ISER Working Paper Series

Income Source Confusion using the SILC

Christopher R. Bollinger
University of Kentucky

Iva V. Tasseva
Department of Social Policy
London School of Economics and Political Science

No. 2022-4
March 2022

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

Economic and Social Research Council

University of Essex

www.iser.essex.ac.uk

Non-Technical Summary

Household survey data on incomes are the main source for analysis of the income distribution. They underpin official statistics and research on the income distribution, poverty and inequality. Household survey data often collect income by source, for example by asking specific questions about labour market earnings and state social security benefits such as for unemployment and old-age. Researchers and analysts add up the incomes from these different sources to get total household income.

It is important to understand the types of error that survey respondents may do when they report their incomes in the data and the implications of these errors for analysis. For example, state benefits are known to be underreported in survey data. But we do not know if survey respondents may confuse benefits with earnings. If so, total income may be correctly reported – if respondents misreport benefits for earnings, their total income will still add up to the same – but the separate income sources will be in the wrong place. Correctly identifying separate income sources is, however, important for many research questions. To study errors in survey incomes, ideally we need linked survey and administrative data, where the administrative data are often assumed to include the true incomes. But it is still rare to have linked survey and administrative data and it is even rarer to have such data for both different types of state benefit and earnings.

We use a unique panel of household survey data – the Austrian version of the European Union Statistics on Income and Living Conditions (SILC) for 2008-2011 – which have been linked to individual administrative records on both state unemployment benefits and earnings.

First, we assess the extent of misreporting across similar benefits and between benefits and earnings. The three sources of income with the highest underreporting in the SILC are Austria's three main unemployment programmes: the Unemployment Insurance, Unemployment Assistance, and Assistance for Covering Living Costs. The programmes are not mutually exclusive and so individuals can be on various combinations of earnings and benefit receipt. We document that many respondents fail to report participation in one or more of the unemployment programmes. Moreover, they inflate earnings for periods when they are unemployed but are actually receiving unemployment compensation.

Second, we show that the implications for misreporting in unemployment programmes go far beyond simply the measurement of unemployment as misreporting is likely to impact other studies through both mismeasured earnings and construction of samples. We demonstrate this by assessing the impact of income source confusion on the extent to which education or job training lead to higher earnings. We find that estimates for the returns to education and training are substantially underestimated when using the survey relative to the administrative data. Our results suggest that interactions between unemployment benefits and any variable or sample construction may be of a significant concern.

Income Source Confusion using the SILC*

Christopher R. Bollinger[1] and Iva V. Tasseva[2]

[1]University of Kentucky
[2]London School of Economics and Political Science

Abstract

We use a unique panel of household survey data – the Austrian version of the European Union Statistics on Income and Living Conditions (SILC) for 2008-2011 – which have been linked to individual administrative records on both state unemployment benefits and earnings. We assess the extent and structure of misreporting across similar benefits and between benefits and earnings. We document that many respondents fail to report participation in one or more of the unemployment programmes. Moreover, they inflate earnings for periods when they are unemployed but receiving unemployment compensation. To demonstrate the impact of income source confusion on estimators we estimate standard Mincer wage equations. Since unemployment is associated with lower education, the reports of unemployment benefits as earnings bias downward the returns to education. Failure to report unemployment benefits also leads to substantial sample bias when selecting on these benefits, as one might in estimating the returns to job training.

JEL codes: D31, C8

Keywords: income source confusion, survey and administrative data, measurement error

*Acknowledgements: We thank Mike Brewer, Michael Fuchs, Katrin Gasior, Stephen Jenkins and John Pepper for valuable comments. We make use of the European Union Statistics on Income and Living Conditions (EU-SILC) for 2008-2011 for Austria provided by Statistics Austria. This research was conducted, in part, while Christopher R. Bollinger was on sabbatical at the Institute for Social and Economic Research at the University of Essex. He thanks the Leverhulme Foundation for financial support. Any errors are our own.

1 Introduction

Survey data often collect income by source, for example asking specific questions about labour market earnings, entitlement payments such as unemployment and retirement, and social safety net programmes. A number of authors (Marquis and Moore 1990; Bollinger and David 1997, 2000; Lynn et al. 2012; Meyer and Mittag 2019; Meyer et al. forthcoming; Celhay et al. 2021) have either documented or suggested that survey respondents may report one income source as another. This type of source confusion can have implications in a number of settings. Angel et al. (2018) examine the implications for poverty using matched survey and administrative data on Austria. They show that while there is substantial error in reporting of the level and receipt of many sources, the overall estimates of poverty are far less biased. While they do not explore the detailed reasons, one possible explanation is that some incomes are reported in the wrong place. That is, cross-reporting may imply that individuals get the specific source wrong but total income is close to correct.

The allocation of income into the correct source, however, is important for many research questions (see Bollinger and David 1997; Lynn et al. 2012). To study errors in survey incomes, ideally we need linked survey and administrative data, where the administrative data are often assumed to be the true incomes. The existing literature using such linked data has studied the data error properties separately in earnings or in state benefits. Mainly for the US, most recently Bollinger et al. (2019) and Hokayem et al. (2015) examine non-response and measurement error in earnings and Meyer and Mittag (2019) and Meyer et al. (2015) in state transfers; while prior work by Bollinger and David (2000, 2005) documents the persistence of response errors over time in food stamps. Mathiowetz and Duncan (1988) examine unemployment spells, but not the benefit amount. For other countries Kapteyn and Ypma (2007) and Jenkins and Rios-Avila (2021a) analyse measurement error in earnings with Swedish and UK data, respectively, and Lynn et al. (2012) study measurement error in UK benefit receipt. But it is still rare to have linked survey and administrative data and it is even rarer to have such data for both different types of state benefit and earnings.

We use data from a unique panel of household survey data — the Austrian version of the European Union Statistics on Income and Living Conditions (SILC) for 2008-2011 — which have been linked to individual administrative records on both Austrian state benefits and earnings. While Angel et al. (2018) and Angel et al. (2019) use the same linked data, Angel et al. (2018) focus on the measurement of poverty, while Angel et al. (2019) examine the relationship between household and survey characteristics and response error. In a different study, Fuchs et al. (2020) use the administrative records to estimate the take-up of social assistance in Austria, noting the underestimation of non-take-up with the SILC survey data due to errors in the reported incomes.

There are various reasons why individuals may misreport income in household surveys. Errors may be due to genuine reporting mistakes or deliberate misreporting (Tourangeau et al. 2000). Genuine reporting mistakes could be due to programme name confusion or confusion of conceptually-related benefits such as benefits for children. Individuals may be in receipt of combinations of benefits and earnings and only know the total. They may also consider certain types of benefit as part of their earnings if those benefits are associated with their employment (past or present). Survey design may also make it easier to report different incomes as one. Deliberate misreporting may occur because some items are sensitive, e.g. benefits targeted to low-income families, and so, individuals may prefer to conceal receipt altogether or to report them as earnings.

Our paper has two goals. The first goal is to assess the extent and structure of misreporting across similar benefits and between benefits and earnings. While Meyer et al. (forthcoming) and Bollinger and David (2000) and others have considered programme confusion, to the best of our knowledge ours is the first paper to examine reporting benefits as earnings. The three sources of income with the highest under-reporting in the SILC are those involving unemployment. Austria has three main unemployment programmes: Unemployment Insurance, Unemployment Assistance, and Assistance for Covering Living Costs. The programmes are not mutually exclusive and so individuals can be on various combinations of the programmes. As we document below, many respondents fail to report participation in one or more of the programmes. Moreover, it appears that they inflate earnings for periods when they are unemployed but receiving unemployment compensation. The second goal is to demonstrate the impact of this kind of income source confusion on estimates typical in the literature. Survey data are often used for estimation of wage equations because of the rich covariates available (which are often unavailable in administrative data). In our first application we document that reports of unemployment as earnings leads to a downward bias in estimates of the returns to education. In our second application, we examine the role of job training for future earnings. We document that the returns to job training can be biased, in some cases quite substantially and with sign changes. In both applications, the misreporting of various unemployment benefits biases the sample, as well as leading to typical types of measurement error bias.

The paper begins by describing the three unemployment benefits in Austria in Section 2. It then describes the data set and the samples used from the data in Section 3. In Section 4, we then document the cross-reporting properties mentioned above and the implications for different research questions. We conclude in Section 5.

2 Unemployment Benefits in Austria

The three main unemployment benefits in Austria, recorded in the SILC, are the Unemployment Insurance (UI, "Arbeitslosengeld"), Unemployment Assistance (UA, "Notstandshilfe") and Assistance for Covering Living Costs (ACLC, "Beihilfe zur Deckung des Lebensunterhaltes"). All three benefits allow for limited labour market participation.

UI is an insurance-based benefit generally provided for 20 up to 52 weeks. UI is the main unemployment programme in Austria. Individuals who have previously worked and paid contributions for unemployment insurance over a sufficiently long period are entitled. The benefit duration depends on the person's age and the length of their insurance. The benefit amount is 55% of previous earnings.[1]

UA is *targeted at low-income groups* only. It extends benefits when low income individuals exhaust their UI benefit. The benefit can be received for up to 12 months but

[1]However, if UI is less than a reference rate (equal to €24.20 per day in 2007, €24.90 in 2008, €25.75 in 2009 and €26.13 in 2010), the benefit amount takes the smaller value between the reference rate and 60% of previous earnings.

follow-up applications are possible. An income-test is applied on the person's and their partner's income. The benefit amount is 95% of UI, if UI does not exceed a reference rate (equal to €24.20 per day in 2007, €24.90 in 2008, €25.75 in 2009 and €26.13 in 2010); it is otherwise 92% of UI, with a minimum of 95% of the reference rate. The income of the partner, after applying allowances based on the age and insurance length of the unemployed person and presence of children, is then subtracted from the benefit amount. Moreover, after 6 months of UA receipt, if the duration of UI was 20 weeks, the amount in UA is capped at the reference rate; and if the UI duration was 30 weeks, the UA amount is capped at a higher rate (of €27.23 per day in 2007, €29.03 in 2008, €30.03 in 2009 and €30.47 in 2010).

While on UI and UA, the labour market agency may require participants to take part in job training programmes or to apply to certain jobs advertised through the agency. If recipients are offered a job, they must accept it as long as it is suitable (e.g. commuting is less than 2 hours per day for a full-time job or the job should not be in conflict with one's religious beliefs). A refusal to take up any of these may result in a (temporary) suspension of the benefit.

The *ACLC* provides income support during job training programmes *for low-income individuals* only and can be received simultaneously with UI or UA. The duration of the programme should be more than one week and at least 16 hours per week. The benefit lasts for the duration of the programme and the amount equals the difference between a guaranteed minimum and the current entitlement to UI/UA.

There are several reasons why respondents may cross report one benefit as another in the SILC. We can rule out similar programme names as a potential reason for programme confusion as the benefit programme names in German (see above) are very different from each other. However, the benefits are similar conceptually, which may lead to programme confusion.[2] These benefits are all paid during periods of unemployment and administered by the same labour market agency ("Arbeitsmarktservice") and the benefit amounts are

[2]For example, based on focus group interviews with survey respondents, Balarajan and Collins (2013) show that respondents from the UK Family Resources Survey report being confused about which benefit they receive and thereby misreport across benefit types.

linked: UA is a function of UI while ACLC is a function of UI/UA. Programme confusion and benefit cross-reporting may also occur if respondents do not know the individual benefit amounts, but know the total they receive; or due to survey design if respondents find it quicker to report everything as one benefit. Fundamentally, these are all issues of salience for respondents.

There are also reasons why respondents may report benefits as earnings in the SILC. First, there may be stigma about receiving the benefits (Currie 2006), especially the UA and ACLC which are targeted at low-income individuals, and so respondents may deliberately misreport them as earnings. Second, individuals may be conceptually thinking of benefits as earnings: UI and UA are linked to previous employment while ACLC pays for job training, and all three benefits allow for limited labour market participation. Third, and as with programme confusion, individuals may not know the individual income amounts but the total they receive; or due to survey design they may find it quicker to report everything as earnings.

