IAB-DISCUSSIONPAPER

Articles on labour market issues

6|2019 Benefit underreporting in survey data and its consequences for measuring non-take-up: new evidence from linked administrative and survey data

Kerstin Bruckmeier, Regina T. Riphahn, Jürgen Wiemers

ISSN 2195-2663

Benefit underreporting in survey data and its consequences for measuring non-take-up: new evidence from linked administrative and survey data

Kerstin Bruckmeier (IAB),
Regina T. Riphahn (Friedrich-Alexander University Erlangen-Nürnberg),
Jürgen Wiemers (IAB)

Content

List of graphics

List of tables

Abstract

The international literature studies non-take-up behavior of eligible populations to evaluate the effectiveness of government programs. A major challenge in this literature is the measurement error regarding benefit take-up. Measurement error is typically addressed by structural assumptions in the modeling framework. In our data, we observe both actual welfare receipt and respondents' survey information on their take-up. This allows us to observe the measurement errors that other researchers must estimate. We describe survey misreporting and investigate how it biases the estimates of the magnitude and patterns of benefit take-up among eligible households. Our findings suggest that the extent of measurement error can be substantial. It varies with the characteristics of the misreporting population and is associated with the drivers of underreporting. This indicates that survey-based analyses of take-up behavior are likely subject to severe biases.

Zusammenfassung

Eine Vielzahl von Studien untersucht die Nicht-Inanspruchnahme von Sozialleistungen, um die Wirksamkeit staatlicher Programme zu bewerten. Eine große Herausforderung in dieser Literatur besteht darin, dass die Messung der Inanspruchnahme in den verwendeten Daten fehlerbehaftet ist. Der Messfehler wird typischerweise durch strukturelle Annahmen in der statistischen Modellierung adressiert. In unseren Daten beobachten wir hingegen sowohl den tatsächlichen Leistungsbezug als auch die Angaben der Befragten zu ihrem Leistungsbezug. So können wir die Messfehler, die üblicherweise geschätzt werden müssen, direkt beobachten. Wir berichten das Ausmaß von falschen Angaben bezüglich des Leistungsbezugs in den von uns verwendeten Surveydaten und untersuchen, wie diese falschen Angaben Schätzungen zu den Determinanten der Inanspruchnahme leistungsberechtigter Haushalte verzerren. Unsere Ergebnisse deuten darauf hin, dass das Ausmaß der Messfehler erheblich sein kann, sodass survey-basierte Analysen des Inanspruchnahmeverhaltens wahrscheinlich mit beträchtlichen Verzerrungen behaftet sind.

JEL-Classification

C81, H75, I32

Keywords

Administrative data, data linkage, misreporting, survey data, take-up, welfare.

Acknowledgments

We thank Mark Trappmann for helpful comments on an earlier version of the manuscript.

1 Introduction

The non-take-up of benefits is an important aspect of state support programs, i.e., if eligible households do not take up transfers, such programs are ineffective and the basic needs of population groups may remain unaddressed. Eurofound (2015) shows that benefit non-take-up is internationally pervasive and frequently affects more than 40 percent of the eligible population. This has motivated a long-standing discussion of the extent and determinants of take-up behavior in the international literature (e.g., Moffitt 1983, Blundell, Fry, and Walker 1988, or Hernandez, Pudney, and Hancock 2007). We contribute to that literature and study the extent to which measurement error biases analyses of take-up behavior. Our unique data provide unusually precise information on true take-up behavior.

A key challenge in the empirical analysis of non-take-up is its measurement. A correct measurement of non-take-up requires valid information on both *program eligibility* and *program take-up*. Most non-take-up studies must rely on survey data to measure both eligibility and program participation. However, several factors can generate measurement errors in the collected survey data.

First, the measurement of program eligibility may be biased if the information provided by respondents on, e.g., income, wealth, or household composition systematically differs from true values. This may be attributable to the phrasing of survey questions, approximation errors, or misreporting. In addition to wrong information, surveys may provide insufficient detail, which similarly renders the simulation of benefit eligibility unreliable (e.g., applying monthly instead of annual values of financial variables).

Second, reported benefit receipt may be observed incorrectly. Recently, Meyer et al. (2015) pointed out that data from household surveys missed the measurement of approximately half of all welfare and food-stamp payments in major household surveys in the U.S. Several reasons contribute to the mismeasurement of benefit receipt (Bound et al. 2001), such as surveys often ask respondents whether they have received benefits during a certain period in the past. Respondents may have completely forgotten past benefit receipt (recall bias), or they may not remember the exact dates of receipt. For example, events can be reported as more recent than they actually occurred, which is known as the “forward telescoping bias” in the survey literature (Bradburn et al. 1994). This form of bias could lead to mismeasurement in the form of benefit under- and overreporting. Additionally, if several benefits are available or benefits can be claimed simultaneously, benefits beneficiaries might report incorrectly, claiming specific benefit(s) that they did not receive while inadvertently omitting the benefit(s) that they did receive, ultimately leading to benefit underreporting and overreporting (Hancock and Barker 2005, Krafft et al. 2015). A final source of benefit underreporting is the “social desirability bias”. In particular, the receipt of means-tested social welfare benefits is often perceived as stigmatizing and thus respondents may underreport their receipt of these benefits.

The literature on non-take-up and program participation has proposed several approaches to handling potential measurement biases. In general, we can distinguish studies that (i) apply structural models to estimate the relevance of errors and mismeasurement from those that focus on (ii) data cleaning or (iii) external validation samples to study measurement errors.

In the first group, the contributions by Duclos (1995, 1997) are seminal. He studies the UK supplementary benefit program for retirees and applies a structural model to describe and estimate the relevance of errors committed by the benefit agency and the analyst. Hernandez and Pudney (2007) refine his contribution and confirm that measurement error strongly affects the estimated effects of benefit entitlement amounts. Pudney (2001) calibrates the effect of measurement errors in income and benefit receipt on the bias of coefficient estimates in the take-up equation, predicted take-up probabilities, and the patterns of claim costs. He finds that even modest measurement errors may generate large biases. In their structural estimation model, Bollinger and David (2001) find a correlation between response error regarding food-stamp participation and nonresponse behavior. The authors point out that response error models may not be stable over time and emphasize the need for validation data.

Hancock and Barker (2005) focus on the effects of careful data handling. The authors study the degree to which ex ante data cleaning can increase the estimates of take-up and affect the correlation patterns of the take-up outcome. However, they find that their efforts have only a small impact.

Finally, external validation samples can be used to assess the extent of measurement errors in the survey data and mitigate their effects in analysis. Studies that employ this approach are scarce because this requires linking survey data to administrative data. Linked data are often not available or only available for specific groups, periods or regions. One example of this approach is Bollinger and David (1997). They take advantage of data with information on true participation and survey responses. Information on response errors in the validation sample is then considered in the likelihood function for the primary sample. The authors find that modeling response errors generate large differences in the estimation of program participation even when the validation data are gathered on a sample that differs from the survey data. Other examples are Mittag (2016) and Meyer and Mittag (2017a). Both investigate different methods to account for misclassification in survey data if linked data are not available to the researcher. They use validation data to evaluate the effectiveness of their formulas for bias reduction.

In this study, we also follow the third approach and link survey data to administrative data, which informs us about the true benefit take-up of survey respondents. This allows us to determine precisely when survey information differs from actual benefit receipt. Thus, we can determine the presence of measurement errors and underreporting directly without invoking assumptions that are required for data cleaning procedures. The study closest to our approach is Meyer and Mittag (2017b), who use linked data to correct survey data on reported benefit receipt but do not consider take-up behavior. They find that the poverty-reducing effect of benefit programs in New York State is nearly doubled using the corrected data. In our study, we use linked data to investigate whether the extent and pattern of program non-take-up differ after correcting for the misreporting of survey respondents. To the best of our knowledge, our study is the first to use linked data to investigate the impact of mismeasurement on program non-take-up.

We consider a general income support program that is available for the working-age population in Germany. This general benefit is less subject to the risk of benefit confusion than specific transfers available, e.g., for retirees only, which have been discussed in the literature (e.g., Duclos 1997, Hancock and Barker 2005). We focus on one specific and important measurement error in survey data, i.e., the underreporting of benefit receipt. Several studies discussed the relevance of underreporting for the reliability of survey data, see, e.g., Meyer et al. (2015), Meyer and Goerge (2011), Taeuber et al. (2004), or Card et al. (2001). We determine the extent and relevance of underreporting for the estimation of participation models. We thus follow up on a recommendation by Duclos (1995, p.414) who suggests that "Richer data (…) would naturally enhance our understanding of the determinants of take-up."

We find that correcting for underreporting in the data significantly modifies the results of take-up regressions. The marginal effects of characteristics associated with benefit take-up change sign and statistical significance and often deviate by more than 30 percent after correcting for measurement error in the outcome. These results are robust to applying different estimators, to using sample weights and to specification changes. We find evidence that the patterns determining benefit underreporting are reflected in the sensitivity of marginal effects to the data correction.

Our results are important for several reasons. First, they show that survey data can yield biased results in the study of take-up behavior based on self-reported information. Our findings are more reliable than prior contributions because they are based on linked survey and administrative data and cover a general and well-known nationwide benefit program. Second, we show that the patterns of underreporting and the estimation biases can be related. The coefficients that are the most biased in take-up equations are those associated with the underreporting groups' characteristics.

We structure our paper in six sections. In the next section, we briefly characterize the benefit program considered in our analysis. We lay out our empirical approach in Section 3. Section 4 describes the nature of our data and provides descriptive statistics. Section 5 presents our empirical results and robustness tests. We draw conclusions in Section 6.

2 Institutional Background

We study the take-up of the German minimum income support program Unemployment Benefit II (UB II). The transfer is available for working-age individuals who are able to work. Alternative programs cover persons who have reached retirement age or are unable to work. UB II eligibility exists if a household's net income is below the legally determined minimum. The minimum income deemed sufficient to guarantee an acceptable minimum living standard for a household is calculated based on the number of household members and – for minors – their age. In 2018, the standard benefit for an adult is 416 Euros per month. Expenses for rent, heating, and health care are paid in addition to the standard benefit; benefits can be higher in special circumstances (e.g., for single-parent families, pregnant women, or those with special food requirements). Households with more than a maximum amount of wealth are not eligible; wealth comprises the value of owned property and assets minus that of liabilities. Eligibility is not conditional on unemployment. In 2016, approximately 41 percent of regular benefit recipients are unemployed and 28 percent receive the benefit to top up (insufficient) earnings from employment. Others are temporarily unable to work, e.g., because of child care obligations (Statistik der Bundesagentur für Arbeit 2018).

The UB II program follows federal regulations and is administered either by the unemployment insurance or by the municipality. Municipalities and the federal government pay the benefit if the beneficiary submits a substantiated claim. In 2016, the program covered approximately 6.2 million individuals in 3.3 million households and paid out approximately 35.2 billion Euros (STBA 2017). Thus, in contrast to some of the literature, we consider a well-known program that is generally available to the entire working-age population capable of working.

Recent studies on take-up of UB II using survey data from the German Socio-economic Panel (SOEP) show that based on monthly data, between 46 and 58 percent of eligible households did not take up the benefit in the years 2005-07 (e.g., Bruckmeier and Wiemers 2012). The authors find that take-up varies with the potential benefit amount and the expected duration of eligibility expressed in proxy variables such as education and region of residence.

3 Theory and Empirical Approach

In recent decades, a large number of empirical studies on the determinants of (non-)take-up have been conducted for a wide range of means-tested benefits (see, e.g., Blundell et al. 1988, Blank and Ruggles 1996, Riphahn 2001, Wilde and Kubis 2005, Whelan 2010, Bruckmeier and Wiemers 2012). All survey-based studies of take-up behavior have to address the problem that the data do not provide information about benefit eligibility. The studies therefore simulate welfare eligibility for every household in the dataset using a microsimulation model. Then, given a model of welfare eligibility, the literature typically defines benefit non-take-up as being eligible according to the simulation model while reporting non-receipt of the benefit in the survey data.

