What did you really Earn Last Year? Explaining Measurement Error in Survey Income Data

Stefan Angel
(Vienna University of Economics and Business)

Franziska Disslbacher
(Vienna University of Economics and Business)

Stefan Humer
(Vienna University of Economics and Business)

Matthias Schnetzer
(Federal Chamber of Labor, Vienna, Austria)

Paper prepared for the 35th IARIW General Conference
Copenhagen, Denmark, August 20-25, 2018
Session 7A-1: New Insights into Economic Behaviour from Households Surveys
Time: Friday, August 24, 2018 [14:00-17:30]

What did you really earn last year? Explaining measurement error in survey income data

Stefan Angel[a], Franziska Disslbacher[a], Stefan Humer[b], and Matthias Schnetzer[c]

[a] *Department of Socioeconomics, Vienna University of Economics and Business, Vienna. Austria.*
[b] *Economics of Inequality, Vienna University of Economics and Business, Vienna, Austria.*
[c] *Department of Economics, Federal Chamber of Labor, Vienna, Austria.*

Abstract

This paper analyses the sources of income measurement error in surveys with a unique dataset. We use the Austrian 2008–2011 waves of EU-SILC which provide individual information on wages, pensions and unemployment benefits from survey interviews and officially linked administrative records. Thus, we do not have to fall back on complex two-sample matching procedures like related studies. We empirically investigate four sources of measurement error, namely (i) social desirability, (ii) socio-demographic characteristics of the respondent, (iii) the survey design, and (iv) the presence of learning effects. We find strong evidence for a social desirability bias in income reporting, while the presence of learning effects is mixed and depends on the income type under consideration. An Owen value decomposition reveals that social desirability is a major explanation of misreporting in wages and pensions, whereas socio-demographic characteristics are most relevant for mismatches in unemployment benefits.

Keywords: Income measurement error; Register data; Response error; Survey data; Survey error; Survey methods
JEL Classifications: C2, C81, C82, C83, D3

1 Introduction

Given the rapidly enhancing possibilities for gathering, storing and providing data, empirical economics is at a crossroads. In particular, the increasing popularity of administrative register-based data has created new standards in volume, velocity, variety and complexity. Data availability for research has enhanced immensely through the digitalisation of traditional paper-and-pencil surveys by official statistical authorities. However, this trend entails a growing importance of data quality and validation to ensure the reliability and accuracy of the gathered information. In this regard, differences between responses in traditional surveys and administrative records are of special interest to assess data quality of either source.

Of all micro-data, income is one of the most comprehensive information since it plays an essential role for a variety of welfare indicators and policy questions. Moreover, public interest in questions of income distribution has been growing considerably in recent years and research on income inequality has rapidly gained momentum. The underlying income information is usually obtained either from household surveys or from administrative records whereby both data sources have their idiosyncratic merits and drawbacks. While policy recommendations are frequently based on survey data, the accuracy of survey responses is often questioned and issues of measurement error are raised. Accordingly, there is still no consensus upon what is the best way to collect income data at the micro-level (Hansen and Kneale, 2013).

Potential differences between survey responses and administrative records have already been addressed in the literature on data quality and measurement error (Moore et al., 2000; Lohmann, 2011). The identification of measurement errors requires by definition a point of reference to judge the accuracy of the information. In the literature, a benchmark dataset is commonly declared to provide the true values in order to analyse deviations from the supposed erroneous data. In most cases, administrative records are proclaimed to be the benchmark even though these data may also be subject to uncertainty, for instance due to contradictory information in merged data sources or owing to the imputation of missing values (Berka et al., 2012; Schnetzer et al., 2015). While we acknowledge that the choice of benchmark data solely reflects the researchers' confidence in the accuracy of a specific dataset, the quality of Austrian administrative data is very high (Asamer et al., 2016). We thus follow the traditional literature and check survey income data against administrative records.

This paper focuses on two questions. What are the reasons for misreporting income in surveys and do these reasons differ with respect to specific types of income, i.e. wages, pensions, and unemployment benefits? The causes why survey responses may deviate from administrative records are manifold (Bound, Brown, and Mathiowetz, 2001; Tourangeau, Rips, et al., 2000). Specifically, we focus on four error sources emphasised in the literature. These are (i) the presence of a social desirability bias in survey responses, (ii) specific socio-demographic characteristics of the respondent, and (iii) the survey design. Finally, we test for (iv) the presence of learning effects in the response behaviour. These reasons may also vary with respect to the income component. For instance, social desirability may have different directions for earned income and for unemployment benefits.

The data requirements for an empirical evaluation of these error sources are extensive. Most previous studies have been forced to combine survey and register records via statistical matching techniques, based on either register data or error-prone self-reported identifiers, such as social security numbers. Additionally, consent of individuals to match survey and register data is usually needed. Since participation in the matched sample often is voluntarily, the sample is biased towards individuals giving more accurate responses (Bollinger, 1998). Briefly speaking, there is evidence of a consent bias for matched data sources. Furthermore, samples based on optional matching are often found to be non-representative for the whole population (Jenkins, Cappellari, et al., 2006; Jenkins, Lynn, et al., 2008; Sakshaug and Eckman, 2016; Sakshaug, K Couper, et al., 2012).

Fortunately, the Austrian 2008–2011 waves of the European Union Statistics on Income and Living Conditions (EU-SILC) do not suffer from such shortcomings. In this dataset, the survey responses are linked directly to administrative records by the National Statistical Institute via unique personal identifiers. Compared to probabilistic methods, this procedure ensures that the administrative income information is linked exactly to the corresponding survey respondent and thus reduces matching uncertainty drastically. Hence, our dataset provides both survey and register income data for the exact same observational units and offers a unique opportunity to evaluate the drivers of income measurement error.

We thus are able to compare survey and administrative data within one dataset which is a considerable advantage compared to most existing research. Even the most prominent studies about income measurement error are based on sophisticated statistical matching procedures of survey and administrative datasets. Other validation studies are restricted to individuals working at a single company (Pischke, 1995). The dataset used in this paper is not subject to any of these limitations. Furthermore, in contrast to papers studying measurement errors in total household income, we are able to analyse different single income components including unemployment benefits, which has rarely been done yet.

The remainder of this article is structured as follows. Section 2 provides a description of potential sources of measurement errors elaborated in the existing literature. In section 3, we discuss the specifics of our EU-SILC dataset and provide descriptive statistics for the structure of errors in our data. We then apply a multinomial logit model to evaluate the effects of the above-mentioned error sources (i) through (iv) in section 4. For a more detailed analysis of error sources (i) and (iv), we conduct panel regressions to detect changes of response behaviour over time. We conclude our analysis quantifying the relative importance of the four error sources based on an Owen value decomposition. Finally, section 5 provides a summary of our results.

2 Reasons for Measurement Error in Survey Income Data

While the use of administrative data in empirical research has rapidly gained momentum (Einav and Levin, 2014), the accuracy of survey information has increasingly been contested during the last years (Meyer et al., 2015). Erroneous information can arise from misreporting by respondents and decrease the overall quality of survey data. Misreporting in surveys is particularly grave when the affected data is, like income, the basis for policy-making.

Following the psychological literature on cognitive processes when answering a survey question, misreporting of income can arise from, first, troubles related to the interpretation and understanding of the question asked, second, problems in retrieving and judging the relevant information as well as its placement in time and, third, difficulties in formulating a response in the required format (Tourangeau, Bradburn, et al., 2010). The theoretical and empirical literature on income measurement errors emphasises four main reasons for misreporting: (i) social desirability, (ii) socio-demographic characteristics of the respondent, (iii) specifics of the survey design, and (iv) the presence of learning effects. However, the existing literature has ignored that those four reasons for misreporting can vary with the income type under consideration.

For instance, Angel et al. (2017) analyse reporting errors for total disposable household income in Austria, which is not observed directly but aggregated ex-post based on a comprehensive inquiry of single income components. Misreporting based on total household income thus disregards potential heterogeneity of the error-generating process between household members and income components. Further, misreporting of income can go into two directions: respondents can overreport or underreport a particular income type. In what follows, we discuss how (i) to (iv) can result in over- or underreporting of wages and then, why (i) to (iv) might lead to different expectations for the misreporting of pensions and unemployment benefits, the two other income components considered in this paper.

First, social desirability bias is probably the most important reason for income misreporting in surveys. Due to the sensitivity of questions about income, social desirability might lead to deliberate misreporting (Moore et al., 2000). It is widely documented that sensitive questions elicit patterns of overreporting (underreporting) for socially (un)desirable behaviour, attitudes, and characteristics (Bound, Brown, and Mathiowetz, 2001). For wages, the resulting hypothesis is that reported values are biased towards the mean, hence reporting errors are expected to be mean-reverting. Respondents at the lower tail of the wage distribution overreport as they feel ashamed of their actual economic conditions, while respondents at the upper tail of the distribution underreport since they do not want to disclose their high wages to an (unknown) interviewer. Such a reporting pattern is consistent with a desire for social comfort in the sense that households tend to locate themselves in the middle of the distribution. Related micro-level validation studies typically confirm the mean-reverting nature of the error in reported earnings (Kreiner et al., 2015; Kim and Tamborini, 2014; Pischke, 1995; Bound, Brown, Duncan, et al., 1994; Bound and Krueger, 1991). As an exception, Hariri and Lassen (2017) find for the Netherlands that respondents at the top of the income distribution overreport their income. In their study, however, income comprises earnings, employers' pension contributions, transfer and capital income, which were collected exclusively via telephone interviews with a one-shot recall question. These results are thus not easily comparable with most other studies that focus on earnings and derive income data from surveys with more complex interview modes.

