ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2019)
182, *Part* 4, *pp.* 1411–1437

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

Stefan Angel, Franziska Disslbacher and Stefan Humer

Vienna University of Economics and Business, Austria

and Matthias Schnetzer

Federal Chamber of Labor, Vienna, Austria

[Received February 2018. Revised March 2019]

Summary. The paper analyses the sources of income measurement error in surveys with a unique data set. We use the Austrian 2008–2011 waves of the European Union 'Statistics on income and living conditions' survey 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 social desirability, sociodemographic characteristics of the respondent, the survey design and the presence of learning effects. We find strong evidence for a social desirability bias in income reporting, whereas the presence of learning effects is mixed and depends on the type of income under consideration. An Owen value decomposition reveals that social desirability is a major explanation of misreporting in wages and pensions, whereas sociodemographic characteristics are most relevant for mismatches in unemployment benefits.

Keywords: Income measurement error; Register data; Response error; Survey data; Survey error; Survey methods

1. Introduction

The rapidly increasing availability of survey and administrative microdata has created immense possibilities for contemporary empirical economics. However, this progress has also raised questions of data quality and validation to ensure the reliability and accuracy of information. In the literature, the analysis of differences between responses in traditional surveys and administrative records has been a fruitful method to assess quality of microdata. Particular attention has been paid to income data since it is essential 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 sources of data have their idiosyncratic merits and drawbacks. Although 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 on what is the best way to collect income data at the microlevel (Hansen and Kneale, 2013).

Potential differences between survey responses and administrative records have already been

Address for correspondence: Stefan Angel, Department of Socioeconomics, Vienna University of Economics and Business, Welthandelsplatz 1, Vienna 1020, Austria.
E-mail: Stefan.Angel@wu.ac.at

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 validation studies, administrative records are frequently used as benchmarks even though this assumption has also been relaxed to a varying extent in the recent literature (Kapteyn and Ypma, 2007; Abowd and Stinson, 2013; Bingley and Martinello, 2017). The choice of the benchmark data reflects the researchers' confidence in the accuracy of a specific data set. We rely on evidence that the quality of Austrian administrative data is very high (Asamer *et al.*, 2016; Schnetzer *et al.*, 2015; Berka *et al.*, 2012).

This paper focuses on two research 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 *et al.*, 2001; Tourangeau *et al.*, 2000; Kapteyn and Ypma, 2007; Abowd and Stinson, 2013; Paulus, 2015). Specifically, we focus on four sources of error emphasized in the literature. These are

(a) the presence of a social desirability bias in survey responses,
(b) specific sociodemographic characteristics of the respondent,
(c) the survey design and
(d) 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 sources of error 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, the consent of individuals to match survey and register data is usually needed. Since participation in the matched sample often is voluntary, the sample is biased towards individuals giving more accurate responses (Bollinger, 1998). Briefly speaking, there is evidence of a consent bias for matched sources of data. Furthermore, samples based on optional matching are often found to be non-representative for the whole population (Jenkins *et al.*, 2006, 2008; Sakshaug and Eckman, 2012, 2016).

Fortunately, the Austrian 2008–2011 waves of the European Union 'Statistics on income and living conditions' (EU SILC) survey do not suffer from such shortcomings. In this data set, the survey responses are linked directly to administrative records by the national statistical institute via unique personal identifiers. Compared with 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 data set 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 can compare survey and administrative data within one data set, which is a considerable advantage compared with most existing research. Even the most prominent studies about income measurement error are based on sophisticated statistical matching procedures of survey and administrative data sets. Other validation studies are restricted to individuals working at a single company (Pischke, 1995). The data set that is used in this paper is not subject to any of these limitations. Furthermore, in contrast with papers studying measurement errors in total household income, we can analyse different single income components including unemployment benefits, which has rarely been done yet.

In summary, we contribute to the literature in multiple ways. We exploit a unique data set providing both administrative records and survey responses to analyse differences in income reports. We can conduct a comparative analysis of three different concepts of income with a panel structure over the period of 4 years. Contrary to other studies that are limited to single companies, our analysis is representative of the whole population. Finally, we apply an Owen value decomposition to assess quantitatively the importance of four specific reasons for misreporting.

The remainder of this paper is structured as follows. Section 2 provides a description of potential sources of measurement errors that have been elaborated in the existing literature. In Section 3, we discuss the specifics of our EU SILC data set 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 sources of error (a)–(d) in Section 4. For a more detailed analysis of sources of error (a) and (d), we conduct panel regressions to detect changes of response behaviour over time. We conclude our analysis quantifying the relative importance of the four sources error of on the basis of an Owen value decomposition. Finally, Section 5 provides a summary of our results.

2. Reasons for measurement error in survey income data

Although 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 recent 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 are, 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 that are 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 *et al.*, 2010). The theoretical and empirical literature on income measurement errors emphasizes four main reasons for misreporting:

(a) social desirability,
(b) sociodemographic characteristics of the respondent,
(c) specifics of the survey design and
(d) the presence of learning effects.

However, the existing literature has ignored that those four reasons for misreporting can vary with the type of income under consideration.

For instance, Angel *et al.* (2018) analysed reporting errors for total disposable household income in Austria, which is not observed directly but aggregated *ex post* on the basis of 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 type of income. In what follows, we discuss how reasons (a)–(d) can result in overreporting or underreporting of wages and, then, why reasons (a)–(d) might lead to different expectations for the misreporting of pensions and unemployment benefits, which are the two other income components that are considered in this paper.

First, social desirability bias is probably the most important reason for income misreporting in surveys. Because of 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 and underreporting for socially desirable and socially undesirable behaviour, attitudes and characteristics respectively (Bound *et al.*, 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, whereas 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 microlevel 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 *et al.*, 1994; Bound and Krueger, 1991). As an exception, Hariri and Lassen (2017) found 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 sociodemographic characteristics of the respondent (Kreiner *et al.*, 2015; Kim and Tamborini, 2014; Bound *et al.*, 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 *et al.* (1994) found 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 that is 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, 2012, 2014; Bound and Krueger, 1991). Two additional sociodemographic characteristics potentially contributing to the accuracy of responses are the number of changes in the employment status during the income reference period, which is 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 employment status can be associated with telescoping errors, which refer to misplacing the receipt of a particular source of income 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 employment status increase reporting errors whereas longer periods of employment lead to less misreporting, since it is easier to recall the remuneration. Kim and Tamborini (2014) found that occupation or industry switchers are more prone to misreport earnings whereas Bound *et al.* (1994) documented 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 urbanization 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 urbanization 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 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 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 (computer-assisted telephone interviewing (CATI)) are considered to be more susceptible to misreporting of wages than a face-to-face setting (computer-assisted personal interviewing (CAPI)). Fessler *et al.* (2017) found that households that are 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. Whereas 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 hypothesize 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.