Following Blundell et al. (1988), we model the take-up decision in a discrete choice framework. This approach assumes that benefit take-up ($P = 1$) will be observed if net utility from claiming a benefit exceeds utility from not claiming the benefit, conditional on being eligible for the benefit (*b>0*) as follows:

$$P = \mathbf{I}\left(U(y+b,\boldsymbol{x}) - C(\boldsymbol{x}) > U(y,\boldsymbol{x}) \mid b > 0\right), \tag{1}$$

where $\mathbf{I}(\cdot)$ is the indicator function, $U(\cdot)$ denotes utility, $b \equiv \max(\bar{b}(\boldsymbol{x}) - NI - t(y^g,\boldsymbol{x}), 0)$ is the level of benefit entitlement determined by the maximum level of benefits for the given household, $\bar{b}(\boldsymbol{x})$, minus nonearned incomes NI and deductible income, $t(y^g,\boldsymbol{x})$, where y^g denotes gross earned income. Net income excluding the benefit b is given by y; $\boldsymbol{x}$ are other observed household characteristics. The cost of claiming the benefit is represented by the function $C(\boldsymbol{x})$. For both the utility and cost of claiming, we assume linearity as follows:

$$\begin{aligned} U(y+b,\boldsymbol{x}) &= \alpha_0 + \alpha_1(y+b) + \boldsymbol{\alpha}_2'\boldsymbol{x} + \varepsilon_{U_1}, \\ U(y,\boldsymbol{x}) &= \alpha_0 + \alpha_1 y + \boldsymbol{\alpha}_2'\boldsymbol{x} + \varepsilon_{U_0}, \\ -C(\boldsymbol{x}) &= \beta_0 + \boldsymbol{\beta}_1'\boldsymbol{x} + \varepsilon_C. \end{aligned} \tag{2}$$

The error terms ε_{U_1}, ε_{U_0}, and ε_C represent unobserved characteristics influencing the take-up decision. Substituting (2) into (1) and assuming a Gaussian distribution for the combined error term $\eta_1 \equiv \varepsilon_{U_1} - \varepsilon_{U_0} + \varepsilon_C$, i.e., $\eta_1 \sim N(0,1)$, the probability of observing take-up is given by the following:

$$\Pr(P=1) = \Pr\left(\eta_1 > -(\beta_0 + \alpha_1 b + \boldsymbol{\beta}_1'\boldsymbol{x})\right) = \Phi(\beta_0 + \alpha_1 b + \boldsymbol{\beta}_1'\boldsymbol{x}), \tag{3}$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution and β_0, α_1, and $\boldsymbol{\beta}_1$ are parameters to be estimated. We follow the literature (e.g., Blundell et al. 1988, Bollinger and David 1997, 2001, Duclos 1995, Pudney 2001) in assuming a standard normal error term distribution. Thus, our first specification for the take-up equation is the (pooled) probit model (3). As a robustness check, we also considered a logistic distribution for the combined error term η_1. We find that the results are robust with respect to the distributional assumption (results available upon request).

In a second specification, we estimate a random effects (RE) probit model of benefit take-up to control for unobserved heterogeneity at the household level. The probability of benefit take-up ($P_{it} = 1$) for household i conditional on eligibility in period t is given by the following:

$$\Pr(P_{it} = 1) = \Pr\big(\eta_{it} > -(\beta_0 + \alpha_1 b_{it} + \boldsymbol{\beta}_1' \boldsymbol{x}_{it} + \nu_i)\big) = \Phi(\beta_0 + \alpha_1 b_{it} + \boldsymbol{\beta}_1' \boldsymbol{x}_{it} + \nu_i), \qquad (4)$$

where η_{it} are i.i.d. Gaussian errors with mean zero and variance normalized to one and assumed independent of the random effects ν_i, which are i.i.d. N(0,σ_ν^2). The share of the total variance contributed by the panel-level variance component is given by $\rho = \sigma_\nu^2/(\sigma_\nu^2 + 1)$.

We build on the existing empirical literature in choosing the household characteristics that influence the take-up decision (Riphahn 2001, Wilde and Kubis 2005, Frick and Groh-Samberg 2007, Whelan 2010, Bruckmeier and Wiemers 2012) to enhance the external validity of our analysis. The most obvious factor affecting utility when claiming UB II is the household's benefit entitlement (see, e.g., Blundell et al. 1988). Because the true level is unobserved for non-take-up households, we use simulated values in our estimations. In addition, utility may vary by household type, i.e., whether it is a single or a couple household and whether or not children are present. Therefore, we consider indicators of household type. We also control for disability of the household head. The costs of claiming a means-tested benefit consist of information costs (e.g., because of insufficient knowledge of entitlement rules, the claiming process, or of the administrative procedures) and stigma costs, i.e., the fear of stigmatization and negative societal attitudes toward welfare dependence (see van Oorschot 1991). Our control variables, i.e., age, education, and disability status, may reflect heterogeneities in information and stigma costs. We expect higher take-up rates in Eastern Germany due to higher unemployment there and thus control for residence in Eastern Germany. As general sociodemographic indicators, we account for age, education, home ownership, and an indicator for migrant status, which takes on the value of one for first- and second-generation immigrants.

4 Data and Sample

4.1 Data

We use data from the household panel study "Labour Market and Social Security" (Panel Arbeitsmarkt und soziale Sicherung, PASS), a survey designed for research on unemployment and poverty (Trappmann et al. 2010, 2013 or Berg et al. (2014) for technical documentation). The first survey of this study interviewed more than 12,000 respondents in 2006-07. The seventh survey wave was completed in 2013. Because the survey instruments and interview program were revised after the first wave (Gebhardt et al. 2009), we only use surveys 2-7.

The data consist of two subsamples. The first subsample considers UB II recipients, while the second subsample covers the overall German population, oversampling those with low socioeconomic status. The UB II sample is randomly drawn from the administrative records of the Federal Employment Agency. To retain a representative character for the population of UB II recipients, subsample one is refreshed each year to include new recipients of UB II (benefit-inflow-sample).

The general population sample is a random draw from a database of addresses of private households in Germany. It is provided by a commercial provider in wave one and is taken from municipality population registers in wave five (refreshment sample). For a detailed description of the sampling design, see Gebhardt et al. (2009).[1] The final weights we use in the descriptive analysis balance distortions arise from the sample design and reflect the entire German population. For all weighted descriptive results, we consider the complex survey design of the PASS for the calculation of point estimators and their standard errors. To achieve this, we use Stata's "svy"-commands and follow the recommendations given in Bethmann et al. (2013).

The PASS data are particularly suitable for our analyses because it focuses on potential beneficiaries living in low-income households. Beste et al. (2018) find that the income distribution in the PASS data (starting with wave 2) is similar to that of two other data sources (SOEP and "Mikrozensus"). Furthermore, PASS interviews respondents about their current welfare receipt, and it allows us to link survey data with administrative records on welfare receipt. Interviewers ask the head of the household whether the household receives UB II. Interviewers determine the head of the household during the household's first participation in the survey as the person who is best informed about the household finances. The PASS gathers information on UB II receipt via "dependent interviewing", i.e., interviewers remind the head of the household of the answer in the previous interview prior to asking about current receipt (Berg et al. 2012). This form of interviewing should result in lower benefit underreporting (Lynn et al. 2012).

In our analysis sample, we consider household observations with realized personal interviews. We drop respondents above age 65 or in receipt of retirement benefits because they are no longer eligible for UB II. We omit students and individuals pursuing apprenticeship training because they benefit from alternative transfer programs. We require that the household responds to the question on current welfare receipt, that there is only one benefit-receiving unit ("community of need") in the household and that there is valid information on earnings. Across waves 2-7, our sample covers 30,878 annual household-level observations overall and approximately 5,000 observations per year. For each household-year observation we simulate UB II benefit eligibility. This yields 17,585 UB II eligible household-year observations (see Appendix A for details).

4.2 Data Linkage

The opportunity to link survey with administrative data is rare in the literature. In particular, we are able to link the PASS survey data to the administrative records of the Federal Employment

[1] The sampling in wave five involved several steps. Step one draws 300 postcodes (regions) as primary sampling units, i.e., households from both populations – UB II recipients and private households – within each postcode. Based on the number of benefit-receiving households (sample 1) and the number of private households (sample 2) in a postcode, each household receives a uniform selection probability. Design weights for the gross sample reflect the selection probability. Logit models for panel participation are the basis to account for the participation probability and to adjust design weights in the second step (see Gebhardt et al. 2009). Finally, both samples were calibrated to official statistics on UB II recipients and private households in Germany.

Agency. The data, originally collected at local job agencies (*"job centers"*), contain information on claims for UB II. The data perfectly reflect official payments. The Institute for Employment Research (IAB) and the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the IAB have access to this administrative data and are responsible for processing, anonymizing and providing it for empirical research. For our analysis, we link the survey data to administrative data of the "Unemployment Benefit II Recipient History" (*Leistungshistorik Grundsicherung – LHG,* Version 11.01.01-150220) of the IAB (Antoni et al. 2016). The administrative UB II data contain information on socioeconomic variables of eligible individuals, household structure, and regional variables.

Because of legal constraints, the survey information can only be linked to the administrative data if the participant consented to linkage in the survey. Therefore, interviewers ask participants for consent to merge their survey data to their administrative data that are available at the IAB (for details please see Appendix B). The consent rate in the PASS is approximately 80 percent, which is comparable to other survey studies (Bethmann et al. 2016; Sakshaug and Kreuter 2012). In our sample of simulated eligible households, we have a consent rate of 83.4 percent (see Table B.1 in Appendix B). Because respondents who do not agree to the data linkage are asked again in the next wave, the proportion of observations for which an approval is available is significantly higher. Overall, we could not use 12 percent of all household-year observations for eligible households because of missing consent to data linkage. This leaves us with 16,874 household-year observations of simulated eligible respondents who agreed to linkage.

Next, we merge these 16,874 observations with a key file generated by the German Record Linkage Center (Antoni and Schnell 2017). This key file is based on the identification of the PASS respondents in administrative data of the IAB. To identify respondents in the administrative data, harmonized information on addresses and personal characteristics from different administrative data sources collected by the Federal Employment Agency are used. Individuals who never worked in dependent employment, who are exempt from social security contributions (e.g., civil servants) or have never been registered as unemployed or benefit recipients are not in the data. The record linkage is based on multilevel deterministic and probabilistic methods for linking datasets (see Sakshaug et al. 2017 for a detailed description and Appendix B). From our sample of 16,874 household-year observations of respondents who agreed to data linkage, we identified 15,925 observations in the administrative data, which amounts to a linkage rate of 94 percent. Of the 15,925 matches, 15,095 were unique matches and 830 were duplicates, which were corrected following a procedure described in Appendix B. As a robustness check, we verify our main findings for a sample without duplicates in Section 5.3. From our sample of 15,925 linked observations, we keep 14,834 observations with no missing values in the covariates for our descriptive results and the regression analysis. Thus, our analysis sample represents 84 percent of the simulated eligible population (17,585).

A potential problem of the data linkage is that results may be biased because of selectivity in either nonconsent or nonidentifiability in the administrative data. With respect to nonconsent, underreporting of benefit receipt might be biased downwards if nonconsent to the data linkage is positively correlated with the underreporting of benefit receipt: households who do not want to admit to receiving UB II might also be reluctant to agree to data linkage if they fear that their misreporting might be discovered. Column 1 of Table B.3 in Appendix B indicates correlation patterns underlying the probability of not giving consent to the data linkage. We find some statistically significant and small effects: immigrants, younger, nondisabled persons, those living in single households, or those living in Western Germany have a higher probability to refuse consent. These findings support the idea that nonconsent might be positively correlated with benefit underreporting. This suggests that our analysis is a rather conservative estimate of underreporting.

With respect to nonidentifiability in administrative data, column 2 of Table B.3 shows the correlation patterns behind the probability that a household cannot be linked to the administrative data for the sample of simulated eligible households with consent to data linkage. Here, we find no significant marginal effects in most sociodemographic and household characteristics except for age groups, migration background, home owners, and the subsample two indicator.

Overall, these results indicate small systematic effects; thus, we conservatively underestimate underreporting and its correction. We will provide two robustness checks in Section 5.3 concerning the potential selectivity of data linkage.

4.3 Over- and Underreporting

Our dependent variable describes whether an eligible household takes up UB II benefits. In our data, 11,265 respondents reported benefit receipt in the survey and – based on administrative records – actually received benefits in the month of the interview (take-up households). Additionally, 2,291 respondents reported not claiming the benefit, which is confirmed by the information from the administrative data (non-take-up households). A group of 904 respondents (7.4 percent of all true recipients, 7.8 percent when survey weights are applied) did not indicate receipt in the survey, but actually received benefits in the month of the interview based on administrative data (underreporting households). In addition, 374 benefit-eligible respondents claimed to receive UB II in the survey, but they did not receive benefits according to the administrative records (overreporting households). Without information from administrative data, these overreporting households would be considered take-up households. This increases the number of take-up households to 11,639. Given the information from administrative data, we can omit the overreporting households from the sample. Alternatively, overreporting households might be truly eligible for UB II. In that case, we should follow the information from administrative data and reclassify overreporters as non-take-up households. Such a scenario may result if respondents are mistaken about the period when they actually received the benefit.

Because we are interested in the effect of underreporting on take-up estimations in a prototypical survey study that does not have access to precise administrative information, we treat overreporting households as take-up households in our baseline analysis. In robustness checks (see Section 5.3), we re-estimate our models first after omitting overreporters from the sample and second after recoding them as non-take-up.

Table 1 shows descriptive statistics of our explanatory variables for the full sample and separately for households with and without benefit take-up and for those underreporting their benefit take-up. The three subgroups differ in their characteristics. Interestingly, we find some similarities between non-take-up households (column 3) and the underreporting households (column 4). Compared to take-up households, the latter two groups have significantly lower simulated benefit entitlements, younger household heads, a higher share of household heads with upper secondary education, a lower share living in Eastern Germany, and a larger share of families with children. This similarity suggests that a take-up regression erroneously classifying underreporting households with the non-take-up households overestimates the heterogeneity between the take-up and non-take-up groups, i.e., after correcting underreporting, we expect that the take-up regression yields coefficients of smaller magnitude.

