On the way to remote access to German official microdata: a glimpse of work in progress

Rainer Lenz

Saarbruecken University of Applied Sciences
Faculty of engineering
Goebenstrasse 40
66117 Saarbruecken
Germany
Rainer.Lenz@htw-saarland.de
http://www.htw-saarland.de

Abstract. A number of methodological and technical preconditions have to be met in order to provide (automated) remote data access. This paper outlines the German approach to achieving this goal with a strong focus on the scientific and methodological challenges, in particular, regarding the access to cross-sectional and longitudinal economic statistics. The paper concentrates mainly on the generation of anonymised data with a structure similar to that of the real data, which are made available to data users. The data are provided in the form of what are called data structure files, which can be produced by using specific variants of microaggregation, multiplicative stochastic noise and multiple imputation. These data structure files are sent to the researcher after he has submitted a request for remote access. They allow checking a program code for syntactic and semantic errors before it can be finally applied to the original data by remote execution.

1 Introduction

The producers of data relating to surveys of economic statistics in Germany have observed a fundamental change in the demand for their products. In early 2000, providing the scientific community with so-called scientific use files (SUFs) – which researchers can use at their own workplaces outside the statistical offices - was considered a way, if not the "royal road", towards giving empirical social and economic research adequate access to official microdata in Germany. Such SUFs have been available for what is called the off-site use of selected and strongly demanded statistics. As regards SUFs of economic statistics, however, these data stocks have not been very well received. Reasons to be mentioned in this respect are the new data perturbating anonymisation methods which are not yet familiar to researchers, too long waiting periods and the excessive effort required to compile the SUFs. Due to the necessary and partly drastic interference in the information structure of the data in