3 Data: The Austrian SILC

The SILC is a household income survey and has a survey design, also used in all EU countries and beyond (Aktinson and Marlier 2010). The Austrian SILC data used in our analysis are for survey years 2008-2011 (income reference period for the calendar years 2007-2010). The sample is based on all private households registered in the central population register ("Zentrales Melderegister"). The data include rich information on individual and household characteristics and are nationally representative. They are widely used by social scientists and serve as the main source for official statistics on income, poverty and inequality, published by the national statistical office (Statistics Austria) and the EU statistical agency (Eurostat). In this paper, we use the national SILC version, including the files with "additional data" ("Zusatzdaten"). These files contain separate variables by types of income source.

The key feature of the data, which allows us to provide novel evidence on income cross-reporting, is that the data include individual-level survey reports and administrative records on both earnings and a range of state benefits. Notably, the survey and administrative data match was successful for more than 96% of the sample (Heuberger et al. 2013). Next, we describe in more detail the main characteristics of the data.

3.1 Linked survey reports and administrative records

From SILC 2012 onwards, to reduce response burden and data editing and improve the quality of the income data, survey participants are no longer asked to provide information on earnings and certain types of state benefit; this information is instead acquired directly from various administrative sources (Heuberger et al. 2013). To provide a continuous data series since 2008 for the measurement of Europe 2020 social inclusion targets, Statistics Austria retrospectively matched the individuals in the survey to their administrative data for the survey years 2008-2011, providing the opportunity to compare individual survey responses with the administrative records on employment earnings and state benefits for a variety of programmes including unemployment. For detailed information on the legislation changes concerning the data collection in Austria and the administrative data sources, see Heuberger et al. (2013) and Angel et al. (2018).

To match individuals from the survey and administrative data, pseudonymised personal identifiers are used. Each identifier has 28 characters and is a function of the person's name, date of birth and sex. These identifiers are available for nearly every person in Austria, in each administrative data source and also in the central population register which is used for the SILC sampling frame (Heuberger et al. 2013). Thus, identifiers exist for all persons registered in the addresses selected in the SILC sample. But there may be also household members who have not been registered at an address, who end up in the SILC sample. Their identifiers need to then be additionally obtained and for some individuals, they may not be found. In fact, identifiers were found for 95.6% of the sample in 2008, 97.7% in 2009, 96.8% in 2010 and 99.4% in 2011 (Statistics Austria 2014).[3] The

[3]In comparison, match rates for other data sets tend to be lower. For example, for linked earnings data the match rate is 86.2% of earners in the US Census Bureau Current Population Survey, 2006–2011 Annual Social and Economic Supplement (Bollinger et al., 2019) and 46% of employed in the UK Family Resources Survey 2011/12 (i.e. 65% of employed who consented to the linkage *times* 71% of those with successful linkage) (Jenkins and Rios-Avila, 2021b); for linked state benefit data, the match rate is 57.3%, based on a pooled matching technique, of respondents in the UK 'Improving Survey Measurement

missing identifiers are due to erroneous data in the survey or administrative data (e.g. incorrect birth date), or are non-existent (e.g. for younger individuals or individuals with non-Austrian citizenship). For detailed information on the linkage, see Heuberger et al. (2013).

Recent work by Kapteyn and Ypma (2007) and Jenkins and Rios-Avila (2021a) suggests that mismatch can play a role. The Austrian system samples from an original frame which already has the identifier, hence the match is generally not ex post. The high rates of deterministic match are evidence that this likely does not play a role.

Separately, work by Jenkins and Rios-Avila (2021a), Bollinger et al. (2019), Bollinger et al. (2018), Paulus (2015) and Kapteyn and Ypma (2007) all suggest that administrative earnings may fail to reflect under the table earnings. Under the table earnings may be playing a role here, but what we show is that this may actually be under-reporting benefits and reporting them as earnings. It is unclear if this is the case in other countries. We note that generally administrative records of social welfare benefits are believed to be quite accurate, see for example Bollinger and David (1997) or Meyer and Mittag (2019).

The administrative data on earnings and benefits are supplied by the relevant authorities. The earnings data derive from wage tax data electronically supplied by employers to the tax authority for the purposes of tax payments and social insurance contributions. The benefit data are provided by the authority administering the benefit payments.

The survey information on earnings and benefits is based on individual-level interviews with adult respondents aged 16 and above.[4] We use the SILC "additional data" files which include detailed information for the level of gross earnings and state benefits and the number of months in receipt for each income. These data give us the necessary level of detail to distinguish between the different types of unemployment benefit and study income cross-reporting. Angel et al. (2018) and Angel et al. (2019) use instead

of Income and Employment' survey in 2003 (Jenkins et al., 2008); while Meyer and Mittag (2019) report that 91% of households in the US Current Population Survey have a member with a personal identification key to facilitate linkage to administrative data.

[4]The SILC questionnaires are available at: `https://ec.europa.eu/eurostat/web/income-and-living-conditions/quality/questionnaires`. The SILC quality reports contain further information on the data, including on sampling design and response errors: `https://ec.europa.eu/eurostat/web/income-and-living-conditions/quality/eu-and-national-quality-reports`.

the harmonised income variables in SILC, which aggregate benefits by function, e.g. all unemployment benefits into a single variable.

Finally, the reference period of the survey and administrative data is the same, i.e. the calendar year before the date of interview. We compare the gross annual income amounts based on the survey versus the administrative data. For earnings, both the survey and administrative data include information for gross annual earnings. For state benefits, while the administrative records include annual benefit amounts, the survey data include monthly benefit amounts. Both data sources include information on the number of months in benefit receipt. Thus, we multiply the monthly benefit amount by the number of months in receipt to derive the gross annual benefit amounts in the survey.[5]

3.2 Longitudinal information

The SILC follows individuals in a 4-year rotating panel. Households are selected and participate in the panel for four years with annual surveys. The individuals are staggered so that each year 25% of the sample is renewed. Hence our sample, covering the four survey years of 2008 through 2011 contains different numbers of waves for different individuals. Individuals initially surveyed in 2008, will appear in all four years of our sample. Individuals initially interviewed in either 2005 or 2011 will appear only once. Table 1 exhibits the number of observations in our sample by wave.

3.3 Sample

The sample in our analysis includes individuals aged 16 and over, with no upper age limit applied (children below the age of 16 are not interviewed). In the next sections, where we cross-tabulate benefit receipt or compare mean benefit amounts in the survey versus administrative data, we exclude missing or imputed values. Later on, in the estimation of ordinary least squares models, we also exclude cases where the outcome variable has missing or imputed values in either data set. Where earnings are used as a continuous

[5] The error in the survey annual benefit amounts can stem from misreporting the benefit duration and/ or the average monthly amount. As we are interested in the total error in benefits, as well as in earnings, we do not attempt to disentangle the two error components. Future research will address this.

control variable, observations with missing or imputed survey or administrative values are excluded. Where a categorical control variable (e.g. occupation) includes missing/ imputed values, the effect of these values on the outcome variable is captured through a dummy variable. Finally, when we assess the impact of the error in income receipts on the returns to education and job training, we further restrict the sample to those aged 19-64.

Table 1 provides summary statistics for the full sample (column 1), men (column 2) and women (column 3), including the few observations with missing/ imputed benefit and earnings values. It shows the mean age and the share (in %) of individuals by different characteristics. The proportion of cases with missing/ imputed UI/ ACLC values in the survey is 0.1%, while the missing/imputed cases in survey earnings and UA are too few to disclose.

There are 44,970 observations in total, with slightly more women (23,696) than men (21,274). The mean age of the sample is 47.4 years. Majority of people (59.7%) have middle-level education though men tend to be more educated than women. Men are also twice more likely to be in full-time, full-year employment than women, but fewer men have earnings – 40.6% of men compared to 49.8% of women – suggesting a larger prevalence of part-time and/or part-year employment among women. Of the full sample, 86% are born in Austria, 5.9% in the rest of the EU-27 and the remaining 8.1% in Turkey, the Western Balkans ('former Yugoslavian countries excluding Slovenia') or other countries. Proxy interviews were carried out for 17.4% of the sample, where the partner (7.8%) or another adult in the household (9.6%) was the proxy. About a fifth of men and women reported to have taken a job training, with 1.4% taking up training provided by the labour market agency which also administers the unemployment benefits. Looking across survey receipt of unemployment compensation, the largest recipient group is of UI (5.1%), followed by UA (1.7%) and ACLC (0.7%). Relatively more men report UI receipt than women – 6% of men versus 4.2% of women – but the proportions by sex are similar for UA (groups with ACLC receipt by sex are too small to disclose). Moreover, among recipients, the majority is observed to have received any of the benefits for a year while around a quarter of UI and UA and 11.4% of ACLC recipients received benefits for multiple years (two up to four years).

[insert Table 1 here]

4 Results

To examine the extent and relationship of misreporting in the three unemployment programmes and the relationship to reports of labour market earnings we employ a number of approaches. First we simply examine cross-tabulations and average benefit and earnings amounts compared to administrative amounts. We then turn to estimation of a series of linear probability models focusing on report of receipt in the survey conditional on having received the benefit in the administrative data. In these models, typical demographic variables are used as controls. Additionally, we include earnings (from the administrative report) and the difference in survey earnings and administrative earnings as well as measures of reporting accuracy (false positive, false negative and true positive) as main variables of interest.

4.1 Benefits underreporting

We document in Table 2 the extent of benefit misreporting in the survey, compared to the administrative data, for each benefit separately and for the total (UI + UA + ACLC). We find that underreporting of benefit receipt – i.e. 'false negative' rates or missing to report the benefit in the survey, conditional on receiving it in the administrative data – is substantial for all three benefits. Furthermore, even conditional on reporting receipt accurately in the survey, benefit recipients make errors in the amount they report.

The first row in Table 2 reports the rate of false negative which is the share of benefit recipients according to the administrative data who do not report the benefit in the survey. The false negative rate is high for all three benefits: 42% for UI, 51.7% for UA and 76.4% for ACLC. In other words, only 58% of UI, around 48% of UA and 24% of ACLC recipients correctly report receipt in the survey. The false negative rate for the total – i.e. the proportion of respondents receiving any unemployment benefit according to the administrative data who falsely claim not to be in the survey – is lower, at 37.7%. Thus, though failing to report individual benefits, the overall receipt of unemployment benefits seems to be better captured in the survey.

The third and fifth rows report two different definitions for the rate of false positive. The first definition refers to the share of benefit non-recipients according to the administrative data who falsely report a receipt in the survey. For each benefit separately as well as the total, the rate is less than 1%. But although the rate is low, as the denominator refers to the large sample of benefit non-recipients in the administrative data, the number of cases is non-negligible. For ACLC, the number of false positive (263) is even higher than that of false negative (185), suggesting both substantial underreporting and overreporting of receipt. Nevertheless, as there are many more UI and UA than ACLC recipients, the issue of underreporting benefit receipt seems to be overall more serious, consistent with findings of underreporting UK and US state benefits in survey data (Lynn et al. 2012; Meyer et al. 2015; Meyer and Mittag 2019).

The second definition of false positive (b) shows the number of benefit non-recipients according to the administrative data who falsely report a receipt in the survey, as a proportion of the number of recipients of the other two benefits in the administrative data. This measure indicates the extent of one benefit being reported as another in the survey (by definition, there is no estimate for the total). The rate is relatively high for UI – 12.5% of UA/ACLC recipients falsely report a UI receipt in the survey, suggesting potential programme confusion. The rates are lower for UA and ACLC. We look in more detail at the extent of benefit cross-reporting in the next section.