Table 1: Descriptive statistics: covariate means

	(1) All	(2) Take-Up Households	(3) Non-Take-Up Households		(4) Underreporting Households	
Simulated Entitlement/100 EUR	6,69	7,25	4,27	***	5,61	***
Female hh	0,55	0,54	0,59	***	0,57	
Hh is immigrant	0,24	0,24	0,23		0,28	***
Age of hh: 15-24 years	0,05	0,05	0,07	***	0,09	***
Age of hh: 25-34 years	0,21	0,20	0,24	***	0,24	***
Age of hh 35-44 years	0,24	0,24	0,28	***	0,26	*
Age of hh: 45-54 years	0,28	0,28	0,27		0,26	
Age of hh: >=55 years	0,22	0,24	0,14	***	0,14	***
Hh is disabled	0,14	0,14	0,12	**	0,11	***
Hh holds no sec. degree	0,08	0,09	0,06	***	0,07	**
Hh holds lower sec. degree	0,39	0,40	0,32	***	0,35	***
Hh holds intermediate sec. degree	0,35	0,34	0,38	***	0,37	
Hh holds upper sec. degree	0,18	0,17	0,23	***	0,21	***
Eastern Germany	0,34	0,35	0,30	***	0,31	*
Household owns home	0,06	0,05	0,11	***	0,05	
Young children in household (age<=4 years)	0,11	0,11	0,10		0,12	
Single person	0,54	0,55	0,47	***	0,53	
Family without children	0,09	0,08	0,11	***	0,07	
Single parents	0,25	0,26	0,22	***	0,25	
Family with children	0,13	0,11	0,21	***	0,15	***
Subsample two	0,09	0,07	0,24	***	0,07	
N	14.834	11.639	2.291		904	

Notes: Asterisks */**/*** denote significantly different means compared to the group of take-up households (column 2) at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample. Unweighted results.

Source: Own calculation based on PASS waves 2-7.

The next section describes our analysis results. First, we describe the extent to which non-take-up as reported in survey data must be corrected once information from administrative data is considered. Then, we look at the effect of correcting the dependent variable on the correlation patterns behind non-take-up behavior and investigate the robustness of these results. We describe the characteristics of those not reporting benefit receipt in the last step.

5 Results

5.1 Descriptive effects of data correction

In Table 2 Group-specific non-take-up rates before and after correction of underreporting we report the simulated group-specific non-take-up rates for the sample that could be linked to administrative records. Column 1 shows the shares before considering corrections for underreporting and column 2 shows the rates after correction. Initially, we observe an overall weighted non-take-up rate of 40 percent (see bottom of column 1 of Table 2) with substantial heterogeneity across subgroups: we observe the highest rate of non-take-up for couples without children (64 percent), while single parent households feature the lowest rates of benefit non-take-up (30 percent). The size of the non-take-up rate and the variation over the subgroups is in line with findings based on other data (see, e.g., Bruckmeier and Wiemers 2012 and the literature cited there).

We use our administrative data on actual benefit receipt to correct for benefit underreporting in the survey and to reclassify non-take-up outcomes.[2] This reduces the overall non-take-up rate from approximately 40 percent to approximately 35 percent (see column 2 of Table 2). Thus, underreporting caused us to overestimate the non-take-up rate by approximately 5 percentage points, or 12 percent. The extent of the correction in the non-take-up rate varies across subgroups (see last two columns of Table 2). The relative decline in non-take-up rates ranges from 6 percent for families without children to 18 percent for households whose head is a first- or second-generation immigrant.

[2] Note that the simulation may erroneously predict benefit eligibility for households that underreport other income than benefits. Because these households are not actual benefit recipients, we would overestimate the non-take-up rate. However, their reported benefit receipt would not be corrected based on linked administrative data. Therefore, these observations are irrelevant to the evaluation of the take-up correction.

Table 2: Group-specific non-take-up rates before and after correction of underreporting

	before corrections	after correcting underreporters	absolute change	relative change
Female hh	0,46	0,40	-5,3%	-11,7%
Hh is immigrant	0,35	0,28	-7,1%	-19,9%
Age of hh: 15-24 years	0,51	0,45	-6,2%	-12,3%
Age of hh: 25-34 years	0,35	0,30	-5,2%	-15,0%
Age of hh 35-44 years	0,43	0,37	-5,9%	-13,8%
Age of hh: 45-54 years	0,42	0,37	-4,9%	-11,8%
Age of hh: >=55 years	0,37	0,34	-3,6%	-9,7%
Hh is disabled	0,41	0,37	-3,5%	-8,6%
Hh holds no sec. degree	0,41	0,37	-4,1%	-10,0%
Hh holds lower sec. degree	0,33	0,29	-4,4%	-13,3%
Hh holds intermediate sec. degree	0,42	0,37	-4,8%	-11,6%
Hh holds upper sec. degree	0,49	0,43	-6,7%	-13,6%
Eastern Germany	0,32	0,27	-5,3%	-16,5%
Young children in household (age<=4 years)	0,36	0,29	-6,1%	-17,1%
Single person	0,36	0,31	-4,8%	-13,2%
Family without children	0,64	0,59	-4,4%	-6,9%
Single parents	0,30	0,25	-4,9%	-16,5%
Family with children	0,60	0,53	-6,5%	-10,9%
All	0,40	0,35	-5,0%	-12,4%

Notes: Hh stands for head of household. Weighted values using cross section sample weights for 14,834 households with simulated entitlements to social assistance. The total non-take-up rate is reduced from 40.1 percent (21.5 percent unweighted) to 35.1 (15.4 percent unweighted) after correcting benefit underreporting. Overreporting households are classified as take-up households both before and after correction of underreporters.

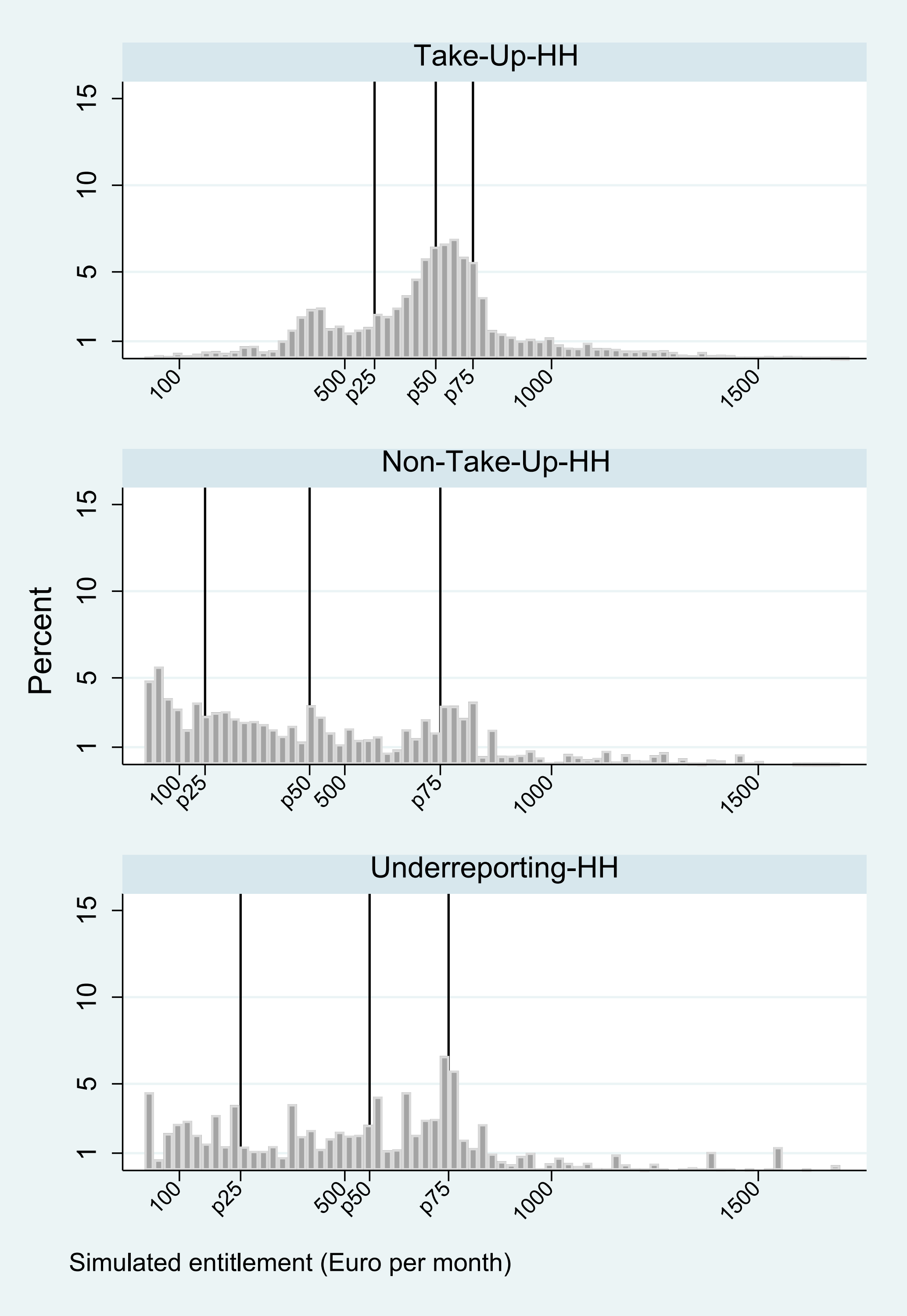
Source: Own calculation based on PASS waves 2-7.

Figure 1 depicts the distribution of simulated benefit entitlements for those taking up the benefit, those not taking up their benefit, and those who took up the benefit but did not report it in the survey. As expected, we observe the highest benefits among households who claim their benefits with a median value of 720 Euros (see the top panel of Figure 1). The distributions of benefit entitlements for the non-take-up and underreporting households yield a large share of households with small claims and median claims of 415 and 560 Euros, respectively.

5.2 Patterns of benefit take-up and the effects of data correction

Table 3 presents the estimation results of our take-up model. We regress an indicator of benefit take-up on household characteristics in a sample of 14,834 pooled observations of benefit-eligible households. In column 1, we present the estimated marginal effects of a pooled probit estimation with cluster robust standard errors; in column 5 we show the estimates of a random effects (RE) probit estimation; and in columns 2 and 6 both estimation approaches are repeated, now using the corrected dependent variable.

Figure 1: Distribution of simulated monthly benefit entitlements



Notes: HH stands for household. 61 outlier observations with monthly entitlements above 1,700 Euros excluded (52 Take-Up-HH, 3 Non-Take-Up-HH, 2 Underreporting-HH). Weighted values using cross-section sample weights.

Source: Own calculation based on PASS waves 2-7.

Table 3: Take-up regression: marginal effects before and after correction of underreporting

Dependent variable: Take-up of UB II	(1) Pooled probit, uncor.	(2) Pooled probit, corrected for underreporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, uncor.	(6) RE probit, corrected for underreporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0449***	.0365***	-.00844***	-18,71%	.0417***	.0317***	-.0100***	-23,98%
	(.00115)	(.00110)	(.000697)		(.00120)	(.00119)	(.000831)	
Female hh	-.013	-.0102	.00285	-21,54%	-.0137	-.0116	.00219	-15,33%
	(.00846)	(.00760)	(.00490)		(.00869)	(.00725)	(.00543)	
Hh is immigrant	-.00561	.0114	.0170***	-303,21%	.000988	.0192**	.0182***	1843,32%
	(.00934)	(.00819)	(.00589)		(.00932)	(.00766)	(.00622)	
Age of hh: 25-34 years	.0793***	.0513***	-.0280**	-35,31%	.0692***	.0395**	-.0297**	-42,92%
(ref.:15-24 years)	(.0178)	(.0162)	(.0121)		(.0179)	(.0160)	(.0125)	
Age of hh: 35-44 years	.0906***	.0611***	-.0295**	-32,56%	.0786***	.0507***	-.0278**	-35,50%
(ref.:15-24 years)	(.0183)	(.0167)	(.0122)		(.0183)	(.0162)	(.0128)	
Age of hh: 45-54 years	.118***	.0815***	-.0364***	-30,93%	.101***	.0667***	-.0345***	-33,96%
(ref.:15-24 years)	(.0181)	(.0164)	(.0121)		(.0182)	(.0160)	(.0127)	
Age of hh: >=55 years	.172***	.120***	-.0516***	-30,23%	.157***	.102***	-.0542***	-35,03%
(ref.:15-24 years)	(.0183)	(.0166)	(.0122)		(.0184)	(.0164)	(.0129)	
Hh is disabled	.00479	-.00628	-.0111*	-231,11%	.0121	.00125	-.0109	-89,67%
	(.0112)	(.0105)	(.00598)		(.0108)	(.00901)	(.00675)	
Hh holds lower sec. degree	-.00809	-.00618	.00191	-23,61%	-.0143	-.0184	-.00408	28,67%
(ref. no sec. degree)	(.0139)	(.0126)	(.00767)		(.0140)	(.0114)	(.00928)	
Hh holds interm. sec. degree	-.0458***	-.0370***	.00874	-19,21%	-.0542***	-.0542***	3.38e-05	0,00%
(ref. no sec. degree)	(.0143)	(.0129)	(.00796)		(.0144)	(.0118)	(.00950)	
Hh holds upper sec. degree	-.0828***	-.0680***	.0149*	-17,87%	-.0948***	-.0832***	.0117	-12,24%
(ref. no sec. degree)	(.0159)	(.0145)	(.00889)		(.0160)	(.0131)	(.0104)	
Eastern Germany	.0531***	.0424***	-.0107**	-20,15%	.0540***	.0397***	-.0144***	-26,48%
	(.00796)	(.00706)	(.00473)		(.00816)	(.00678)	(.00513)	
Household owns home	.00852	-.00178	-.0103	-120,89%	-.00476	-.0175	-.0127	267,65%
	(.0145)	(.0122)	(.00731)		(.0147)	(.0119)	(.00840)	
Young children in household	.0517***	.0464***	-.00527	-10,25%	.0524***	.0554***	.00292	5,73%
(age<=4 years)	(.0116)	(.00989)	(.00744)		(.0120)	(.00959)	(.00845)	
Family without children	-.0845***	-.0717***	.0128	-15,15%	-.0833***	-.0685***	.0148	-17,77%
(ref. single person)	(.0158)	(.0144)	(.00855)		(.0155)	(.0134)	(.00934)	
Single parents	-.0127	-.00777	.00495	-38,82%	-.00833	-.00362	.00471	-56,54%
(ref. single person)	(.0105)	(.00926)	(.00605)		(.0108)	(.00897)	(.00683)	
Family with children	-.234***	-.199***	.0349***	-14,96%	-.214***	-.186***	.0281***	-13,08%
(ref. single person)	(.0172)	(.0167)	(.0104)		(.0171)	(.0155)	(.0107)	
Subsample two	-.211***	-.227***	-.0157**	7,58%	-.245***	-.267***	-.0219***	8,98%
	(.0169)	(.0165)	(.00666)		(.0170)	(.0156)	(.00790)	
N	14.834	14.834			14.834	14.834		
Log Likelihood	-6.189	-4.898,10			-5.754,50	-4.369,40		
AIC	12.426	9.844,20			11.559	8.788,81		
rho					.62***	.77***		

Notes: Asterisks */**/*** denote statistically significant results using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. For the RE models, “rho” denotes the share of the total variance contributed by the panel-level variance component. "Subsample two" indicates whether an observation belongs to the second, nationally representative subsample. Survey wave indicators are included in all estimation.