Whereas some variables might have similar effects on wages, pensions and unemployment benefits, we expect diverging effects for others. For instance, wages (and pensions) are attached to the labour market and tied to (past) individual effort, whereas unemployment benefits are often stigmatized as charity despite actually being an insurance. The socially desirable behaviour is thus to downplay the level received leading to a general underreporting of unemployment benefits. Regarding the sociodemographic characteristics of the respondent, we expect males to underreport unemployment benefits more often and stronger than females since receiving benefits contrasts the common male breadwinner model. Being unemployed might be associated with a higher social stigma for the better educated. Therefore, we expect underreporting or overreporting of unemployment benefits to be respectively an increasing or decreasing function of education. Although 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 who are born abroad might be particularly prone to downplay the level of received state transfers.

With respect to the number of changes in employment status, some particularities must be mentioned. Typically, respondents retire only once and thus might be better informed about the level of pension that they will receive. Precisely for this reason, we expect to find less misreporting

Table 1. Expected effects of misreporting by income types†

	Expectations for wages		*Expectations for pensions*		*Expectations for unemployment benefits*	
	$S<A$	$S>A$	$S<A$	$S>A$	$S<A$	$S>A$
(a) Social desirability						
Income decile (*increasing*)	+	−	+	−	+	−
(b) Sociodemographic characteristics						
Gender (*reference: female*)	−	+	−	+	+	−
Education (*reference: compulsory school*)	−	−	−	−	−	−
Age (*increasing*)	−	−	−	−	−	−
Country of birth (*reference: Austria*)	+	+	+	+	+	−
Health status (*reference: very bad*)	−	−	−	−	+	+
Degree of urbanization (*reference: <10000 inhabitants*)	−	−	−	−	−	−
Changes in employment status (*reference: none*)	+	+	−	−	+	+
Months in corresponding employment status (*reference: 12 months*)	+	+	+	+	∼	∼
(c) Survey setting						
Mode of interview (*reference: CAPI*)	+	+	+	+	+	+
Type of interview (*reference: personal*)	+	+	+	+	+	+
Month of interview (*reference: March–May*)	+	+	+	+	+	+
(d) Learning effect						
Wave of interview (*reference: first*)	−	−	−	−	−	−

†The table summarizes our expectations about the likelihood to overreport and underreport income conditionally on different sources of errors. $S<A$, underreporting in the survey; $S>A$, overreporting in the survey; +, an increasing probability of falling in that specific reporting category; −, decreasing probability; ∼, an ambiguous relationship.

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 spells of unemployment might be associated with less misreporting since respondents are better informed about the actual transfer due to the singularity of the situation. However, respondents might care less about the level of received unemployment benefits, the shorter the time that is spent in unemployment.

Table 1 summarizes our expectations regarding the direction of the effect of the relevant variables that are related to social desirability, sociodemographic characteristics, the survey setting and the learning effect, for overreporting 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 survey, which 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 years or older and is conducted partly with CAPI and partly with 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 utilize Austrian EU SILC data is a unique feature compared with 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 could merge administrative income data with 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 the EU SILC panel enables us 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, and the 2006 and 2010 cohorts in two waves. In contrast with Eurostat's EU SILC user database, the Austrian national SILC data set provided by Statistics Austria delivers cross-sectional and longitudinal information in one single database. In this integrated data set, there are one cross-sectional and three longitudinal weights for those households in the rotational sample that have already been interviewed repeatedly. By virtue of permanent household and individual identifiers, we can track the changes in household responses compared with changes in the administrative records over time. It should be noted in this respect that, in contrast with the first-time interrogation, follow-up interviews were predominantly accomplished via telephone (CATI).

Survey income data are collected retrospectively in the EU SILC panel and correspond to the calendar year before 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 that case, the specific income is estimated on the basis of 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. Although the amount of imputations is generally rather low in the EU SILC data, 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, 2014a). 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, pensions and unemployment benefits for which both survey responses and administrative data are available in the EU SILC panel. Table 2 provides descriptive statistics for the number of observations, the interview mode and the share of imputations for the selected income types.

Register information for the types of income that is used in our study is obtained from various administrative sources whereof the most important are the Austrian social security database, the wage tax register and the transfers data set by Public Employment Service Austria. We briefly describe these data sets below whereas a detailed documentation of administrative sources that are used in the EU SILC panel is presented in the on-line data appendix A and in Statistics Austria (2014b). The Austrian social security database provides the social security status, e.g. whether individuals are employees, pensioners or eligible for unemployment benefits. The wage tax register contains information on all taxable earnings of employees and pensioners. This

Table 2. Observations, interview mode and share of imputations—SILC 2008–2011†

Year	*Number of observations*		*Share of interview mode (%)*		*Imputation share for wages (%)*		*Imputation share for pensions (%)*		*Imputation share for unemployment benefits (%)*	
	Households	*Persons*	*CAPI*	*CATI*	*Survey*	*Administrative*	*Survey*	*Administrative*	*Survey*	*Administrative*
2008	5707	10946	70.0	30.0	4.6	2.6	1.5	0.5	0.2	0.8
2009	5878	11056	57.1	42.9	5.9	1.9	2.0	0.3	0.2	0.8
2010	6188	11493	59.6	40.4	4.8	2.4	1.5	0.4	0.3	0.9
2011	6187	11475	57.3	42.7	3.4	2.6	0.5	0.5	0.3	1.1

†The six rightmost columns show the share of imputed values for selected income types. Source, SILC 2008–2011; own calculations.

data set includes wages, public pensions (retirement benefits), paid maternity leave and sickness benefits. The transfers data set contains the beginning and ending date of spells of unemployment and the respective benefits on a daily basis.

In Austria, employers must report wages directly to the tax authorities, retain payroll taxes and social security contributions, and transfer the residual net income to the employees. Since this is a standardized electronic procedure and there are regular tax inspections, the quality of the data is significantly higher than of self-reported income tax returns. Concerning flaws in administrative data due to tax avoidance, there is empirical evidence that this is comparatively less of an issue in Austria than in other countries (Alm and Torgler, 2006; Hassan and Schneider, 2016). Information on pensions and unemployment benefits is provided straight by public authorities and correspond to actual payments to the entitled individuals. The probability of measurement errors in this data set is very low since reports are effectively linked to payments that are also administrated by the data holders. All in all, recently carried out quality reports on Austrian register data certify high confidence in its reliability (Asamer *et al.*, 2016; Statistics Austria, 2014b).

