

**Parallel Session 2B:** “Surveys and Data Combination” (Chair Stephen P. Jenkins)

**Title:** *Extending the Empirical Basis for Wealth Inequality Research using Statistical Matching of Administrative and Survey Data*

**Authors:**

Anika RASNER<sup>a</sup>, DRV-Bund and MPI Rostock, rasner@demogr.mpg.de  
Joachim R. Frick, DIW Berlin, TU Berlin and IZA Bonn, jfrick@diw.de  
Markus M. Grabka, DIW Berlin, TU Berlin, mgrabka@diw.de

**Abstract**

Over the last decades, employment paths in Germany have vastly changed by becoming less continuous and increasingly heterogeneous. These *new* employment biographies are characterized by reduced working hours, employment in atypical and marginal jobs and most importantly, more frequent periods of unemployment. All of these are likely to reduce the individual’s entitlements in the earnings-related old-age pension scheme. At the same time, there is increasing pressure on the overall performance of the German pay-as-you-go pension system. This is - at least partly – due to far-reaching demographic changes resulting from an unprecedented rise in life-expectancy as well as decreases in fertility rates. These demographic pressures have prompted a sequence of reforms in the public pension system. The interaction of both, changes in employment paths as well as reforms of the public pension system, will have significant repercussions on economic inequality among future retirees.

Quantifying these distributional consequences is of innate importance, however difficult because of the lack of adequate micro data: While administrative data are appreciated for the large number of observations and high precision, they typically don’t cover the entire population. From a welfare economics perspective even more problematic, these data only include the individual’s pension-relevant income. Information on income above the maximum contribution ceiling as well as income at the level of the household and needs unit is not available. In contrast, population surveys might be representative for the entire population. At the same time, they typically suffer from small numbers of observations as well as missing lifecycle earnings and employment information.

Ideally, administrative and survey data could be directly merged using an unique identifier such as the individual’s Social Security Number (*record linkage*) in order to bring the *best of both worlds* together. In Germany, this approach is infeasible for data protection reasons. Without the explicit informed consent on behalf of survey respondents, record linkage is not permitted. Hence,

---

<sup>a</sup> Corresponding Author: Anika Rasner, Deutsche Rentenversicherung Bund, Ruhrstrasse 2, 10704 Berlin.  
Email: rasner@demogr.mpg.de.

*statistical matching* of administrative and survey data is the only way to overcome the drawbacks of both data sources.

In this paper, we present such an approach by complementing data from the population representative German Socio-Economic Panel Study (SOEP) with information from the Sample of Active Pension Accounts (*Versichertenkontenstichprobe*) provided by the German Social Security Administration (*Deutsche Rentenversicherung Bund*). Target variable of our analysis is the individual's public pension entitlement as of 2007. This is the most recent year of data available in both survey and register data. Using retrospective life history data as well as concurrently collected data from up to 24 annual interviews in the SOEP our matching variables include education, (un-)employment and income history (as well as fertility information for women). We separate our sample by sex, region and immigrant status taking into account pronounced regional and gender differences in pension entitlements. In a first step, we evaluate the appropriateness of our matching approach for the population of retirees. For this population, we can compare the current amount of pension payments as collected in survey data with the *true* information from the register match. Checking for robustness, we apply several approaches and eventually identify the technique based on Mahalanobis distances performing best for the data at hand. Results indicate that matches were better for persons who never divorced than for persons who divorced reflecting that the property settlement in case of a divorce (*Versorgungsausgleich*) can have an effect on individual pension entitlements in Germany. Thus, an additional match with the *Divorce Settlement Statistics* of the German Social Security Administration (GSSA), i.e., matching two datasets within the GSSA which have not been linked in such a way before, helped to correct for this bias. Adding information about the maintenance settlement improved the results for divorcees significantly.

This matching strategy was then applied to the entire population in order to provide an augmented database for the analysis of economic well-being for future retirees in Germany using a measure of net worth (based on housing and financial wealth) from SOEP data in conjunction with pension wealth entitlements in the earnings-related public pension scheme derived from the administrative records. Preliminary results indicate a very strong impact on wealth levels and wealth aggregates (plus 60% of non-pension wealth), on the portfolio structure as well as on wealth inequality (the Gini coefficient is reduced by more than 20 points) which exemplifies the dominant role of the public pension system in Germany. Such an augmented wealth measure may not only be useful for cross-national comparative research on wealth inequality but it may also reduce the respective bias *within* Germany resulting from different incentive structures to invest in old-age provision depending on the individual's employment status (self-employed, public servants and dependent employees).