Source: Own calculation based on PASS waves 2-7.

A comparison of the pooled and random effects probit models in columns 1 and 5 reveals the importance of controlling for unobserved heterogeneity at the household level. The results of the random effects estimation in column 5 allow us to reject the pooled model of take-up in column 1: the share of the total variance contributed by household-level variance reaches 62 percent, which is highly statistically significant at the one percent level. Therefore, we prefer the RE estimations when interpreting the marginal effects of the determinants of take-up.

We start by briefly discussing the estimated marginal effects for the models prior to correcting the dependent variable for underreporting (see columns 1 and 5). In general, the signs of the estimated marginal effects meet our expectations; for example, the propensity of benefit take-up is positively correlated with benefit entitlements, lower education, the advanced age of the head of household, or the presence of young children in the household. The size of the effects is broadly in line with other studies on the determinants of taking up UB II (e.g., Bruckmeier and Wiemers 2012, 2017). The marginal effect of the simulated benefit entitlement in the uncorrected probit model (column 1) implies that raising the entitlement by 100 Euros per month increases the probability of take-up by 4.5 percentage points. The RE probit (column 5) results in a slightly smaller marginal effect (4.2 percentage points).

Overall, the results across the two estimation approaches (see columns 1 and 5) are somewhat similar: younger household heads claim their benefits significantly less often than older ones. Interestingly, disability status is not correlated with benefit take-up. Take-up is less likely with an increasing level of education. Those residing in Eastern Germany have a higher propensity to take-up benefits than their counterparts in the West, which may reflect the relatively poor labor market situation and lower earnings expectations. Bruckmeier and Wiemers (2017) find similar relationships based on data from the S. We observe significant marginal effects of household composition. Ceteris paribus, the likelihood of claiming the benefit is significantly lower for couples with or without children compared to single person households, i.e., the reference group. This result might reflect the importance of a potential second earner for the take-up decision in couple households. Additionally, the presence of young children below age four strongly increases the probability of take-up.

Next, we study the results when we consider a corrected indicator of benefit take-up. We now correct all eligible households who report no receipt in the survey data but who actually receive benefits according to administrative records as take-up households. Columns 2 and 6 of Table 3 present the estimated marginal effects, columns 3 and 7 show the absolute differences in marginal effects and their statistical significance and columns 4 and 8 present the relative change in individual marginal effects when the corrected instead of the original take-up measure is used.

In the pooled and the panel estimation, some of the marginal effects change in economically and statistically significant ways when we correct the dependent variable (columns 3 and 7). For both estimators, the largest absolute change results for the indicators of advanced age vs. young age and for living in a family with children vs. living as a single person. We find the largest relative change in marginal effects for the immigration indicator, which doubles in size and becomes statistically significant after correcting for underreporting in the RE model. For both the pooled probit and the RE estimations, several statistically significant effects change by more than 30 percent (e.g., age and single parent status).

Thus, correcting for underreporting affects not only the level of non-take-up but also the impact of the correlates of the take-up decision. Therefore, we empirically corroborate the findings of Pudney's (2001) Monte Carlo simulations that even moderate measurement errors in reported benefit receipt can lead to strong biases in estimates coefficients.

5.3 Robustness tests

We offer several types of robustness tests. First, we consider the two alternative approaches of correcting for overreporting described in Section 4. So far, we have coded benefit-eligible overreporting households as take-up households because that is how they would be treated in a study that only uses survey data. In Table C.1 and Table C.2 in Appendix C we present the estimation results that we obtain when overreporters are instead coded as non-take-up households – which they are based on our administrative data – or are dropped from the sample, respectively. A comparison of Table C.1 and Table 3 shows that coding the overreporters as non-take-up households instead of take-up households reduces the large relative effect of recoding for the marginal effect of the immigrant indicator. The large and significant changes in the effects of age after recoding are robust to the alternative coding of the overreporting households. In Table C.2 we present the results when the subsample of overreporters is omitted from the estimation sample in the correction steps in columns 2 and 6. The outcomes are very close to the baseline results in Table 3. Overall, the robustness checks support our central finding that correcting for underreporting significantly affects the estimates of the take-up regression.

In the second robustness check, we repeat the estimations in Table 3 using sample weights. We use Stata's "svy"-commands to account for the complex survey design of the PASS following the recommendations given in Bethmann et al. (2013). The direction of almost all the statistically significant marginal effects in Table C.3 is robust to adding sample weights compared to Table 3. Some marginal effects increase in magnitude (e.g., immigrant, age, household composition). Our key interest is the effect of correcting the dependent variable. Most results from Table 3 are robust, although the absolute changes are generally smaller (see columns 3 and 7 in Table C.3).

Next, we address concerns about the potential selectivity arising from a) missing consent to the data linkage and b) failed linkage because not all households that agreed to the data linkage could actually be linked (see Section 4.2 and Appendix B). We reestimate our take-up equation for the full sample of eligible households before data linkage and thus without corrections in a third robustness check. Table C.4 in Appendix C shows the take-up estimation results for the full sample of households that are simulated as eligible to UB II. Thus, the results in Table C.4 are independent of the linkage procedure and are comparable to the typical situation in which administrative data are not available. Columns 1 and 3 show the results for all eligible households, while columns 2 and 4 again show the results of the linked sample from Table 3. With the exception of the marginal effect for the female indicator, the magnitude of the marginal effects changes only slightly and the statistical significance of all variables stays the same. Thus, we conclude that the linkage procedure does not introduce relevant selectivity effects.

In two further robustness checks, we analyze whether false matches in the linked data may bias our estimates. In the fourth robustness check, we only keep observations that were identified in the matching procedure based on the "gold standard" or deterministic procedure, i.e., the most reliable matches (see Appendix B).[3] False matches based on these two procedures are very un-

[3] The "gold standard" matches use an exact match of the household identifier, name, sex, and date of birth. Observations, which cannot be matched by the gold standard linkage, are matched based on "deterministic linkage," which uses first name, last name, zip code, city, street name, house number, sex, and the birth cohort indicator. Both gold standard and deterministic linkage should result in highly reliable results.

likely. We show the results for the take-up regression for the restricted sample of gold standard/deterministic-matches in Table C.5 in Appendix C. The comparison with our main findings depicted in Table 3 shows that the statistical significances, the signs and the magnitude of the marginal effects and the effects of the data correction change only slightly between the two models. In the fifth robustness check, we dropped all nonunique matches, which were corrected after matching (see Section 4.2 and Appendix B). We present the results of our take-up estimation for the sample without these observations in Table C.6 in Appendix C. Again, we find only minor differences compared to the main results, which are based on all matched observations (Table 3). This indicates that the inclusion of the corrected duplicates does not affect our results.

5.4 Patterns of underreporting

Finally, we offer a multivariate characterization of those households who underreported their benefit receipt going beyond the descriptive statistics of Table 1and Table 2. Using our baseline specification, we estimate probit models for the outcome "underreporting" with pooled and random effect models. Table 4 presents the results with pooled and RE models in columns 1 and 2, respectively.

Table 4 presents the marginal effects. We find large and statistically significant effects, particularly for the indicators of the age of the household head, where younger heads are most likely to underreport, and for families with children, who are more likely to underreport than individuals in single person households. Furthermore, households with smaller entitlements or whose head of household is higher educated or an immigrant are more likely to underreport. Finally, we find a lower probability of underreporting benefits for households from Eastern Germany.

Table 4: Regression of underreporting on household characteristics: marginal effects

Dependent variable: Underreporting of UB II	(1) Pooled probit	(2) RE-probit
Monthly simulated SA-entitlement	-.0159***	-.0162***
(in 100 €)	(.00109)	(.00110)
Female hh	.00255	.00329
	(.00585)	(.00616)
Hh is immigrant	.0186***	.0160**
	(.00685)	(.00693)
Age of hh: 25-34 years	-.0420***	-.0416***
(ref.:15-24 years)	(.0145)	(.0150)
Age of hh: 35-44 years	-.0453***	-.0448***
(ref.:15-24 years)	(.0148)	(.0153)
Age of hh: 45-54 years	-.0566***	-.0558***
(ref.:15-24 years)	(.0147)	(.0153)
Age of hh: >=55 years	-.0781***	-.0799***
(ref.:15-24 years)	(.0147)	(.0152)
Joint signif. of age (p-values)	*0,0000*	*0,0000*
Hh is disabled	-.0112	-.0113
	(.00725)	(.00759)
Hh holds lower sec. degree	.00531	.00259
(ref. no sec. degree)	(.00836)	(.00943)
Hh holds interm. sec. degree	.0167*	.0133
(ref. no sec. degree)	(.00874)	(.00978)
Hh holds upper sec. degree	.0278***	.0271**
(ref. no sec. degree)	(.0079)	(.0148)
Joint signif. of educ. (p-values)	*.0109*	*.02421*
Eastern Germany	-.0180***	-.0202***
	(.00548)	(.00568)
Household owns home	-.0170*	-.0149
	(.00989)	(.0107)
Young children in household	-.0113	-.0110
(age<=4 years)	(.00830)	(.00880)
Family without children	.0223**	.0245**
(ref. single person)	(.0112)	(.0118)
Single parents	.00526	.00604
(ref. single person)	(.00699)	(.00741)
Family with children	.0817***	.0783***
(ref. single person)	(.0143)	(.0143)
Joint signif. of hh (p-values)	*0,0000*	*0,0000*
Subsample two	.0105	.0146
(ref.: SA-recipients-sample)	(.0104)	(.0117)
N	12,169	12,169
(Pseudo)log-likelihood	-2,956.5	-2,894.1
AIC	5,961.1	5,838.2
rho		.439***

Notes: Asterisks */**/*** denote statistically significant results using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. For the RE models, "rho" denotes the share of the total variance contributed by the panel-level variance component. "Subsample two" indicates whether an observation belongs to the second, nationally representative subsample. Survey wave indicators are included in all estimation.

Source: Own calculation based on PASS waves 2-7.

A comparison of the characteristics of underreporting households on the one hand (see Table 4) and of the underreporting-induced bias of marginal effects of given characteristics on the other hand (see Table 3) yields similar patterns. We find both strong age dependence in underreporting and a strong response of underreporting correction in the marginal effect of age. Similarly, we observe large absolute changes in the marginal effects of living in a family with children when correcting the take-up measure. These outcomes are also directly correlated with the underreporting of benefit take-up. Clearly, the estimation of take-up regressions is more reliable for those groups for whom the outcome is measured correctly. Thus, the bias in survey-based estimations of take-up equations may vary depending on the extent to which an analysis framework correlates with the propensity to underreport.

Take-up analyses are often determined by an interest in distributional effects of government transfers, e.g., regarding household income. If the well-being indicators of certain parts of the distribution are more likely mismeasured due to their misreporting of benefit receipt, the results for these groups tend to be biased. We are among the first to show such patterns empirically, which are important for the correct interpretation of distributional analyses of government benefits.

6 Conclusions

This study contributes to the literature on benefit non-take-up behavior. Because this literature relies on survey data, it suffers from measurement error if respondents do not reveal their true use of welfare benefits. We inspect this issue for the case of a general welfare program using linked representative survey and administrative data. Approximately 84 percent of the simulated eligible households could be linked to our administrative data. For the linked households, we simulate a non-take-up rate based on survey information of only 40 percent, which is in line with results found for comparable benefit programs in other countries (see, e.g., Eurofound 2015). The data linkage yields a benefit underreporting rate of 7.8 percent in the survey data. Correcting the survey responses on benefit receipt for underreporting reduces the simulated non-take-up rate to 35 percent.

We use the information on benefit underreporting to test whether the results of take-up regressions differ depending on the treatment of underreporting households, i.e., depending on whether we code them as take-up or non-take-up households in our outcome measure. We estimate pooled and panel probit models and calculate marginal effects. When we compare the estimation results obtained with corrected and uncorrected dependent variables, we find that the absolute difference in marginal effects is statistically significant and often large. In relative terms, many marginal effects change by at least 30 percent after the correction. These results are robust both to various changes in the estimation approach and to alternative treatments of overreporting households.