The merging process between survey and administrative data is accomplished reliably with a branch-specific personal identification number (PIN) for official statistics which serves as a unique identifier in both sources of data. These 172-digit PINs were introduced to protect privacy in the communication between public authorities. The PINs are created by the Austrian Data Protection Commission and are used to identify individuals in the EU SILC survey and in 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 2008–2011. 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, between 95.6% (2008) and 99.4% (2011) of the respondents in the EU SILC survey could be identified with a PIN to assign the register information (Statistics Austria, 2014b). The residual population in the EU SILC survey could not be found in the administrative sources and thus no PIN was available. These individuals are mainly younger than 40 years old, non-Austrian citizens and not registered at their main residence. Another major advantage of the accurate linkage is that the income reference periods for the survey and the administrative records overlap exactly and no adjustments to ensure comparability between the sources of data had to be made.

Reporting income is a two-stage process in the EU SILC survey. At the first stage, respondents must indicate whether a certain income component was received during the reference period. Only in a second step must the amount of income received from a particular source be reported.

Table 3. Reporting of income types—SILC 2008–2011†

Year	*Results for wages (%)*			*Results for pensions (%)*			*Results for unemployment benefits (%)*		
	Survey	*Administrative*	*Difference*	*Survey*	*Administrative*	*Difference*	*Survey*	*Administrative*	*Difference*
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

†The table shows the share of respondents reporting a specific type of income. Source, SILC 2008–2011; own calculations.

Consequently, a mismatch between survey and register data can result at either stage. Table 3 shows the percentage of respondents reporting the three income types involved. Wages are consistently underreported in the survey data by 2.4–3.2 percentage points. By contrast, the number of individuals reporting pension income is slightly higher in the survey data compared with 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 note a prevailing underreporting of receipt of income 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 ($S \times 0$) or, vice versa, not report a particular type of income even though it was received according to register data ($0 \times A$). 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 (0×0). 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 to fulfil our operational definition of correct answers. We define a categorical variable $\Pr(Y_{i,k} = j)$ for mismatch types j, individuals $i = 1, \ldots, N$ and income component $k \in [1, 3]$ as

$$\Pr(Y_{i,k} = j) = \begin{cases} \Pr(Y_{i,k} = 0 \times 0) & \text{if no income in the survey and administrative data,} \\ \Pr(Y_{i,k} = 0 \times A) & \text{if only administrative record income report} \\ & \quad \text{(false negative),} \\ \Pr(Y_{i,k} = S < A) & \text{if a survey underreporter,} \\ \Pr(Y_{i,k} = S = A) & \text{if the survey value corresponds to the} \\ & \quad \text{administrative value,} \\ \Pr(Y_{i,k} = S > A) & \text{if a survey overreporter,} \\ \Pr(Y_{i,k} = S \times 0) & \text{if only survey income report (false positive).} \end{cases}$$

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 types of income, the shares of overreporters and underreporters are very stable

Table 4. Structure of mismatch in income reporting—SILC 2008–2011†

	Results for wages			*Results for pensions*			*Results for unemployment benefits*		
	Observations (%)	*Absolute (P50, €)*	*Relative (P50, %)*	*Observations (%)*	*Absolute (P50, €)*	*Relative (P50, %)*	*Observations (%)*	*Absolute (P50, €)*	*Relative (P50, %)*
2008									
0×0	40.3	0.0	0.0	73.5	0.0	0.0	88.6	0.0	0.0
$0\times A$	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
$S\times 0$	3.0	8459.3	∞	2.4	10907.1	∞	1.2	4200.0	∞
2009									
0×0	39.4	0.0	0.0	73.4	0.0	0.0	88.6	0.0	0.0
$0\times A$	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
$S\times 0$	2.6	7420.0	∞	1.9	10904.8	∞	1.2	4200.0	∞
2010									
0×0	39.3	0.0	0.0	72.5	0.0	0.0	86.7	0.0	0.0
$0\times A$	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
$S\times 0$	2.8	5716.7	∞	2.2	11237.4	∞	0.7	2800.0	∞
2011									
0×0	38.5	0.0	0.0	74.2	0.0	0.0	86.6	0.0	0.0
$0\times A$	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
$S\times 0$	2.5	4419.5	∞	1.2	11540.6	∞	0.6	5220.0	∞

†The 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. 0×0, no income in survey and administrative data; $0\times A$, income only in administrative data; $S<A$, reported income is below the value in administrative data; $S=A$, survey income corresponds to administrative data; $S>A$, reported income is above the value in administrative data; $S\times 0$, only income in survey data. Source, SILC 2008–2011; own calculations.

over the years. There are roughly 17% overreporters and 23% underreporters for wages, which is a high number compared with only 11–13% of respondents with matching values (apart from the 40% with no wage income in both sources of data). 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 overreporters 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

(a) social desirability,
(b) sociodemographic characteristics,
(c) aspects of the survey design and
(d) 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 reasons (a)–(d) for the *direction* of misreporting, we distinguish between those with practically correct information, overreporters and underreporters, and consider both the positive and the negative mismatch. In contrast, when analysing the effect 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 (a)–(d) for misreporting. The presence of (a) 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 (b) the sociodemographic characteristics of the respondent are gender, educational attainment according to the 'International standard classification of education', age and a categorical variable referring to health status. We also include the country of birth, the degree of urbanization 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 employment status during the income reference period should capture the stability of employment. Finally, depending on the type of income that is investigated, we include the number of months being either full- or part-time employed, retired or unemployed. The distribution of overreporters and underreporters across the sociodemographic characteristics is shown in Table B.1 in the on-line supporting information.

The explanatory variables related to (c) the survey setting comprise a dummy variable for the interview mode (CAPI or CATI), a dummy variable related to the reporting status (proxy *versus* self-reported income) and a categorical variable specifying the month of the interview (March–May, June–August or September–November). The motivation for this 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 B.1. The indicator for (d) the learning effect is a dummy variable for the interview wave ranging from 1 to 4. Individuals participating in more than one wave 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 effect of the single reasons on the direction of mismatch and thus the probability to overreport 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 sources of error for misreporting with an Owen decomposition. In what follows, we describe each of the three steps briefly.

