies: Angrist and Imbens' contributions to causal identification
joint with Yannick Guyonvarch – *Revue d'économie politique*, 133(6), pp. 845-905, 2023

[Link to journal article](#)

Takeaway: We provide a comprehensive review of Angrist and Imbens' Nobel Prize-winning contributions to causal inference through the notion of Local Average Treatment Effect and of their subsequent research.

Abstract: In the 1990s, Joshua Angrist and Guido Imbens studied the causal interpretation of Instrumental Variable estimates (a widespread methodology in economics) through the lens of potential outcomes (a classical framework to formalize causality in statistics). Bridging a gap between those two strands of literature, they stress the importance of treatment effect heterogeneity and show that, under defensible assumptions in various applications, this method recovers an average causal effect for a specific subpopulation of individuals whose treatment is affected by the instrument. They were awarded the Nobel Prize primarily for this notion of Local Average Treatment Effect (LATE). The first part of this article presents this methodological contribution in depth: its origins in earlier applied articles, the different identification results and extensions, and related debates on the relevance of LATEs for public policy decisions. The second part reviews the authors' main contributions beyond LATE. Joshua Angrist has pursued the search for informative and varied empirical research designs in several fields, particularly in education. Guido Imbens has complemented the toolbox for treatment effect estimation in many ways, notably through propensity score reweighting, matching, and, more recently, adapting machine learning procedures.

segresmall: A command to estimate segregation in the presence of small units
joint with Xavier D'Haultfoeuille and Roland Rathelot – *Stata Journal*, 21(1), pp. 152-179, 2021

[Link to journal article](#)

Takeaway: We provide a Stata command to apply methods that correct the upward bias in segregation indices when units (neighborhoods, classrooms, etc.) are small, thus enabling accurate measurement and comparisons across contexts.

Abstract: Consider a population of interest, composed of a minority and a majority group, that is allocated into units, which can be neighborhoods, workplaces, classrooms, etc. Qualitatively, there is some segregation whenever the allocation results in a concentration of minority individuals in some units more than in others. Quantitative measures of segregation have struggled with the small-unit bias. When units contain few individuals, indices based on minority shares are upward-biased. For instance, they would point to a positive amount of segregation even when allocation is strictly

random. The Stata command “segresmall” implements three recent methods correcting for such bias: the nonparametric, partial identification approach of D’Haultfœuille and Rathelot (2017, *Quantitative Economics*), the parametric model of Rathelot (2012, *Journal of Business and Economic Statistics*), and the linear correction of Carrington and Troske (1997, *Journal of Business and Economic Statistics*). The package also allows for conditional analyses, namely, measures of segregation that account for characteristics of individuals or units.

Work in progress

Average marginal effects in nonparametric instrumental regressions based on RKHS

joint with Elia Lapenta

Status: Preparing for submission

Takeaway: We develop a computationally efficient single-tuning-parameter method for estimating average treatment effects in nonparametric instrumental variable models, with theoretical guarantees and good small-sample performance.

Abstract: We propose an estimator for average marginal effects in nonparametric instrumental regressions based on Reproducing Kernel Hilbert Spaces that requires a single regularization parameter. We obtain the consistency and asymptotic normality of our estimator. Since the asymptotic distribution depends on an intricate asymptotic covariance, we propose a weighted Bayesian bootstrap method to obtain the critical values. In simulations, our test exhibits good small-sample performances. Two empirical applications show its implementation in real-data contexts.

Measures of several dimensions of residential segregation in France between 1968 and 2019 based on the Labor Force Survey clusters

Status: Working paper ([link](#))

Takeaway: We document that residential segregation has remained stable over five decades in France but varies substantially across dimensions, being strongest for nationality and ethnicity and weakest for labor market status.

Abstract: This article takes advantage of the sampling scheme of the French Labor Force Survey, which draws clusters of around thirty adjacent housing units, to study several dimensions of residential segregation in France over 1968-2019. Such clusters form relevant neighborhoods to study residential segregation, provided the indices account for the small-unit bias so that they can be compared over time or across different dimensions of segregation (French versus non-French people, jobseekers versus employed, college graduates versus non-graduates, white-collar versus blue-collar workers, etc.). Applying the methodology developed in D’Haultfœuille and Rathelot (*Quantitative Economics*, 2017) and Rathelot (*Journal of Business and Economic Statistics*, 2012), we estimate annual segregation indices for different specifications of the “minority” and “majority” groups, aiming to quantify several dimensions of residential segregation and compare them. The results suggest two main conclusions. First, whatever the dimension under study (ethnicity, immigrant status, nationality, occupational category, labor market status, education), the estimated indices do not reveal significant evolution over time: within each dimension, the magnitude of residential segregation has remained globally constant over the past decades. Second, they reveal that the magnitude of segregation differs across the different dimensions according to the following decreasing ranking: nationality, ethnicity (using as a proxy parents’ country of birth), being an immigrant, social status (occupational category or college education), and labor market status (unemployed or employed). A conditional analysis, separating neighborhoods that belong to urban areas of 200,000 inhabitants or more from neighborhoods belonging to smaller urban areas, complements the unconditional analysis.