We find that the patterns of the changed marginal effects agree with the correlation patterns underlying underreporting behaviors: households of foreign origin, with small benefit claims, or with young household heads and with children are particularly likely to underreport their benefit receipt. The marginal effects of just these characteristics changed the most when we corrected our take-up outcome measure. This agrees well with the literature showing that households close to the labor market or with a risk of benefit confusion tend to underreport or to not take-up their benefit (Bruckmeier and Wiemers 2012, 2017, Bruckmeier et al. 2014). The mechanisms of these groups' underreporting are different and individually plausible. Because German naturalization rules require that applicants should be able to support themselves and do not rely on social transfers or means-tested benefits, there may be a high perceived cost connected to admitting benefit receipt for immigrants (see, e.g., Riphahn and Saif 2018). Households who are close to the labor market may suffer from (perceived) stigma effects and work the hardest to avoid transfer dependence. Those receiving several social transfers at the same time may not be able to keep track of the specific transfer programs from which they benefit, particularly if transfer eligibility changes in short intervals. Overall, distributional analyses of government transfer programs may therefore be at risk of measurement-induced errors for specific population groups.

In sum, the analyses based on our linked data suggest that research concerned with take-up and its determinants needs to account for potential misreporting of benefit receipt. The marginal effects in regressions of benefit take-up may well be biased unless the outcome measures can be corrected for misreporting.

Bibliography

Antoni, M.; Ganzer, A.; vom Berge; P. (2016): Sample of integrated labour market biographies (SIAB) 1975-2014. FDZ-Datenreport, 04/2016 (en)), Nürnberg.

Antoni, Manfred; Schnell, Rainer (2017): The past, present and future of the German Record Linkage Center (GRLC). Journal of Economics and Statistics, online first.

Antoni, M.; Dummert, S.; Trenkle, S. (2017): PASS-Befragungsdaten verknüpft mit administrativen Daten des IAB (PASS-ADIAB) 1975-2015. FDZ-Datenreport, 06/2017 (de), Nürnberg.

Bargain, O.; Immervoll, H.; Viitamäki, H. (2012): No claim, no pain. Measuring the non-take-up of social assistance using register data, The Journal of Economic Inequality 10, 375-395.

Berg, M.; Cramer, R.; Dickmann, C.; Gilberg, R.; Jesske, B.; Kleudgen, M.; Bethmann, A.; Fuchs, B.; Huber, M.; Trappmann, M.; (2014): Codebook and documentation of the panel study “Labour market and social security” (PASS). Datenreport wave 7. Nürnberg.

Berg, M.; Cramer, R.; Dickmann, C.; Gilberg, R.; Jesske, B.; Kleudgen, M.; Bethmann, A.; Fuchs, M.; Trappmann, M.; Wurdack, A. (2012): Codebook and documentation of the panel study “Labour market and social security” (PASS). Datenreport wave 5. Nürnberg.

Beste, J.; Grabka, M. M.; Goebel, J. (2018): Armut in Deutschland – Ein Vergleich zwischen den beiden Haushaltspanelstudien SOEP und PASS, AStA Wirtschafts- und Sozialstatistisches Archiv 12, 27-62.

Bethmann, A.; Fuchs, B.; Huber, M.; Trappmann, M.; Reindl, A.; Berg, M.; Cramer, R.; Dickmann, C.; Gilberg, R.; Jesske, B.; Kleudgen, M.; (2016): Codebook and documentation of the panel study “Labour market and social security” (PASS). Datenreport wave 9. Nürnberg.

Bethmann, A.; Fuchs, B.; Wurdack, A. (2013): User Guide “Panel Study labour Market and Social Security” (PASS) – Wave 6, Nürnberg.

Blank, R.; Ruggles, P. (1996): When do women use aid to families with dependent children and food stamps? The dynamics of eligibility versus participation. Journal of Human Resources, Vol. 31, 57–89

Blundell, R.; Fry, V.; Walker, I. (1988): Modelling the take-up of means-tested benefits: the case of housing benefits in the UK. Economic Journal 98, 58-74.

Bollinger, C.R.; David, M.H. (1997): Modeling discrete choice with response error: food stamp participation, Journal of the American Statistical Association 92, 827-835.

Bollinger, C.R.; David, M.H. (2001): Estimation with response error and nonresponse: food-stamp participation in the SIPP, Journal of Business and Economic Statistics 19, 129-141.

Bound, J., Brown,C.; Mathiowetz,N. (2001): Measurement error in survey data. In: Handbook of Econometrics, vol. 5 (eds J. J. Heckman and E. Leamer), pp. 3705–3843. Amsterdam: Elsevier Science.

Bradburn N.M.; Huttenlocher J.; Hedges L. (1994): Telescoping and Temporal Memory. In: Autobiographical Memory and the Validity of Retrospective Reports (eds N. Schwarz and S. Sudman). Springer, New York, NY.

Bruckmeier, K.; Müller, G.; Riphahn, R.T. (2014): Who misreports welfare receipt in surveys? Applied Economics Letters 21, 812-816, DOI: 10.1080/13504851.2013.877566

Bruckmeier, K.; Wiemers, J. (2012): A new targeting: a new take-up? Non-take-up of social assistance in Germany after social policy reforms, Empirical Economics 43, 565-580.

Bruckmeier, K.; Wiemers, J. (2017): Differences in welfare take-up between immigrants and natives – a microsimulation study. International Journal of Manpower 38(2), 226-241.

Bundesagentur für Arbeit (2008): Grundsicherung für Arbeitsuchende: Bedarfe, Leistungen und Haushaltsbudget. Bericht der Statistik der Bundesagentur für Arbeit, Nürnberg.

Card, David, Andrew K. G. Hildreth and Lara D. Shore-Sheppard (2001): The Measurement of Medicaid Coverage in the SIPP: Evidence from California 1990-1996, NBER Working Paper No. 8514, Cambridge, Mass.

Duclos, J.-Y. (1995): Modelling the take-up of state support, Journal of Public Economics 58, 391-415.

Duclos, J.-Y. (1997): Estimating and Testing a Model of Welfare Participation: The Case of Supplementary Benefits in Britain, Economica 64, 81-100.

Eurofound (2015): Access to social benefits: Reducing non-take-up, Publications Office of the European Union, Luxembourg.

Frick, J. R.; Groh-Samberg, O. (2007): To Claim or Not to Claim: Estimating Non-Take-Up of Social Assistance in Germany and the Role of Measurement Error. SOEP Papers on Multidisciplinary Panel Data Research, No. 53, DIW, Berlin.

Gebhardt, D.; Müller, G.; Bethmann, A.; Trappmann, M.; Christoph, B.; Gayer, C.; Müller, B.; Tisch, A.; Siflinger, B.; Kiesl, H.; Huyer-May, B.; Achatz, J.; Wenzig, C.; Rudolph, H.; Graf, T.; Biedermann, A. (2009): Codebook and documentation of the panel study "Labour Market and Social Security" (PASS) * Volume I: Introduction and overview. Wave 2 (2007/2008). FDZ-Datenreport, 06/2009.

Hancock, R.; Barker, G. (2005): The Quality of Social Security Benefit Data in the British Family Resources Survey: Implications for Investigating Income Support Take-up by Pensioners. Journal of the Royal Statistical Society. Series A (Statistics in Society) 168, 63-82.

Hernandez, M., Pudney, S. E. (2007): Measurement error in models of welfare participation. Journal of Public Economics 91, 327-341.

Hernandez, M; Pudney, S.; Hancock, R. (2007): The welfare cost of means-testing: pensioner participation in income support. Journal of Applied Econometrics 22, 581-598.

Krafft, C.; Davis, E. E.; Tout, K. (2015): Elizabeth E. Davis, Kathryn Tout: The Problem of Measurement Error in Self-Reported Receipt of Child-Care Subsidies: Evidence from Two States. Social Service Review 89, 686-726.

Lynn, P., Jäckle, A.; Jenkins, S. P.; 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 175, 289-308.

Meyer, B. D.; Mittag, N. (2017a): Misclassification in binary choice models. Journal of Econometrics 200, 295-311.

Meyer, B. D.; Mittag, N. (2017b): Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net, IZA Discussion Papers, No. 10943, Institute of Labor Economics (IZA), Bonn.

Meyer, B. D.; Goerge, R. M. (2011): Errors in survey reporting and imputation and their effects on estimates of food stamp program participation, CES (Center for Economics Studies) Working Paper 11-14, U.S. Census Bureau, Washington.

Meyer, B. D.; Wallace, K. C. M.; Sullivan, J. X. (2015): Household Surveys in Crisis, Journal of Economic Perspectives 29, 199-226.

Mittag, N. (2016): Correcting for Misreporting of Government Benefits, IZA Discussion Papers No. 10266, IZA Bonn.

Moffitt, R. (1983): An Economic Model of Welfare Stigma. American Economic Review 73, 1023-1035.

Pudney, S. (2001): The Impact of Measurement Error in Probit Models of Benefit Take-Up (mimeo) University of Leicester.

Riphahn, R. T. (2001): Rational Poverty of Poor Rationality? – The Take-up of Social Assistance Benefits. Review of Income and Wealth 47(3), 379–398.

Riphahn, R. T. and S. Saif (2018): Naturalization and the labor market performance of immigrants in Germany, FAU Erlangen-Nürnberg, mimeo.

Sakshaug, J. W.; Antoni, M.; Sauckel, R. (2017): The quality and selectivity of linking federal administrative records to respondents and nonrespondents in a general population survey in Germany. Survey Research Methods 11(1), 63-80.

Sakshaug, J. W.; Kreuter, F. (2012): Assesing the Magnitude of Non-Consent Bias in Linked Survey and Administrative Data. Survey Research Methods 6, 113-122.

Statistik der Bundesagentur der Arbeit (2018): Tabellen, Strukturen der Grundsicherung SGB II (Zeitreihe Monats- und Jahreszahlen ab 2005), Nürnberg, Mai 2018.

STBA (Statistisches Bundesamt) (2017): Statistisches Jahrbuch 2017, Wiesbaden.

Taeuber, C.; Dean M. Resnick, D. M.; Love, S. P.; Staveley, J.; Wilde, P.; Larson, R. (2004): Differences in the Estimates of Food Stamp Program Participation Between Surveys and Administrative Records, Working Paper, U.S. Census Bureau, Washington.

Trappmann, M.; Beste, J.; Bethmann A.; Müller, G. (2013): The PASS panel survey after six waves, Journal of Labor Market Research 46, 275-281.

Trappmann, M.; Gundert, S.; Wenzig, C.; Gebhardt, D. (2010): PASS – A Household Panel Survey for Research on Unemployment and Poverty, Schmollers Jahrbuch (Journal of Applied Social Science Studies) 130, 609-622.

van Oorschot, W. (1991): Non-Take-Up of Social Security Benefits in Europe. Journal of European Social Policy 1(1), 15–30.

Whelan, S. (2010): The take-up of means-tested income support. Empirical Economics 39(3), 847–875.

Wilde, J.; Kubis, A. (2005): Nichtinanspruchnahme von Sozialhilfe – Eine empirische Analyse des Unerwarteten. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 225, 347–373.

Appendix A: Simulation of welfare benefit eligibility / simulation of entitlement

We base our simulation on waves 2 to 7 of the Panel Study "Labour Market and Social Security" (PASS). In these six waves, 55,069 household interviews were realized. In general, the simulation of welfare benefits using survey data requires several sample selection steps (see Table A.1).

We first restrict the sample to households for which both household and personal interviews are available. We further drop households headed by individuals over the age of 65 because they are not eligible according to the rules of the German minimum income support program Unemployment Benefit II (UB II). If households do not answer the question on benefit receipt, we are not able to distinguish between take-up and non-take-up. Therefore, we omit household observations with missing information on benefit receipt. In the next step, we drop students and apprentices because they must apply for different benefits. Furthermore, we drop households in which not all members participated in the survey. Households in which more than one "community of need" exists are also omitted because we cannot assign household income to different household members. The "community of need" is the legally relevant unit in the means test; in most cases, it is identical with the household. The "community of need" consists of singles and their partner and their children up to age 24. Since we cannot assign household incomes and household wealth across different "communities of need", we keep only those households in the sample that consist of exactly one such community.

For the computation of household income, which is relevant to the means test, we need information on household incomes as reported in the household questionnaire and on personal incomes of household members as reported in individual questionnaires. Thus, we must drop households with missing information on these variables. After these sample selection steps, the simulation sample is composed of 30,878 household-year observations.

Table A.1: Sample selection for the simulation of unemployment benefit II (number of cases)

	Wave2	Wave3	Wave4	Wave5	Wave6	Wave7	All
All households	8.429	9.535	7.848	10.235	9.513	9.509	55.069
-Only household interview available	140	190	109	50	58	49	596
-Interviews of persons over 65 years	847	939	1.164	1.766	1.781	1.840	8.337
-Missing informationon receipt of unemployment benefitII	58	45	68	15	13	17	216
-Households of students and apprentices	368	434	352	468	421	414	2.457
-Households, In which not all individuals responded	1.854	2.332	1.292	1.851	1.525	1.562	10.416
-More than one "community of needs" in the household	168	179	164	231	226	203	1.171
-Missing values in income variables	346	351	257	18	17	9	998
Final simulation sample	4.648	5.065	4.442	5.836	5.472	5.415	30.878

Source: Own calculation based on PASS waves 2-7.