3.2.1. Multinomial logit

With a multinomial logit model, we search for factors that help us to understand better why self-reported incomes are above or below their corresponding administrative record values. 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 (0×0) have no mismatch, this group of observations is not of interest for our main research questions and has therefore been discarded from the analysis. Additionally, because of the very low number of observations without income information either in the survey ($0 \times A$) or in the administrative register ($S \times 0$), we restrict our attention to overreporters, underreporters and the consistent group. We provide estimates for linear probability models in Table C.1 in the on-line appendix. For each type of income, three linear probability models were estimated: one for those reporting less ($S<A$), one for those reporting the same ($S=A$) and one for those reporting more ($S>A$) income in the survey. In each model the dependent variable is a dummy variable indicating whether a certain mismatch type is observed.

By estimating a multinomial logit model via maximum likelihood, we explicitly allow the estimated coefficients to vary across the mismatch categories. Thereby, we can identify the determinants of mismatch separately for overreporters and underreporters. This is a considerable advantage compared with ordinary least square (OLS) regression, since it may be very difficult 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 also positive and negative errors could potentially cancel out and leave us with statistically insignificant estimates. In contrast, the effect of for example gender or education on the probability of misreporting is allowed to differ between overreporters 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.

3.2.2. Panel regression

Although we consider an extensive and diverse set of control variables, we cannot rule out that our estimate lacks relevant but unobservable determinants. To check the robustness of our findings, we thus make use of the longitudinal dimension of the EU SILC panel from 2008 to 2011 and employ fixed effect estimates. The dependent variable is the difference between the survey and administrative records for each person-year. 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. In Table B.9 in the on-line supplementary materials we report the number and the proportion of observations in our unbalanced panel sample that jumped over a given number of income deciles. For wages we find that 3.348 individuals (37.9%) experienced a change of at least one decile, in the case of old age benefits 1.174 people (25.2%) and in the case of unemployment benefits 491 (76.8%) units changed at least one decile across the years. We conclude that the observed variation is sufficient to identify the corresponding coefficients. We prefer a linear OLS specification as this provides a clear interpretation of marginal effects on the original scale. In contrast, for longitudinal (non-linear) 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 for example Pforr (2014) for a more detailed discussion).

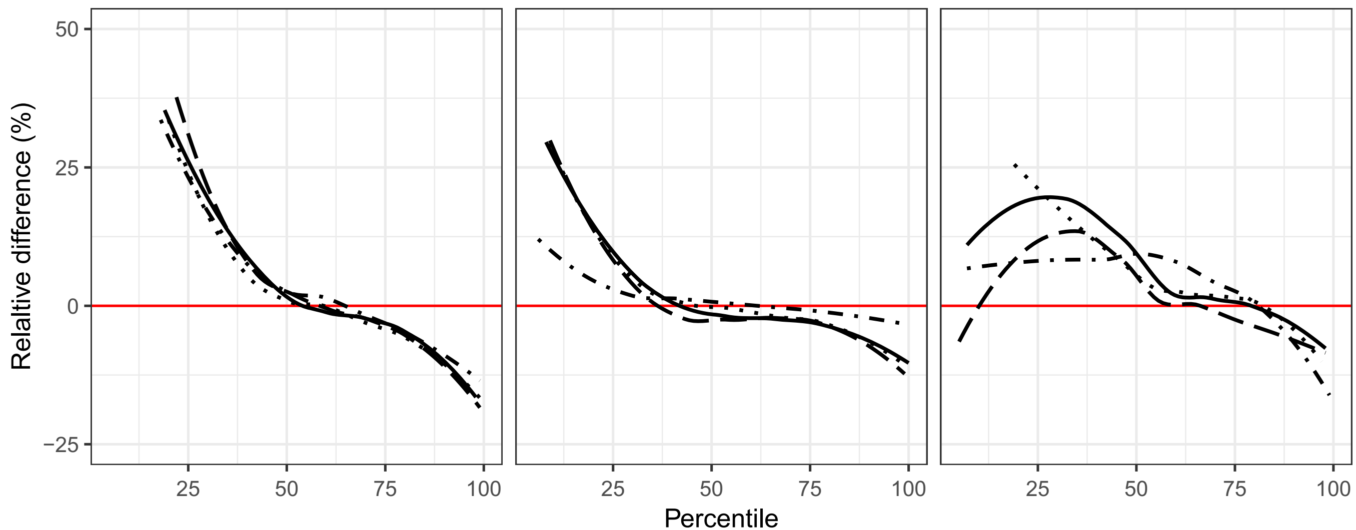
In the first panel specification, our primary interest is the influence of social desirability on the difference between the survey report and the administrative value. We expect a clear pattern across income deciles, with a positive mismatch (i.e. overreporting) in the lower parts of the distribution 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 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 between adults.

3.2.3. Owen value decomposition

Finally, we are interested in the relative importance of error sources (a)–(d) and apply an Owen value decomposition of the explained variance (R^2) in a pooled cross-sectional regression. This procedure enables us to estimate the marginal contribution of each group of explanatory variables to the total R^2. The Owen value decomposition is a generalization 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 and Sunder, 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.



(a) (b) (c)

Fig. 1. Social desirability—SILC 2008–2011: the graphs show the mean relative difference between survey responses and register information on (a) wages, (b) old age and (c) unemployment benefits by percentile of corresponding income distribution and year of the interview (———, 2008; · · · · · ·, 2009; — — —, 2010; ·—·—·, 2011); positive values in the lower half of the distribution imply structural overreporting, whereas negative values above the median indicate underreporting in surveys at the top; source, SILC 2008–2011; own calculations

Fig. 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 type of income report values that are higher than their register record values and vice versa for respondents at the upper part of the distribution. Fig. 1 is cut off between a relative mismatch of +50% and −25%. The pattern is most distinctive for wages and least pronounced for unemployment benefits, but in general corresponds to the expected social desirability behaviour.

Fig. 2 illustrates the learning effect over the four survey waves. The covers depict the absolute logarithmic difference between the two sources of data in the mean, the median, the 25th and the 75th percentile. For wages, we find a slight reduction in the differences for all observed income quantiles after the first wave. A small reduction in error over time is also present in the 25th quantile and the median of unemployment benefits. In contrast, for pension income, the error does not seem to decrease over time. Thus, the data display no or, if anything, a very small learning effect. Fig. B.1 in the on-line appendix replicates Fig. 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.

In what follows, we shall 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. changes in employment status), a significant effect of mean reverting errors would emphasize the role of social desirability as an important source of error.

4.1. Likelihood of reporting more or less

4.1.1. *Social desirability*

For ease of interpretation, Table 5 displays average marginal effects that are derived from the multinomial logit models. For all three types of personal income, the estimates for the corresponding income deciles—with the fifth decile as reference category—confirm the mean reverting error. Thus, compared with 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 levels of income in both sources ($S=A$), lower income groups tend to report correct wages less often than do 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 estimates 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.