Second, misreporting of income might vary with socio-demographic characteristics of the respondent (Kreiner et al., 2015; Kim and Tamborini, 2014; Bound, Brown, and Mathiowetz, 2001; Tamborini and Kim, 2013). We expect to find a higher propensity to overreport wages for males, due to a desire to demonstrate social status and to comply with the male-breadwinner model. Existing research suggests different misreporting patterns by sex, where males are found to overreport earnings more often than females (Bound and Krueger, 1991; Bollinger, 1998; Micklewright and Schnepf, 2010; Pedace and Bates, 2000; Kim and Tamborini, 2014). Kim and Tamborini (2014) and Bound, Brown, Duncan, et al. (1994) find that higher educated workers report earnings more correctly. Respondents with higher education might display lower rates of misreporting as this group is potentially more likely to be familiar with the purpose and relevance of households surveys. The positive correlation between education and financial literacy documented in the literature (Lusardi and Mitchell, 2014) may also explain some of these findings.

The relationship between misreporting and age is a priori unclear. On the one hand, cognitive abilities to answer the questionnaire decrease with age. On the other hand, older respondents receive wages for a longer period of time and are in more stable employment. Since the vast majority of validation studies has found a negative relationship between misreporting and age, we adopt these findings for the expectations for our Austrian sample (Kim and Tamborini, 2014; Kim and Tamborini, 2012; Bound and Krueger, 1991). Two additional socio-demographic characteristics potentially contributing to the accuracy of responses are the number of changes in the employment status during the income reference period, related to receiving income from different sources and second, the number of months spent in a specific employment status during the reference period. Changes in the employment status can be associated with telescoping errors, which refer to misplacing the receipt of a particular income source in time. A stable employment status is associated with less variation in the level of income received. With respect to wages, it is reasonable to assume that changes in the employment status increase reporting errors while longer employment periods lead to less misreporting, since it is easier to recall the remuneration. Kim and Tamborini (2014) find that occupation or industry switchers are more prone to misreport earnings while Bound, Brown, Duncan, et al. (1994) document a negative relationship between job tenure (years with current employer) and response error.

What the literature has largely ignored is the relationship between misreporting and health, the degree of urbanisation, and the country of birth. Healthier individuals may give more accurate answers because they are in better mental conditions and therefore are less likely to make recall or response errors. Further, we expect to find less misreporting, the higher the degree of urbanisation at the respondent's place of residence is. The rationale of this argument is rooted in the anonymity of cities, whereas in rural areas, mistrust in unknown interviewers might be more pronounced. Additionally, misreporting is expected be related to the respondent's country of birth. Being foreign-born can serve as an indicator of language skills and familiarity with institutional settings. As both factors are relevant for the comprehension of the questions and the correct allocation of total income across income types, we expect to find more misreporting of wages for foreign-born respondents.

Third, and regarding the survey design and setting, a wide range of variables is likely to influence the response behaviour. We focus on the role of the interview mode (Lynn et al., 2012), the time span between the income reference period and the interview, and proxy responses. Regarding the interview mode, telephone interviews are considered to be more susceptible for misreporting of wages than a face to face setting. Fessler et al. (2017) find that households interviewed via telephone report higher incomes on average than those interviewed personally. Furthermore, it is crucial whether the respondent provides the required income information personally or via an entitled third party. While some studies have found little proxy bias (Bound and Krueger, 1991; Mellow and Sider, 1983), others suggest a downward bias (Tamborini and Kim, 2013; Reynolds and Wenger, 2012). We expect more misreporting of wages for proxy responses, resulting from a lack of information. Finally, we hypothesise that the larger the time span between the interview and the income reference period, the larger the error in reported wages because of recall and memory problems.

The fourth, and last, central issue with reporting errors is the presence of learning effects. If present, reporting errors are supposed to attenuate with cumulated survey experience. Learning effects are related to recall and retrieval strategies of respondents and are best explained in the context of panel surveys, where the same households are interviewed repeatedly. In the first wave of participation, respondents are unexperienced regarding the survey setting and unprepared to answer the questionnaire. In the follow-up waves, however, respondents might be equipped with income tax documents and other relevant files. Therefore, we expect to find misreporting of wages to decrease with accumulated survey experience. Likewise, a learning effect can also be expected for pensions and unemployment benefits.

While some variables might have similar effects on wages, pensions, and unemployment benefits, we anticipate diverging effects for others. For instance, wages (and pensions) are attached to the labor market and tied to (past) individual effort, whereas unemployment benefits are often stigmatised as charity despite actually being an insurance. The socially desirable behaviour is thus to downplay

Table 1: Expected effects of misreporting by survey income types

	Wages		Pension		Unemp. ben.	
	S<A	S>A	S<A	S>A	S<A	S>A
(i) Social Desirability						
Income Decile *(increasing)*	+	−	+	−	+	−
(ii) Socio-demographic characteristics						
Gender *(Ref: Female)*	−	+	−	+	+	−
Education *(Ref: Compulsory school)*	−	−	−	−	−	−
Age *(increasing)*	−	−	−	−	−	−
Country of birth *(Ref: Austria)*	+	+	+	+	+	−
Health status *(Ref: Very bad)*	−	−	−	−	+	+
Degree of urbanisation *(Ref: <10,000 inhabitants)*	−	−	−	−	−	−
Changes in employment status *(Ref: None)*	+	+	−	−	+	+
Months in corresponding emp. status *(Ref: 12 months)*	+	+	+	+	∼	∼
(iii) Survey setting						
Mode of interview *(Ref: CAPI)*	+	+	+	+	+	+
Type of interview *(Ref: Personal)*	+	+	+	+	+	+
Month of interview *(Ref: March-May)*	+	+	+	+	+	+
(iv) Learning effect						
Wave of interview *(Ref: First)*	−	−	−	−	−	−

Notes: This table summarises our expectations about the likelihood to over- and underreport income conditional on different sources of errors. S<A: Underreporting in Survey; S>A: Overreporting in Survey; + stands for an increasing probability to fall in that specific reporting category, − symbols a decreasing probability, ∼ implies an ambiguous relationship.

the level received leading to a general underreporting of unemployment benefits. Regarding (ii) the socio-demographic characteristics of the respondent, we expect males to underreport unemployment benefits more often and stronger than females since receiving benefits contrasts the common male bread-winner model. Being unemployed might be associated with a higher social stigma for the better educated. Therefore, we expect underreporting (overreporting) of unemployment benefits to be an increasing (decreasing) function of education. While being foreign-born might lead to a higher misreporting of wages and pension income, we expect to find less overreporting of unemployment benefits since those born abroad might be particularly prone to downplay the level of received state transfers.

With respect to the number of changes in the employment status, some particularities have to be mentioned. Typically, respondents only retire once and thus might be better informed about the level of pension they are going to receive. Precisely for this reason, we expect to find less misreporting of pensions, if the employment status for those receiving pension income has changed in the reference period. The relationship between the number of months spent in unemployment and misreporting of unemployment benefits is ambiguous. Short unemployment spells might be associated with less misreporting since respondents are better informed about the actual transfer due to the singularity of the situation. On the other hand, respondents might care less about the level of received unemployment benefits, the shorter the time spent in unemployment.

Table 2 summarises our expectations regarding the direction of the effect of the relevant variables related to social desirability, socio-demographic characteristics, the survey setting, and the learning effect, for over- and underreporting wages, pension income and unemployment benefits.

3 Data, variables, and method

3.1 Data and variables

For the assessment of income measurement errors in surveys, we make use of the Austrian EU-SILC (European Union Statistics on Income and Living Conditions). EU-SILC is a rotational household panel with a quarter of respondents being exchanged each year. The sample is drawn from all private households with a main residence in Austria according to the central population register. The main questionnaire is aimed at household members aged 16 or older and is conducted partly with computer-assisted personal interviews (CAPI) and partly with telephone interviews (CATI). In certain cases, proxy interviews with other household members were carried out instead of personal interviews (e.g. roughly 11% in 2011).

The primary motivation to utilise Austrian EU-SILC data is a unique feature compared to other household surveys: for four consecutive years, it provides combined income information from personal interviews and administrative sources. Up to the year 2011, incomes on both personal and household levels were obtained via conventional survey interrogation. From 2012 on, income data gathered from administrative records have replaced survey data for certain components of disposable household income, like wages, pensions, and unemployment benefits. Fortunately, Statistics Austria was able to merge administrative income data to the full EU-SILC survey sample from 2011 back to 2008. The detailed merging process of register and survey data is described further below.

The rotational character of EU-SILC allows to track households for as long as four consecutive years. With the data at hand, this maximum period of observation applies to the survey cohort that first participated in 2008 and remained in the survey until 2011. The 2007 and 2009 cohorts are each covered in three waves, the 2006 and 2010 cohorts in two waves respectively. In contrast to Eurostat's EU-SILC User Database (UDB), the Austrian national SILC provided by Statistics Austria delivers cross-sectional and longitudinal information in one single database. By virtue of permanent household and individual identifiers, we are able to track the changes in household responses compared to changes in the administrative records over time. It has to be noted in this respect that, in contrast to the first-time interrogation, follow-up interviews were predominantly accomplished via telephone (CATI).

Survey income data are collected retrospectively in EU-SILC and correspond to the calendar year prior to the interview. For income components with unequal net and gross values, respondents were asked to report either one or both values. When refusing to deliver a precise figure, interviewees could also report an income bandwidth. In the latter case, the specific income is estimated based on the empirical distribution of the corresponding income component. In the event of item non-response for single income categories, missing values are derived partly from socio-economic characteristics like sex, education, and age in an econometric exercise, and partly from statutory regulations like collective wage agreements. While the amount of imputations is generally rather low in EU-SILC, the application of such estimation methods may be an important source of measurement error. For instance, roughly 0.5% of the records had to be completely imputed in 2011 (Statistics Austria, 2014b). Since we focus on error generating processes in personal survey responses, we exclude imputed values from the analysis. We thus restrict our study on net income from wages, pension and unemployment benefits for which both survey responses and administrative data are available in EU-SILC. Table 2 provides descriptive statistics for the number of observations, the interview mode, and the share of imputations for the selected income types.