In addition to misreporting of benefit receipt, we also provide evidence for substantial misreporting of benefit amounts. In the remaining rows, Table 2 includes statistics on benefit amounts in the administrative and survey data, conditional on true positive or that is a positive benefit receipt in both data sets. The discrepancy between the mean administrative and survey amount of ACLC is relatively large and ACLC is on average overreported in the survey, consistent with the high number of false positive cases. For

UI, UA the total though, the mean administrative and survey amounts are notably similar – UA is only slightly underreported, while UI and the total are slightly overreported in the survey on average. However, this masks the high prevalence and size of the error in amount in the survey. Around 3/4 of UI and UA recipients misreport the benefit amount in the survey by at least 10% of the administrative amount. We find a similar pattern for the total benefit amount too. There is also a substantial number of people – around 1/5 of UI and UA recipients – who make very large errors of 70% or more of the administrative amount. The proportion of people with such large errors is somewhat lower, at 17%, if we consider the total benefit amount – this may suggest that respondents make fewer mistakes when reporting the total if they misreport one benefit as another, consistent with the results for the second definition of false positive. (Groups with different size of error in ACLC are too small to disclose.) The errors are also reflected in the relatively low correlation of administrative and survey amounts and large standard deviation of the error which stands at 65% of UI, 52% of UA, substantial 100% of ACLC and 55% of the total mean administrative amount. We examine the misreporting in benefit amounts in more detail in Section 4.2.2, where we estimate the errors separately on the samples of benefit true positive, false negative and false positive.

[insert Table 2 here]

4.2 Benefits and earnings cross-reporting

4.2.1 Cross-reporting at the extensive margin

As a departure point, we examine the extent of benefit and earnings cross-reporting at the extensive margin. We show that false negatives are predominantly reported as earnings. Furthermore, falsely reporting receipt of one benefit (e.g. UI) lowers the probability of reporting another benefit (e.g. UA), conditional on receiving it, providing evidence for programme confusion.

Table 3 provides basic descriptive statistics focusing on combinations of survey income receipt, given administrative receipt. The diagonal includes the true positive rates, e.g.

38.3% of 'UI only' recipients in the administrative data correctly report a positive receipt of UI only in the survey. The remaining cells indicate misreporting of income receipt.

Most notably, the fourth column 'Earnings only' demonstrates the extent of failing to report a benefit and reporting it as earnings. Starting with the first row, among true recipients of UI only, 11.2% fail to report UI but report earnings instead in the survey. Similarly, among true recipients of UA only, 6.4% fail to report UA and report earnings instead. Moreover, while 51.1% of recipients of both UI and earnings report receipt of both UI and earnings, 38% falsely report earnings only in the survey. Among recipients of both UA and earnings, only 18.7% report the receipt of both incomes, while 27.3% report receiving earnings only. This pattern of misreporting is also evident among recipients of UI, UA and earnings as well as recipients of ACLC and other income: 23.4% and 31.4%, respectively, incorrectly report earnings only. Thus, earnings are inflated in the survey while benefits are underreported. This suggests that respondents may be better at reporting their total income (unemployment benefits + earnings) than the individual income components. Though this type of bias does not matter for analysis based on total income, such as trends in poverty or inequality in household net income, it matters for analysis by income source and measurements of programme participation. In the later sections, we study in particular the implications of this error for different earnings outcomes.

The findings for benefit and earnings cross-reporting, which have not been previously documented, point to a serious issue in the survey data. Potential explanations, such as benefit stigma, thinking of benefits as earnings rather than state support or knowing better the total rather than the components of one's income, are not specific to the SILC or the Austrian context and can apply to other country and data settings, underlining the importance of the findings.

Moreover, Table 3 indicates that benefit cross-reporting is a serious issue too. Among recipients of UA only, 11% fail to report UA but falsely report UI. Moreover, 14.3% of UA + earnings and 28.8% of UI + UA + earnings recipients report receipt of UI + earnings in the survey, hence falsely putting some of UA in UI or reporting it all as UI and earnings.

While UI is paid universally to unemployed who have contributed to the insurance system, UA is restricted to lower-income individuals. UA is also paid after entitlement to UI is exhausted. Thus, people may deliberately misreport UA due to stigma or think of UA as a part of UI rather than a new benefit. While respondents may get the total amount of unemployment benefits right, there may still be implications of this type of misreporting on important outcomes such as the poverty-reducing impact of short-term (unemployment insurance) versus long-term (unemployment assistance) state support.

[insert Table 3 here]

Cross-reporting by programme

While the tabulations of the previous section clearly indicate cross-reporting, to further understand the issue we estimate linear probability models for the probability of reporting a benefit receipt in the survey, conditional on administrative receipt. These models investigate false negative reports and allow us to control for myriad demographic characteristics, earnings and receipt and report of the other programmes. Table 4 presents results for UI. In column (1) we condition on the logarithm of the administrative amount of UI; the administrative amount as well as the discrepancy between the survey and administrative amounts of earnings (in thousand); and dummies for false positive, false negative and true positive in earnings, UA and ACLC with true negative being omitted. In column (2) we control for a rich set of covariates, including sex, benefit duration in the administrative data, job training, proxy interview, age, education, number of children and adults in the household, self-reported health, region, occupation, industry, if a civil servant, country of birth, month of interview, survey wave, interaction between wave and year, interview type, and if the same interviewer as last year.

Table 4 presents evidence for misreporting UI as earnings. We find that the higher the administrative earnings and the more earnings are overstated in the survey, compared to the administrative data, the lower the likelihood of reporting UI in the survey, by 0.008 and 0.012, respectively. This is consistent with respondents putting UI into earnings. Based on this, we may also expect that a false positive in earnings lowers the probability of reporting UI. However, a false positive in earnings, relative to a true negative or that is reporting zero earnings in both the survey and administrative data, makes it more likely to report UI in the survey (by 0.173). A positive effect (0.278) is also shown for a true positive in earnings. This suggests that respondents may be reporting UI twice in the survey – once as earnings and then again as UI – which can arise if respondents think of UI as earnings. In the survey, people are first asked about their earnings (question number 42 in the SILC 2008 questionnaire), so they may report UI there, and then about benefit entitlements (question number 70), where they could declare UI again, without going back to correct their earnings answer.

Table 4 also presents clear evidence for benefit cross-reporting of UI and UA. A false positive in UA, compared to a true negative, makes it less likely (by 0.145) to report UI in the survey, suggesting programme confusion. Though the coefficient for a false positive in ACLC is imprecisely estimated, the sign is positive. UI and ACLC can be received simultaneously and UI recipients can be asked by the labour market agency to participate in job training. If they do, they may be thinking of some of the UI amount as a payment/ earnings towards the training and report it as ACLC which supports individual's incomes while on training.

Salience in the UI amount matters: the effect of the administrative amount of UI is 0.13, so a 1% increase in UI increases the probability of reporting a UI receipt by 0.1%. Participating in job training paid by the labour market agency, which pays out the unemployment benefits and potentially making UI receipt more salient, also increases the likelihood of reporting UI in the survey (by 0.116). On the other hand, holding constant the UI amount, the longer one receives UI, the less likely to report its receipt. This result potentially suggests benefit stigma. It is possible that because of limits on participation, those who participated for a longer period in the reference year (interviews refer to the prior calendar year), may no longer be participating and thus fail to recall. We do find that the later in the interview year respondents are interviewed, so the larger the time gap between the income reference and data collection period, the less likely to accurately report UI receipt in the survey (see Table A.1). However, stigma is still likely to be playing some role here.

Women respondents are less likely than men to report UI receipt in the survey, conditional on receiving it in the administrative data, by 0.036. We come back to the differences by sex at the end of this section.

Survey characteristics are important too. Having a proxy interview reduces the probability of reporting receipt, suggesting lower reliability of proxy responses versus self-reports (consistent with e.g. Bollinger et al. 2019). Table A.1 contains the full results from the regression analysis.

[insert Table 4 here]

Repeating the analysis for UA reveals it is likely UA is also being misreported as earnings and other benefits. Table 5 presents estimates for the probability of reporting UA in the survey, conditional on an administrative receipt. As with UI, the higher the administrative earnings and the more earnings are overstated in the survey relative to the administrative data, the less likely it is for respondents to report UA receipt. Furthermore, the probability of reporting UA in the survey is reduced with a false positive in earnings, by 0.219. These results suggest that respondents are putting UA into earnings, which may be driven by e.g. benefit stigma; thinking of UA as earnings given that UA entitlement is linked to previous contributions while working; and/ or knowing one's total income better than the separate components. There is also clear evidence for reporting UA as UI. Relative to a true negative, a false positive in UI makes it less likely to report UA. The effect (-0.31) is bigger than the effect on reporting UI of a false positive in UA (-0.145 in Table 4). Thus, it seems more likely for respondents to confuse UA with UI than the other way around. This is not surprising as UA is paid out *after* UI receipt is exhausted.

Salience in the UA amount matters, consistent with the results for UI. A 1% increase in the UA amount increases the probability of reporting UA in the survey by 0.033%. Job training by the labour market agency also makes it more likely to accurately report UA receipt (by 0.18). A longer benefit duration increases the probability of reporting UA too. Unlike with UI, UA recipients can send follow-up applications to extend benefit duration and hence, be on UA for long spells. For a given UA amount, a longer benefit duration can mean respondents are less likely to confuse UA with UI. For the full results from the regression analysis, see Table A.2.[6]

[insert Table 5 here]

We also estimated a model with ACLC receipt as the outcome variable but cannot report the results due to small sample size. Nevertheless, we want to re-highlight the relationship between ACLC and the other benefits. In our results, we find no evidence to counter what we have documented so far.

Differences by sex

Finally, we note that being a woman lowers the probability of reporting each of the unemployment benefits in the survey conditional on a receipt in the administrative data (see Table 4 and Table 5). To explore these differences by sex and understand differences in male and female reporting patterns, Table A.3 in the Appendix repeats the analysis in Table 2 by sex. We also repeat the regression analysis separately for men and women – Table A.4 and Table A.5 in the Appendix show results for the probability of reporting UI and UA, respectively, in the survey conditional on an administrative receipt.

Both men and women have reporting errors but on average women are slightly better than men at reporting the benefits if they receive them. This can be seen from the rates of false negative in Table A.3 which are substantial for both men and women but consistently lower for women – for UI, UA and the total of the three benefits in particular though less so for ACLC. For example, the false negative for UI is 43.9% for men compared to the lower 39.2% for women.

Relative to men, women are more prone to confusion between UI and UA. Misreporting of both UI and UA is positively associated with both the level of administrative earnings and the difference between survey and administrative earnings. But compared to men, women are less likely to report a benefit receipt twice, e.g. once as earnings and then again as a benefit. For example, a false positive in UI reduces the probability to report UA in the survey by 0.291 for men versus 0.348 for women (columns 2 and 4, respectively,

[6] We note here that, though Table A.1 and Table A.2 suggest a negative correlation between older age groups and reporting UI and UA receipt, respectively, in the survey, only 1.6% of UI and 2.7% of UA recipients in the administrative data are aged 60+. Other things such as education and country of birth do not seem to matter much either.

in Table A.5). Hence, women are more likely than men to confuse UA for UI. But, a false positive in earnings makes it more likely to report UI receipt in the survey by 0.364 for men, suggesting reporting UI receipt twice, while we find a small negative effect for women (columns 2 and 4, respectively, in Table A.4). Relatedly, a false positive in ACLC increases the probability of reporting UI/ UA for men, suggesting reporting UI/ UA receipt twice, but it reduces the probability of reporting the benefits for women, potentially due to programme confusion (Table A.4 and Table A.5).

These results imply that compared to men women may be less prone to inflate their total income but more likely to confuse benefits and report them in the wrong category. We explore further the implications of these differences by sex in the later sections.

4.2.2 Cross-reporting at the intensive margin

In this section, we examine benefit and earnings cross-reporting at the intensive margin, i.e. the errors respondents do in the amounts of income they report in the survey. We show that on average people make mistakes in the amount they report for different income sources. But respondents seem to be putting benefits and earnings together, providing further support for the hypothesis that they seem to know their overall income better than the amounts by income source.

We begin with Table 6 which compares the mean survey and administrative amounts (columns 2 and 3) separately by benefit type, earnings and total income (all unemployment benefits + earnings) shown in different rows. These comparisons are done for different samples, conditional on a benefit true positive, false negative or false positive. Table 6 also presents the mean error (survey – administrative amount in column 4), also expressed in % of the mean administrative income (column 5), the standard deviation of the errors (column 6) and the correlation between the survey and administrative amounts (column 7).