Identification and estimation of a polarization index in large choice sets, with an application to U.S. Congress speech (1873-2016)

joint with Xavier D’Haultfœuille and Roland Rathelot

Status: Working paper available upon request

Takeaway: We develop a method to measure speech polarization in congressional discourse, showing that partisan language differences reached historic highs in 2008-2016 after rising since the 1980s-1990s.

Abstract: We quantify the extent to which Democrats and Republicans use different words during debates in the U.S. Congress. Such reliable measures are a prerequisite to study the causes or consequences of the polarization of political discourse. To do so, we propose a partially testable statistical framework, formal identification results for a speech polarization index, and easily computable estimators and confidence intervals for this index. The results suggest an increase in speech polarization starting in the 1980s-1990s, reaching its highest historical level between 2008 and 2016. However, this phenomenon is not entirely new as the analysis also reveals a period of relatively high polarization during the first decade of the XXth century and, to a lesser extent, around the 1930s.

Talks

Conferences and workshops

2026 ADRES Conference, Centre d'économie de la Sorbonne (planned)
 2025 Paris Econometrics Workshop
 2022 CIREQ Montreal Econometrics Conference
 2019 Gender, digital platform and NLP method workshop, Paris-Saclay University
 2018 10th French Econometrics Conference (poster session), Paris School of Economics

Seminars

2025 Center for Economics at Paris-Saclay; Paris-Saclay University
 2022 University Paris 1 – Paris School of Economics
 2021 CREST Microeconometrics seminar
 2020 Paris-Saclay University
 2019 CREST Microeconometrics seminar
 2018 CREST Ph.D. students seminar
 2017 CREST Ph.D. students seminar

Professional Service

Referee: *Annals of Economics and Statistics, Eastern European Economics, Statistica.*

Organizer: CREST Microeconometrics seminar (2018-19)

Other: Ph.D. candidates representative, GENES Board of Directors (2018-20)

Teaching

Teaching philosophy

I am passionate about teaching and believe in active learning. In lectures, I prioritize writing on the board over slides to maintain student focus, encourage questions, and emphasize key concepts and intuitions. I complement this approach with comprehensive lecture notes and practice quizzes to support student learning. Examples of my teaching materials can be found [here](#).

Main lectures

École Polytechnique “Maths in practice: Mathematical Tools for Economics” (introduction to statistics and econometrics) in the Bachelor of Sciences (undergraduate) in 2024-25 (12h), 2025-26 (12h)

ENSAE Paris “Econometrics 2” (causal inference and non-linear models) in the Engineering Master Program (graduate) in 2025-26 (18h)

ENS Paris-Saclay “Intermediate Econometrics” (linear models and causal inference) in the Master of Economics (graduate) in 2024-25 (34h)

Tutorial sessions

As a teaching assistant at ENSAE Paris for

Graduate courses in econometrics, mathematics, and statistics

- “Econometrics 1” by Xavier D’Haultfœuille in 2019-20 (18h), 2020-21 (18h), 2021-22 (18h)
- “Econometrics 2” by Xavier D’Haultfœuille in 2017-18 (18h), 2018-19 (18h), by Michael Visser in 2020-21 (18h), 2021-22 (18h)
- “Econometrics crash course” by Bertrand Garbinti in 2020-21 (12h)
- “Linear Time Series” by Christian Francq in 2021-22 (18h)
- “Introduction to Stochastic Processes” by Nicolas Chopin in 2023-24 (15h)

- “Statistics 1” by Nicolas Chopin in 2017-18 (18h), 2018-19 (18h), by Arnak Dalalyan in 2019-20 (18h), 2020-21 (18h), 2021-22 (18h)
- “Mathematical Statistics” by Guillaume Lecué in 2018-19 (12h), 2019-20 (12h), 2020-21 (12h), 2021-22 (12h)

Undergraduate courses in mathematics, microeconomics, and statistics

- “Analysis and Topology” by Laurent Decreusefond in 2020-21 (24h), by Marie Touahir in 2021-22 (24h), by Rim Essifi in 2022-23 (27h), by Etienne Donier-Meroz in 2023-24, 2024-25 (27h)
- “Measure Theory and Mathematical Foundations of Probability” by Anna Korba in 2023-24 (12h), by Cristina Butucea in 2025-26 (18h)
- “Introduction to Microeconomics” by Thibaud Vergé in 2020-21 (24h), by Laurent Linnemer and Thibaud Vergé in 2021-22 (24h)
- “Introduction to Statistics” by Marco Cuturi in 2017-18 (24h), by Arnak Dalalyan in 2018-19 (24h)

Miscellaneous

Languages

- French (native language)
- English (full professional proficiency)
- Spanish (DELE level C1 obtained in 2015)

Programming and Software

- Expert: R, Stata, LaTeX
- Proficient: Python, Matlab
- Packages (with co-authors): BoundEdgeworth (R), NAVAECI (R), segregsmall (Stata)

Distinctions

- Full doctoral scholarship – CREST (2017-20)
- Dean’s Honor List – HEC Paris (2013-17)

Other activities

- Sports: badminton (ranked at the regional level), tennis, running, cycling, triathlon
- Reading