Table 2: Observations, Interview Mode, Share of Imputations – SILC 2008-2011

	Obs.		**Int. Mode (%)**		**Wages (%)**		**Pensions (%)**		**Unemp. (%)**	
	HH	Persons	CAPI	CATI	Svy.	Adm.	Svy.	Adm.	Svy.	Adm.
2008	5,707	10,946	70.0	30.0	4.6	2.6	1.5	0.5	0.2	0.8
2009	5,878	11,056	57.1	42.9	5.9	1.9	2.0	0.3	0.2	0.8
2010	6,188	11,493	59.6	40.4	4.8	2.4	1.5	0.4	0.3	0.9
2011	6,187	11,475	57.3	42.7	3.4	2.6	0.5	0.5	0.3	1.1

Notes: Columns 5–10 show the share of imputed values for selected income types. Source: SILC 2008-2011, own calculations.

Register information for the income types used in our study is obtained from various administrative sources. The most important data providers are the Austrian social security database, the payroll tax register, the Austrian federal pension fund, and the Austrian transfers dataset. Since the quality of administrative benchmark data is crucial, some papers on measurement error in surveys simultaneously recognise weaknesses of administrative data (Abowd and Stinson, 2013; Kapteyn and Ypma, 2007; Paulus, 2015). However, the overall quality of Austrian administrative income data is very high (Asamer et al., 2016). Wages are directly reported to the tax authorities by employers, while information on pensions and unemployment benefits is provided straight by government authorities. The probability of measurement errors in this data is thus expected to be negligible. Concerning flaws in administrative data due to tax avoidance, there is some evidence that this is comparatively less of an issue in Austria than in other OECD countries (Alm and Torgler, 2006; Hassan and Schneider, 2016).

The merging process between survey and administrative data is accomplished reliably with a branch-specific personal identification number for official statistics (bPIN OS) which serves as a unique ID in both data sources. These 172-digit PINs were introduced to protect privacy in the communication among public authorities via e-government. The PINs are created by the Austrian Data Protection Commission and used to identify individuals in the EU-SILC survey and the administrative sources. Unlike previous studies, we thus do not have to fall back on two-sample matching processes or the like, since survey responses and retrospective administrative information are already combined for the years 2008 to 2011. Studies on measurement errors usually also depend on consent to link survey responses to administrative records which often leads to small sample sizes (Kreuter et al., 2010). In our study, more than 95% percent of the respondents in the EU-SILC survey could be identified with a PIN in order to assign the register information (Statistics Austria, 2014a). Further, the income reference periods for the survey and the various administrative records used overlap exactly and no adjustments to ensure comparability between the data sources had to be made.

Reporting income is a two-stage process in EU-SILC. At the first stage, respondents have to indicate whether a certain income component was received during the reference period. Only in a second step, the amount of income received from a particular source has to be reported. Consequently, a mismatch between survey and register data can result at either stage. Table 3 shows the percentage of respondents reporting the three involved income types. Wages are consistently underreported in

Table 3: Reporting of Income Types – SILC 2008-2011

	Wages (%)			**Pensions (%)**			**Unemp. ben. (%)**		
	Svy.	Adm.	Diff.	Svy.	Adm.	Diff.	Svy.	Adm.	Diff.
2008	53.8	56.6	-2.8	24.9	24.1	0.8	7.1	10.2	-3.1
2009	54.8	58.0	-3.2	25.0	24.7	0.4	7.4	10.2	-2.8
2010	55.4	57.9	-2.4	25.9	25.3	0.6	9.1	12.6	-3.5
2011	55.8	59.0	-3.2	24.0	24.6	-0.6	9.2	12.7	-3.5

Notes: This table shows the share of respondents reporting a specific income type.
Source: SILC 2008-2011, own calculations.

the survey data by 2.4 to 3.2 percentage points. By contrast, the number of individuals reporting pension income is slightly higher in the survey data compared to the administrative records with the exception of 2011. The share of survey respondents with declared unemployment benefits is generally lower than indicated by the official statistics. The deviations range between 2.8 and 3.5 percentage points. Overall, we notice a prevailing underreporting of income receipt in survey responses with the exception of old-age benefits from 2008 to 2010.

With regard to the level of reported income, we distinguish the following possible cases of mismatch. Respondents can report a particular source of income in the survey even though it was not received according to register data (`Sx0`) or vice versa, not report a particular type of income even though it was received according to register data (`0xA`). Further, the survey report can positively or negatively deviate from the register data, which corresponds to overreporting (`S>A`) and underreporting (`S<A`), respectively. No mismatch occurs if a respondent reports the amount that corresponds to the register entry within a narrow range of $\pm 5\%$ (`S=A`) or if a specific type of income was not received according to both survey and register data (`0x0`). Since reporting an amount in the survey that corresponds exactly to the register value is almost impossible, we allow the survey report to deviate marginally from the register entry in order to fulfil our operational definition of correct answers. We define a categorical variable $Pr(Y_{i,k} = j)$ for mismatch types j, individuals $i = 1, ..., N$, and income component $k \in [1, 3]$ as

$$Pr(Y_{i,k} = j) = \begin{cases} Pr(Y_{i,k} = \texttt{0x0}) & \text{if no income in survey and admin} \\ Pr(Y_{i,k} = \texttt{0xA}) & \text{if only admin income report (false negative)} \\ Pr(Y_{i,k} = \texttt{S<A}) & \text{if survey underreporter} \\ Pr(Y_{i,k} = \texttt{S=A}) & \text{if survey corresponds to admin} \\ Pr(Y_{i,k} = \texttt{S>A}) & \text{if survey overreporter} \\ Pr(Y_{i,k} = \texttt{Sx0}) & \text{if only survey income report (false positive)} \end{cases}$$

Table 4 gives an overview of the structure of mismatch in Austrian EU-SILC data from 2008 to 2011. In this summary, we display the shares of observations in the respective reporting categories and the median of the absolute and relative deviation between survey and administrative responses. For all income types, the shares of over- and underreporters are very stable over the years. There are roughly 17% overreporters and 23% underreporters for wages which is a high number compared to only 11 to 13% of respondents with matching values (apart from the 40% with no wage income in

Table 4: Structure of mismatch in income reporting – SILC 2008-2011

	Wages			Pensions			Unemp. benefits		
	Obs. (%)	Abs. (P50, €)	Rel. (P50, %)	Obs. (%)	Abs. (P50, €)	Rel. (P50, %)	Obs. (%)	Abs. (P50, €)	Rel. (P50, %)
2008									
`0x0`	40.3	0.0	0.0	73.5	0.0	0.0	88.6	0.0	0.0
`0xA`	5.9	−1566.2	−100.0	1.6	−4474.5	−100.0	4.3	−1703.1	−100.0
`S<A`	22.5	−3257.7	−18.1	8.0	−2334.2	−13.5	2.6	−882.4	−32.6
`S=A`	11.0	−78.7	−0.7	8.8	−110.6	−0.9	0.7	−38.3	−1.4
`S>A`	17.3	3314.3	27.0	5.6	3474.9	23.8	2.5	1137.4	48.1
`Sx0`	3.0	8459.3	*Inf*	2.4	10907.1	*Inf*	1.2	4200.0	*Inf*
2009									
`0x0`	39.4	0.0	0.0	73.4	0.0	0.0	88.6	0.0	0.0
`0xA`	5.8	−1386.2	−100.0	1.6	−4666.3	−100.0	4.0	−1398.7	−100.0
`S<A`	22.9	−3255.7	−17.8	8.3	−2580.0	−14.4	2.6	−833.8	−29.4
`S=A`	11.7	−56.7	−0.5	8.9	−81.9	−0.6	0.7	−31.4	−0.7
`S>A`	17.7	3177.2	25.9	5.9	3046.7	21.9	2.9	908.2	42.5
`Sx0`	2.6	7420.0	*Inf*	1.9	10904.8	*Inf*	1.2	4200.0	*Inf*
2010									
`0x0`	39.3	0.0	0.0	72.5	0.0	0.0	86.7	0.0	0.0
`0xA`	5.2	−1368.6	−100.0	1.5	−4860.6	−100.0	4.2	−1802.5	−100.0
`S<A`	22.9	−3197.6	−16.8	9.1	−2465.8	−13.4	4.0	−1003.7	−32.1
`S=A`	12.4	−86.1	−0.7	9.0	−114.0	−1.0	1.2	−1.2	0.0
`S>A`	17.3	3093.5	26.0	5.6	3497.2	25.7	3.2	923.2	34.6
`Sx0`	2.8	5716.7	*Inf*	2.2	11237.4	*Inf*	0.7	2800.0	*Inf*
2011									
`0x0`	38.5	0.0	0.0	74.2	0.0	0.0	86.6	0.0	0.0
`0xA`	5.7	−1281.0	−100.0	1.9	−5632.1	−100.0	4.2	−1589.2	−100.0
`S<A`	22.7	−3083.0	−16.9	1.1	−3455.7	−17.6	3.8	−1020.9	−30.8
`S=A`	13.3	−121.2	−0.8	20.6	0.0	0.0	1.1	−10.6	−0.5
`S>A`	17.2	3178.2	26.3	1.0	5249.3	84.9	3.6	1019.9	34.3
`Sx0`	2.5	4419.5	*Inf*	1.2	11540.6	*Inf*	0.6	5220.0	*Inf*

Notes: This table shows the proportion and the median difference (both in absolute and relative terms) of reported survey income and the income recorded in administrative sources per income type and survey year. (1) `0x0`: no income in survey and administrative data. (2) `0xA`: only income in administrative data. (3) `S<A`: reported income is below the value in administrative data (4) `S=A`: survey income corresponds to administrative data. (5) `S>A`: reported income is above the value in administrative data. (6) `Sx0`: only income in survey data. Source: SILC 2008-2011, own calculations.

both data sources). The median relative deviation from the administrative information lies around -17% for underreporters and roughly 26% for overreporters.