We show that, while on average there is a gap between the survey and administrative amounts, the errors by income source rather than reinforcing tend to at least partly offset each other and, thereby, reduce the overall error in income. Furthermore, the errors in the UI/UA true positive groups are the smallest, suggesting that on average respondents who correctly report receipt of at least one of their benefits tend to make fewer mistakes in the incomes they report.

In more detail, in the UI true positive group UI and ACLC are slightly overstated in the survey, compared to the administrative data, with a mean error of €185 in UI and €77 in ACLC; while UA and earnings are understated by €156 and €149, respectively. The positive and negative errors offset each other, suggesting benefit and earnings cross-reporting, so the total mean error is –€44 or only –0.26% of the average administrative income. Furthermore, while the standard deviation of the error is quite high for earnings, equal to 82% of mean administrative earnings, it is lower for total income, equal to 65% of the administrative amount. A similar pattern can be seen for the group with UA true positive, where the total mean error in income is just 1.64% of the administrative income. Moreover, while the standard deviation of the error as a % of the mean administrative amount is large for the separate income sources, it is much lower for total income.

The mean errors in the false negative and false positive groups are larger than for the true positives but still the errors by income source tend to reduce each other. In all false negative groups, earnings and at least one of the benefits are overstated on average suggesting misreporting the benefit as earnings or a different benefit. Furthermore, respondents with false positives understate their earnings (in the UA and ACLC group) and at least one of the unemployment benefits (in all groups), on average. Again, this suggests putting benefits into the wrong benefit category or bundling together benefits and earnings.

Looking closer at the different samples, in the UI false negative group the mean survey UI amount is zero by definition and so, UI is understated in the survey. The average survey and administrative amounts in UA are close though the survey amount is lower. In comparison, ACLC and earnings in particular are overstated with the mean error in earnings equal to 21.03% of the administrative income. This result is consistent with the coefficient in Table 4 showing that the more earnings are overstated in the survey compared to the administrative data, the less likely it is for respondents to report UI in the survey. And as the errors by income source partly compensate each other the error in total income is lower than that in earnings only, equal to 10.55%.

The UI false positive group contains some of the largest errors. The average UI amount in the survey is €3,643 which is large compared to the people who are actually on the programme and amounts to 44.44% of the mean income in the administrative data. But this error is partly compensated by underreporting the UA amount by €1,828 on average. Respondents in this group, however, overstate their earnings and ACLC, so the total error of 53% is high and by far the worst among all groups.

The patterns in the remaining groups for UA and ACLC also support the evidence for offsetting errors. In the UA false negative group, UA is understated in the survey by 30.56% of the mean administrative income but all other sources are overstated, especially earnings by 44.47% on average. The mean total error in income is thus lower, at 21.28%. Similarly, in the ACLC false negative group, the negative error in ACLC of –€635 is offset by the positive error in earnings of +€1,079, so the total error in income is +€531 or 5.64% of the administrative amount. Respondents in the UA false positive group overstate substantially UA, on average, by 55.31% of mean administrative income but this is being compensated by the negative errors in UI (-5.11%) and earnings (-39.46%). The errors by source in the ACLC false positive group are large but notably the mean error in total income is just over 3%.

[insert Table 6 here]

4.3 Implications of income source confusion

We examine the implications of income source confusion on estimates of some standard models in economics. While much literature has focused on the implications of measurement error in right hand side variable, we focus on two aspects here. First, we note that since misreporting of earnings is correlated with variables related to unemployment, coefficients on these variables can be biased. We highlight this using estimates of the returns to education in standard Mincer wage equation. Since unemployment is associated with lower education, reporting unemployment compensation in earnings inflates earnings disproportionately for that group.

Our second focus is on construction of a sample. When a key variable in sample choice is mismeasured, and that mismeasurement is related to the dependent variable, substantial bias can occur. We highlight this in estimation of the returns to job training. As is often the case, sample selection hinges on being unemployed in one time period. This combined with subsequent misallocation of earnings for those with persistent unemployment leads to substantial bias in the estimates of job training.

4.3.1 Bias in the returns to education

In Table 7 we document the distribution of education by benefit status: 'recipient' if receiving any of UI, UA or ACLC and 'non-recipient' if not receiving unemployment benefits. Column 1 shows education level: **low** (lower secondary or less, i.e. compulsory schooling), **middle** (secondary education including high school) and **high** (tertiary education including craftsman education and university degree). Columns 2-5 and 6-9 report the education shares in % among non-recipients and recipients, respectively, separately for men and women and by survey versus administrative data.

The key finding is that compared to *non-recipients*, both men and women *recipients* tend to be less educated, suggesting that unemployment is associated with lower education. In particular, there are relatively more low-educated and fewer high-educated persons among recipients than non-recipients. Looking at men first, among non-recipients only 9-10% are low-educated while almost 1/4 are high-educated. These shares are reversed for recipients: 1/4 or less (depending on the dataset) are low-educated while only 11% are high-educated. While the difference is less stark, 28-29% (depending on the dataset) of female recipients are low-educated, while only 18% of non-recipients are. Similarly, 17% of female non-recipients are high-educated, while only 11-12% of recipients fall into this category. Women on the whole tend to be less educated than men.[7]

As unemployment is associated with lower education and unemployment compensation gets misreported as earnings as established in the previous sections, it follows that the

[7]Table A.6, Table A.7 and Table A.8 in the Appendix include detailed results by UI, UA and ACLC status, respectively. The findings by benefit type are consistent with Table 7.

earnings of lower educated individuals will therefore be inflated. Looking at the total number of recipients by dataset (last row in Table 7), it is worth re-iterating that women are relatively more likely to report a receipt of at least one unemployment benefit in the survey than men, suggesting that women may be less likely than men to misreport benefits as earnings: number of women recipients is 30% lower in the survey than in the administrative data (1,280 survey versus 1,837 administrative cases), compared to the larger gap of 36% for men recipients (1,510 versus 2,345).

[insert Table 7 here]

Given the differences in education by benefits status, we test if the misreporting of unemployment compensation as earnings biases the returns to education. We focus on the sample of full-year and full-time workers including those who were unemployed but would have worked the full-year otherwise.[8] We construct two samples: one using the survey measures of earnings, UI, UA and ACLC to determine who is full-year, full-time worker and one using the administrative measures on earnings and unemployment benefits. We estimate standard Mincer wage equations separately for men and women based on survey versus administrative earnings. Table 8 presents the results. We control for the level of education – low, middle, high –, age, region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interaction between wave and year, interview type, and if the same interviewer as last year.

For *men*, the returns to high education, relative to low, are 55.7% in the survey versus the larger 62.6% in the administrative data. The returns to middle education, relative to low, are 28.6% in the survey compared to 32% in the administrative data. Using the Wald test, we jointly test the differences in the coefficients for middle and high education based on the two datasets and reject the null hypothesis of equivalence. Based on the results above, this is due to low education individuals reporting unemployment benefits as earnings, thus overstating their earnings.

[8] *Full-year and full-time workers* are identified based on receiving earnings for 12 months and if full-time employment is reported as the main economic activity in the survey for all 12 months of the reference period (information on working hours during the reference period is not available in the SILC). *Unemployed who would have worked full-year otherwise* are identified if the number of months in full-time employment and unemployment as the main economic activity add up to 12 or if receiving unemployment benefits (UI, UA or ACLC) for the entire year.

For *women*, we also find differences in the returns to education by dataset, although the differences are smaller than for men. Returns to high education, relative to low, for women are 62.2% in the survey compared to 65.8% in the administrative data. Testing these differences fails to reject the null hypothesis of equivalence. This suggests that women on the whole may be making fewer mistakes than men in reporting benefits versus earnings in the survey.

[insert Table 8 here]

The bias in the returns to education is the result of two effects: differences in the survey versus administrative sample and differences in the reports to earnings. For both men and women the survey sample in Table 8 is bigger than the administrative one. We find that there are many individuals who are unemployed, but do not report this in the survey. The sample is contaminated by workers who did not work the full year. Additionally, those workers report the UI and other unemployment benefits as earnings. Thus, the annual labour market earnings are mismeasured as well. The two problems re-enforce each other, leading to bias.

To find out which of the two is the bigger issue, we repeat the analysis in Table 9 by restricting the sample to the same people in both datasets, i.e. those with positive earnings in both the survey and administrative data. For women, most of the difference in the returns to education by dataset goes away, suggesting that differences in the sample dominate the results. In comparison, for men the differences mostly remain, so it is differences in the reports to earnings which mainly cause the bias in the returns to education. Using the Wald test, we again reject the null hypothesis of equivalence for the returns to education for men but not for women.

[insert Table 9 here]

4.3.2 Bias in the returns to job training

Job training programmes are of long term interest to social scientists and policy makers. In particular the efficacy of job training for unemployed workers is of interest as policy makers often want to tie unemployment programmes to job training.

Our sample is constructed of anyone who received unemployment benefits or reported being unemployed for at least one month in year t.[9] The outcome variables are earnings and employment in $t+1$, i.e. the year after receiving training. Two samples are constructed: one using only survey measures of UI, UA and ACLC to establish unemployment and one using administrative measures of UI, UA and ACLC.[10]

We estimate a standard Mincer wage regression with indicators for three types of job training: self-funded, employer-funded, and funded by the labour market agency. We control for the level of education, age, region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interview type, and if the same interviewer as last year. As sample size is relatively small, we cannot estimate models separately by sex. Hence, we estimate models for the combined sample of men and women.

Table 10 presents the results for earnings. We focus our discussion on the coefficient on labour market agency training, as this has the most policy relevance. The coefficient in the survey data is a robust 13.3% gain from training provided by the labour market agency. However, in the administrative data, this rises to 22.7%. Due to the under-reporting of UI, UA and ACLC, the survey sample is only 1,023 individuals, while the administrative sample is 1,435. The loss of nearly a third of the potential sample is clearly an issue. Likely as well, individuals who remain or become unemployed again in year $t+1$ are reporting benefits as earnings. We also note substantial bias in the coefficients on self-funded job training (-13.2% to -21.2%) and employer-funded training (-3.7% to 16.4%).

[insert Table 10 here]

Table 11 show results based on OLS estimation of being employed (having job market earnings) in $t+1$. The samples here are larger, but nearly the same difference between the administrative and survey sample (412 in table 10 and 397 in table 11).[11] We again focus on the coefficient on labour market agency. In the survey data, this coefficient is

[9] We exclude those who have reported for any month of the year to be a student, retired, with caring responsibilities, with a disability, in military service or inactive for any other reason.

[10] We also use the survey measure on main economic activity for each month of the year, indicating unemployment spells.

[11] The larger samples in the earner regressions arise naturally because this includes individuals who failed to obtain earnings in year $t+1$.

not statistically significant and is 2.9%. In the administrative data the coefficient rises to 5.7% and is now statistically significant at the 10% level. The larger positive coefficient here also helps to explain the differences found in the earnings regression of Table 10: those with job training in year t were more likely to be employed in year $t + 1$. We note that in Table 4 and Table 5 those who receive job training are more likely to report their UI and UA receipt. We also note that since they are employed in year $t + 1$, they are less likely to be receiving UI or UA, and thus less likely to be biasing their earnings value. This is consistent with recall bias found in Mathiowetz and Duncan (1988). Thus, the comparison for earnings is biased.

[insert Table 11 here]

In Table A.9 and Table A.10 we repeat the analysis for earnings and being employed in *t+1*, respectively, by restricting the sample to the same individuals in both the survey and administrative data. Most of the difference by dataset in the labour market agency coefficient goes away, confirming that it is the difference in the samples which leads to bias in the returns to training.

5 Conclusions

Administrative links to survey data have long been used to document measurement error in single variables (often earnings). Few (Bollinger and David 2000; Marquis and Moore 1990; Angel et al. 2018, 2019; Lynn et al. 2012; Meyer and Mittag 2019) studies have examined the possibility of programme confusion. We document both programme confusion in three Austrian unemployment benefit programmes, and also demonstrate that unemployment benefits may be reported as earnings. Many researchers have speculated that survey respondents may report benefits as earnings, or take other short cuts to reduce survey burden. Angel et al. (2018) show, using the same data, that while reporting errors in different sources of income lead to biases in income from those sources, the errors offset so the poverty estimate based on total income is far less biased.

