

**THE BUCHAREST UNIVERSITY OF ECONOMIC
STUDIES**



**DOMAIN: CYBERNETICS AND STATISTICS
DOCTORAL SCHOOL: ECONOMIC CYBERNETICS AND
STATISTICS**

**ABSTRACT OF THE
HABILITATION THESIS**

**The power of data: advanced statistical and machine
learning methods for economic analysis**

Candidate: Prof. univ. dr. Bogdan OANCEA

Abstract

This habilitation thesis synthesizes a sustained academic and research trajectory at the intersection of statistical science, econometrics, machine learning, and computational methods, with the objective of advancing quantitative methodologies for data-intensive societal applications. The central perspective is process-aware quantitative inference, in which data are analyzed together with the mechanisms of measurement, aggregation, and uncertainty propagation that underpin reliable evidence for policy analysis.

Scientifically, the thesis is organized around three interconnected pillars. The first pillar develops machine learning and advanced modelling for economic and social measurement problems, including automated product classification for price statistics, forecasting of core macroeconomic variables (e.g., inflation and GDP), and predictive modelling in applied domains. Across these studies, the contributions are framed in a comparative and “fit-for-purpose” logic: classical econometric structures are systematically confronted with modern learning approaches, with explicit attention to interpretability, robustness under structural change, and computational feasibility through efficient implementations and high-performance computing practices.

The second pillar addresses income distribution and inequality using administrative and fiscal data. The thesis contributes models and empirical analyses that clarify how income components—especially capital income—shape overall inequality and how distributional regularities can be summarized through parametric and entropy-based measures and decompositions. The results are positioned to support evidence-based evaluation, while maintaining the methodological discipline required for consistent production of distributional indicators over time.

The third pillar advances a modular framework for official-statistics production from complex and partially observed data-generating systems. This perspective is exemplified by digit-based statistical signatures, structural metrics for trade, and a fully probabilistic end-to-end pipeline for deriving population statistics from mobile network data. The latter integrates hidden Markov models, Bayesian de-duplication of devices, distributional aggregation, and hierarchical inference into a coherent probabilistic chain designed for auditability, uncertainty propagation, and technological evolvability (e.g., across telecom generations).

These contributions emphasize not only the use of advanced quantitative techniques, but also their operational integration into reproducible analytical systems. A recurring theme throughout the thesis is the need to move beyond isolated model performance and toward complete methodological pipelines in which data generation, preprocessing, model estimation, validation, computational scalability, and uncertainty assessment are treated as inseparable components of scientific inference. This orientation is particularly relevant in contemporary economic and statistical analysis, where administrative records, digital traces, scanner data, web-scraped information, and mobile network data increasingly complement traditional survey-based sources. By combining statistical rigor with computational implementation, the thesis shows how machine learning, econometric modelling, Bayesian inference, entropy-based measurement, and high-performance computing can jointly support robust evidence production in contexts characterized by large data volumes, structural heterogeneity, and

evolving technological environments. In this sense, the work contributes both to methodological development and to the practical modernization of quantitative evidence systems for economics, statistics, and policy-relevant empirical research.

Importantly, the thesis concludes with a dedicated chapter on academic and research career development, documenting the progression from methodological contributions to sustained scientific leadership. This chapter synthesizes the author's long-term teaching activity and curriculum development, the supervision and mentoring of students and junior researchers, and sustained scholarly dissemination through peer-reviewed publications and conference participation. It also highlights institutional and professional service (including participation in scientific and statistical communities), leadership roles in interdisciplinary collaborations, and the development of reproducible software artefacts that operationalize the thesis' methods. The scientific results and the career-development synthesis position the thesis as a mature contribution aligned with European expectations for senior academic qualification: methodological originality, demonstrable impact, reproducibility, and the capacity to lead research, training, and knowledge transfer in data-driven domains.