Concerning pensions, we see stable shares of reporting types except for 2011 where the percentage of correct responses takes a sudden jump to roughly 20%. Between 2008 and 2010, the share of underreporters exceeds the number of overreporters, however, the median deviation is considerably higher for the latter group. With regard to unemployment benefits, the share of correct survey answers is very low and consistently smaller than the shares of over- and underreporters. The median respondents with lower survey than administrative values report approximately 30% less income. The median relative deviation for overreporters ranges between 34 and 48%.

As we have shown that differences between administrative records and survey responses are relevant for various income components, we aim to ascribe the occurring mismatches to the above-mentioned reasons for misreporting. Thus, we are interested in the effect of (i) social desirability, (ii) socio-demographic characteristics, (iii) aspects of the survey design, and (iv) learning effects on the presence, direction and extent of misreporting of three components of total disposable household income (wages, pensions, and unemployment benefits).

In the empirical analysis, we use different approaches to shed light on these issues. When studying the reasons (i) to (iv) for the *direction* of misreporting, we distinguish between those with practically correct information, over-, and underreporters, and consider both the positive and negative mismatch. In contrast, when analysing the impact of these reasons on the *extent* of misreporting, we focus on the metric *difference* (survey minus register).

The right hand side in our econometric specification comprises the variables describing reasons (i) to (iv) for misreporting. The presence of (i) social desirability bias is indicated by the respondent's position in the respective income distribution specified by the income decile in the register data. The explanatory variables referring to (ii) the socio-demographic characteristics of the respondent are gender, educational attainment according to the ISCED classification, age, and a categorical variable referring to the health status. We also include the country of birth, the degree of urbanisation at the place of residence, and the employment status with the following options: full-time employed, part-time employed, full-time entrepreneur, part-time entrepreneur, unemployed, retired, domestic worker, student or other. Additionally, dummy variables indicating the number of changes in the employment status during the income reference period should capture the stability of employment. Finally, depending on the type of income investigated, we include the number of months being either full- or part-time employed, retired, or unemployed. The distribution of over- and underreporters across the socio-demographic characteristics is shown in Table A.1.

The explanatory variables related to (iii) the survey setting comprise a dummy variable for the interview mode (CAPI/CATI), a dummy variable related to the reporting status (proxy vs. self-reported income), and a categorical variable specifying the month of the interview (March to May, June to August, September to November). The motivation for the latter variable is straightforward: the earlier the interview took place, the shorter is the time span between the income reference period and the income reporting. The distribution of the response categories across the survey setting variables is given in the lower part of Table A.1. The indicator for (iv) the learning effect is a dummy variable for the interview wave ranging from one to four. Individuals participating in more than one waves are a priori expected to have more experience with income surveys and tend to provide more reliable responses.

3.2 Method

Our empirical strategy is a three-step procedure. First, we apply a multinomial logit regression to assess the impact of the single reasons on the direction of mismatch and thus the probability to over- or underreport income. Second, we enrich the analysis with panel regressions to estimate the effect of the single reasons on the extent of misreporting while controlling for unobserved individual characteristics. Third, we determine the relative importance of the error sources for misreporting with an Owen decomposition. In the following, we describe each of the three steps briefly. For a more elaborate discussion on the technical details of the methods used, we refer to standard econometric textbooks like Maddala (1983) and Hsiao (2014).

Multinomial logit. With a multinomial logit model, we search for factors that help us to better understand why self-reported incomes are above or below their corresponding administrative records. Our dependent variable is the mismatch category $Pr(Y_{i,k} = j)$ and we calculate the probabilities for reporting less (`S<A`), the same (`S=A`), or more (`S>A`) than his or her true income. Although strictly speaking, respondents who do not report income in both sources (`0x0`) have no mismatch, this group of observations is not of interest for our main research questions and therefore discarded from the analysis. Additionally, due to the very low number of observations without income information either in the survey (`0xA`) or in the administrative register (`Sx0`), we restrict our attention to overreporters, underreporters, and the consistent group.

By estimating a multinomial logit model via maximum likelihood, we explicitly allow the estimated coefficients to vary across the mismatch categories. Thereby, we are able to identify the determinants of mismatch separately for overreporters and underreporters. This is a considerable advantage compared to OLS regression, since it may be very hard to defend the assumption that the variables influencing underreporting are similarly affecting the probability of overreporting. Even more, in the standard OLS framework all reporting errors are pooled together. Consequently, assuming parameter homogeneity across mismatch categories could not only lead to misleading interpretations, but positive and negative errors could potentially cancel out and leave us with statistically insignificant estimates. In contrast, the effect of e.g. gender or education on the probability of misreporting is allowed to differ between over- and underreporters in the multinomial logit model. This flexibility enables us to draw a more detailed picture of the factors that influence the response behaviour of individuals in income surveys.

Panel regression. Although we consider an extensive and diverse set of control variables, we cannot rule out that our estimation lacks relevant but unobservable determinants. To check the robustness of our findings, we thus make use of the longitudinal dimension of EU-SILC from 2008 to 2011 and employ fixed-effect estimations. The focus on within-individual changes makes it possible to control for individual characteristics that are unobserved but supposedly constant over time, such as the cognitive ability to answer interviews or past experience with surveys. To purge these unobservable characteristics in two related specifications, we apply panel OLS regression models with individual and time fixed effects. The dependent variable is the difference between survey and administrative records for each person-year. [1].

In the first panel specification, our primary interest is the influence of social desirability on the difference between the survey report and the administrative value. We expect a clear pattern across income deciles, with a positive mismatch (i.e. overreporting) in the lower parts of the distribution and a growing negative mismatch (i.e. underreporting) as we approach the top income earners. In the second panel specification, we are particularly interested in the learning effect where only

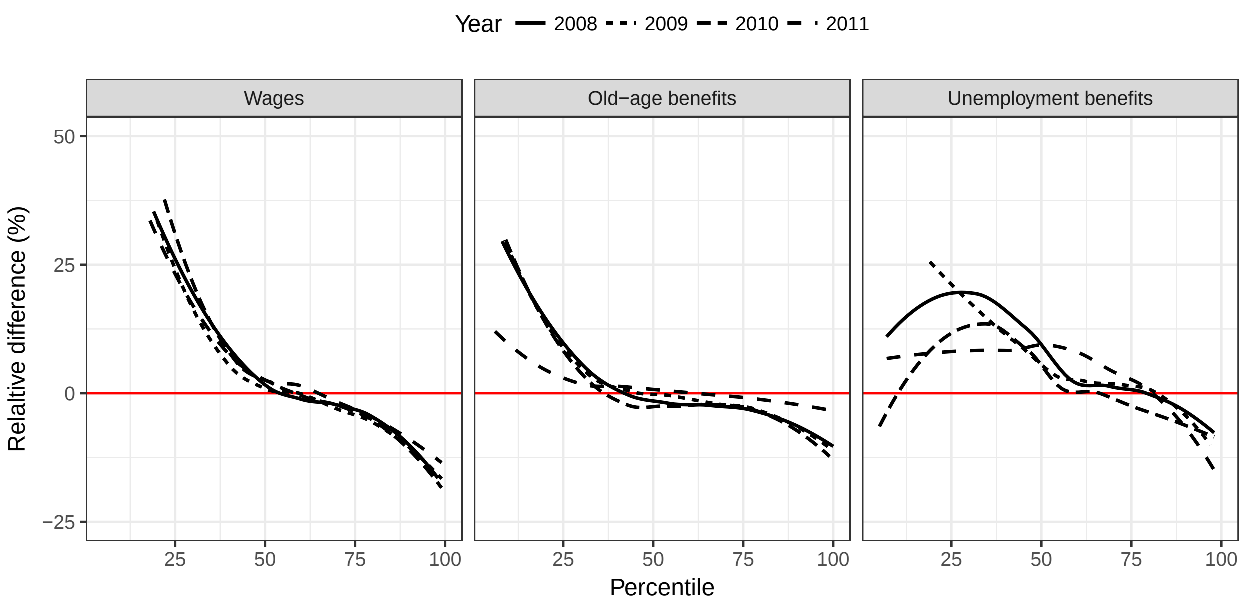
[1]We prefer a linear OLS specification as this provides a clear interpretation of marginal effects on the original scale. In contrast, for longitudinal (nonlinear) binary and multinomial logit response models with fixed effects, the intuitive interpretation of estimates as predicted probabilities (or various types of marginal effects) is not a viable option because the unobserved heterogeneity vector of person fixed effects is not estimated (see e.g. Pforr, 2014 for a more detailed discussion)

the absolute mismatch is relevant. The question is whether individuals participating in multiple survey waves tend to decrease reporting errors and repeated interrogations are associated with a statistically significant learning effect over time. The set of explanatory variables resembles the multinomial models apart from gender and the educational level, which both show negligible within-variation among adults.

Owen value decomposition. Finally, we are interested in the relative importance of error sources (i) to (iv) and apply an Owen value decomposition of the explained variance (R^2) in a pooled cross-sectional regression. This procedure allows to estimate the marginal contribution of each group of explanatory variables to the total R^2. The Owen value decomposition is a generalisation of the Shapley value decomposition and is suitable to assert the relative importance of groups of regressors for explaining the variance of the dependent variable (Huettner, 2012). To the best of our knowledge, such an assessment has not yet been done in the literature on income measurement error. To assess both time-varying and time-constant explanatory variables, the decomposition is based on pooled cross-sectional models for all four years with the difference between survey and register data as the dependent variable.

4 Results

Before turning to the results of the econometric exercise, we study the unconditional existence of a social desirability bias and a learning effect in the data.

