

# What Did you Really Earn Last Year: Measurement Error in Survey Income Data

Stefan Humer  
INEQ - WU Vienna  
[shumer@wu.ac.at](mailto:shumer@wu.ac.at)

Stefan Angel  
WU Vienna

Franziska Disslbacher

Matthias Schnetzer  
Chamber of Labor

Of all micro-data, income is very likely to be the most pervasive information gathered by data providers, since income plays an essential role for a variety of welfare indicators, inequality measures and policy questions. Research on income inequality has rapidly gained momentum in recent years and the public interest in questions of income distribution has been growing considerably. The underlying income information for research and the public debate is usually obtained either from household surveys or from administrative records whereby both data sources have their idiosyncratic advantages and disadvantages. While the use of administrative data in empirical research has rapidly gained momentum, the accuracy of survey data has increasingly been contested during the last years. Administrative data not only represents an alternative source for income data, it can also be applied as a validation source for survey income data.

This paper provides an empirical evaluation of reasons for income measurement error resulting from inaccurate survey responses by using a unique data-set. We take advantage of the fact that due to a legal requirement, the Austrian 2008-2011 waves of EU-SILC provide income information from both survey and administrative register data for the same individuals and a wide range of components of total disposable household income. There are several causes why survey reports may deviate from administrative records. Specifically, we focus on error sources that correspond to (i) social desirability bias in survey responses; (ii) socio-demographic characteristics of the respondent, and (iii) the survey design. Last, we test for (iv) the presence of learning effects in the response behavior. Social desirability bias is probably the most important source of income misreporting. Due to the sensitivity of questions about income, social desirability might lead to deliberate misreporting of income. For income reporting, the resulting hypotheses is that reporting incomes is biased towards the mean. Second, misreporting of income might vary with the socio-demographic characteristics of the respondent. However, here a wide range of factors, such as gender, age, (changes in) labor market status, education and job tenure have been highlighted. With respect to the survey design and setting, a wide range of variables is likely to influence the response behavior: the interview mode, the time span between the income reference period and the interview, proxy responses.

## Learning Effect

Most previous studies have been forced to combine survey and register data via matching techniques entailing additional uncertainty. In our dataset, however, information is linked directly by the National Statistical Institute via personal identifiers derived by the Austrian Data Protection Commission. This procedure ensures that the income information from survey and register data are linked exactly to the corresponding individuals. We thus are able to compare survey and administrative data within one dataset. This is a considerable advantage compared to most existing research that is based on sophisticated matching procedures of survey and administrative dataset. Even the most prominent studies concerning income measurement error in surveys suffer from serious data limitations. Participation in the matched sample often is voluntarily, resulting in a sample that is biased towards individuals giving more accurate responses. Other prominent validation studies are restricted to individuals working at a single company. The dataset used in this paper is not subject to any of these shortcomings. Additionally, we are able to evaluate income measurement error for various components of total disposable household income in the same dataset.

In our application we make use of two related techniques, multinomial logit and panel regressions, which are perfectly suited to scrutinize our hypothesis on socio-demographic and survey related determinants of income measurement errors. First, we search for factors that help us to achieve a better understanding why self-reported income of individuals is above or below their administrative income. Therefore, our dependent variable is the mismatch category and we model the probability that in individual reports either less, the same or more than his or her true income jointly in a multinomial logit model. By estimating a multinomial logit model, we explicitly allow the estimated coefficients to vary across mismatch groups. Thereby we are able to identify different mismatch determinants for over-reporters and under-reporters. Although we consider an extensive and diverse set of control variables, it can not be ruled out that our regression results suffer from the omission of relevant but unobservable determinants. To check the robustness of our findings, we secondly make use of the longitudinal dimension of EU-SILC and employ fixed-effects estimations. The focus on within individual changes over time makes it possible to control for individual characteristics that are not observed but supposed to be constant over time, such as the cognitive ability to answer interviews or past experience with surveys. We apply panel regression models with individual and time-fixed effects in two related specifications. First, the dependent variable is the absolute difference between the survey report and the administrative value. In this setting our primary interest is in the influence of social desirability bias. Using the raw absolute difference should result in a clear pattern across income quintiles. In lower parts of the distribution we are expecting on average a positive mismatch (i.e. over-reporting), whereas the parameters should turn negative (i.e. under-reporting) as we approach the top income earners. Second, we transform the mismatch variable and project all negative values into the positive range by altering their sign. In this examination we focus on the magnitude of the error and study if repeated interrogations are associated with a statistically significant learning effect over time.