Alternative Ways of Assessing Confidentiality

Currently, Statistics Canada, as many other statistical organizations in the world, uses a set of rules (statistical algorithms) to determine whether data cells are deemed confidential and need to be supressed. They determine whether a sufficient number of businesses are represented by a data cell and whether any of the businesses in the data cell has a dominant effect. These rules are premised on the Statistics Act which prohibits the disclosure of any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual business. The suppression rules were developed at a time when most economic statistics were directly estimated from survey data.

Data users of economic statistics are increasingly demanding access to very detailed aggregate economic statistics, whether it be aggregate statistics by industry or geography or activity. Given this, as well as the fact that economic statistics are now compiled through a fairly sophisticated set of estimation techniques using a combination of survey and administrative information it is worth investigating whether the simple application of the disclosure rules should still be the only way of assessing confidentiality. This paper will look at a number of case studies to examine whether current statistical procedures (indexing, modelling, editing, imputation, allocation, estimation, calibration, integration of administrative and survey information) alter the information collected from businesses to such an extent that a third party (a party other than Statistics Canada or the responding business) cannot with certainty relate the data cell proposed to be released to the particulars obtained by Statistics Canada from any individual return. It will argue that in some of these cases data cell disclosure control based on simply assessing minimum number of businesses and dominance is overly cautious and leads to over-suppression.