We show that all three Austrian unemployment benefit programmes, UI, UA and ACLC are underreported. We also show that UA and ACLC are the most underreported and are associated with false positives for UI (the main program). Further, we show that even when participation is correctly reported, benefits are still misreported.

We show that earnings reports are higher compared to administrative data when any of the three programmes are underreported, suggesting that respondents know how much income they received, and report most (or all) of it as labour market earnings.

The implications for misreporting in unemployment programmes go far beyond simply measurement of unemployment as misreporting is likely to impact other studies through both mismeasured earnings and construction of samples. We demonstrate the bias in earnings regressions with two examples. In the first, we show that because unemployment is concentrated among workers with lower educational attainment, returns to education are biased down. This occurs because low education workers report unemployment as earnings, raising their earnings relative to those with higher education and thus fewer unemployment spells. In our second example, we examine the implications for estimation of the impact of job training programs. As is typical in this literature, we select our sample based on those experiencing unemployment spells. We show that failure to report unemployment benefits can lead to substantial sample bias when selecting on these benefits. The bias is further exacerbated by the fact that those with persistent unemployment may misreport continued unemployment as earnings in periods after training occurs. Our results suggest that interactions between unemployment spells and any variable or sample construction may be of significant concern.

Moreover, the implications for our study go well beyond the Austrian unemployment programmes or the SILC data. If survey respondents provide accurate reports of their total income, but by doing so misclassify benefit income as earnings, this can bias both estimates of programme participation (Bollinger and David 1997) and bias earnings regressions. Further, these types of misreporting can bias estimates of how programmes such as UA help address poverty.

References

Aktinson, A. B., and Marlier, E. (Eds.) (2010). *Income and Living Conditions in Europe.* Luxembourg: Publications Office of the European Union.

Angel, S., Disslbacher, F., Humer, S., and Schnetzer, M. (2019). "What Did You Really Earn Last Year?: Explaining Measurement Error in Survey Income Data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(4), 1411–1437.

Angel, S., Heuberger, R., and Lamei, N. (2018). "Differences Between Household Income from Surveys and Registers and How These Affect the Poverty Headcount: Evidence from the Austrian SILC." *Social Indicators Research*, *138*, 575–603.

Balarajan, M., and Collins, D. (2013). "A Review of Questions Asked About Receipt of State Benefits on the Family Resources Survey." *Department for Work and Pensions Working paper No 115*.

Bollinger, C. R., and David, M. H. (1997). "Modeling Discrete Choice With Response Error: Food Stamp Participation." *Journal of the American Statistical Association*, *92*(439), 827–835.

Bollinger, C. R., and David, M. H. (2000). "Differential Reporting of Food Stamps and AFDC: Explanations and Conjectures." *Proceedings of the Survey Methods Sections of the American Statistical Association*, 256–260.

Bollinger, C. R., and David, M. H. (2005). "I Didn't Tell, and I Won't Tell: Dynamic Response Error in the SIPP." *Journal of Applied Econometrics*, *20*(4), 563–569.

Bollinger, C. R., Hirsh, B. T., Hokayem, C. M., and Ziliak, J. P. (2018). "The Good, The Bad and The Ugly: Measurement Error, Non-response andAdministrative Mismatch in the CPS." *draft*.

Bollinger, C. R., Hirsh, B. T., Hokayem, C. M., and Ziliak, J. P. (2019). "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch." *Journal of Political Economy*, *127*(5), 2143–2185.

Celhay, P. A., Meyer, B. D., and Mittag, N. (2021). "Errors in Reporting and Imputation of Government Benefits and Their Implications." *NBER Working Paper 29184*, *11*(2), 176–204.

Currie, J. (2006). "The Take-Up of Social Benefits." In A. J. Auerbach, D. Card, and J. M. Quigley (Eds.), *Public Policy and the Income Distribution*, 80–148, New York: Russell Sage Foundation.

Fuchs, M., Gasior, K., Premrov, T., Hollan, K., and Scoppetta, A. (2020). "Falling through the Social Safety Net? Analysing Non-take-up of Minimum Income Benefit and Monetary Social Assistance in Austria." *Social Policy & Administration*, *54*(5), 827–843.

Heuberger, R., Glaser, T., and Kafka, E. (2013). "The Use of Register Data in the Austrian SILC Survey." In M. Jäntti, V.-M. Törmälehto, and E. Marlier (Eds.), *The Use of Registers in the Context of EU–SILC: Challenges and Opportunities*, 141–152, Luxembourg: Publications Office of the European Union.

Hokayem, C., Bollinger, C., and Ziliak, J. P. (2015). "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association*, *110*(511), 935–945.

Jenkins, S. P., Lynn, P., Jäckle, A., and Sala, E. (2008). "The Feasibility of linking household survey and administrative record data: New evidence for Britain." *International Journal of Social Research Methodology*, *11*(1), 29–43.

Jenkins, S. P., and Rios-Avila, F. (2021a). "Measurement Error in Earnings Data: Replication of Meijer, Rohwedder, and Wansbeek's Mixture Model Approach to Combining Survey and Register Data." *Journal of Applied Econometrics*, 1–10.

Jenkins, S. P., and Rios-Avila, F. (2021b). "Reconciling Reports: Modelling Employment Earnings and Measurement Errors Using Linked Survey and Administrative Data." *IZA DP No. 14405.*

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

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." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(1), 289–308.

Marquis, K. H., and Moore, J. C. (1990). "Measurement Errors in SIPP Program Reports." *Proceedings of the Bureau of the Census Annual Research Conference*, 721–745.

Mathiowetz, N. A., and Duncan, G. J. (1988). "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment." *Journal of Business & Economic Statistics*, *6*(2), 221–229.

Meyer, B. D., and Mittag, N. (2019). "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net." *American Economic Journal: Applied Economics.*

Meyer, B. D., Mittag, N., and Goerge, R. M. (forthcoming). "Errors In Survey Reporting and Imputation and Their Effects On Estimates Of Food Stamp Program Participation." *Journal of Human Resources*, *11*(2), 176–204.

Meyer, B. D., Mok, W. K. C., and Sullivan, J. X. (2015). "Household Surveys in Crisis." *Journal of Economic Perspectives*, *29*(4), 199–226.

Paulus, A. (2015). "Tax Evasion and Measurement Error: An Econometric Analysis of Survey Data Linked with Tax Records." *ISER Working Paper No. 2015-10.*

Statistics Austria (2014). "Methodenbericht zur Rückrechnung von EU-SILC 2008-2011 auf Basis von Verwaltungsdaten."

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

6 Tables

Table 1: Sample characteristics: SILC 2008-2011

	All	Men	Women
Age	47.4	46.6	48.1
Education:			
Low	23.4	17.3	29.0
Middle	59.7	61.9	57.7
High	16.9	20.8	13.4
With partner	62.5	66.1	59.2
Country of birth:			
Austria	86.0	86.8	85.4
EU-27	5.9	5.1	6.6
Other	8.1	8.1	8.0
Full-time, full-year employed	30.8	41.8	21.0
With survey earnings	45.7	40.6	49.8
Proxy interview:			
No	82.6	80.4	84.5
Partner	7.8	11.0	4.8
Someone else	9.6	8.5	10.6
Job training:			
Did not take	67.6	67.1	68.0
Mostly paid with own resources	5.1	4.7	5.4
Employer	13.8	16.3	11.5
Labour market agency	1.4	1.3	1.5
Other institutions	.8	.8	.8
Missing/ n/a	11.4	9.8	12.8
UI in the survey:			
Missing/ imputed	.1	.2	.1
No receipt	94.8	93.8	95.6
With receipt:	5.1	6.0	4.2
1 year	74.2	71.9	77.0
Multiple years	25.8	28.1	23.0
UA in the survey:			
No receipt	98.3	98.2	98.5
With receipt:	1.7	1.8	1.5
1 year	73.5	70.1	76.6
Multiple years	26.5	29.9	23.4
ACLC in the survey:			
Missing/ imputed	.1	.	.
No receipt	99.2	.	.
With receipt:	.7	.	.
1 year	88.6	.	.
Multiple years	11.4	.	.
Observations:			
1 wave only	8,903	4,231	4,672
2 waves	12,559	5,951	6,608
3 waves	15,461	7,302	8,159
4 waves	8,047	3,790	4,257
All observations	44,970	21,274	23,696

Notes: The sample includes individuals aged 16+. The table shows the mean age and the share (in %) of individuals by different characteristics. Groups with missing/ imputed cases in survey earnings and UA, and by ACLC receipt for men versus women are too small to disclose. *Source:* Own calculations with the SILC.

Table 2: Misreporting of unemployment benefits

		UI	UA	ACLC	Total
False negative	%	42.0	51.7	76.4	37.7
	n	1,399	683	185	1,517
a) False positive	%	.8	.2	.6	.6
	n	317	105	263	245
b) False positive	%	12.5	1.3	5.0	.
	n	151	44	202	.
Conditional on true positive:					
Mean administrative amount		2,847	4,419	2,138	3,646
Mean survey amount		3,030	4,397	2,690	3,784
Absolute error in % of administrative amount:					
$<$ 10%		23.4	28.1	.	27.0
10-30%		29.8	25.9	.	29.6
30-50%		17.8	14.9	.	17.1
50-70%		9.6	8.6	.	9.2
$\geq$ 70%		19.4	22.4	.	17.0
SD of error		1,840	2,307	2,130	1,990
Correlation administrative and survey amounts		.74	.68	.64	.78

Notes: Total = UI + UA + ACLC. False negative in % is the share of benefit recipients according to the administrative data who do not report the benefit in the survey. Two definitions of false positive are considered: a) in % equals the share of benefit non-recipients according to the administrative data who report the benefit in the survey; b) in % equals the number of benefit non-recipients according to the administrative data who report a receipt in the survey, as a share of the number of recipients of the other two benefits in the administrative data (by definition, there is no estimate for the total). Mean amounts and errors are based on the sample of true positives, i.e. those who receive benefits in both the administrative and survey data. The error in amount equals the difference between the survey and administrative amount. Groups with different size of error in ACLC too small to disclose. Observations with missing/ imputed benefit values are excluded. Sample is restricted to those aged 16+. *Source:* Own calculations with the SILC.

Table 3: Combinations of income receipt, given administrative receipt

	survey income									
	UI only	UA only	Earn. only	UI + earn.	UA+ earn.	UI + UA	UI+ UA + earn.	ACLC + others	none	*total n*
admin data										
UI only	38.3	.	11.2	10.2	.	.	.	.	31.1	206
UA only	11.0	48.6	6.4	.	.	.	.	.	22.8	346
Earn. only	.	.	91.0	.3	.	.	.	.1	8.5	21,042
UI + earn.	2.7	.	38.0	51.1	.	.	.	2.4	2.9	2,490
UA + earn.	.	12.0	27.3	14.3	18.7	.	6.7	7.7	.	300
UI + UA	.	.	.	.	.	16.8	.	18.5	16.8	119
UI + UA + earn.	.	.	23.4	28.8	5.4	.	16.6	6.8	.	368
ACLC + others	.	.	31.4	8.7	.	.	.	22.3	29.8	242
none	.1	.2	5.4	.3	.	.	.	.1	93.8	18,395
total n	304	328	22,348	1,652	112	73	112	317	19,623	

Notes: UI = unemployment insurance benefit; UA = unemployment assistance; ACLC = assistance for covering living costs; earn. = earnings. Missing values (.) indicate cells with too few cases (cannot be disclosed). Numbers on the diagonal are highlighted in gray. Otherwise, numbers with a value of 10% or more are highlighted in light red. Observations with missing/ imputed administrative/ survey values are excluded. Sample is restricted to those aged 16+. *Source:* Own calculations with the SILC.