Our simulation procedure implements the eligibility rules for UB II benefits. A household is eligible if the household's total need exceeds the income and the household's wealth remains below the household-specific maximum. Additionally, we simulate eligibility for means-tested housing allowance and supplementary child allowance ("Kinderzuschlag"), which must be claimed prior to claiming UB II benefits. Including these prioritized benefits in the simulation procedure is important to assess UB II eligibility correctly: the household can only claim UB II if benefits from the housing allowance (possibly in combination with the supplementary child allowance) do not cover its basic needs.

In the first step of our simulation procedure, we calculate the total needs of each household. We determine total needs as the legally defined regular personal needs of household members, additional needs, and housing costs. Next, we consider an additional national standardized benefit for single parents, which varies with the age of their children. The PASS data provide information on the household type and the age of children in the household. Households also report their housing costs. We employ the reported monthly rent for tenants and the reported monthly repayment of mortgage loans for homeowners. Furthermore, we consider reported housing costs for all households. We do not consider additional needs that can be claimed by households in "special circumstances". "Special circumstances" include certain disabilities, pregnancy, and special dietary needs for health reasons. The incidence of these special benefits is very low: in 2007, approximately 5 percent of all households received one of these benefits, and the average benefit was 50 Euros per month (Bundesagentur für Arbeit 2008). Therefore, we can assume that ignoring them will not alter our results considerably.

In the second step of our simulation, we determine the household income relevant to the means test. Household income consists of the sum of all the individual incomes of household members. We apply the earned income exemption rules, which depend on family type and vary with gross wages. All other reported types of income – capital income, rental income, public, and private transfers – reduce the benefit by 100 percent of their amount.

In the third step, we check whether the household's wealth exceeds the allowable wealth. Each household reports its total financial wealth in each survey wave. The answers are coded in wealth brackets. We assume that the mean of the reported wealth bracket represents the household's wealth and compare it with the individual maximum allowable wealth. The calculation of total allowable maximum wealth depends on household structure and age only.

In step four, we calculate entitlements to the housing allowance and the supplementary child allowance and compare the results with those calculated for UB II. If the combined amount of the former two benefits exceeds the UB II benefit, the household is ineligible for UB II.

Following these steps, our simulation classifies 17,585 out of 30,878 households as eligible for UB II. One possibility to assess the quality of the simulation is to consider the number of households reporting benefit receipt in the survey, but for whom the simulation model fails to simulate eligibility, i.e., the type II or beta error (see, e.g., Bargain et al. 2012, Frick and Groh-Samberg 2007). The number of misclassified households provides an upper bound of the simulation error because beta errors can also be caused by administrative errors in the assessment of eligibility or false answers provided by the respondents in the survey. Our simulation yields a beta error rate of 3.9 percent (weighted) or 4.6 percent (unweighted). We interpret these small beta error rates as indicative of a high simulation quality. For comparison, based on the German Socio-Economic Panel, Frick and Groh-Samberg (2007) report a beta-error rate of 12.6 percent.

In addition to quantifiable errors, our simulations may be subject to nonquantifiable errors. These may result if the simulations mispredict transfer eligibility or if the information on household income and wealth is incorrect. If, e.g., respondents underreport household finances, then a simulated benefit eligibility may be wrong and the observed non-take-up rate is overestimated. However, this will not affect our evaluation of household misreporting on transfer receipt, since these households will not appear in the administrative data as benefit recipients.

Appendix B: Data Linkage

In the person interview, the PASS asks respondents aged 15 to 65 for their consent to link administrative data from the Federal Employment Agency to their survey data. The question reads as follows (English translation, see Sakshaug and Kreuter 2012):

> "To keep the interview as brief as possible, the Institute for Employment Research in Nuremberg could merge the study results with data about your times of employment, unemployment or participation in measures by the employment office (Arbeitsamt). For the results of this study it would be a great advantage. For reasons of data protection this cannot be done without your agreement, which I kindly ask you to provide. This is of course just as voluntary as the interview you are so kind as to give us. Of course, you may withdraw your consent at any time. It goes without saying that all rules of data protection and of the de-personalization of the results reported apply to these additional data as well.
>
> So, may I write down your answer: Do you agree to the use of this additional data?"

Overall, in the PASS, the share of respondents who agree to merge their data is approximately 80 percent (Berg et al. 2014). Our sample of 17,585 simulated UB II eligible household-year observations consists of 8,318 different individual households. Table B.1 shows for the sample of 8,318 UB II eligible households how often the consent question was asked in each wave and how many respondents agreed. The number of times the consent question was asked (9,349) exceeds the number of eligible household observations, as participants who did not agree to the data linkage during their initial interview are asked once more in the next wave. We find a high average consent rate of approximately 83 percent in our sample.

Table B.1: Data linkage consent by wave

Wave	Number of respondents with simulated UB II entitlements who were asked for consent to the data linkage	Number of respondents with simulated UB II entitlements who agreed to the data linkage	Share of respondents who agreed to the data linkage (in percent)
Wave 1	3.622	3.022	83,4
Wave 2	1.161	906	78,0
Wave 3	908	742	81,7
Wave 4	676	595	88,0
Wave 5	1.565	1.355	86,6
Wave 6	828	686	82,9
Wave 7	589	488	82,9
All	9.349	7.794	83,4

Note: Respondents who agreed to linkage in wave t were linked in all subsequent waves. Respondents who did not agree to linkage in wave t where asked once again in the subsequent wave. If they declined again, they were not asked about linkage again.

Source: Own calculation based on PASS waves 2-7.

In the sample that we use for the analysis, we can link 16,874 household-year observations of simulated eligible respondents to the administrative data, i.e., 96 percent of our 17,585 UB II eligible household-year observations. Thus, our rate of linkage is substantially higher than the average consent rate. This happens for two reasons: first, a consent given once holds for all future and past waves of the PASS. Second, participants who do not agree to the data linkage are asked again in the next wave. Only if households refuse to give their consent in two consecutive waves is the question no longer repeated in future waves.

In the next step, we merge the observations of respondents who agreed to the linkage to a key file, which identifies respondents in the administrative data. The German Record Linkage Center provides this file, which utilizes several administrative data sources collected by the Federal Employment Agency (BA) (Antoni et al. 2016).

The matching variables used in the linkage are a person's first name, last name, zip code, city, street name, house number, sex, and an indicator for the birth cohort (Antoni et al. 2017). These variables are available in the sampling data and in the administrative data. For the PASS sample drawn from UB II recipients, an additional household identifier is available. The linkage follows a stepwise procedure with variation across the number of matching-variables and record linkage processes. Antoni et al. (2017) and Sakshaug et al. (2017) describe the linkage processes. They label a match "gold standard linkage" if it is based on an exact match of the household identifier, name, sex, and date of birth. This highest-quality match is possible only for households in the UB II sample. Observations that cannot be matched by the gold standard linkage are matched based on "deterministic linkage". This procedure uses first name, last name, zip code, city, street name, house number, sex, and the birth cohort indicator. Both gold standard and deterministic linkage should result in highly reliable results. For observations that could not be linked using these two procedures, distance-based and probabilistic linkage procedures are used, which match based on comparison functions using first name, last name, zip code, city, street name, house number, sex, and birth cohort.

Table B.2 shows the frequency of linkage procedures for our sample of simulated UB II eligible households. From our 15,925 matched observations, 13,089 observations (82 percent) are linked by the gold standard match. Adding the 2,073 observations which are linked by the deterministic match, our overall share of highly reliable matches (gold standard and deterministic) exceeds 95 percent. Because of this high share of reliable matches, we consider the overall match quality to be excellent.

Only 574 observations, mainly from the population sample, are matched based on the distance-based procedure. Finally, 189 observations are valid matches, but the type of match is recorded as missing in the data. Since there might be concerns about the reliability of these latter two types of matches, we provide a robustness check (Table C.5 in Appendix C) in which we keep only the gold standard and deterministic matches in our estimation sample.

Table B.2: Linkage procedures by sample and reported benefit receipt (household-year observations)

	Observations		Sample		UB II receipt	
			UB II	Popula-tion	reported	not re-ported
Gold standard match	13.089	82,2%	13.089	0	2.521	10.568
Deterministic match	2.073	13,0%	1.021	1.052	628	1.445
Distance-based/probabilistic match	574	3,6%	132	442	247	327
Valid match, unknown match type	189	1,2%	142	47	51	138
All	15.925	100,0%	14.384	1.541	3.447	12.478

Note: Linkage procedures for the sample of 15,925 simulated UB II eligible household-year observations with consent to data linkage by sample type (columns 3 and 4) and reported UB II receipt (columns 5 and 6).

Source: Own calculation based on PASS waves 2-7 linked to administrative data.

In some instances, the matching procedure generated duplicate matches, i.e., a survey observation can have more than one valid match in the administrative data and vice versa. In our linked data, duplicate matches are resolved by choosing one of the duplicate observations based on gender, year of birth and highest level of education. This affected 830 cases for which two survey respondents were assigned to the same person in the administrative data and 77 cases for which two persons in the administrative data were assigned to one survey respondent. As a robustness check, we reestimated the key results presented in this paper based on a sample in which we dropped all observations with ambiguous, i.e., duplicate matches. All results proved to be robust against this selection step (see Table C.6 in Appendix C).

One potential problem with the data linkage is that results may be biased because of selectivity in nonconsent and nonidentifiability in the administrative data. For the sample of simulated eligible households with consent to data linkage, Table B.3 shows the correlates of the probability of not giving consent to data linkage and the probability that a household cannot be linked to the administrative data. The results indicate only a minor selection bias concerning the composition of simulated eligible households.

Table B.3: Regression of UB II-eligible nonconsent (1) and nonlinkable (2) household-year observations: marginal effects

Dependent variable Model	(1) No consent to data linkage Pooled probit	(2) Consent to data linkage but not linkable Pooled probit
Monthly simulated SA-entitlement	-.000106	-8.57e-05
(in 100 €)	(.000672)	(.000812)
Female hh	.00368	.00773
	(.00508)	(.00662)
Hh is immigrant	.0156***	.0126*
	(.00559)	(.00747)
Age of hh: 25-34 years	-.0194*	-.0233
(ref.:15-24 years)	(.0103)	(.0154)
Age of hh: 35-44 years	-.0117	-.0334**
(ref.:15-24 years)	(.0107)	(.0156)
Age of hh: 45-54 years	-.0226**	-.0270*
(ref.:15-24 years)	(.0105)	(.0156)
Age of hh: >=55 years	-.0218**	-.0290*
(ref.:15-24 years)	(.0107)	(.0160)
Hh is disabled	-.0142***	.000576
	(.00498)	(.00799)
Hh holds lower sec. degree	-.00708	-.0160
(ref. no sec. degree)	(.00763)	(.0117)
Hh holds interm. sec. degree	-.00178	-.00656
(ref. no sec. degree)	(.00796)	(.0124)
Hh holds upper sec. degree	.0142	-.00641
(ref. no sec. degree)	(.00906)	(.0128)
Eastern Germany	-.0115**	-.00964
	(.00466)	(.00663)
Household owns home	.00932	.0259**
	(.0108)	(.0122)
Young children in household	-.00447	-.00969
(age<=4 years)	(.00712)	(.00902)
Family without children	-.0226***	.0190
(ref. single person)	(.00658)	(.0122)
Single parents	-.0219***	6.22e-05
(ref. single person)	(.00617)	(.00876)
Family with children	-.0227***	.00259
(ref. single person)	(.00705)	(.0101)
Subsample two	-.00605	.0511***
(ref.: SA-recipients-sample)	(.00662)	(.0121)
N	16.357	15.707
(Pseudo)log-likelihood	-2.658,69	-3.290,48
AIC	5.365,37	6.628,96

Notes: Asterisks */**/*** denote statistically significant results using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample. Survey wave indicators are included in all estimation. Column (1) shows the estimates for the consent to data linkage (with 0 = no consent (N = 15,707), 1 = consent (N = 650)) for the sample of 17,585 eligible household-year observations reduced by 1,228 observations with missing values in some covariates. Column (2) shows the estimates for being a nonlinkable household-year observation (with 0 = linkable (N = 14,834) and 1 = not linkable (N = 873)), for the sample of 16,874 household-year observations that agreed to the data linkage reduced by 1,167 observations with missing values in some covariates.

Source: Own calculation based on PASS waves 2-7.

