SHARE WORKING PAPER SERIES

Measurement error in the Survey of Health, Ageing and Retirement in Europe: A validation study with administrative data for education level, income and employment

Paul Bingley and Alessandro Martinello

Working Paper Series 16-2014

Measurement error in the Survey of Health, Ageing and Retirement in Europe: A Validation Study with Administrative Data for Education Level, Income and Employment

Paul Bingley[1] and Alessandro Martinello[1,3]

[1]*SFI - The Danish National Centre for Social Research, Copenhagen, Denmark*
[3]*University of Copenhagen, Copenhagen, Denmark*

January 17, 2014

Summary

We link the Survey of Health, Ageing and Retirement in Europe (SHARE) to Danish Administrative Registers, comparing schooling, retirement status and income. We are able to retrieve administrative records for 1670 out of the original 1707 respondents from the first survey wave in 2004. We compare individual linked records in an analysis of measurement error. Overall, we find only minor non-random misclassification of schooling, but otherwise SHARE provides reliable data for socio-economic analysis of schooling, income and retirement. SHARE Denmark overestimates the proportion of individuals with higher education: the probability of misclassification is higher for lower educated, richer individuals. Labour market status is precisely reported, and misclassification probability decreases with age. Average gross household income is not statistically different in SHARE and register data, and we show that measurement error is classical.

Keywords: Validation study; Measurement error; Misclassification; Survey; Administrative records

1 Introduction

Surveys are an important data source for socioeconomic analysis. However, their reliability as empirical evidence depends on both the precision of respondent's assessments and the distribution of errors originating in the interviewing process. The magnitude and the characteristics of measurement errors are crucial for the consistency of empirical analysis conducted on the basis of survey data. Knowing the nature of measurement error helps to significantly improve the robustness of empirical models (Fuller, 1987).

In general researchers can accommodate for measurement errors in the variables of statistical models and adjust estimates accordingly. However, in practice social scientists often impose restrictions on their distribution, such that the model can easily handle the resulting complications. Typically, models assume that measurement errors are additive, are normally distributed with zero mean and constant variance, and are independent of the observed variable of interest, the other variables in the model, and their associated measurement error. If these properties hold, the literature refers to this as classical measurement error. A number of validation studies of other surveys (see Bound et al. (2001) for an overview) find that measurement errors are often non-classical, and are potentially problematic for inference.

The Survey of Health, Ageing and Retirement in Europe (SHARE) is the most comprehensive international panel study of senior individual health and social status. It is ongoing and forms the basis of an ever-growing number of empirical studies on the European 50+ population. It is managed by a dedicated international research team that pays particular attention to accuracy and consistency of data. However, as with any other survey, SHARE still largely relies on the precision of personal assessments by individual respondents. This paper performs a validation study on the Danish SHARE subsample, and characterizes measurement error with a particular focus on first-party (or proxy) response error (Biemer et al., 2004).

Validation studies can be classified as either external or internal in nature. External validation studies typically compare two data sources measured with error. Examples are cross-reports from an employer (Barron et al., 1997) or personnel records from firm payrolls (Duncan and Hill, 1985; Bound et al., 1994) compared with ad-hoc surveys. Applying results from external validation studies such as these requires assuming that the measurement error model, i.e. the distribution of measurement errors, is the same in both validation data and the survey of interest.

Internal validation studies typically link a subsample of a survey to an official validation source which is assumed to be measured without error. In this case, the underlying assumption is that the validated subsample has to be as good as random and no selection must take place. While some internal validation studies such as those for the U.S. Current Population Survey (CPS) or the Survey of Income and Program Participation (SIPP) validate a large proportion of the sample and check carefully for selection mechanisms, they use income from the Social Security Administration as a validation source and this is censored at the maximum taxable level (Bound and Krueger, 1991; Pedace and Bates, 2000). Others rarely link more than 50% of the observations (Bricker and Engelhardt, 2008), and whether the validated sample is as good as random is hard to assess.