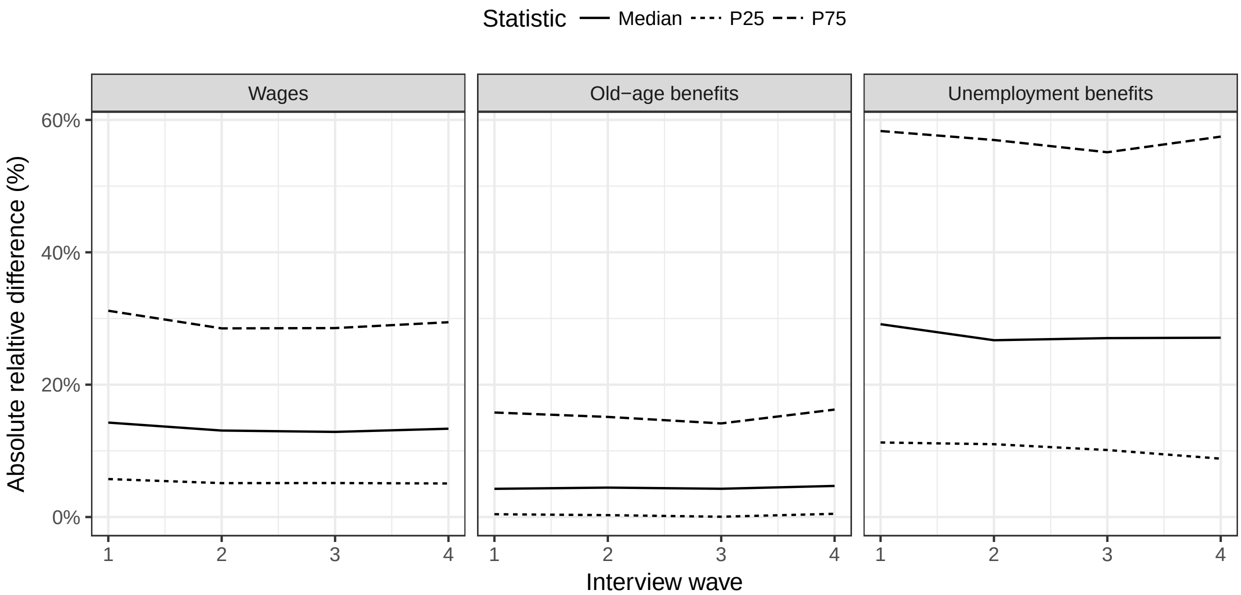


Notes: This graph shows the mean relative difference between survey responses and register information on wages, old-age and unemployment benefits by percentile of corresponding income distribution and year of the interview. Positive values in the lower half of the distribution imply structural over-reporting, whereas negative values above the median indicate under-reporting in surveys at the top. Source: SILC 2008-2011, own calculations.

Figure 1: Social desirability – SILC 2008-2011

Figure 1 shows the average relative difference between the survey response and the register entry per percentile of the distribution of wages, pensions, and unemployment benefits. Overall, the typical pattern of a mean-reverting error is visible. Thus, respondents in lower income percentiles of a specific income type report values that are higher than their register records and vice versa for respondents at the upper part of the distribution.[2] The pattern is most distinctive for wages and least pronounced for unemployment benefits, but in general corresponds to the expected social desirability behaviour.

Figure 2 illustrates the learning effect over the four survey waves. The lines depict the absolute logarithmic difference between the two data sources in the mean, the median, the 25^{th}, and the 75^{th} percentile. For wages, we find a slight reduction in the differences for all observed income quantiles after the first wave. A small error reduction over time is also present in the 25^{th} quantile and the median of unemployment benefits. In contrast, for pension income, the error does not seem to decrease over time. Thus, the data displays no or, if anything, a very small learning effect.[3]



Notes: This graph shows the median as well as the 25^{th} and 75^{th} percentile of the absolute values of relative difference between survey responses and register information on wages, old-age and unemployment benefits by wave of the interview (i.e. negative values have been multiplied by -1). A reduction of absolute errors over time would be evidence for a learning effect in repeated interrogations, however the data display no strong support for this hypothesis. Source: SILC 2008-2011, own calculations.

Figure 2: Learning effect – SILC 2008-2011

In what follows, we will test whether the observed patterns of mean-reverting errors and the learning effect remain valid in a multivariate model. Given that we control for a variety of variables capturing the complexity of the annual income stream (e.g. employment status changes), a significant effect of mean-reverting errors would emphasise the role of social desirability as an important error source.

4.1 Likelihood of reporting more or less

Social desirability. For ease of interpretation, table 5 displays average marginal effects derived from the multinomial logit models.[4] For all three types of personal income, the estimates for the

[2]The figure is cut between a relative mismatch of $+50\%$ and -25%.

[3]Figure A.1 in the Appendix replicates figure 2, however, using only observations that remained in the panel in all four waves, i.e. from 2008 to 2011. With this sample the outcome for pensions and unemployment benefits resembles that of the unbalanced sample. For wages, the learning effect almost disappears.

[4]Full tables with the estimates expressed in log(odds) are available from the authors upon request.

corresponding income deciles—with the 5^{th} decile as reference category—confirm the mean-reverting error. Thus, compared to the middle of the wage distribution, the likelihood of underreporting is significantly lower in the bottom deciles and significantly higher in the top deciles. Vice versa, the probability of overreporting increases by up to 51 percentage points in the first decile and is significantly lower in higher deciles.

For pensions, this pattern is very similar although less pronounced since the average marginal effects of income deciles on both underreporting and particularly overreporting are smaller. Finally, recipients of unemployment benefits are no exception from the general pattern. Higher unemployment benefits correspond to a higher likelihood of underreporting and a lower likelihood of overreporting. Effects are statistically significant at the 5% level for almost all deciles. If we look at the probability of roughly similar income levels in both sources ($S = A$), lower income groups tend to report correct wages less often than middle-income groups. To a smaller extent, this also applies to the highest deciles. For pensions, changes in the probability of $S = A$ are more symmetrically spread around the middle whereas for unemployment benefits, income deciles generally do not have strong statistically significant effects.

Summing up, we interpret these results as evidence of an income-mean-reverting type of social desirability. The estimations also reveal that the mean-reverting pattern does not exclusively apply to wages but seems to be present also for non-market income and transfer payments.

Socio-demographic characteristics. Males have a significantly higher/lower tendency of overreporting/underreporting all three types of income. Overall, the gender-specific effect is largest for wages. Related to that, men display a slightly lower propensity to report matching values for wages. These results could reflect some underlying male breadwinner/masculinity norm which ceteris paribus renders men to overreport market income more often. Based on this argument, we also expected rather men to conceal receipt of non-market transfers. However, our estimates for unemployment benefits do not provide support for the latter hypothesis. Concerning education, we find significant differences between respondents with higher educational attainments compared to compulsory education for wages and pensions. Underreporting decreases whereas overreporting increases with the educational level. For the point estimates it also does not make much difference whether respondents hold a post-secondary or a tertiary degree. Contrary to what we have expected for unemployment benefits, there is no evidence that underreporting is an increasing function of education. All education dummies are statistically insignificant in this model. In line with the results for social desirability (income percentiles) and gender, it seems that there is no big social stigma related to levels of unemployment transfers (assuming that the other control variables capture cognitive errors sufficiently). Instead, it is possible that these transfers are generally regarded as legally acquired insurance payments. Age only exerts a statistically significant, albeit very small negative effect on the likelihood of underreporting unemployment benefits.

