

# **Robust estimation for two-phase sampling designs in the presence of influential units**

David Haziza\*

Université de Montréal, Montreal, Canada david.haziza@umontreal.ca

Jean-François Beaumont

Statistics Canada, Ottawa, Canada jean-francois.beaumont@statcan.gc.ca

Cyril Favre-Martinoz

École nationale de statistique et de l'analyse de l'information, Bruz, France

Cyril.favremartinoz@ensai.fr

Influential units, also called outliers, often occur in practice, especially in business surveys due to the skewness of the distribution of economic variables collected by this type of surveys. A unit is said to be influential when its inclusion or exclusion from the sample has an important impact on the magnitude of survey statistics (e.g., estimated totals). In this presentation, we consider the problem of robust estimation in the context of two-phase sampling designs. In this context, the total error of an estimator can be expressed as the sum of the error due to the first phase and that due to the second phase. An influential unit may potentially have an impact on both errors. We define the concept of conditional bias attached to a unit with respect to both phases and show it can be viewed as a measure of its contribution to the total error. Following Beaumont, Haziza and Ruiz-Gazen (2012), we propose a class of robust estimators for two-phase sampling designs. In the presence of unit nonresponse, the set of respondents is often viewed as a second phase of selection. The proposed method can thus be naturally extended to that case. Results of a limited simulation study will be shown.

**Key Words:** Conditional bias, influential unit, nonresponse, robust estimation, two-phase sampling designs.