Table 4: Probability of reporting the unemployment insurance (UI) benefit in the survey, conditional on receiving it

	(1)	(2)
Constant	-.384***	-.024
	(.069)	(.153)
Woman		-.036*
		(.020)
Ln admin UI	.118***	.130***
	(.008)	(.012)
Admin earnings (in thousand)	-.008***	-.008***
	(.001)	(.001)
Survey–admin earnings (in thousand)	-.013***	-.012***
	(.001)	(.001)
Earnings: true –	ref	ref
false +	.161**	.173***
	(.064)	(.064)
false –	-.048	-.004
	(.042)	(.043)
true +	.270***	.278***
	(.034)	(.036)
UA: true –	ref	ref
false +	-.143*	-.145*
	(.075)	(.075)
false –	-.002	-.004
	(.029)	(.029)
true +	.064*	-.006
	(.037)	(.038)
ACLC: true –	ref	ref
false +	.043	.019
	(.044)	(.047)
Admin benefit duration (in months)		-.017***
		(.006)
Job training: did not take		ref
Labour market agency		.116***
		(.032)
Mostly paid with own resources		.053
		(.040)
Employer		.029
		(.031)
Other institutions		.026
		(.078)
No proxy		ref
partner is proxy		-.100***
		(.032)
someone else is proxy		-.105***
		(.031)
Controls	No	Yes
R-squared	.137	.181
N	3155	3146

Notes: This table shows an estimation of a linear probability model. The dependent variable equals 1 if the benefit amount is positive in both the survey and administrative data; and 0 if the survey amount is 0 while the administrative amount is positive. 'True +' implies positive income amounts in both the survey and administrative data; 'false +' means positive amount in the survey and zero in the administrative data; 'false –' means zero in the survey and positive amount in the administrative data; and 'true –' means zero amounts in both the survey and administrative data. Column (2) adds controls for: age group (in 5-year age bands), number of children in the household (1, 2, 3+), number of adults in the household (1, 2, 3+), the highest achieved education level (low, middle, high), health status (5 categories), region (Vienna, borough with more than 100,000 residents, borough with 10,000-100,000 residents, borough with less than 10,000 residents), occupation (10 categories), industry (25 categories), being a civil servant, country of birth (6 categories), wave (interviewed for the 1st, 2nd, 3rd, 4th time), interaction between wave and year (2008 to 2011), interview type (in person or by phone), same interviewer as last year. Sample is restricted to those aged 16+. Observations with missing/ imputed administrative/ survey UI or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: UA with missing/ imputed values; ACLC false -, true + and with missing/ imputed values; and job training with missing or n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses. *Source:* Own calculations with the SILC.

Table 5: Probability of reporting the unemployment assistance (UA) in the survey, conditional on receiving it

	(1)	(2)
Constant	-.109	.730***
	(.097)	(.186)
Woman		-.021
		(.029)
Ln admin UA	.094***	.033**
	(.011)	(.016)
Admin earnings (in thousand)	-.012***	-.009***
	(.003)	(.003)
Survey–admin earnings (in thousand)	-.015***	-.013***
	(.002)	(.002)
Earnings: true –	ref	ref
false +	-.279***	-.219***
	(.063)	(.064)
false –	.023	.020
	(.040)	(.042)
true +	.011	.009
	(.040)	(.042)
UI: true –	ref	ref
false +	-.312***	-.310***
	(.042)	(.042)
false –	-.107***	-.053
	(.039)	(.041)
true +	-.058*	-.060
	(.035)	(.039)
missing/imputed	-.039	-.096**
	(.043)	(.045)
ACLC: true –	ref	ref
false +	.177***	.042
	(.044)	(.049)
Admin benefit duration (in months)		.017***
		(.006)
Job training: did not take		ref
Labour market agency		.180***
		(.036)
Mostly paid with own resources		.014
		(.070)
Employer		-.039
		(.072)
No proxy		ref
partner is proxy		-.096
		(.062)
someone else is proxy		.024
		(.051)
Controls	No	Yes
R-squared	.234	.302
N	1262	1247

Notes and Source: See Table 4. Observations with missing/ imputed administrative/ survey UA or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: ACLC false -, true + and with missing/ imputed values; and job training by other institutions and with n/a values.

Table 6: Mean survey and administrative amounts by benefits true positive, false negative and false positive

	Survey	Admin	Δ	... in % of Admin Income	SD of Δ	ρ
Conditional on UI true +						
UI	2,987	2,802	185	1.09	1,804	.74
UA	166	323	-156	-.92	756	.69
ACLC	81	4	77	.45	571	.00
Earnings	13,660	13,809	-149	-.88	11,268	.64
Income	16,894	16,938	-44	-.26	11,075	.63
Observations	1,801					
Conditional on UI false −						
UI	0	1,946	-1,946	-11.00	2,068	.
UA	309	342	-34	-.19	1,247	.57
ACLC	127	3	125	.71	1,044	-.01
Earnings	19,126	15,405	3,722	21.03	11,437	.67
Income	19,562	17,696	1,867	10.55	11,351	.65
Observations	1,340					
Conditional on UI false +						
UI	3,643	0	3,643	44.44	2,903	.
UA	442	2,271	-1,828	-22.30	2,933	.30
ACLC	122	10	112	1.36	689	-.01
Earnings	8,336	5,917	2,419	29.51	11,219	.35
Income	12,543	8,198	4,345	53.00	11,051	.27
Observations	299					
Conditional on UA true +						
UI	1,115	1,034	81	.98	1,710	.62
UA	4,500	4,558	-58	-.70	2,295	.70
ACLC	165	3	162	1.96	678	-.01
Earnings	2,613	2,664	-50	-.61	2,696	.84
Income	8,393	8,258	135	1.64	3,009	.79
Observations	514					
Conditional on UA false −						
UI	1,903	1,279	623	6.56	3,036	.29
UA	0	2,904	-2,904	-30.56	2,770	.
ACLC	80	2	78	.82	527	-.01
Earnings	9,541	5,316	4,225	44.47	9,765	.53
Income	11,524	9,501	2,022	21.28	9,408	.53
Observations	599					
Conditional on UA false +						
UI	1,226	1,578	-351	-5.11	2,412	.53
UA	3,807	0	3,807	55.31	3,062	.
ACLC	119	0	119	1.73	462	.
Earnings	2,588	5,304	-2,716	-39.46	8,669	.25
Income	7,740	6,882	859	12.48	8,300	.50
Observations	94					
Conditional on ACLC false −						
UI	916	808	108	1.14	1,111	.78
UA	37	57	-20	-.21	431	-.02
ACLC	0	635	-635	-6.74	884	.
Earnings	8,996	7,918	1,079	11.45	4,233	.89
Income	9,949	9,418	531	5.64	4,391	.87
Observations	41					
Conditional on ACLC false +						
UI	2,103	2,388	-285	-2.67	3,086	.46
UA	1,416	1,787	-371	-3.47	2,132	.69
ACLC	2,522	0	2,522	23.62	3,034	.
Earnings	4,966	6,501	-1,535	-14.38	17,798	.63
Income	11,007	10,675	331	3.10	17,835	.62
Observations	211					

Notes: Income = UI + UA + ACLC + Earnings. Δ = mean survey – mean administrative amount. SD of Δ = standard deviation of (survey amount – administrative amount). ρ = correlation between survey and administrative amounts. Group with ACLC true positives is too small to disclose. Sample includes those aged 16+. Observations with missing/ imputed earnings or benefits (UI, UA or ACLC) are excluded.

Table 7: Education shares (%) by unemployment benefits status

	Non-recipients				Recipients			
	Men		**Women**		**Men**		**Women**	
Education	Survey	Admin	Survey	Admin	Survey	Admin	Survey	Admin
low	9.6	9.3	18.4	18.3	24.7	21.7	29.0	27.8
middle	67.1	66.6	64.6	64.4	64.0	67.5	58.9	61.0
high	23.3	24.1	17.0	17.2	11.3	10.8	12.1	11.3
total n	14,434	13,639	16,029	15,570	1,510	2,345	1,280	1,837

Notes: Recipient status based on a receipt of at least one of UI, UA or ACLC. Education refers to the highest level of education achieved, following on the International Standard Classification of Education (ISCED): low (lower secondary or less, i.e. compulsory schooling); middle (secondary education including high school); high (tertiary education including craftsman education and university degree). Observations with missing/ imputed administrative/ survey benefits are excluded. Sample is restricted to those aged 19-64.

Table 8: Returns to education: log-earnings regression

	Men		Women	
	Survey (1)	**Admin** (2)	**Survey** (3)	**Admin** (4)
Constant	10.135***	10.125***	9.884***	9.874***
	(.048)	(.059)	(.066)	(.088)
Education: low	ref	ref	ref	ref
middle	.286***	.320***	.319***	.333***
	(.020)	(.024)	(.024)	(.031)
high	.557***	.626***	.622***	.658***
	(.022)	(.027)	(.029)	(.038)
Controls	Yes	Yes	Yes	Yes
R-squared	.272	.325	.298	.339
N	10017	9564	5570	5307

Notes: OLS estimation of log-earnings regressions. Sample is restricted to those aged 19-64; and based on survey versus administrative information for full-year and full-time employed or unemployed, but who would have worked full-year otherwise. Education refers to the highest level of education achieved, following on the International Standard Classification of Education (ISCED): low (lower secondary or less, i.e. compulsory schooling); middle (secondary education including high school); high (tertiary education including craftsman education and university degree). Columns (1) and (3) are based on survey earnings and columns (2) and (4) on administrative earnings. Observations with imputed administrative/ survey earnings, UI, UA or ACLC are excluded. The controls include: age group (5-year bands), region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interaction between wave and year, interview type, and if the same interviewer as last year. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses.

Table 9: Returns to education: log-earnings regression
(sample with both positive administrative and survey earnings)

	Men		Women	
	Survey (1)	**Admin** (2)	**Survey** (3)	**Admin** (4)
Constant	10.156***	10.152***	9.834***	9.789***
	(.048)	(.054)	(.065)	(.072)
Education: low	ref	ref	ref	ref
middle	.265***	.283***	.315***	.309***
	(.019)	(.022)	(.023)	(.026)
high	.527***	.590***	.610***	.613***
	(.022)	(.025)	(.029)	(.031)
Controls	Yes	Yes	Yes	Yes
R-squared	.267	.295	.284	.301
N	9376	9376	5131	5131

Notes and Source: See Table 8. Sample is further restricted to those with both positive administrative and survey earnings.

Table 10: Returns to job training: log-earnings regression

	Survey (1)	**Admin** (2)
Constant	9.398***	8.899***
	(.165)	(.178)
Job training: did not take	ref	ref
Labour market agency	.133*	.227***
	(.069)	(.086)
Mostly paid with own resources	-.132	-.212**
	(.104)	(.107)
Employer	-.037	.164*
	(.093)	(.089)
Job-training-year earnings (in thousand)	.034***	.038***
	(.002)	(.002)
Controls	Yes	Yes
R-squared	.481	.475
N	1023	1435

Notes: OLS estimation with log-earnings in *t+1* as the outcome variable. Sample is restricted to those aged 19-64; based on survey versus administrative information on being unemployed in *t*, i.e. received unemployment benefits (UI, UA or ACLC) or reported they were unemployed for parts of the year; and with positive survey versus administrative earnings in *t+1*. Column (1) is based on survey and column (2) on administrative earnings while the indicator for job training is always based on the survey. The controls include: education, age group (5-year bands), region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interview type, and if the same interviewer as last year. Observations with imputed administrative/ survey earnings, UI, UA or ACLC are excluded. Cells with too few observations cannot be disclosed and are not shown: i.e. job training by other institutions and with missing/ n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses.