Table 5: Multinomial logit regressions

	Wages			Pensions			Unemp. benefits		
	S<A	S=A	S>A	S<A	S=A	S>A	S<A	S=A	S>A
(i) Social desirability									
Realtive income (Ref: 5. Decile)									
1. Decile	**-28.3**	**-21.5**	**49.8**	**-15.2**	**-16.8**	**32.0**	**-22.4**	-5.0	**27.4**
2. Decile	**-17.8**	**-17.0**	**34.7**	-4.3	*-5.8*	**10.1**	**-11.0**	-3.0	**14.0**
3. Decile	**-11.2**	**-12.1**	**23.3**	-4.2	-1.5	**5.7**	-0.3	*-6.5*	6.7
4. Decile	-2.0	**-7.7**	**9.7**	0.9	-3.9	3.0	-1.0	*-5.5*	6.5
6. Decile	**8.5**	-0.8	**-7.7**	**8.7**	**-6.5**	-2.2	*10.4*	1.8	**-12.1**
7. Decile	**14.2**	-1.3	**-12.9**	**8.1**	**-7.6**	-0.5	*8.9*	-0.5	-8.4
8. Decile	**18.6**	-3.3	**-15.3**	**9.5**	**-7.9**	-1.7	**17.1**	1.0	**-18.1**
9. Decile	**27.1**	**-7.4**	**-19.6**	**12.7**	**-8.0**	**-4.7**	**20.8**	*7.6*	**-28.3**
10. Decile	**38.6**	**-14.7**	**-23.9**	**24.2**	**-19.1**	**-5.1**	**31.4**	5.6	**-37.1**
(ii) Sociodemographic characteristics									
Gender (Ref: Female)									
Male	**-10.7**	**-3.7**	**14.5**	**-4.8**	-0.8	**5.6**	*-5.3*	-2.3	**7.6**
Age									
Age	0.0	0.0	0.0	-0.2	0.0	0.2	*-0.2*	**0.2**	0.0
Education (Ref: Compulsory)									
Upper secondary	**-8.1**	0.7	**7.5**	**-5.4**	*2.5*	**2.8**	4.2	-0.4	-3.8
Post-secondary	**-8.8**	-0.3	**9.0**	**-6.9**	2.6	*4.2*	-5.6	-0.5	6.1
1^{st} stage tertiary	**-11.0**	1.2	**9.8**	-4.4	-0.5	4.9	2.3	-5.2	2.9
2^{nd} stage tertiary	-4.1	-3.8	*7.9*	1.4	1.8	-3.2	-7.2	7.3	-0.2
Country of birth (Ref: AUT)									
EU15	-4.1	0.2	3.9	-2.8	**-8.5**	**11.3**	-0.4	0.6	-0.2
CEE	-2.7	0.4	2.3	0.4	-1.3	0.9	10.5	-7.4	-3.1
Turkey	*6.6*	-1.0	-5.6	4.1	-0.8	-3.2	-4.7	0.4	4.4
Yugosphere	**5.1**	**-5.1**	-0.1	4.9	-6.1	1.2	*6.7*	*-5.9*	-0.8
other	**7.9**	**-5.2**	-2.7	0.1	2.4	-2.6	-1.5	-1.1	2.5
Health status (Ref: Very bad)									
Bad	-5.1	2.5	2.7	-2.5	*5.7*	-3.2	10.7	0.5	-11.2
Fair	-4.2	0.1	4.1	-2.7	**7.8**	*-5.1*	2.9	4.8	-7.7
Good	-8.0	1.7	6.3	-4.3	**9.9**	*-5.6*	6.8	1.7	-8.5
Very good	*-12.9*	1.8	*11.1*	**-8.8**	**10.8**	-2.0	5.1	5.0	-10.1
Degree of urbanisation (Ref: < 10 000 inhabitants)									
> 10000 & < 100000	-1.6	*2.1*	-0.5	*-4.1*	*3.5*	0.5	-3.7	1.6	2.0
>100 000 inhabitants	**-4.6**	1.6	**3.1**	**-4.8**	*3.1*	1.7	**-8.7**	3.1	*5.6*
Changes in employment status (Ref: None)									
Once	*4.6*	*-4.4*	-0.2	**30.8**	**-24.0**	**-6.8**	8.6	-5.3	-3.2
Twice	*4.7*	-0.7	-4.0	**30.5**	**-29.1**	-1.4	9.2	*-6.2*	-3.0
Thrice or more	-0.6	-6.3	*6.9*	15.7	-9.7	-6.1	*13.1*	*-8.1*	-5.0
Months in corresponding employment status (Ref: 12 months)									
<6 months	**27.4**	**-8.1**	**-19.3**	4.0	*-6.5*	2.5	**40.8**	3.5	**-44.3**
6-8 months	**11.7**	*-5.6*	**-6.1**	-9.1	2.2	6.9	**16.4**	3.4	**-19.8**
9-11 months	3.5	**-6.6**	3.1	4.0	-12.0	8.0	*10.7*	3.6	*-14.4*
(iii) Survey setting									
Mode of interview (Ref: CAPI)									
CATI	-2.2	**2.9**	-0.7	-2.1	2.8	-0.7	-5.6	*7.5*	-1.8
Interviewer									
Same interviewer	0.9	1.1	-2.0	0.1	1.6	-1.7	*-8.4*	6.9	1.5
Type of interview (Ref: Personal)									
Proxy	**7.9**	**-5.8**	*-2.2*	*3.8*	**-5.6**	1.8	0.3	-0.7	0.4
Month of interview (Ref: March to May)									
June to Aug.	**2.3**	**-2.9**	0.6	1.3	*-2.5*	1.2	-0.2	0.2	0.0
Sept. to Nov.	2.2	**-5.7**	*3.5*	**8.6**	**-9.6**	1.0	0.2	0.6	-0.8
Year of interview (Ref: 2008)									
2009	-0.2	-0.8	1.0	0.3	-2.7	2.4	-2.3	-1.7	4.0
2010	0.5	0.1	-0.6	2.8	*-3.7*	0.9	2.1	3.6	-5.7
2011	-0.7	-0.2	0.9	**-30.4**	**46.9**	**-16.5**	1.1	1.5	-2.6
(iv) Learning effect									
Wave of interview (Ref: 1^{st})									
2^{nd}	-1.4	0.4	1.0	*3.0*	-3.1	0.1	*7.4*	-5.8	-1.5
3^{rd}	-1.8	0.8	1.0	2.0	*-4.5*	2.5	4.2	*-5.9*	1.8
4^{th}	-2.8	2.0	0.7	2.1	-2.3	0.2	4.0	-1.7	-2.3

Notes: This table shows the estimated average marginal effects of multinomial regressions per income type (wages, pensions and unemployment benefits) in three categories: (1) S<A: reported income is below the value in administrative data (2) S=A: survey income corresponds to administrative data. (3) S>A: reported income is above the value in administrative data. Estimates with an associated p–value below 0.01 are depicted in bold, values below 0.05 in italics. Insignificant estimates (p–value above 0.05) are given in gray. Source: SILC 2008-2011, own calculations.

Being born in a CEE[5] country does not make a significant difference for the likelihood of misreporting wages compared to Austria, whereas we find a higher probability of underreporting wages for the Yugosphere[6], Turkey and other countries of birth (and a corresponding lower propensity to report matching wages). For the remaining two income components, the country of birth is less relevant with two exceptions: being born in the EU15 significantly raises the likelihood of overreporting pensions by 11%-points. Furthermore, natives from the Yugosphere bear a higher probability of underreporting unemployment benefits. Summing up, the evidence hints to remaining problems of correctly understanding the data collection process by non-natives but this is primarily an issue for the underreporting of wages.

Health problems are expected to hinder income reporting. The strongest evidence for a decreasing probability of misreporting with improved health is found for pensions. Related to that, respondents tend to provide correct pension incomes the healthier they are. Better health also reduces the likelihood of underreporting wages but this is only significant when we compare those in *very good* health to those with *very bad* health. However, this pattern is completely reversed for overreporting wages. Moreover, there is no evidence for a relationship between health and the misreporting of unemployment benefits.

For all three types of income, underreporting is less prevalent, if the respondent's place of residence is in a highly urbanised region. In the case of overreporting, this relation is insignificant for pensions and inverted for wages and unemployment benefits. As we expected *any* kind of misreporting to be reduced with rising degree of urbanisation, the latter result could e.g. be due to differences in types of jobs and the associated wage structures in urbanised regions.

As an indicator for income stability, we include variables capturing the number of status changes and the number of months spent in a specific employment status. As expected, Table 5 exhibits that status changes correspond to higher probabilities of misreporting and to lower chances to report correct incomes. In general, the effects show significantly higher probabilities for underreporting, particularly for pensions and unemployment benefits. Furthermore, we find that overreporting unemployment benefits is 41%-points more likely if the spell was shorter than 6 months and 16%-points more likely if the spell lasted between 6 and 8 months during the income reference period. An explanation for this finding could be that particularly short spells of unemployment are associated with recall and telescoping errors. There are no significant average marginal effects of the number of months on the probability of a match between the two data sources. Overreporting unemployment benefits is 44%-points less likely for short-term recipients than for long-term recipients. This general pattern is similar for underreported wages but almost nonexistent for pension incomes.

Survey setting. Against what we have expected, average marginal effects of telephone interviews (CATI) on both types of misreporting are generally negative, however, statistically not significant. For wages, responses from CATI are 3%-points more likely to match administrative records than from CAPI. Proxy interviews increase the probability of underreporting wages and pensions. For wages, there is also a marginally negative effect on overreporting. Proxy interviews neither have an effect on overreporting of pensions nor on any type of misreporting of unemployment benefits. Moreover, having the same interviewer as in the previous year does not have a major

[5] Bulgaria, Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, and Slovenia

[6] Bosnia and Herzegovina, Croatia, Serbia and Montenegro, and Macedonia

impact on misreporting. Only unemployment benefits are slightly less likely to be underreported.

Looking at the month of the interview, we find that reporting errors for wages are more likely, the later in the year the interview is conducted, i.e. the larger the time span between the income reference period and the interview. For pensions, however, this applies only to more distant interview months (>8 months) and is only statistically significant for underreporting. For unemployment benefits, there is no significant relationship at all. Finally, the interview year generally does not play any significant role. The large and significant time effects for pensions in 2011 partly appear by design because (only) for this year, Statistics Austria already derived a great share of pension incomes from registers before they back-calculated the Austrian SILC.

Learning effect. As already indicated by the descriptive illustration in figure 2, the multinomial logit model based on pooled cross-sections does not yield strong evidence for learning effects. In fact, we find slightly lower probabilities to provide correct answers for pensions and unemployment benefits for respondents in advanced survey waves. In the next step, we will apply panel regressions which additionally purge unobservable individual fixed effects, for this question and test whether this preliminary result holds.

To sum up the results of the logit models, the average marginal effects suggest that the income level and socio-demographic characteristics are far more relevant than the interview context for explaining reporting errors. Relative importance, however, will be investigated more systematically further below. Moreover, the general patterns found for social desirability (income percentiles) and for gender, age, education among the sociodemographic factors are quite similar for all three income types. Concerning the survey setting, effects for wages and pensions are close, whereas unemployment benefits are hardly influenced by this set of variables.

4.2 Extent of misreporting

In a next step, we employ individual fixed effects OLS panel regressions to gain further insights on the presence of social desirability bias and learning effects. The results of the panel estimation for the two subjects are displayed in tables 6 and 7, respectively. Although we present only the estimated coefficients relevant for these two errors sources, the estimations were done using the full set of controls, corresponding to table 5. The remainder of the estimates is contained in tables A.2 and A.3 in the Appendix.

Social desirability. The dependent variable in the panel regression focusing on the evaluation of social desirability is the negative or positive absolute difference between the survey report and the administrative record. We find that the difference between questionnaire and register wages increases with rising distance from the middle of the distribution. This effect is more pronounced at higher percentiles as compared to lower quantiles. For instance, in the lowest decile the average overreporting is 7,400 Euros above the error in the 5^{th} decile, whereas the difference amounts to 11,600 Euros in the top decile. A similar pattern is present for differences in pensions, although only statistically significant at both ends of the distribution. In the case of unemployment benefits, point estimates roughly indicate a mean-reverting pattern in the longitudinal perspective. Only some quantiles show statistically significant differences from the average error in the 5^{th} decile and indicate underreporting at the top.