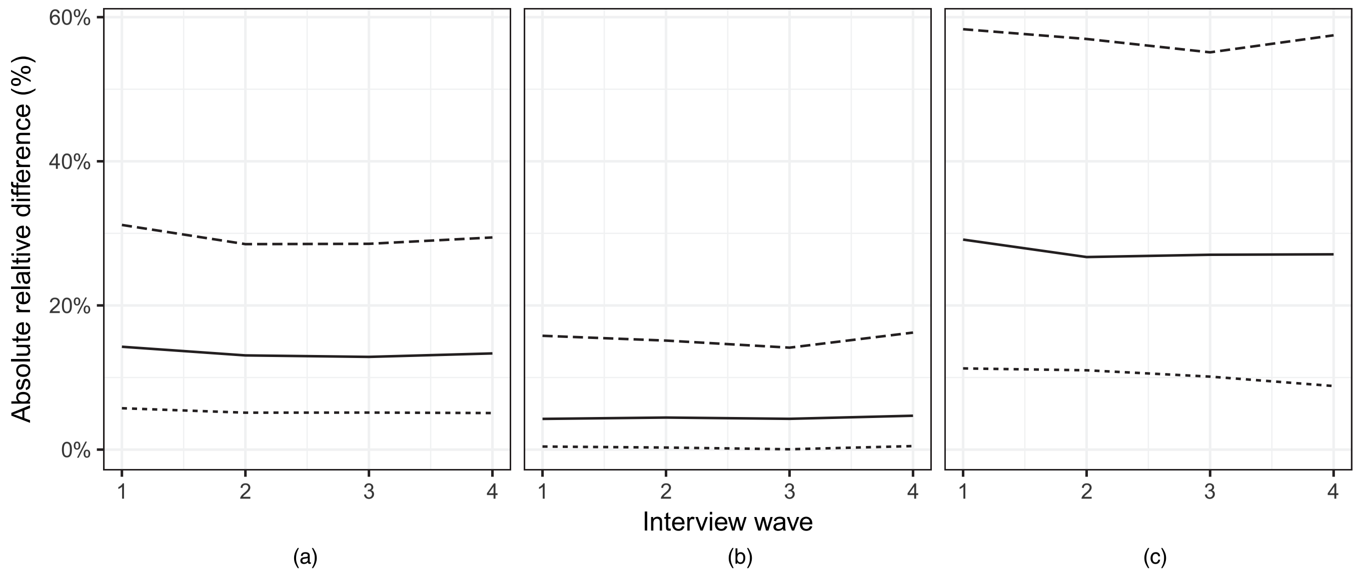


Fig. 2. Learning effect—SILC 2008–2011: the graphs show the median (———) as well as the 25th (------) and 75th percentile (– – –) of the absolute values of relative difference between survey responses and register information on (a) wages, (b) old age and (c) unemployment benefits by wave of the interview (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

4.1.2. Sociodemographic characteristics

Males have a significantly higher or lower tendency of respectively overreporting or 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 or masculinity norm which *ceteris paribus* renders men to overreport market income more often. On the basis of this argument, we also expected men rather to conceal receipt of non-market transfers. However, our estimates for unemployment benefits do not provide support for this hypothesis. Concerning education, we find significant differences between respondents with higher educational attainments compared with 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 quantiles) 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. Only age exerts a statistically significant, albeit very small, negative effect on the likelihood of underreporting unemployment benefits.

Being born in a central and eastern European country (Bulgaria, Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia) does not make a significant difference for the likelihood of misreporting wages compared with Austria, whereas we find a higher probability of underreporting wages for the Yugosphere (Bosnia and Herzegovina, Croatia, Serbia and Montenegro, and Macedonia), 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 EU 15 countries significantly raises the likelihood of overreporting pensions by 11 percentage points. Furthermore, natives from the Yugosphere bear a higher probability of underreporting unemployment benefits. Summing up, the evidence hints at 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 with 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 urbanized region. In the case of overreporting, this relationship 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 urbanization, this result could for example be due to differences in types of jobs and the associated wage structures in urbanized regions.

As an indicator for income stability, we include variables capturing the number of changes in status and the number of months spent in a specific employment status. As expected, Table 5 exhibits that changes in status correspond to higher probabilities of misreporting and to lower chances of reporting correct incomes. In general, the effects show significantly higher probabilities for underreporting, particularly for pensions and unemployment benefits. Furthermore, we

Table 5. Multinomial logit regressions†

	Results for wages			Results for pensions			Results for unemployment benefits		
	S<A	*S=A*	*S>A*	*S<A*	*S=A*	*S>A*	*S<A*	*S=A*	*S>A*
(a) Social desirability									
Relative income (reference: 5th decile)									
1st decile	−28.3‡	−21.5‡	49.8‡	−15.2‡	−16.8‡	32.0‡	−22.4‡	−5.0	27.4‡
2nd decile	−17.8‡	−17.0‡	34.7‡	−4.3	−5.8§	10.1‡	−11.0‡	−3.0	14.0‡
3rd decile	−11.2‡	−12.1‡	23.3‡	−4.2	−1.5	5.7‡	−0.3	−6.5§	6.7
4th decile	−2.0	−7.7‡	9.7‡	0.9	−3.9	3.0	−1.0	−5.5§	6.5
6th decile	8.5‡	−0.8	−7.7‡	8.7‡	−6.5‡	−2.2	10.4§	1.8	−12.1‡
7th decile	14.2‡	−1.3	−12.9‡	8.1‡	−7.6‡	−0.5	8.9§	−0.5	−8.4
8th decile	18.6‡	−3.3	−15.3‡	9.5‡	−7.9‡	−1.7	17.1‡	1.0	−18.1‡
9th decile	27.1‡	−7.4‡	−19.6‡	12.7‡	−8.0‡	−4.7‡	20.8‡	7.6§	−28.3‡
10th decile	38.6‡	−14.7‡	−23.9‡	24.2‡	−19.1‡	−5.1‡	31.4‡	5.6	−37.1‡
(b) Sociodemographic characteristics									
Gender (reference: 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 (reference: 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
1st stage tertiary	−11.0‡	1.2	9.8‡	−4.4	−0.5	4.9	2.3	−5.2	2.9
2nd stage tertiary	−4.1	−3.8	7.9§	1.4	1.8	−3.2	−7.2	7.3	−0.2
Country of birth (reference: Austria)									
EU 15 countries	−4.1	0.2	3.9	−2.8	−8.5‡	11.3‡	−0.4	0.6	−0.2
Central and eastern European country	−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 (reference: 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 urbanization (reference: <10000 inhabitants)									
>10000 and <100000	−1.6	2.1§	−0.5	−4.1§	3.5§	0.5	−3.7	1.6	2.0
>100000 inhabitants	−4.6‡	1.6	3.1‡	−4.8‡	3.1§	1.7	−8.7‡	3.1	5.6§