Table 11: Returns to job training: being an earner regression

	Survey (1)	**Admin** (2)
Constant	.654***	.763***
	(.073)	(.067)
Job training: did not take	ref	ref
Labour market agency	.029	.057*
	(.029)	(.031)
Mostly paid with own resources	-.037	-.002
	(.047)	(.041)
Employer	-.033	.013
	(.047)	(.037)
Job-training-year earnings (in thousand)	.007***	.006***
	(.001)	(.001)
Controls	Yes	Yes
R-squared	.382	.292
N	1549	1946

Notes: OLS estimation with being an earner in *t+1* (yes=1 and no=0) as the outcome variable. Sample is restricted to those aged 19-64; based on survey versus administrative information on being unemployed in *t*, i.e. received unemployment benefits (UI, UA or ACLC) or reported they were unemployed for parts of the year. Column (1) is based on survey and column (2) on administrative earnings while the indicator for job training is always based on the survey. The controls include: education, age group (5-year bands), region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interview type, and if the same interviewer as last year. Observations with missing/ imputed administrative/ survey earnings, UI, UA or ACLC are excluded. Cells with too few observations cannot be disclosed and are not shown: i.e. job training by other institutions and with missing/ n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses.

Supplementary Materials

A Tables

Table A.1: Probability of reporting the unemployment insurance (UI) benefit in the survey, conditional on receiving it

	(1)	(2)
Constant	-.384***	-.024
	(.069)	(.153)
Woman		-.036*
		(.020)
Ln admin UI	.118***	.130***
	(.008)	(.012)
Admin earnings (in thousand)	-.008***	-.008***
	(.001)	(.001)
Survey–admin earnings (in thousand)	-.013***	-.012***
	(.001)	(.001)
Earnings: true –	ref	ref
false +	.161**	.173***
	(.064)	(.064)
false –	-.048	-.004
	(.042)	(.043)
true +	.270***	.278***
	(.034)	(.036)
UA: true –	ref	ref
false +	-.143*	-.145*
	(.075)	(.075)
false –	-.002	-.004
	(.029)	(.029)
true +	.064*	-.006
	(.037)	(.038)
ACLC: true –	ref	ref
false +	.043	.019
	(.044)	(.047)
Admin benefit duration (in months)		-.017***
		(.006)
Job training: did not take		ref
Labour market agency		.116***
		(.032)
Mostly paid with own resources		.053
		(.040)
Employer		.029
		(.031)
Other institutions		.026
		(.078)
No proxy		ref
partner is proxy		-.100***
		(.032)
someone else is proxy		-.105***
		(.031)
Age: 35-39		ref
16-19		.009
		(.054)
20-24		-.080**
		(.035)
25-29		-.019
		(.032)
30-34		-.088***
		(.033)
40-44		-.044
		(.031)
45-49		-.074**
		(.033)
50-54		-.088**
		(.036)
55-59		-.147***
		(.043)
60-64		-.156**

	(.074)
Education: low	ref
middle	-.018
	(.022)
high	.016
	(.034)
Country of birth: Austria	ref
EU15/EFTA	.039
	(.048)
new EU12	.001
	(.051)
former Yugosl.	-.001
	(.031)
Turkey	.025
	(.044)
other	-.065
	(.044)
In a couple	.020
	(.027)
No children in the hh	ref
1 child	.038*
	(.022)
2 children	-.026
	(.028)
3+ children	.042
	(.041)
1 adult in the hh	ref
2 adults	-.084***
	(.032)
3+ adults	-.064**
	(.031)
Region: Vienna	ref
100,000+ residents	.054
	(.035)
10,000-100,000 residents	.012
	(.029)
less than 10,000 residents	-.010
	(.025)
Civil servant: no	ref
missing	-.213
	(.469)
Occupation: elementary	ref
senior officials and managers	.010
	(.075)
professionals	-.016
	(.058)
associate prof. and technical	.004
	(.039)
clerks (admin and secretarial)	.009
	(.044)
service and sales workers	.034
	(.035)
skilled agricultural	-.030
	(.077)
craft and trades workers	-.023
	(.034)
plant and machine operators	.008
	(.040)
n/a	.122
	(.475)
missing	.024
	(.054)
Industry: manufacturing	ref
construction	-.071
	(.057)
trade	-.064
	(.050)
transportation	-.172**
	(.077)
accommodation and food	-.038
	(.057)
info and communication	-.162*
	(.099)

admin and support services	-.108
	(.073)
public admin., defence etc	-.112
	(.091)
education	-.004
	(.101)
health services	.025
	(.094)
residential care and social work	-.098
	(.100)
n/a as hasn't worked	-.236***
	(.073)
Health: very bad	ref
very good	-.043
	(.079)
good	-.032
	(.078)
fair	.015
	(.079)
bad	.029
	(.083)
Month of interview: Mar	ref
Apr	-.071**
	(.033)
May	-.093***
	(.031)
Jun	-.087***
	(.033)
Jul	-.069*
	(.036)
Aug	-.143***
	(.041)
Sep	-.136***
	(.046)
Oct	-.130
	(.097)
Interview in person	ref
interview by phone	.132***
	(.045)
Same interviewer: yes	ref
no	-.008
	(.043)
missing	.465
	(.464)
n/a	.008
	(.051)
don't know	.497*
	(.263)
Wave 1	ref
wave 2	.025
	(.047)
wave 3	.002
	(.051)
wave 4	.050
	(.053)
Wave 1 × year 2008	ref
wave=1 × year=2009	.003
	(.072)
wave=1 × year=2010	.010
	(.071)
wave=1 × year=2011	-.040
	(.074)
wave=2 × year=2009	-.029
	(.060)
wave=2 × year=2010	.013
	(.061)
wave=2 × year=2011	-.043
	(.062)
wave=3 × year=2009	.001
	(.068)
wave=3 × year=2010	.030
	(.065)
wave=3 × year=2011	-.012

		(.066)
wave=4 × year=2009		-.133*
		(.072)
wave=4 × year=2010		-.022
		(.066)
wave=4 × year=2011		-.093
		(.071)
R-squared	.137	.181
N	3155	3146

Notes: This table shows an estimation of a linear probability model. The dependent variable equals 1 if the benefit amount is positive in both the survey and administrative data; and 0 if the survey amount is 0 while the administrative amount is positive. 'True +' implies positive income amounts in both the survey and administrative data; 'false +' means positive amount in the survey and zero in the administrative data; 'false –' means zero in the survey and positive amount in the administrative data; and 'true –' means zero amounts in both the survey and administrative data. Sample is restricted to those aged 16+. Observations with missing/ imputed administrative/ survey UI or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: UA with missing/ imputed values; ACLC false -, true + and with missing/ imputed values; job training with missing or n/a values; age group 65-69; being a civil servant; industry equal to agriculture, mining and quarrying, electricity, gas, steam and air conditioning supply, water supply, sewage, waste management and remediation, finance, real estate, science, other professional, scientific and technical activities, arts, entertainment and recreation, other services, activities of households as employers, undifferentiated goods- and services-producing activities of households for own use and with no answer; health status with missing values; and month of interview equal to November. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses. *Source:* Own calculations with the SILC.

Table A.2: Probability of reporting the unemployment assistance (UA) in the survey, conditional on receiving it

	(1)	(2)
Constant	-.109	.730***
	(.097)	(.186)
Woman		-.021
		(.029)
Ln admin UA	.094***	.033**
	(.011)	(.016)
Admin earnings (in thousand)	-.012***	-.009***
	(.003)	(.003)
Survey–admin earnings (in thousand)	-.015***	-.013***
	(.002)	(.002)
Earnings: true –	ref	ref
false +	-.279***	-.219***
	(.063)	(.064)
false –	.023	.020
	(.040)	(.042)
true +	.011	.009
	(.040)	(.042)
UI: true –	ref	ref
false +	-.312***	-.310***
	(.042)	(.042)
false –	-.107***	-.053
	(.039)	(.041)
true +	-.058*	-.060
	(.035)	(.039)
missing/imputed	-.039	-.096**
	(.043)	(.045)
ACLC: true –	ref	ref
false +	.177***	.042
	(.044)	(.049)
Admin benefit duration (in months)		.017***
		(.006)
Job training: did not take		ref
Labour market agency		.180***
		(.036)
Mostly paid with own resources		.014
		(.070)
Employer		-.039
		(.072)
No proxy		ref
partner is proxy		-.096
		(.062)
someone else is proxy		.024
		(.051)

Age: 35-39	ref
16-19	-.113
	(.094)
20-24	-.019
	(.058)
25-29	-.024
	(.050)
30-34	-.030
	(.047)
40-44	.064
	(.047)
45-49	-.011
	(.050)
50-54	-.003
	(.051)
55-59	-.091*
	(.053)
60-64	-.012
	(.090)
Education: low	ref
middle	-.050*
	(.029)
high	-.026
	(.051)
Country of birth: Austria	ref
new EU12	-.097
	(.069)
former Yugosl.	-.060
	(.049)
Turkey	-.043
	(.052)
other	.023
	(.055)
In a couple	-.095**
	(.038)
No children in the hh	ref
1 child	.047
	(.036)
2 children	.053
	(.043)
3+ children	.145***
	(.056)
1 adult in the hh	ref
2 adults	.018
	(.041)
3+ adults	-.096**
	(.045)
Region: Vienna	ref
100,000+ residents	-.021
	(.050)
10,000-100,000 residents	.019
	(.038)
less than 10,000 residents	-.042
	(.034)
Occupation: elementary	ref
professionals	-.005
	(.096)
associate prof. and technical	-.006
	(.074)
clerks (admin and secretarial)	-.005
	(.080)
service and sales workers	-.074
	(.061)
craft and trades workers	.027
	(.072)
plant and machine operators	-.010
	(.082)
n/a	-.186**
	(.084)
missing	-.125
	(.099)
Industry: manufacturing	ref
construction	-.014

	(.058)
trade	-.107**
	(.052)
transportation	-.089
	(.085)
accommodation and food	-.014
	(.060)
admin and support services	-.116
	(.075)
public admin., defence etc	-.104
	(.102)
health services	-.059
	(.104)
other services	-.097
	(.089)
n/a as hasn't worked	-.331***
	(.080)
Health: very bad	ref
very good	.000
	(.070)
good	.059
	(.067)
fair	.069
	(.065)
bad	.017
	(.067)
Month of interview: Mar	ref
Apr	-.036
	(.049)
May	.006
	(.048)
Jun	-.092*
	(.051)
Jul	-.027
	(.054)
Aug	-.094
	(.063)
Sep	-.004
	(.070)
Interview in person	ref
interview by phone	.130*
	(.068)
Same interviewer: yes	ref
no	-.123**
	(.056)
n/a	-.105
	(.077)
Wave 1	ref
wave 2	-.064
	(.069)
wave 3	-.087
	(.074)
wave 4	-.010
	(.082)
Wave 1 × year 2008	ref
wave=1 × year=2009	-.100
	(.106)
wave=1 × year=2010	-.060
	(.105)
wave=1 × year=2011	.062
	(.106)
wave=2 × year=2009	-.020
	(.086)
wave=2 × year=2010	.043
	(.085)
wave=2 × year=2011	-.016
	(.087)
wave=3 × year=2009	.102
	(.094)
wave=3 × year=2010	.073
	(.091)
wave=3 × year=2011	.027
	(.093)

wave=4 × year=2009		-.154
		(.102)
wave=4 × year=2010		-.169*
		(.098)
wave=4 × year=2011		-.084
		(.108)
R-squared	.234	.302
N	1262	1247

Notes and Source: See Table A.1. Observations with missing/ imputed administrative/ survey UA or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: ACLC false -, true + and with missing/ imputed values; job training by other institutions and with n/a values; age group 65-69; country birth equal to EU15/EFTA; being a civil servant; industry equal to agriculture, mining and quarrying, electricity, gas, steam and air conditioning supply, water supply, sewage, waste management and remediation, information and communication, finance, real estate, other professional, scientific and technical activities, education, residential care, arts, entertainment and recreation, activities of households as employers, undifferentiated goods- and services-producing activities of households for own use and with no answer; occupation equal to senior officials and managers and skilled agricultural; and month of interview equal to October and November.

