

Combining the German Ageing Survey with the Sample of the Insured Population Pension Records via statistical matching as a source for the analysis of life courses and old age incomes

Julia Simonson, Laura Romeu Gordo, Nadiya Titova

Abstract

In social and economic research statistical matching is an increasingly applied technique for combining information from different data sources when no one-to-one correspondence by an identifier is possible. The basic idea of such a procedure is to find similar cases in relevant characteristics in different data bases and to link the information referring to these cases. Statistical matching avoids the attrition rate linked to the informed consent requirements for direct matching based on a unique identifier. Thereby it provides substantial opportunities for research and allows for analyses that would be impossible from one input data source alone.

This paper examines the combination of the German Ageing Survey (DEAS) with administrative pension insurance data, the so-called Sample of the Insured Population Pension Records (VSKT), via statistical matching. The framework of this matching is the project 'Life Course, Ageing and Well-Being', which is funded by the Volkswagen Foundation. The aims of the project are threefold: first, we want to analyse to what extent the life courses of those birth cohorts currently in the middle of their lives show increased pluralism and inhomogeneity, compared to retirees today; and second we want to investigate how these changes will affect the lives of the elderly in the future - especially the protection by public and occupational pensions as well as private forms of old age security. Furthermore, it is inquired how the old age security systems should be structured and re-structured in order to accommodate non-traditional types of life courses. For this we want to combine different longitudinal data sets (namely the VSKT, the German Socio-Economic Panel Study (GSOEP) and the DEAS).

The paper deals with the combination of the DEAS and the VSKT as an example for matching survey and register data. Considering them separately both the survey and the insurance data have their advantages and drawbacks. The main advantages of the insurance data are their large number of cases and exact recording of entire biographies of pension-related circumstances. Process produced data from the pension funds' records incorporate the opportunity to examine current pensions and estimate future pensions most closely related to complex regulations of the pension law. On the other hand, the survey data is rich in information about subjective indicators such as plans of individuals in mid and advanced adulthood and other outcomes like health. By using the employment history information of the insurance data and the retrospective information about the life course of the survey data we can combine both data sets so that the strengths of both data sources are brought together.

The paper outlines the matching procedure used to combine the data as well as the required preparatory steps. Furthermore, challenges of the matching resulting from the characteristics of the data are examined as well as outcomes related to the goodness of matching and to the information we get about life courses and old age incomes from the joint data. Finally, some practical remarks on the statistical matching of survey and register data deduced from the experience with the German Ageing Survey and the administrative pension insurance data are given.

Julia Simonson, German Centre of Gerontology, Manfred-von-Richthofen-Str. 2, 12101 Berlin, Germany, Email: julia.simonson@dza.de

Laura Romeu Gordo, German Centre of Gerontology, Manfred-von-Richthofen-Str. 2, 12101 Berlin, Germany, Email: laura.romeu-gordo@dza.de

Nadiya Titova, German Centre of Gerontology, Manfred-von-Richthofen-Str. 2, 12101 Berlin, Germany, Email: nadiya.titova@dza.de