(*continued*)

Table 5 (*continued*)

	Results for wages			*Results for pensions*			*Results for unemployment benefits*		
	$S<A$	$S=A$	$S>A$	$S<A$	$S=A$	$S>A$	$S<A$	$S=A$	$S>A$
(b) Sociodemographic characteristics									
Changes in employment status (reference: 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 (reference: 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§
(c) Survey setting									
Mode of interview (reference: 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 (reference: personal)									
Proxy	7.9‡	−5.8‡	−2.2§	3.8§	−5.6‡	1.8	0.3	−0.7	0.4
Month of interview (reference: March–May)									
June–August	2.3‡	−2.9‡	0.6	1.3	−2.5§	1.2	−0.2	0.2	0.0
September–November	2.2	−5.7‡	3.5§	8.6‡	−9.6‡	1.0	0.2	0.6	−0.8
Year of interview (reference: 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
(d) Learning effect									
Wave of interview (reference: 1st)									
2nd	−1.4	0.4	1.0	3.0§	−3.1	0.1	7.4§	−5.8	−1.5
3rd	−1.8	0.8	1.0	2.0	−4.5§	2.5	4.2	−5.9§	1.8
4th	−2.8	2.0	0.7	2.1	−2.3	0.2	4.0	−1.7	−2.3

†The table shows the estimated average marginal effects of multinomial regressions per type of income (wages, pensions and unemployment benefits) in three categories: $S<A$, reported income is below the value in administrative data; $S=A$, survey income corresponds to administrative data; $S>A$, reported income is above the value in administrative data. Source: SILC 2008–2011; own calculations.
‡p-value less than 0.01.
§p-value less than 0.05.

find that overreporting unemployment benefits is 41% percentage 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 sources of data. Overreporting unemployment benefits is 44% percentage points less likely for short-term recipients than for long-term recipients. This general pattern is similar for underreported wages but almost non-existent for pension incomes.

4.1.3. Survey setting

Against what we have expected, average marginal effects of telephone interviews (CATI) on both types of misreporting are generally negative; however, they are statistically not significant. For wages, responses from CATI are 3 percentage 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 have an effect neither 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 effect 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 (more than 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 backcalculated the Austrian SILC.

4.1.4. Learning effect

As already indicated by the descriptive illustration in Fig. 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 of providing correct answers for pensions and unemployment benefits for respondents in advanced survey waves. In the next step, we shall 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 sociodemographic 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 that were found for social desirability (income quantiles) and for gender, age and education among the sociodemographic factors are quite similar for all three types of income. 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 the next step, we employ individual fixed effects OLS panel regressions to gain further insights into 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 that are relevant for these two sources of errors, the estimates

Table 6. Panel regression results—social desirability†

	Results for wages	*Results for pensions*	*Results for unemployment benefits*
Social desirability			
Relative income (reference: 5th decile)			
1st decile	7387.25 (1020.64)‡	3536.36 (764.73)‡	192.52 (1048.41)
2nd decile	5765.24 (776.71)‡	1775.13 (684.73)§	516.38 (441.90)
3rd decile	3524.27 (698.17)‡	1098.89 (471.01)‡	−33.88 (834.57)
4th decile	1631.32 (362.68)‡	−136.74 (372.11)	182.32 (471.82)
6th decile	−1449.64 (509.38)‡	−349.92 (330.50)	−625.39 (661.51)
7th decile	−2972.27 (653.50)‡	−1389.06 (982.90)	−799.96 (937.23)
8th decile	−4425.72 (841.99)‡	−2143.49 (1184.84)	−1387.76 (701.87)§
9th decile	−8118.60 (1319.65)‡	−3420.77 (773.14)‡	−1554.50 (705.41)§
10th decile	−11598.02 (1696.99)‡	−6968.90 (1769.55)‡	−2839.05 (831.64)‡
Other controls			
Individual fixed effects	Yes	Yes	Yes
Sociodemographic characteristics	Yes	Yes	Yes
Survey setting	Yes	Yes	Yes
Learning effect	Yes	Yes	Yes
Number of observations	20372	10105	2470
R^2 (full model)	0.70	0.72	0.81
Adjusted R^2 (full model)	0.32	0.41	0.30

†The table shows the results of unbalanced panel regressions with positive and negative reporting errors as dependent variable (i.e. negative values correspond to underreporting and positive values to overreporting). Cluster robust standard errors are given in parentheses. As identification in fixed effect models relies on sufficiently large within variation, the variables gender, education and country of birth have been removed from the baseline specification. See the full list of estimated coefficients in the on-line Table B.2. Source, SILC 2008–2011; own calculations.
‡p-value less than 0.01.
§p-value less than 0.05.

were done using the full set of controls, corresponding to Table 5. The remainder of the estimates is contained in Tables B.2 and B.3 in the on-line appendix.

4.2.1. 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 compared with lower quantiles. For instance, in the lowest decile the average overreporting is €7400 above the error in the fifth decile, whereas the difference amounts to €11600 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 fifth decile and indicate underreporting at the top.

4.2.2. 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,

Table 7. Panel regression results—learning effect†

	Results for wages	*Results for pensions*	*Results for unemployment benefits*
Learning effect			
Wave of interview (reference: 1st)			
2nd	−164.75 (631.32)	−1622.45 (264.10)‡	191.71 (300.39)
3rd	−353.89 (770.74)	−2546.06 (520.99)‡	231.58 (433.76)
4th	−478.33 (959.25)	−3397.04 (821.01)‡	144.55 (568.08)
Other controls			
Individual fixed effects	Yes	Yes	Yes
Social desirability	Yes	Yes	Yes
Sociodemographic characteristics	Yes	Yes	Yes
Survey setting	Yes	Yes	Yes
Number of observations	20372	10105	2470
R^2 (full model)	0.74	0.71	0.82
Adjusted R^2 (full model)	0.40	0.39	0.31

†The table shows the results of unbalanced panel regressions with absolute values of the reporting errors as dependent variable (i.e. negative values (underreporting) have been multiplied by −1). Cluster robust standard errors are given in parentheses. As identification in fixed effect models relies on sufficiently large within variation, the variables gender, education and country of birth have been removed from the baseline specification. See the full list of estimated coefficients in the on-line Table B.3. Source, SILC 2008–2011; own calculations.
‡p-value less than 0.01.

we do not distinguish between overreporting and underreporting and thus negative values of the dependent variable are multiplied by −1. 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 in 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.