Table 6: Panel regression results – Social desirability

	Wages	Pensions	Unempl. Benefits
(i) Social desirability			
Relative income (Ref: 5. Decile)			
1. Decile	**7387.25** (1020.64)	**3536.36** (764.73)	192.52 (1048.41)
2. Decile	**5765.24** (776.71)	**1775.13** (684.73)	516.38 (441.90)
3. Decile	**3524.27** (698.17)	**1098.89** (471.01)	−33.88 (834.57)
4. Decile	**1631.32** (362.68)	−136.74 (372.11)	182.32 (471.82)
6. Decile	**−1449.64** (509.38)	−349.92 (330.50)	−625.39 (661.51)
7. Decile	**−2972.27** (653.50)	−1389.06 (982.90)	−799.96 (937.23)
8. Decile	**−4425.72** (841.99)	−2143.49 (1184.84)	**−1387.76** (701.87)
9. Decile	**−8118.60** (1319.65)	**−3420.77** (773.14)	**−1554.50** (705.41)
10. Decile	**−11598.02** (1696.99)	**−6968.90** (1769.55)	**−2839.05** (831.64)
Other controls			
Individual fixed effects	yes	yes	yes
(ii) Sociodemographic characteristics	yes	yes	yes
(iii) Survey setting	yes	yes	yes
(iv) Learning effect	yes	yes	yes
Num. obs.	20372	10105	2470
R^2 (full model)	0.70	0.72	0.81
Adj. R^2 (full model)	0.32	0.41	0.30

Notes: This table shows the results of unbalanced panel regressions with positive and negative reporting errors as dependent variable (i.e. negative values correspond to under-reporting and positive values to over-reporting). Cluster robust standard errors are given in parentheses. Gender, education and country of birth have been removed from the baseline specification due to negligible within-variation. See full list of estimated coefficients in Table A.2. Source: SILC 2008-2011, own calculations.

Table 7: Panel regression results – Learning effect

	Wages	Pensions	Unempl. Benefits
(iv) Learning effect			
Wave of interview (Ref: 1^{st})			
2^{nd}	−164.75 (631.32)	**−1622.45** (264.10)	191.71 (300.39)
3^{rd}	−353.89 (770.74)	**−2546.06** (520.99)	231.58 (433.76)
4^{th}	−478.33 (959.25)	**−3397.04** (821.01)	144.55 (568.08)
Other controls			
Individual fixed effects	yes	yes	yes
(i) Social desirability	yes	yes	yes
(ii) Sociodemographic characteristics	yes	yes	yes
(iii) Survey setting	yes	yes	yes
Num. obs.	20372	10105	2470
R^2 (full model)	0.74	0.71	0.82
Adj. R^2 (full model)	0.40	0.39	0.31

Notes: This table shows the results of unbalanced panel regressions with absolute values of the reporting errors as dependent variable (i.e. negative values (under-reporting) have been multiplied by −1). Cluster robust standard errors are given in parentheses. Gender, education and country of birth have been removed from the baseline specification due to negligible within-variation. See full list of estimated coefficients in Table A.3. Source: SILC 2008-2011, own calculations.

Learning effect. Table 7 presents additional evidence on the presence of learning effects. Note that we define the learning effect as a decline in the absolute reporting errors over multiple survey waves. In this case, we do not distinguish between over- and underreporting and thus negative values of the dependent variable are multiplied by minus one. We find mixed evidence for learning effects, which crucially depend on the type of income under consideration. For wages and unemployment benefits, there is no statistically significant reduction of reporting errors with increasing panel duration, whereas such a pattern clearly emerges for pension incomes. For wages, our results are in line with previous literature finding a positive, although not statistically significant, serial correlation of misreporting.

Robustness checks. We applied two robustness checks for our baseline panel specification: (1) estimations based on the 4-wave balanced sample using longitudinal weights (see tables A.4 & A.5), (2) specifications with the difference in log income between register and survey data as dependent variable (see tables A.6 & A.7). The panel regressions for the balanced sample broadly confirm the social desirability effects although with slightly smaller point estimates. Surprisingly, the already small learning effect for pension income disappears completely in the balanced 4-wave panel. In the second check, we test whether our conclusions remain valid when studying the *relative* deviations of survey answers from administrative records. Overall, the models with the difference of the natural logs of incomes have a higher model fit. Mean-reverting errors found in the baseline model are observed again and have a similar pattern of statistical significance. Again, learning effects are only present for pensions where the model predicts a reduction of the difference between survey and register data of approximately 19% in the last panel wave.

4.3 Relative importance of error sources

Finally, we apply an Owen value decomposition in order to assess the relative importance of the four error sources under consideration. The decomposition aims to assign a proportion of the explained variance to groups of the explanatory variables. We consider two connected settings. First, reporting errors and income variables enter the regressions transformed via the inverse hyperbolic sine function. This transformation is closely related to the well known logarithmic transformation, however, it is also defined for negative and zero values. In the context of a significant mass of zeros and negative values among the reporting errors, this is a desirable property as it allows us to consider the same number of observations as in the preceding calculations. The results of this exercise are given in table 8.

The first row contains the adjusted R^2 for each cross-sectional model. The total variance explained is clearly highest for reporting mismatch in unemployment benefits, where 46% can be traced back to the model variables. In contrast, for wage and pension differences the corresponding figures amount to 16% and 11% respectively. Noteworthy, especially for unemployment benefits, but also in the case of wages and pensions, the magnitude of explained variance in our regression is comperatively high (Kim and Tamborini, 2012). The remaining rows show the group sums of Owen values as percentage of the overall R^2. While the patterns for wages and pensions are rather similar, unemployment benefits show a quite different picture. For the former two, around 30% of the explained variance can be attributed to the group of socio-demographic variables, whereas social desirability turns out to be of the highest relative importance (around 60%). For wage differences, the survey setting and variables measuring the panel participation (and thus learning effects) virtually do not contribute to the total R^2 at all. Learning effects also play a minor role for pensions whereas the survey setting contributes roughly 15%.

Outcomes are considerably different for unemployment benefits. With a share of 65%, the group of socio-demographic variables is most relevant for overall R^2. Compared to the models for wages and pensions, social desirability is substantially less important whereas the socio-demographic characteristics gain relevance. Thus, misreporting of unemployment benefits does not so much depend on the level of unemployment benefits but rather on socio-demographic characteristics of the recipients and is also more sensitive to the interview context and mode.

Table 8: Decomposition of explained variance – SILC 2008-2011

	Wages	Pensions	Unempl. Benefits
Proportion of variance explained	15.8	11.4	46.4
Relative importance:			
(i) Social desirability	65.1	58.8	23.7
(ii) Sociodemographic characteristics	34.0	24.5	64.5
(iii) Survey setting	0.8	14.5	9.2
(iv) Learning effect	0.0	2.2	2.6

Notes: This table shows the goodness-of-fit of OLS regressions and its decomposition to four error sources, i.e. four groups of explanatory variables. We quantify the relative importance of (i) social desirability, (ii) socio-demographic characteristics, (iii) aspects of the survey design, and (iv) learning effects on the basis of separate regressions for wages, pensions and unemployment benefits. Reporting errors are regressed on the same set of explanatory variables that were used before (see section 3), using all available pooled cross-sections. Error and income variables are transformed via the inverse hyperbolic sine function, which facilitates a log-log interpretation in the context of a significant mass of zeros and negative values among the errors. Source: SILC 2008-2011, own calculations.

Additionally, we repeat the same procedure without transforming the input variables (see table A.8 in the online appendix). Whereas the results on old-age and unemployment benefits are almost identical, the R^2 of the wage regression drops by two thirds. However, our estimates on the relative importance of the four error sources are hardly affected, which strengthens our confidence in the robustness of our findings.

5 Conclusions

Income is very likely one of the most pervasive information in micro datasets, since it plays an essential role for a wide range of welfare indicators and policy questions. The traditional way of collecting income information are household surveys, however, the accuracy of survey data has increasingly been contested during the last years. A main factor behind this critique is the suspected presence of measurement error in surveys, resulting from (un)intentional misreporting. The identification of data errors requires by definition some point of reference. We follow the traditional literature and check survey data against administrative records using a unique dataset: the Austrian 2008-2011 waves of EU-SILC. We make use of the fact that due to a legal initiative, the Austrian SILC provides both survey and register income data for the same observational units for four consecutive years.

While the vast majority of existing research assesses measurement error in income data for US households, there is virtually no research using European panel data for various income types. EU-SILC is a key dataset for social policy issues since it provides the main indicators for evaluating the Europe 2020 strategy. Given its importance as reference source for comparative statistics on income distribution and social inclusion, data quality is a crucial matter. Compared to previous literature, using the Austrian EU-SILC for assessing income measurement error has two main advantages. First, we do not have to fall back on two-sample matching processes since agreement from respondents concerning data linkage was not legally required. This advantage helps to avoid selection bias and, given that the Austrian EU-SILC is representative for the national population residing in private households, ensures high external validity. Second, we are able to evaluate income measurement error for various components of total disposable household income in the very same dataset.

We elaborate four major reasons for misreporting discussed in the literature: social desirability, socio-demographic characteristics of the respondent, specifics of the survey design and the presence of learning effects for three types of personal income (wages, pensions and unemployment benefits).

The main findings are the following. For personal income and in line with the existing literature, statistically significant mean-reverting errors are revealed in both, cross-sectional and panel regression models. We find significantly lower probabilities of underreporting at the bottom of the wage distribution and vice versa, higher likelihoods of overreporting at the top tail. By including a broad range of control variables in order to capture the complexity of the annual income stream, we interpret this result as evidence of social desirability in reporting wages. Although the effects are generally less pronounced and sometimes even statistically insignificant, similar patterns occur for pension income and unemployment benefits.

Concerning socio-demographic characteristics, males are found to have a significantly higher tendency of overreporting wages, pensions, and unemployment benefits. Additionally, there is a significant relationship between health and misreporting, for instance, good health conditions correlate with more correct survey responses particularly for pensions. The higher chances of underreporting for respondents being born outside the EU–15 hint to the presence of language and comprehension problems, despite the fact that interviews were also conducted in other languages if requested by respondents. We find consistent evidence that a complex income and employment context with many changes during the income reference period hinders recall, and thus increases misreporting. Multiple changes in the employment status have a strong effect on reporting errors of pensions, a shorter status duration increases errors particularly for unemployment benefits and wages.

