



Statistical Learning for Unit Nonresponse Treatment in Surveys Using Riesz Representers

Supervisor

David Haziza
University of Ottawa
dhaziza@uottawa.ca

Co-Supervisor

Mehdi Dagdoug
McGill University
mehdi.dagdoug@mcgill.ca

Abstract

National statistical offices increasingly use machine learning to address unit nonresponse with inverse probability weighting. Machine learning may be used to model complex, high-dimensional relationships but introduces challenges such as variance estimation and model selection. We plan to investigate the Riesz Representer approach, combined with cross-fitting, to ensure consistent variance estimation and obtain square-root consistent estimators. This proposal will also explore hyperparameter selection and adaptive weight trimming to improve stability. This research aims to enhance the reliability of statistical inference in surveys with nonresponse.

Interdisciplinary/Applied experience

In the past decade, machine learning has become integral to survey estimation strategies, including to address nonresponse. However, the properties and theoretical foundations of these methods remain underexplored. In particular, variance estimation poses significant challenges when inverse probability weighting is combined with machine learning algorithms. Accurate uncertainty quantification is essential for constructing valid confidence intervals and ensuring robust statistical inference. Addressing these questions can lead to more efficient survey estimation methods, enabling researchers to produce more reliable and representative statistics. Such improvements are crucial for understanding social, economic, and health trends, identifying disparities, and informing inclusive policies.

We plan for the postdoctoral fellow to spend part of the second year at Statistics Canada. This experience will allow for a real-world assessment and validation of the proposed methodology using real data from Statistics Canada.



David Haziza's expertise in survey sampling and Mehdi Dagdoug's knowledge of statistical learning provide a strong combination to achieve the project's goals. This project will leverage techniques from both fields to develop robust, practical solutions for survey estimation.

Teaching/Training/Education

Teaching: The postdoc will teach one three-credit course at McGill University in Year 1 and one at the University of Ottawa in Year 2. In Year 1, the PDF will be supported by the Faculty and Teaching Development team from the Teaching and Learning Services at McGill University, who will provide the postdoc with professional development training, guidance, and support with course design, including effective teaching and assessment strategies. In Year 2, the PDF will be supported by the Teaching and Learning Support Services at the University of Ottawa.

Support: The fellow will be encouraged to participate in a CANSSI-sponsored summer training program. In Year 1, the PDF will be encouraged to participate in workshops offered by the McGill Career Planning Service, which assists the PDF in career development, project management, networking, and job search. In Year 2, the fellow will be encouraged to take workshops related to career development provided by the Career Development Center at the University of Ottawa.

Mentoring

There will be bi-weekly Zoom meetings with both supervisors and the PDF to present research updates. Given the short distance between Montreal and Ottawa, we will meet jointly in person at least twice a year.

The PDF will be involved in the (informal) co-supervision of a graduate student and a summer undergraduate student (NSERC-USRA).

In Year 2, we expect the PDF to spend some time at Statistics Canada. This collaboration will introduce the PDF to practical challenges, enhance their communication skills, and allow them to share their expertise with non-specialists (data users at Statistics Canada). The PDF fellows will showcase their research at both national (e.g., the annual conference of the Statistical Society of Canada) and international conferences (e.g., the joint statistical meetings), which will also serve as platforms for networking.

Supervisor: David Haziza (Full Professor, University of Ottawa) will provide guidance on survey sampling and variance estimation.



Co-supervisor: Mehdi Dagdoug (Assistant Professor, McGill University) will provide guidance on Riesz representers, crossfitting, and their use in the statistical literature.

There are weekly statistics seminars at McGill University and the University of Ottawa. Further, monthly colloquiums are held jointly between the four Montreal universities and the other Quebec universities. The PDF will also be able to participate in the ongoing reading groups within each department. M. Dagdoug is closely involved in organizing these events.

Schedule

Year 1: The PDF will spend the first year of the fellowship at McGill University in Montreal.

Research: The PDF will leverage the Riesz representation theorem and crossfitting to develop consistent variance estimators for inverse probability-weighted survey estimators. During the first year, the postdoc will be expected to prepare at least one peer-reviewed research article focusing on the methodology of the proposed technique. The PDF will also attend conferences and workshops.

Teaching: Teach a statistics course in the Department of Mathematics and Statistics at McGill University.

Supervision: The PDF will help co-supervise an undergraduate summer student at McGill University.

Year 2: The PDF will spend the second year of the fellowship at the University of Ottawa in Ottawa, ON.

Research: The PDF will focus on model selection and hyperparameter tuning. The PDF will also apply the proposed methods using data from Statistics Canada during a 2–3-month internship. The research outcomes are expected to result in 1–2 additional peer-reviewed research articles focusing on theoretical, computational, and applied issues. The PDF will present their work at conferences.

Teaching: Teach a statistics course in the Department of Mathematics and Statistics at the University of Ottawa.

Supervision: The PDF will help co-supervise an MSc student at the University of Ottawa.

List of qualifications

PhD in Statistics or a closely related field.

Knowledge and experience in one (or more) of the following areas: statistical learning, causal inference, and survey sampling.

Proficiency with programming languages, including R or Python.

Ability to teach courses in statistics at the undergraduate level in English (or French, as the University of Ottawa is a bilingual university).

Strong communication skills and willingness to learn.