4.2.3. Robustness checks

We applied two robustness checks for our baseline panel specification:

(a) estimates based on the four-wave balanced sample by using longitudinal weights (see Tables B.4 and B.5 in the on-line appendix);
(b) specifications with the difference in log-income between register and survey data as dependent variable (see Tables B.6 and B.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 four-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 logarithms of incomes have a higher model fit. Mean reverting errors that were 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 in the difference between survey and register data of approximately 19% in the last panel wave.

Table 8. Decomposition of explained variance—SILC 2008–2011†

	Wages (%)	*Pensions (%)*	*Unemployment benefits (%)*
Proportion of variance explained	15.8	11.4	46.4
Relative importance			
(a) Social desirability	65.1	58.8	23.7
(b) Sociodemographic characteristics	34.0	24.5	64.5
(c) Survey setting	0.8	14.5	9.2
(d) Learning effect	0.0	2.2	2.6

†The table shows the goodness of fit of OLS regressions and its decomposition to four sources of error, i.e. four groups of explanatory variables. We quantify the relative importance of (a) social desirability, (b) sociodemographic characteristics, (c) aspects of the survey design and (d) 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 as 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 0s and negative values among the errors. Source: SILC 2008–2011; own calculations.

4.3. *Relative importance of sources of error*

Finally, we apply an Owen value decomposition to assess the relative importance of the four sources of error 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 0-values. In the context of a significant mass of 0s 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 of Table 8 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 comparatively high (Kim and Tamborini, 2012). The remaining rows show the group sums of Owen values as percentages of the overall R^2. Whereas the patterns for wages and pensions are quite similar, unemployment benefits show quite a different picture. For wages and pensions, around 30% of the explained variance can be attributed to the group of sociodemographic 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%.

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

Additionally, we repeat the same procedure without transforming the input variables (see Table B.8 in the on-line 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 sources of error are hardly affected, which strengthens our confidence in the robustness of our findings.

5. Conclusions

Income is very likely to be one of the most pervasive pieces of information in micro data sets, since it plays an essential role for a wide range of welfare indicators and policy questions. The traditional way of collecting income information is household surveys; however, the accuracy of survey data has increasingly been contested during recent 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 by using a unique data set: the Austrian 2008–2011 waves of the EU SILC survey. We make use of the fact that, because of a legal initiative, the Austrian SILC provides both survey and register income data for the same observational units for four consecutive years.

Whereas 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 types of income. The EU SILC data are a key data set for social policy issues since they provide 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 with previous literature, using the Austrian EU SILC data set 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 of the national population residing in private households, ensures high external validity. Second, we can evaluate income measurement error for various components of total disposable household income in the very same data set.

We elaborate four major reasons for misreporting discussed in the literature: social desirability, sociodemographic 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 as follows. 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 higher probabilities of overreporting at the bottom of the wage distribution and, vice versa, higher likelihoods of underreporting at the top tail. By including a broad range of control variables 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 sociodemographic characteristics, males are found to have a significantly higher tendency to overreport 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 countries hint at 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 employment status have a strong effect on reporting errors in pensions; a shorter status duration increases errors particularly for unemployment benefits and wages.

For survey designers, the Owen value decomposition might be of particular interest since it reveals that social desirability is a major explanation for misreporting wages and pension income. For unemployment benefits, sociodemographic characteristics of the respondents seem to play the major role for reporting errors. The survey setting is relatively less important for explaining misreporting, whereas learning effects are hardly noticeable. Our findings from the Owen value decomposition suggest that data producers should be even more aware of social desirability when constructing interview questionnaires. The order and wording of questions on income could incite or inhibit erroneous income reports. It is crucial that survey responses should be validated by actual income proofs, such as payslips. If wages of employees are available on line in web portals of the financial authorities (like in Austria), data producers could possibly lobby for cross-checking with these sources. Moreover, proxy interviews show significantly higher probabilities of misreporting than do personal interviews. 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 effect is mixed and depends crucially on the type of income under consideration.

Survey users may be interested in the various types of bias that result from measurement error in linear models. In Table B.10 of the on-line supplementary materials, we present estimates of three types of bias: bias in OLS estimators due to measurement error in the explanatory variables, bias in instrumental variable estimators due to measurement error in the explanatory variables, and bias due to measurement error in the dependent variable. Quantifying the bias that is associated with misreporting gives data users and practitioners some guidance on what to expect when they use (or read publications using) income data from SILC in regression models. For instance, we find that all results estimating returns (e.g. of education) on wages are underestimated by about 25%. However, we advise against interpreting these results as general attenuation bias of survey income data. The reasons for and structure of misreporting might vary considerably across countries.

For policy makers, our results point to the fact that for some sociodemographic 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 policies. 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 who have more than one of the three income components at once in a given year and check whether those who underreport one source are also more likely to underreport another. It could also be worthwhile to replicate our analysis on the basis of 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 a right-hand-side variable (health, wellbeing, etc.) are altered when survey data are replaced by register data.

Acknowledgements

We thank Wilfried Altzinger, Stephen P. Jenkins, Dean Jolliffe, Branko Milanovic, the participants of various seminars, the Joint Editor and Associate Editor and two referees for helpful comments on earlier versions of the paper. Stefan Humer's research is supported by funds from the Oesterreichische Nationalbank (Oesterreichische Nationalbank anniversary fund, project 16728).