For data producers, our results reveal that notably both wages and unemployment benefits are reported more accurately in telephone interviews than in personal interviews. Proxy interviews are associated with significantly higher probabilities of misreporting. Similarly, the larger the time span between the income reference period and the interview, the more likely is misreporting. Evidence on the presence of a learning effects is mixed and depends crucially on the type of income under consideration.

Finally, an Owen value decomposition reveals that social desirability is a major explanation for misreporting wages and pension income. For unemployment benefits, socio-demographic characteristics of the respondents seem to play the major role for reporting errors. The survey setting is relatively less important for explaining misreporting, while learning effects are hardly noticeable.

For policy makers, our results point to the fact that for some socio-demographic groups — including those most relevant for policy makers— survey income data may potentially be an infirm ground for decision-making. Thus, investments in the development, maintenance, advancement, and accessibility of public administrative data, as an alternative, may pay-off for better targeted polices. Having both survey and register data at hand for the same units opens up several perspectives for further research. For instance, a next step could be to look at the joint distribution of errors for respondents that have more than one of the three income components at once in a given year and check if those who underreport one source are also more likely to underreport another. It could also be worthwhile to replicate our analysis based on gross income values. Another fruitful avenue for further inquiries could be to train a statistical model that allows survey producers to correct the income information based on other unobservable characteristics of the respondents. Moreover, income is a key explanatory variable in social science research. As in Hariri and Lassen (2017), one could therefore check how conclusions for prominent regressions where income is used as right-hand-side variable (health, well-being, etc.) are altered when survey data is substituted by register data.

Acknowledgements

We thank Wilfried Altzinger, Stephen P. Jenkins, Arthur Kennickel and participants of various seminars for helpful comments on earlier versions of the paper. Stefan Humer's research is supported by funds of the Oesterreichische Nationalbank (OeNB Anniversary Fund, project number: 16728).

References

Abowd, J. M. and M. H. Stinson (2013): "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data". In: *Review of Economics and Statistics* 95 (5), pp. 1451–1467.

Alm, J. and B. Torgler (2006): "Culture differences and tax morale in the United States and in Europe". In: *Journal of Economic Psychology* 27 (2), pp. 224–246.

Angel, S., N. Lamei, and R. Heuberger (2017): "Differences between household income from surveys and registers and how these affect the poverty headcount. Evidence from the Austrian SILC". In: *Social Indicators Research (forthcoming)*).

Asamer, E., F. Astleithner, P. Cetkovic, S. Humer, M. Lenk, M. Moser, and H. Rechta (2016): "Quality Assessment for Register-based Statistics - Results for the Austrian Census 2011". In: *Austrian Journal of Statistics* 45 (2), pp. 3–14.

Berka, C., S. Humer, M. Moser, M. Lenk, H. Rechta, and E. Schwerer (2012): "Combination of Evidence from Multiple Administrative Data Sources: Quality Assessment of the Austrian Register-Based Census 2011". In: *Statistica Neerlandica* 66 (1), pp. 18–33.

Bollinger, C. R. (1998): "Measurement Error in the Current Population Survey". In: *Journal of Labor Economics* 16 (3), pp. 576–594.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data". In: *Journal of Labor Economics* 12 (3), pp. 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data". In: *Handbook of Econometrics*. Ed. by J. Heckman and E. Leamer. Vol. 5. Elsevier, pp. 3705–3843.

Bound, J. and A. B. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" In: *Journal of Labor Economics* 9 (1), pp. 1–24.

Einav, L. and J. Levin (2014): "Economics in the age of big data". In: *Science* 346 (6210), p. 715.

Fessler, P., M. Kasy, and P. Lindner (2017): *Survey mode effects on income inequality measurement*. Working Paper 48766. Harvard University.

Hansen, K. and D. Kneale (2013): "Does How You Measure Income Make a Difference to Measuring Poverty? Evidence from the UK". In: *Social Indicators Research* 110 (3), pp. 1119–1140.

Hariri, J. G. and D. D. Lassen (2017): "Income and Outcomes". In: *Public Opinion Quarterly* 81 (2), pp. 564–576.

Hassan, M. and F. Schneider (2016): "Size and Development of the Shadow Economies of 157 Worldwide Countries: Updated and New Measures from 1999 to 2013". In: *Journal of Global Economics* 4 (3), pp. 1–14.

Hsiao, C. (2014): *Analysis of Panel Data*. 3rd ed. Econometric Society Monographs. Cambridge University Press.

Huettner Frank; Sunder, M. (2012): "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values." In: *Electronic Journal of Statistics* 6), pp. 1239–1250.

Jenkins, S. P., L. Cappellari, P. Lynn, and E. Sala (2006): "Patterns of Consent: Evidence from a General Household Survey". In: *Journal of the Royal Statistical Society. Series A* 169 (4), pp. 701–722.

Jenkins, S. P., A. J. Lynn, and E. Sala (2008): "The Feasibility of Linking Household Survey and Administrative Record Data: New Evidence for Britain". In: *International Journal of Social Reserach Methodology* 11 (1), pp. 29–43.

Kapteyn, A. and J. Y. Ypma (2007): "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data". In: *Journal of Labor Economics* 25 (3), pp. 513–551.

Kim, C. H. and C. Tamborini (2012): "Do Survey Estimate Earnings Inequality Correctly? Measurement Errors Among Black and White Male Workers". In: *Social Forces* 90 (4), pp. 1157–1181.

Kim, C. and C. R. Tamborini (2014): "Response Error in Earnings An Analysis of the Survey of Income and Program Participation Matched With Administrative Data". In: *Sociological Methods & Research* 43 (1), pp. 39–72.

Kreiner, C. T., D. D. Lassen, and S. Leth-Petersen (2015): "Measuring the Accuracy of Survey Responses using Administrative Register Data: Evidence from Denmark". In: *Improving the Measurement of Consumer Expenditures*. Ed. by C. D. Carroll, T. F. Crossley, and J. Sabelhaus. National Bureau of Economic Research Studies in Income and Wealth 74.

Kreuter, F., G. Muller, and M. Trappmann (2010): "Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data". In: *Public Opinion Quarterly* 74 (5), pp. 880–906.

Lohmann, H. (2011): "Comparability of EU-SILC Survey and Register Data: The Relationship among Employment, Earnings and Poverty". In: *Journal of European Social Policy* 21 (1), pp. 37–54.

Lusardi, A. and O. S. Mitchell (2014): "The Economic Importance of Financial Literacy: Theory and Evidence". In: *Journal of Economic Literature* 52 (1), pp. 5–44. URL: http://www.aeaweb.org/articles?id=10.1257/jel.52.1.5.

Lynn, P., S. P. Jenkins, and E. Sala (2012): "The Impact of Questioning Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study". In: *Journal of the Royal Statistical Society. Series A* 175 (1), pp. 289–308.

Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. 1st ed. Econometrics Society Monographs. Cambridge University Press.

Mellow, W. and H. Sider (1983): "Accuracy of Response in Labor Market Surveys: Evidence and Implications." In: *Journal of Labor Economics* 1), pp. 331–344.

Meyer, B. D., K. M. Wallace, and J. Sullivan (2015): "Household Surveys in Crisis". In: *Journal of Economic Perspectives* 29 (4), pp. 199–226.

Micklewright, J. and S. V. Schnepf (2010): "How Reliable are Income Data Collected with a Single Question?" In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173 (2), pp. 409–429.

Moore, J. C., L. L. Stinson, and E. J. Welniak Jr. (2000): "Income Measurement Error in Surveys: A Review". In: *Journal of Official Statistics* 16 (4), pp. 331–362.

Paulus, A. (2015): *Tax evasion and measurement error: An econometric analysis of survey data linked with tax data*. Technical Paper ISER Working Paper Series 2015-10. Institute for Social and Economic Research (ISER), University of Essex.

Pedace, R. and N. Bates (2000): "Using Administrative Data to Assess Earnings Reporting Error in the Survey of Income and Program Participation". In: *Journal of Economic and Social Measurement* 26 (3,4), pp. 173–192.

Pforr, K. (2014): "femlogit - Implemenation of the multinomial logit model with fixed effects". In: *The Stata Journal* 14 (4), pp. 847–862.

Pischke, J.-S. (1995): "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study". In: *Journal of Business & Economic Statistics* 13 (3), pp. 305–314.

Reynolds, J. and J. Wenger (2012): "He Said, She Said: The Gender Wage Gap According To Self and Proxy Reports in the Current Population Survey." In: *Social Science Research* 41 (2), pp. 392–411.

Sakshaug, J. W. and S. Eckman (2016): "Are Survey Nonrespondents Willing to Provide Consent to use Administrative Records?" In: *Public Opinion Quarterly* 81 (2), pp. 495–522.

Sakshaug, J. W., M. P. K Couper, M. B. Ofstedal, and D. R. Weir (2012): "Linking Survey and Administrative Records: Mechanisms of Consent". In: *Sociological Methods and Reserach* 41 (4), pp. 535–569.

Schnetzer, M., F. Astleithner, P. Cetkovic, S. Humer, M. Lenk, and M. Moser (2015): "Quality Assessment of Imputations in Administrative Data". In: *Journal of Official Statistics* 31 (2), pp. 231–247.

Statistics Austria (2014a): *Methodenbericht zur Rückrechnung von EU-SILC 2008-2011 auf Basis von Verwaltungsdaten*. Tech. rep. Statistics Austria.

— (2014b): *Standard-Dokumentation Metainformationen. (Definitionen, Erläuterungen, Methoden, Qualität) zu EU-SILC 2011*. Tech. rep. Statistics Austria.

Tamborini, C. R. and C. Kim (2013): "Are Proxy Interviews Associated with Biased Earnings Reports? Marital Status and Gender Effects of Proxy". In: *Social Science Research* 42 (2), pp. 499–512.

Tourangeau, R., N. M. Bradburn, P. V. Marsden, and J. D. Wright (2010): "The Psychology of Survey Response". In: *Handbook of Survey Research* 2), pp. 315–346.

Tourangeau, R., L. J. Rips, and K. Rasinski (2000): *The Psychology of Survey Response*. Cambridge University Press.