Appendix C: Regression Results – Robustness Checks

Table C.1: Take-up regression: marginal effects after correction of underreporting and after correction of under- and overreporting (overreporters recoded to non-take-up households)

Dependent variable: Take-up of UB II	(1) Pooled probit, corrected for underreporting	(2) Pooled probit, corrected for under-/ over-reporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, corrected for under- reporting	(6) RE probit, corrected for under-/ over-reporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0365***	.0376***	.00118**	3,01%	.0317***	.0327***	.000956*	3,15%
	(.00110)	(.00119)	(.000531)		(.00119)	(.00125)	(.000552)	
Female hh	-.0102	-.00874	.00145	-14,31%	-.0116	-.0101	.00149	-12,93%
	(.00760)	(.00856)	(.00419)		(.00725)	(.00773)	(.00342)	
Hh is immigrant	.0114	.00602	-.00536	-47,19%	.0192**	.0156*	-.00357	-18,75%
	(.00819)	(.00942)	(.00516)		(.00766)	(.00826)	(.00389)	
Age of hh: 25-34 years	.0513***	.0353**	-.0160**	-31,19%	.0395**	.0292*	-.0104*	-26,08%
(ref.:15-24 years)	(.0162)	(.0173)	(.00697)		(.0160)	(.0165)	(.00584)	
Age of hh: 35-44 years	.0611***	.0496***	-.0115*	-18,82%	.0507***	.0438***	-.00693	-13,61%
(ref.:15-24 years)	(.0167)	(.0175)	(.00681)		(.0162)	(.0167)	(.00612)	
Age of hh: 45-54 years	.0815***	.0757***	-.00583	-7,12%	.0667***	.0645***	-.00220	-3,30%
(ref.:15-24 years)	(.0164)	(.0172)	(.00655)		(.0160)	(.0165)	(.00599)	
Age of hh: >=55 years	.120***	.109***	-.0110	-9,17%	.102***	.0929***	-.00947	-8,92%
(ref.:15-24 years)	(.0166)	(.0176)	(.00709)		(.0164)	(.0169)	(.00668)	
Hh is disabled	-.00628	-.0154	-.00908	145,22%	.00125	-.00556	-.00680	-544,80%
	(.0105)	(.0117)	(.00612)		(.00901)	(.00960)	(.00514)	
Hh holds lower sec. degree	-.00618	-.00215	.00403	-65,21%	-.0184	-.0150	.00337	-18,48%
(ref. no sec. degree)	(.0126)	(.0146)	(.00852)		(.0114)	(.0127)	(.00700)	
Hh holds interm. sec. degree	-.0370***	-.0249*	.0121	-32,70%	-.0542***	-.0434***	.0108	-19,93%
(ref. no sec. degree)	(.0129)	(.0148)	(.00829)		(.0118)	(.0131)	(.00694)	
Hh holds upper sec. degree	-.0680***	-.0564***	.0116	-17,06%	-.0832***	-.0670***	.0162**	-19,47%
(ref. no sec. degree)	(.0145)	(.0165)	(.00895)		(.0131)	(.0143)	(.00717)	
Eastern Germany	.0424***	.0456***	.00322	7,55%	.0397***	.0407***	.00101	2,52%
	(.00706)	(.00794)	(.00389)		(.00678)	(.00730)	(.00329)	
Household owns home	-.00178	-.00306	-.00128	71,91%	-.0175	-.0214	-.00392	22,29%
	(.0122)	(.0134)	(.00479)		(.0119)	(.0130)	(.00526)	
Young children in household	.0464***	.0464***	7.09e-06	0,00%	.0554***	.0605***	.00512	9,21%
(age<=4 years)	(.00989)	(.0116)	(.00643)		(.00959)	(.0103)	(.00480)	
Family without children	-.0717***	-.0688***	.00292	-4,04%	-.0685***	-.0714***	-.00295	4,23%
(ref. single person)	(.0144)	(.0150)	(.00584)		(.0134)	(.0136)	(.00626)	
Single parents	-.00777	-.0121	-.00434	55,73%	-.00362	-.00484	-.00122	33,70%
(ref. single person)	(.00926)	(.0106)	(.00527)		(.00897)	(.00979)	(.00467)	
Family with children	-.199***	-.199***	.000199	0,00%	-.186***	-.184***	.00213	-1,08%
(ref. single person)	(.0167)	(.0173)	(.00705)		(.0155)	(.0157)	(.00673)	
Subsample two	-.227***	-.225***	.00200	-0,88%	-.267***	-.265***	.00209	-0,75%
	(.0165)	(.0173)	(.00597)		(.0156)	(.0160)	(.00492)	
N	14.834	14.834			14.834	14.834		
Log Likelihood	-4.898,10	-5.686,94			-4.369,40	-5.005,06		
AIC	9.844,20	11.421,88			8.788,81	1.060,12		
rho					.77***	.78***		

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample.

Source: Own calculation based on PASS waves 2-7.

Table C.2: Take-up regression: marginal effects after correction of underreporting and after correction of under- and overreporting (overreporters dropped from the sample)

Dependent variable: Take-up of UB II	(1) Pooled probit, corrected for underreporting	(2) Pooled probit, corrected for under-/ over-reporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, corrected for under- reporting	(6) RE probit, corrected for under-/ over-reporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0365***	.0373***	.000851***	2,19%	.0317***	.0322***	.000447**	1,58%
	-.0011	-.00112	-.000111		-.00119	-.00125	-.00018	
Female hh	-.0102	-.00960	.000590	-5,88%	-.0116	-.0113	.000297	-2,59%
	-.0076	-.00779	-.000672		-.00725	-.00726	-.000756	
Hh is immigrant	.0114	.0111	-.000252	-2,63%	.0192**	.0198***	.000630	3,13%
	-.00819	-.00838	-.000756		-.00766	-.00767	-.000794	
Age of hh: 25-34 years	.0513***	.0496***	-.00170	-3,31%	.0395**	.0368**	-.00276*	-6,84%
(ref.:15-24 years)	-.0162	-.0165	-.00147		-.016	-.0161	-.00155	
Age of hh: 35-44 years	.0611***	.0610***	-8.96e-05	-0,16%	.0507***	.0501***	-.000609	-1,18%
(ref.:15-24 years)	-.0167	-.0169	-.00137		-.0162	-.0163	-.00159	
Age of hh: 45-54 years	.0815***	.0814***	-9.26e-05	-0,12%	.0667***	.0672***	.000475	0,75%
(ref.:15-24 years)	-.0164	-.0166	-.00139		-.016	-.0161	-.00148	
Age of hh: >=55 years	.120***	.121***	.000646	0,83%	.102***	.102***	-.000171	0,00%
(ref.:15-24 years)	-.0166	-.0168	-.00143		-.0164	-.0165	-.00163	
Hh is disabled	-.00628	-.00865	-.00238**	37,74%	.00125	-.000253	-.00150	-120,24%
	-.0105	-.0109	-.00104		-.00901	-.00915	-.00125	
Hh holds lower sec. degree	-.00618	-.00568	.000498	-8,09%	-.0184	-.0179	.000535	-2,72%
(ref. no sec. degree)	-.0126	-.013	-.00119		-.0114	-.0116	-.00127	
Hh holds interm. sec. degree	-.0370***	-.0363***	.000655	-1,89%	-.0542***	-.0536***	.000616	-1,11%
(ref. no sec. degree)	-.0129	-.0133	-.00116		-.0118	-.012	-.00124	
Hh holds upper sec. degree	-.0680***	-.0677***	.000206	-0,44%	-.0832***	-.0821***	.000999	-1,32%
(ref. no sec. degree)	-.0145	-.0149	-.00133		-.0131	-.0132	-.00141	
Eastern Germany	.0424***	.0434***	.000974*	2,36%	.0397***	.0403***	.000655	1,51%
	-.00706	-.00722	-.000588		-.00678	-.00679	-.000684	
Household owns home	-.00178	-.00172	5.47e-05	-3,37%	-.0175	-.0174	3.10e-05	-0,57%
	-.0122	-.0124	-.00116		-.0119	-.0119	-.00123	
Young children in household	.0464***	.0465***	2.68e-05	0,22%	.0554***	.0574***	.00203	3,61%
(age<=4 years)	-.00989	-.0102	-.00101		-.00959	-.00977	-.00138	
Family without children	-.0717***	-.0744***	-.00269*	3,77%	-.0685***	-.0729***	-.00437*	6,42%
(ref. single person)	-.0144	-.0148	-.00154		-.0134	-.0133	-.00252	
Single parents	-.00777	-.00823	-.000464	5,92%	-.00362	-.00327	.000346	-9,67%
(ref. single person)	-.00926	-.00949	-.000772		-.00897	-.00907	-.000988	
Family with children	-.199***	-.202***	-.00354**	1,51%	-.186***	-.189***	-.00262	1,61%
(ref. single person)	-.0167	-.0171	-.00168		-.0155	-.0155	-.00248	
Subsample two	-.227***	-.229***	-.00285*	0,88%	-.267***	-.270***	-.00290	1,12%
	-.0165	-.0168	-.00165		-.0156	-.0154	-.00197	
N	14.834	14.460			14.834	14.460		
Log Likelihood	-4.898,10	-4.833,02			-4.369,40	-4.286,60		
AIC	9.844,20	9.741,05			8.788,81	8.623,19		
rho					.77***	.79***		

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample.

Source: Own calculation based on PASS waves 2-7.

Table C.3: Take-up regression: marginal effects before and after correction of underreporting (weighted results)

Dependent variable: Take-up of UB II	(1) Pooled probit, uncor.	(2) Pooled probit, corrected for underreporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, uncor.	(6) RE probit, corrected for underreporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0457***	.0389***	-.00683***	-14,88%	.0391***	.0319***	-.00724***	-18,41%
	(.00251)	(.00260)	(.00135)		(.00257)	(.00294)	(.00156)	
Female hh	-.0561***	-.0489**	.00722	-12,83%	-.0547***	-.0468***	.00785	-14,44%
	(.0189)	(.0190)	(.00879)		(.0193)	(.0172)	(.00935)	
Hh is immigrant	.0267	.0524**	.0257***	96,25%	.0357*	.0585***	.0228***	63,87%
	(.0221)	(.0225)	(.00900)		(.0214)	(.0196)	(.00865)	
Age of hh: 25-34 years	.0998***	.0821***	-.0177	-17,74%	.0886***	.0675***	-.0211	-23,81%
(ref.:15-24 years)	(.0324)	(.0315)	(.0135)		(.0276)	(.0247)	(.0130)	
Age of hh: 35-44 years	.104***	.101***	-.00274	-2,88%	.106***	.0986***	-.00751	-6,98%
(ref.:15-24 years)	(.0314)	(.0300)	(.0172)		(.0286)	(.0264)	(.0167)	
Age of hh: 45-54 years	.114***	.101***	-.0136	-11,40%	.0990***	.0791***	-.0199	-20,10%
(ref.:15-24 years)	(.0318)	(.0313)	(.0157)		(.0294)	(.0275)	(.0152)	
Age of hh: >=55 years	.173***	.146***	-.0267*	-15,61%	.149***	.115***	-.0340**	-22,82%
(ref.:15-24 years)	(.0324)	(.0328)	(.0147)		(.0305)	(.0293)	(.0148)	
Hh is disabled	-.0314	-.0397*	-.00828	26,43%	-.00639	-.0146	-.00825	128,48%
	(.0224)	(.0232)	(.0106)		(.0206)	(.0183)	(.0118)	
Hh holds lower sec. degree	.0340	.0332	-.000806	-2,35%	.0133	.00105	-.0122	-92,11%
(ref. no sec. degree)	(.0347)	(.0353)	(.0121)		(.0334)	(.0299)	(.0145)	
Hh holds interm. sec. degree	-.0178	-.0135	.00431	-24,16%	-.0441	-.0531*	-.00901	20,41%
(ref. no sec. degree)	(.0335)	(.0346)	(.0134)		(.0320)	(.0287)	(.0155)	
Hh holds upper sec. degree	-.0586*	-.0435	0,0151	-25,77%	-.0900***	-.0845***	.00548	-6,11%
(ref. no sec. degree)	(.0345)	(.0350)	(.0164)		(.0334)	(.0291)	(.0182)	
Eastern Germany	.0625***	.0613***	-.00122	-1,92%	.0674***	.0670***	-.000372	-0,59%
	(.0213)	(.0210)	(.00825)		(.0227)	(.0187)	(.0108)	
Household owns home	-.0142	-.0330	-.0188	132,39%	-.0399	-.0604**	-.0205	51,38%
	(.0307)	(.0313)	(.0149)		(.0301)	(.0275)	(.0163)	
Young children in household	.0960***	.105***	.00871	9,37%	.0992***	.110***	.0109	10,89%
(age<=4 years)	(.0249)	(.0274)	(.0174)		(.0250)	(.0247)	(.0162)	
Family without children	-.154***	-.149***	.00404	-3,25%	-.142***	-.135***	.00618	-4,93%
(ref. single person)	(.0402)	(.0404)	(.0106)		(.0370)	(.0324)	(.0132)	
Single parents	.0231	.0183	-.00487	-20,78%	.0134	.00713	-.00624	-46,79%
(ref. single person)	(.0242)	(.0248)	(.0108)		(.0296)	(.0288)	(.0123)	
Family with children	-.289***	-.252***	.0371*	-12,80%	-.264***	-.230***	.0342*	-12,88%
(ref. single person)	(.0364)	(.0412)	(.0205)		(.0372)	(.0387)	(.0192)	
Subsample two	-.419***	-.436***	-.0175*	4,06%	-.447***	-.468***	-.0212**	4,70%
	(.0290)	(.0306)	(.00954)		(.0277)	(.0277)	(.0107)	
N	14.834	14.834			14.834	14.834		
Number of strata	69	69			69	69		
Number of PSUs	397	397			397	397		

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample. Weighted estimation accounts for the complex survey design of the PASS by using Stata’s “svy”-commands.

Source: Own calculation based on PASS waves 2-7.