References

Abowd, J. M. and Stinson, M. H. (2013) Estimating measurement error in annual job earnings: a comparison of survey and administrative data. *Rev. Econ. Statist.*, **95**, 1451–1467.
Alm, J. and Torgler, B. (2006) Culture differences and tax morale in the United States and in Europe. *J. Econ. Psychol.*, **27**, 224–246.
Angel, S., Lamei, N. and Heuberger, R. (2018) Differences between household income from surveys and registers and how these affect the poverty headcount: evidence from the Austrian SILC. *Socl Indic. Res.*, **138**, 575–603.
Asamer, E., Astleithner, F., Cetkovic, P., Humer, S., Lenk, M., Moser, M. and Rechta, H. (2016) Quality assessment for register-based statistics—results for the Austrian Census 2011. *Austrn J. Statist.*, **45**, 3–14.
Berka, C., Humer, S., Moser, M., Lenk, M., Rechta, H. and Schwerer, E. (2012) Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based census 2011. *Statist. Neerland.*, **66**, 18–33.
Bingley, P. and Martinello, A. (2017) Measurement error in income and schooling and the bias of linear estimators. *J. Lab. Econ.*, **35**, 1117–1148.
Bollinger, C. R. (1998) Measurement error in the Current Population Survey. *J. Lab. Econ.*, **16**, 576–594.
Bound, J., Brown, C., Duncan, G. J. and Rodgers, W. L. (1994) Evidence on the validity of cross-sectional and longitudinal labor market data. *J. Lab. Econ.*, **12**, 345–368.
Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement error in survey data. In *Handbook of Econometrics*, vol. 5 (eds J. Heckman and E. Leamer), pp. 3705–3843. New York: Elsevier.
Bound, J. and Krueger, A. B. (1991) The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *J. Lab. Econ.*, **9**, 1–24.
Einav, L. and Levin, J. (2014) Economics in the age of big data. *Science*, **346**, 715.
Fessler, P., Kasy, M. and Lindner, P. (2017) Survey mode effects on income inequality measurement. *Working Paper 48766*. Harvard University, Cambridge.
Hansen, K. and Kneale, D. (2013) Does how you measure income make a difference to measuring poverty?: Evidence from the UK. *Socl Indic. Res.*, **110**, 1119–1140.
Hariri, J. G. and Lassen, D. D. (2017) Income and outcomes. *Publ. Opin. Q.*, **81**, 564–576.
Hassan, M. and Schneider, F. (2016) Size and development of the shadow economies of 157 worldwide countries: updated and new measures from 1999 to 2013. *J. Globl Econ.*, **4**, 1–14.
Huettner, F. and Sunder, M. (2012) Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values. *Electron. J. Statist.*, **6**, 1239–1250.
Jenkins, S. P., Cappellari, L., Lynn, P. and Sala, E. (2006) Patterns of consent: evidence from a general household survey. *J. R. Statist. Soc.* A, **169**, 701–722.
Jenkins, S. P., Lynn, A. J. and Sala, E. (2008) The feasibility of linking household survey and administrative record data: new evidence for Britain. *Int. J. Socl Res. Methodol.*, **11**, 29–43.
Kapteyn, A. and Ypma, J. Y. (2007) Measurement error and misclassification: a comparison of survey and administrative data. *J. Lab. Econ.*, **25**, 513–551.
Kim, C. and Tamborini, C. R. (2012) Do surveys estimate earnings inequality correctly?: Measurement errors among black and white male workers. *Socl Forces*, **90**, 1157–1181.
Kim, C. and Tamborini, C. R. (2014) Response error in earnings: an analysis of the survey of income and program participation matched with administrative data. *Sociol. Meth. Res.*, **43**, 39–72.
Kreiner, C. T., Lassen, D. D. and Leth-Petersen, S. (2015) Measuring the accuracy of survey responses using administrative register data: evidence from Denmark. In *Improving the Measurement of Consumer Expenditures* (eds C. D. Carroll, T. F. Crossley and J. Sabelhaus). Cambridge: National Bureau of Economic Research.
Kreuter, F., Muller, G. and Trappmann, M. (2010) Nonresponse and measurement error in employment research: making use of administrative data. *Publ. Opin. Q.*, **74**, 880–906.
Lohmann, H. (2011) Comparability of EU-SILC survey and register data: the relationship among employment, earnings and poverty. *J. Eur. Socl Poly*, **21**, 37–54.

Lusardi, A. and Mitchell, O. S. (2014) The economic importance of financial literacy: theory and evidence. *J. Econ. Lit.*, **52**, 5–44.
Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2012) The impact of questioning method on measurement error in panel survey measures of benefit receipt: evidence from a validation study. *J. R. Statist. Soc.* A, **175**, 289–308.
Mellow, W. and Sider, H. (1983) Accuracy of response in labor market surveys: evidence and implications. *J. Lab. Econ.,* **1**, 331–344.
Meyer, B. D., Wallace, K. M. and Sullivan, J. (2015) Household surveys in crisis. *J. Econ. Perspect.*, **29**, 199–226.
Micklewright, J. and Schnepf, S. V. (2010) How reliable are income data collected with a single question? *J. R. Statist. Soc.* A, **173**, 409–429.
Moore, J. C., Stinson L. L. and Welniak, Jr, E. J. (2000) Income measurement error in surveys: a review. *J. Off. Statist.*, **16**, 331–362.
Paulus, A. (2015) Tax evasion and measurement error: an econometric analysis of survey data linked with tax data. *Working Paper 2015-10*. Institute for Social and Economic Research, University of Essex, Colchester.
Pedace, R. and Bates, N. (2000) Using administrative data to assess earnings reporting error in the Survey of Income and Program Participation. *J. Econ. Socl Measmnt*, **26**, 173–192.
Pforr, K. (2014) femlogit—implemenation of the multinomial logit model with fixed effects. *Stata J.*, **14**, 847–862.
Pischke, J. (1995) Measurement error and earnings dynamics: some estimates from the PSID Validation Study. *J. Bus. Econ. Statist.*, **13**, 305–314.
Reynolds, J. and Wenger, J. (2012) He said, she said: the gender wage gap according to self and proxy reports in the Current Population Survey. *Socl Sci. Res.*, **41**, 392–411.
Sakshaug, J. W., Couper, M. P. K., Ofstedal, M. B. and Weir, D. R. (2012) Linking survey and administrative records: mechanisms of consent. *Sociol. Meth. Res.*, **41**, 535–569.
Sakshaug, J. W. and Eckman, S. (2016) Are survey nonrespondents willing to provide consent to use administrative records? *Publ. Opin. Q.*, **81**, 495–522.
Schnetzer, M., Astleithner, F., Cetkovic, P., Humer, S., Lenk, M. and Moser, M. (2015) Quality assessment of imputations in administrative data. *J. Off. Statist.*, **31**, 231–247.
Statistics Austria (2014a) Standard-Dokumentation Metainformationen (Definitionen, Erläuterungen, Methoden, Qualität) zu EU-SILC 2011. *Technical Report*. Statistics Austria, Vienna.
Statistics Austria (2014b) Methodenbericht zur Rückrechnung von EU-SILC 2008-2011 auf Basis von Verwaltungs-daten. *Technical Report*. Statistics Austria, Vienna.
Tamborini, C. R. and Kim, C. (2013) Are proxy interviews associated with biased earnings reports?: Marital status and gender effects of proxy. *Socl Sci. Res.*, **42**, 499–512.
Tourangeau, R., Bradburn, N. M., Marsden, P. V. and Wright, J. D. (2010) The psychology of survey response. In *Handbook of Survey Research* (eds P. V. Marsden and J. D. Wright), 2nd edn, pp. 315–346. Bingley: Emerald.
Tourangeau, R., Rips, L. J. and Rasinski, K. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Online supplementary material’.