Table A.3: Misreporting of unemployment benefits by sex

		UI		UA		ACLC		Total	
		Men	Women	Men	Women	Men	Women	Men	Women
False negative	%	43.9	39.2	53.5	49.5	77.3	76.0	39.6	35.1
	n	859	540	378	305	58	127	926	591
a) False positive	%	.9	.6	.2	.3	.6	.6	.6	.6
	n	179	138	44	61	122	141	118	127
b) False positive	%	14.1	10.5	.	.	4.1	6.2	.	.
	n	94	57	.	.	95	107	.	.
Conditional on true positive:									
Mean admin amount		2,999	2,649	5,116	3,668	.	.	3,911	3,304
Mean survey amount		3,205	2,802	5,061	3,682	.	.	4,070	3,416
Absolute error in % of admin amount:									
< 10%		20.8	26.8	25.8	30.7	.	.	25.3	29.1
10-30%		29.2	30.6	23.9	28.1	.	.	29.7	29.5
30-50%		18.4	16.9	16.0	13.9	.	.	17.2	17.0
50-70%		10.4	8.7	9.5	7.6	.	.	9.5	8.9
≥ 70%		21.2	17.0	24.8	19.8	.	.	18.2	15.5
SD of error		2,001	1,607	2,647	1,876	.	.	2,194	1,691
Correlation admin and survey amounts		.73	.75	.64	.70	.	.	.77	.80

Notes: Total = UI + UA + ACLC. False negative in % is the share of benefit recipients according to the administrative data who do not report the benefit in the survey. Two definitions of false positive are considered: a) in % equals the share of benefit non-recipients according to the administrative data who report the benefit in the survey; b) in % equals the number of benefit non-recipients according to the administrative data who report a receipt in the survey, as a share of the number of recipients of the other two benefits in the administrative data (by definition, there is no estimate for the total). Mean amounts and errors are based on the sample of true positives, i.e. those who receive benefits in both the administrative and survey data. The error in amount equals the difference between the survey and administrative amount. Groups with UA false positive (b) and ACLC true positive too small to disclose. Observations with missing/ imputed benefit values are excluded. Sample is restricted to those aged 16+.

Table A.4: Probability of reporting the unemployment insurance (UI) benefit in the survey, conditional on receiving it

	Men		Women	
	(1)	(2)	(3)	(4)
Constant	-.535***	-.121	-.231**	.118
	(.094)	(.202)	(.102)	(.242)
Ln admin UI	.123***	.128***	.111***	.133***
	(.010)	(.017)	(.012)	(.019)
Admin earnings (in thousand)	-.008***	-.008***	-.010***	-.011***
	(.001)	(.001)	(.002)	(.002)
Survey–admin earnings (in thousand)	-.012***	-.012***	-.015***	-.015***
	(.001)	(.001)	(.002)	(.002)
Earnings: true –	ref	ref	ref	ref
false +	.345***	.364***	-.058	-.027
	(.084)	(.086)	(.099)	(.101)
false –	.038	.083	-.123**	-.069
	(.061)	(.064)	(.058)	(.061)
true +	.380***	.397***	.186***	.192***

	(.050)	(.055)	(.047)	(.051)
UA: true –	ref	ref	ref	ref
false –	.059	.058	-.082*	-.085*
	(.039)	(.040)	(.044)	(.044)
true +	.047	-.017	.083	.000
	(.052)	(.054)	(.053)	(.056)
ACLC: true –	ref	ref	ref	ref
false +	.117*	.066	-.007	-.010
	(.066)	(.070)	(.060)	(.066)
Admin benefit duration (in months)		-.013		-.024***
		(.008)		(.009)
Job training: did not take		ref		ref
Labour market agency		.196***		.067
		(.049)		(.045)
Mostly paid with own resources		.010		.092
		(.055)		(.062)
Employer		.019		.056
		(.039)		(.055)
No proxy		ref		ref
partner is proxy		-.088**		-.179***
		(.038)		(.066)
someone else is proxy		-.113***		-.063
		(.043)		(.048)
Controls	No	Yes	No	Yes
R-squared	.146	.187	.127	.172
N	1856	1852	1299	1294

Notes: This table shows an estimation of a linear probability model. The dependent variable equals 1 if the benefit amount is positive in both the survey and administrative data; and 0 if the survey amount is 0 while the administrative amount is positive. 'True +' implies positive income amounts in both the survey and administrative data; 'false +' means positive amount in the survey and zero in the administrative data; 'false –' means zero in the survey and positive amount in the administrative data; and 'true –' means zero amounts in both the survey and administrative data. Columns (2) and (4) add controls for: age group (in 5-year age bands), number of children in the household (1, 2, 3+), number of adults in the household (1, 2, 3+), the highest achieved education level (low, middle, high), health status (5 categories), region (Vienna, borough with more than 100,000 residents, borough with 10,000-100,000 residents, borough with less than 10,000 residents), occupation (10 categories), industry (25 categories), being a civil servant, country of birth (6 categories), wave (interviewed for the 1st, 2nd, 3rd, 4th time), interaction between wave and year (2008 to 2011), interview type (in person or by phone), same interviewer as last year. Sample is restricted to those aged 16+. Observations with missing/ imputed administrative/ survey UI or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: UA false + and with missing/ imputed values; ACLC false -, true + and with missing/ imputed values; and job training by other institutions and with missing or n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses. *Source:* Own calculations with the SILC.

Table A.5: Probability of reporting the unemployment assistance (UA) in the survey, conditional on receiving it

	Men		Women	
	(1)	(2)	(3)	(4)
Constant	-.267*	.375	.037	1.160***
	(.141)	(.269)	(.137)	(.281)
Ln admin UA	.108***	.059**	.082***	-.004
	(.016)	(.024)	(.016)	(.025)
Admin earnings (in thousand)	-.013***	-.008*	-.011*	-.010
	(.004)	(.004)	(.006)	(.006)
Survey–admin earnings (in thousand)	-.014***	-.011***	-.018***	-.017***
	(.002)	(.002)	(.004)	(.004)
Earnings: true –	ref	ref	ref	ref
false +	-.296***	-.242***	-.239**	-.166
	(.082)	(.088)	(.097)	(.102)
false –	.122**	.127**	-.116*	-.114*
	(.052)	(.057)	(.063)	(.067)
true +	.043	.039	-.026	.006
	(.056)	(.062)	(.058)	(.063)
UI: true –	ref	ref	ref	ref
false +	-.287***	-.291***	-.347***	-.348***
	(.052)	(.055)	(.068)	(.072)
false –	-.044	-.009	-.178***	-.127**
	(.053)	(.059)	(.059)	(.064)
true +	-.073	-.058	-.023	-.059
	(.048)	(.053)	(.054)	(.062)
missing/imputed	-.073	-.079	-.018	-.086
	(.068)	(.075)	(.059)	(.063)

ACLC: true –	ref	ref	ref	ref
false +	.227***	.103	.108	-.010
	(.058)	(.069)	(.066)	(.076)
Admin benefit duration (in months)		.018**		.020**
		(.008)		(.010)
Job training: did not take		ref		ref
Labour market agency		.154***		.202***
		(.053)		(.055)
Mostly paid with own resources		-.005		.008
		(.098)		(.108)
No proxy		ref		ref
partner is proxy		-.112		-.160
		(.079)		(.112)
someone else is proxy		.054		-.080
		(.069)		(.084)
Controls	No	Yes	No	Yes
R-squared	.282	.344	.191	.269
N	678	669	584	578

Notes and Source: See Table A.4. Observations with missing/ imputed administrative/ survey UA or earnings are excluded. Cells with too few observations cannot be disclosed and are not shown: ACLC false -, true + and with missing/ imputed values; and job training by employer, other institutions and with missing or n/a values.

Table A.6: Education shares (%) by unemployment insurance (UI) status

	Non-recipients				**Recipients**			
	Men		**Women**		**Men**		**Women**	
Education	Survey	Admin	Survey	Admin	Survey	Admin	Survey	Admin
low	10.2	10.0	18.8	18.7	21.3	18.9	25.4	23.7
middle	66.8	66.4	64.4	64.2	66.4	69.8	61.6	64.6
high	23.0	23.7	16.9	17.1	12.3	11.2	13.0	11.7
total n	14,693	14,030	16,275	15,881	1,224	1,887	941	1,335

Notes: Education refers to the highest level of education achieved, following on the International Standard Classification of Education (ISCED): low (lower secondary or less, i.e. compulsory schooling); middle (secondary education including high school); high (tertiary education including craftsman education and university degree). Observations with missing/ imputed administrative/ survey benefits are excluded. Sample is restricted to those aged 19-64.

Table A.7: Education shares (%) by unemployment assistance (UA) status

	Non-recipients				**Recipients**			
	Men		**Women**		**Men**		**Women**	
Education	Survey	Admin	Survey	Admin	Survey	Admin	Survey	Admin
low	10.6	10.2	18.9	18.8	37.1	31.9	41.4	36.4
middle	67.0	67.1	64.4	64.4	55.5	58.9	47.2	53.1
high	22.4	22.7	16.7	16.8	7.4	9.1	11.4	10.4
total n	15,660	15,335	17,126	16,882	364	689	360	604

Notes and Source: See Table A.6.

Table A.8: Education shares (%) by assistance for covering living costs (ACLC) status

	Non-recipients		**Recipients**	
Education	Survey	Admin	Survey	Admin
low	15.3	15.3	31.0	34.9
middle	65.4	65.4	56.2	52.4
high	19.3	19.2	12.8	12.7
total n	33,227	33,328	290	189

Notes and Source: See Table A.6.

Table A.9: Returns to job training: log-earnings regression
(restricted to the same sample in the survey and administrative data)

	Survey (1)	**Admin** (2)
Constant	9.282***	9.320***
	(.172)	(.190)
Job training: did not take	ref	ref
Labour market agency	.172**	.198**
	(.077)	(.085)
Mostly paid with own resources	-.145	-.103
	(.105)	(.115)
Employer	-.022	.108
	(.093)	(.101)
Job-training-year earnings (in thousand)	.033***	.037***
	(.002)	(.003)
Controls	Yes	Yes
R-squared	.474	.481
N	879	879

Notes: OLS estimation with log-earnings in *t+1* as the outcome variable. Sample is restricted to those aged 19-64; based on survey versus administrative information on being unemployed in *t*, i.e. received unemployment benefits (UI, UA or ACLC) or reported they were unemployed for parts of the year; with both positive survey and administrative earnings in *t+1*; and non-missing/ non-imputed survey and administrative job-training-year earnings. Column (1) is based on survey and column (2) on administrative earnings while the indicator for job training is always based on the survey. The controls include: education, age group (5-year bands), region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interview type, and if the same interviewer as last year. Observations with imputed administrative/ survey earnings, UI, UA or ACLC are excluded. Cells with too few observations cannot be disclosed and are not shown: i.e. job training by other institutions and with missing/ n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses.

Table A.10: Returns to job training: being an earner regression
(restricted to the same sample in the survey and administrative data)

	Survey (1)	**Admin** (2)
Constant	.678***	.787***
	(.076)	(.080)
Job training: did not take	ref	ref
Labour market agency	.051	.046
	(.033)	(.034)
Mostly paid with own resources	-.053	.006
	(.048)	(.050)
Employer	-.027	.029
	(.049)	(.050)
Job-training-year earnings (in thousand)	.007***	.006***
	(.001)	(.001)
Controls	Yes	Yes
R-squared	.380	.303
N	1425	1425

Notes: OLS estimation with being an earner in *t+1* (yes=1 and no=0) as the outcome variable. Sample is restricted to those aged 19-64; based on survey versus administrative information on being unemployed in *t*, i.e. received unemployment benefits (UI, UA or ACLC) or reported they were unemployed for parts of the year; with both non-missing/ non-imputed survey and administrative outcome variable and job-training-year earnings. Column (1) is based on survey and column (2) on administrative earnings while the indicator for job training is always based on the survey. The controls include: education, age group (5-year bands), region, if a civil servant, country of birth, proxy interview, month of interview, survey wave, interview type, and if the same interviewer as last year. Observations with missing/ imputed administrative/ survey earnings, UI, UA or ACLC are excluded. Cells with too few observations cannot be disclosed and are not shown: i.e. job training by other institutions and with missing/ n/a values. Significance levels indicated as * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ and standard errors shown in parentheses.