Table C.4: Take-up regression: marginal effects of linked sample vs. unlinked sample

Dependent variable: Take-up of UB II	(1) Pooled probit, linked	(2) Pooled probit, not linked	(3) RE probit, linked	(4) RE probit, not linked
Simulated Entitlement/100 EUR	.0449***	.0448***	.0417***	.0413***
	(.00115)	(.00110)	(.00120)	(.00115)
Female hh	-.0130	-.0152*	-.0137	-.0169**
	(.00846)	(.00806)	(.00869)	(.00827)
Hh is immigrant	-.00561	.000873	0,000988	.00897
	(.00934)	(.00882)	(.00932)	(.00876)
Age of hh: 25-34 years	.0793***	.0726***	.0692***	.0624***
(ref.:15-24 years)	(.0178)	(.0170)	(.0179)	(.0169)
Age of hh: 35-44 years	.0906***	.0884***	.0786***	.0762***
(ref.:15-24 years)	(.0183)	(.0174)	(.0183)	(.0172)
Age of hh: 45-54 years	.118***	.113***	.101***	.0965***
(ref.:15-24 years)	(.0181)	(.0172)	(.0182)	(.0171)
Age of hh: >=55 years	.172***	.170***	.157***	.154***
(ref.:15-24 years)	(.0183)	(.0174)	(.0184)	(.0174)
Hh is disabled	.00479	.00698	.0121	.0138
	(.0112)	(.0106)	(.0108)	(.0102)
Hh holds lower sec. degree	-.00809	-.0101	-.0143	-.0145
(ref. no sec. degree)	(.0139)	(.0131)	(.0140)	(.0132)
Hh holds interm. sec. degree	-.0458***	-.0475***	-.0542***	-.0544***
(ref. no sec. degree)	(.0143)	(.0134)	(.0144)	(.0136)
Hh holds upper sec. degree	-.0828***	-.0886***	-.0948***	-.1000***
(ref. no sec. degree)	(.0159)	(.0149)	(.0160)	(.0150)
Eastern Germany	.0531***	.0577***	.0540***	.0594***
	(.00796)	(.00763)	(.00816)	(.00781)
Household owns home	.00852	-.00963	-.00476	-.0231
	(.0145)	(.0142)	(.0147)	(.0142)
Young children in household	.0517***	.0539***	.0524***	.0561***
(age<=4 years)	(.0116)	(.0112)	(.0120)	(.0115)
Family without children	-.0845***	-.0885***	-.0833***	-.0875***
(ref. single person)	(.0158)	(.0151)	(.0155)	(.0149)
Single parents	-.0127	-.0135	-.00833	-.00911
(ref. single person)	(.0105)	(.0101)	(.0108)	(.0104)
Family with children	-.234***	-.227***	-.214***	-.206***
(ref. single person)	(.0172)	(.0164)	(.0171)	(.0162)
Subsample two	-.211***	-.233***	-.245***	-.267***
	(.0169)	(.0160)	(.0170)	(.0159)
N	14.834	16.357	14.834	16.357
Log Likelihood	-6.189,00	-6.881,59	-5.754,50	-6.387,01
AIC	12.426,00	13.811,17	11.559,00	12.824,03
rho			.62***	.64***

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample. Columns 1 and 3 correspond to columns 1 and 5 in Table 3 and are repeated for convenience. Source: Own calculation based on PASS waves 2-7.

Table C.5: Take-up regression: marginal effects when using only gold standard/deterministic links

Dependent variable: Take-up of UB II	(1) Pooled probit, uncor.	(2) Pooled probit, corrected for underreporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, uncor.	(6) RE probit, corrected for underreporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0448***	.0372***	-.00751***	-16,96%	.0417***	.0326***	-.00906***	-21,82%
	(.00118)	(.00112)	-0,001		(.00122)	(.00121)	-0,001	
Female hh	-.0142*	-.00932	0,005	-34,37%	-.0150*	-.0107	0,004	-28,67%
	(.00863)	(.00778)	-0,005		(.00886)	(.00747)	-0,005	
Hh is immigrant	-.00277	.0161*	.0189***	-681,23%	.00326	.0225***	.0193***	590,18%
	(.00951)	(.00833)	-0,006		(.00947)	(.00785)	-0,006	
Age of hh: 25-34 years	.0804***	.0486***	-.0318***	-39,55%	.0715***	.0374**	-.0341***	-47,69%
(ref.:15-24 years)	(.0182)	(.0165)	-0,012		(.0183)	(.0164)	-0,013	
Age of hh: 35-44 years	.0925***	.0597***	-.0329***	-35,46%	.0796***	.0470***	-.0326**	-40,95%
(ref.:15-24 years)	(.0188)	(.0170)	-0,013		(.0188)	(.0166)	-0,013	
Age of hh: 45-54 years	.122***	.0817***	-.0401***	-33,03%	.105***	.0645***	-.0409***	-38,57%
(ref.:15-24 years)	(.0185)	(.0166)	-0,012		(.0186)	(.0163)	-0,013	
Age of hh: >=55 years	.174***	.120***	-.0544***	-31,03%	.161***	.103***	-.0583***	-36,02%
(ref.:15-24 years)	(.0188)	(.0169)	-0,012		(.0188)	(.0166)	-0,013	
Hh is disabled	.00665	-.00490	-.0115**	-173,68%	.0115	.000944	-0,011	-91,79%
	(.0115)	(.0109)	-0,006		(.0110)	(.00927)	-0,007	
Hh holds lower sec. degree	-.0104	-.00413	0,006	-60,29%	-.0191	-.0171	0,002	-10,47%
(ref. no sec. degree)	(.0142)	(.0130)	-0,008		(.0142)	(.0118)	-0,009	
Hh holds interm. sec. degree	-.0480***	-.0362***	0,012	-24,58%	-.0576***	-.0520***	0,006	-9,72%
(ref. no sec. degree)	(.0146)	(.0134)	-0,008		(.0146)	(.0122)	-0,009	
Hh holds upper sec. degree	-.0861***	-.0655***	.0206**	-23,93%	-.0992***	-.0807***	.0185*	-18,65%
(ref. no sec. degree)	(.0163)	(.0149)	-0,009		(.0162)	(.0135)	-0,010	
Eastern Germany	.0507***	.0417***	-.00899*	-17,75%	.0524***	.0396***	-.0127**	-24,43%
	(.00815)	(.00724)	-0,005		(.00834)	(.00697)	-0,005	
Household owns home	.00702	-.00236	-0,009	-133,62%	-.00746	-.0177	-0,010	137,27%
	(.0148)	(.0125)	-0,007		(.0150)	(.0122)	-0,008	
Young children in household	.0514***	.0448***	-0,007	-12,84%	.0509***	.0519***	0,001	1,96%
(age<=4 years)	(.0120)	(.0103)	-0,008		(.0123)	(.0101)	-0,009	
Family without children	-.0789***	-.0637***	.0152*	-19,26%	-.0788***	-.0629***	.0159*	-20,18%
(ref. single person)	(.0159)	(.0144)	-0,009		(.0156)	(.0136)	-0,010	
Single parents	-.00971	-.00643	0,003	-33,78%	-.00594	-.00257	0,003	-56,73%
(ref. single person)	(.0107)	(.00949)	-0,006		(.0111)	(.00928)	-0,007	
Family with children	-.239***	-.211***	.0285***	-11,72%	-.219***	-.195***	.0241**	-10,96%
(ref. single person)	(.0179)	(.0175)	-0,010		(.0177)	(.0163)	-0,011	
Subsample two	-.222***	-.237***	-.0152**	6,76%	-.257***	-.277***	-.0208***	7,78%
	(.0175)	(.0170)	0,001		(.0175)	(.0161)	0,001	
N	14.002	14.002			14.002	14.002		
Log Likelihood	-5.786,92	-4.607,21			-5.386,57	-4.114,75		
AIC	11.621,83	9.262,43			10.823,13	8.279,50		
rho					.62***	.77***		

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample.

Source: Own calculation based on PASS waves 2-7.

Table C.6: Take-up regression: marginal effects after correction of underreporting and after correction of under- and overreporting (sample without corrected duplicates)

Dependent variable: Take-up of UB II	(1) Pooled probit, uncor.	(2) Pooled probit, corrected for underreporting	(3) Abs. diff. (2)-(1)	(4) Rel. diff. (2)/(1)	(5) RE probit, uncor.	(6) RE probit, corrected for underreporting	(7) Abs. diff. (6)-(5)	(8) Rel. diff. (6)/(5)
Simulated Entitlement/100 EUR	.0445***	.0358***	-.00871***	-19,55%	.0417***	.0312***	-.0105***	-25,18%
	(.00119)	(.00113)	-0,001		(.00123)	(.00121)	-0,001	
Female hh	-.00680	-.00542	0,001	-20,29%	-.00818	-.00772	0,000	-5,62%
	(.00853)	(.00761)	-0,005		(.00883)	(.00739)	-0,006	
Hh is immigrant	-.00923	.0104	.0196***	-212,68%	-.00367	.0174**	.0211***	-574,11%
	(.00944)	(.00820)	-0,006		(.00950)	(.00780)	-0,006	
Age of hh: 25-34 years	.0786***	.0488***	-.0298**	-37,91%	.0677***	.0361**	-.0316**	-46,68%
(ref.:15-24 years)	(.0180)	(.0162)	-0,012		(.0182)	(.0162)	-0,013	
Age of hh: 35-44 years	.0913***	.0593***	-.0321**	-35,05%	.0784***	.0486***	-.0298**	-38,01%
(ref.:15-24 years)	(.0185)	(.0167)	-0,013		(.0186)	(.0164)	-0,013	
Age of hh: 45-54 years	.118***	.0807***	-.0370***	-31,61%	.0999***	.0654***	-.0345***	-34,53%
(ref.:15-24 years)	(.0183)	(.0164)	-0,012		(.0185)	(.0162)	-0,013	
Age of hh: >=55 years	.173***	.120***	-.0533***	-30,64%	.157***	.103***	-.0543***	-34,39%
(ref.:15-24 years)	(.0185)	(.0165)	-0,013		(.0187)	(.0165)	-0,013	
Hh is disabled	.0112	.000538	-.0107*	-95,20%	.0145	.00270	-.0118*	-81,38%
	(.0111)	(.0103)	-0,006		(.0109)	(.00917)	-0,007	
Hh holds lower sec. degree	-.00581	-.00428	0,002	-26,33%	-.0129	-.0178	-0,005	37,98%
(ref. no sec. degree)	(.0141)	(.0128)	-0,008		(.0143)	(.0116)	-0,010	
Hh holds interm. sec. degree	-.0423***	-.0336**	0,009	-20,57%	-.0501***	-.0507***	-0,001	1,20%
(ref. no sec. degree)	(.0146)	(.0132)	-0,008		(.0147)	(.0120)	-0,010	
Hh holds upper sec. degree	-.0810***	-.0665***	0,015	-17,90%	-.0929***	-.0822***	0,011	-11,52%
(ref. no sec. degree)	(.0162)	(.0147)	-0,009		(.0163)	(.0134)	-0,011	
Eastern Germany	.0480***	.0374***	-.0105**	-22,08%	.0479***	.0343***	-.0136**	-28,39%
	(.00804)	(.00706)	-0,005		(.00828)	(.00685)	-0,005	
Household owns home	.0292**	.0169	-0,012	-42,12%	.0196	.00453	-.0150*	-76,89%
	(.0143)	(.0118)	-0,008		(.0145)	(.0114)	-0,009	
Young children in household	.0517***	.0453***	-0,006	-12,38%	.0525***	.0554***	0,003	5,52%
(age<=4 years)	(.0117)	(.00987)	-0,008		(.0121)	(.00945)	-0,009	
Family without children	-.0816***	-.0682***	0,013	-16,42%	-.0803***	-.0642***	0,016	-20,05%
(ref. single person)	(.0160)	(.0146)	-0,009		(.0158)	(.0137)	-0,010	
Single parents	-.0166	-.0101	0,007	-39,16%	-.0141	-.00753	0,007	-46,60%
(ref. single person)	(.0106)	(.00924)	-0,006		(.0110)	(.00912)	-0,007	
Family with children	-.234***	-.198***	.0363***	-15,38%	-.217***	-.188***	.0287**	-13,36%
(ref. single person)	(.0179)	(.0174)	-0,011		(.0177)	(.0161)	-0,011	
Subsample two	-.201***	-.213***	-0,012	5,97%	-.233***	-.251***	-.0184*	7,73%
	(.0196)	(.0190)	-0,008		(.0200)	(.0185)	-0,010	
N	14.139	14.139			14.139	14.193		
Log Likelihood	-5.813,61	-4.543,12			-5.427,46	-4.076,69		
AIC	11.675,58	9.315,60			11.093,80	8.392,29		
rho					.61***	.76***		

Notes: Asterisks */**/*** denote statistically significant effects using cluster-robust standard errors at the significance level of 0.1/0.05/0.01. Hh stands for head of household. “Subsample two” indicates whether an observation belongs to the second, nationally representative subsample. Sample of simulated eligible households with consent to data linkage, uniquely identified in the administrative data and no missing values in covariates.

Source: Own calculation based on PASS waves 2-7.

Imprint

IAB-Discussion Paper 6|2019

Publication date

2 April 2019

Editorial address

Institute for Employment Research (IAB)
of the Federal Employment Agency (BA)
Regensburger Straße 104
90478 Nuremberg
Germany

Download of this Discussion Paper

http://doku.iab.de/discussionpapers/2019/dp0619.pdf

All publications in the series "IAB-Discussion Paper" can be downloaded from

https://www.iab.de/en/publikationen/discussionpaper.aspx

Website

www.iab.de

ISSN

2195-2663

DOI

12.3456/123456

For further inquiries contact the authors

Kerstin Bruckmeier
Phone 0911 179-4432
E-Mail Kerstin.Bruckmeier@iab.de

Jürgen Wiemers
Phone 0911 179-8671
E-Mail Juergen.Wiemers@iab.de