In this paper, using an individual linkage with Danish administrative registers, we perform an internal validation study of SHARE Denmark data for education, labour market status and gross household income. These measures are respectively ordered, multinomial and continuous, and present a range of measurement error challenges. Our validation study has two strengths over those conducted for other surveys:

(a) We use public administrative data as a validation source. Danish administrative data are an uncensored and precise data source (see Jensen and Rasmussen, 2011, for schooling, Browning and Leth-Petersen, 2003, for income and Leth-Sørensen, 1993, for labour market status).

(b) We link 97% of the SHARE Denmark sample. Thus, there is virtually no selection taking place on the validated subsample with respect to the originally surveyed sample.

As in most validation studies, throughout the paper we maintain the crucial assumption that our validation data is measured without error. This assumption might not hold in general, which can have consequences for our results. We relax and test this assumption for income and schooling in a separate paper (Bingley and Martinello, 2014).

The remainder of the paper is organized as follows. Section 2 provides an overview of the most relevant literature on validation studies of socio-economic surveys, focusing on earnings and other income-related variables. Section 3 describes the linkage procedure, the data used in our analysis and a comparison of distributions of variables from survey, population register and overlap samples.

Section 4 provides the results from our validation studies at the individual level for each of the three variables considered, and their consequences for econometric analysis. Section 5 concludes.

2 Other validation studies of social surveys

The need for reliable validation studies is well grounded (see Keating et al. (1950)). However, early reports focus primarily on establishing the presence of measurement error and on the estimation of significant average discrepancies between respondent reports and validation sources, especially concerning labour market status, earnings, wages, and other income-related variables Miller and Paley (1958); Mellow and Sider (1983); Greenberg and Halsey (1983). In a measurement error setting, those papers primarily test the zero-mean hypothesis (mostly rejecting it, albeit marginally), and pay little attention to the consequences for estimation and inference on the basis of incorrectly reported data.

Duncan and Hill (1985) fill this gap, using an ad-hoc survey validated with administrative records in order to study the distribution of measurement errors in a number of labour related variables. The authors show that reporting errors in earnings follow a non-normal distribution with a high concentration around zero. Errors are correlated with job tenure, thus causing a bias in the estimation of the coefficients of a typical wage-experience model. Bound and Krueger (1991) and Bound et al. (1994) study measurement errors in the CPS and in the Panel Study of Income Dynamics Validation Study (PSID-VS) respectively, focusing specifically on estimation biases, and present the econometric background necessary to study the consequences of a non-classical distribution of measurement errors.

While Bound et al. (1994) use an external validation survey for Panel Study of Income Dynamics (PSID) where respondents come from a single large firm that shared its payroll records, and therefore lacks in generality, Bound and Krueger (1991) perform an internal validation study on CPS data with Social Security Administration data as validation source. However, the validation data are censored at the upper tax bracket threshold, resulting in incomplete data for almost 50% of the male sample. Moreover, because respondents had to report their social security number to the interviewer, only one third of the original CPS sample was successfully matched, potentially introducing selection if, for example, people willing to reveal their social security number are also better and more precise reporters. Nonetheless, those studies provide important insight into the distributions of measurement errors.

In particular, they show that measurement errors in earnings are negatively related to the true value of the variable and that, as Bollinger (1998) confirms in PSID-VS using a nonparametric technique, low-income respondents typically over-report their earnings. For earnings, the ratio of measurement error variance to total variance, or unreliability ratio, is 0,276 for men and 0,089 for women in the CPS, and respectively 0,302 and 0,133 in the PSID-VS. Both papers conclude that measurement errors are non-classical, and that the negative correlation observed between the true value of earnings and its measurement error causes a negative bias when earnings appear as the dependent variable in an OLS regression.

While Pedace and Bates (2000) perform an almost complete internal validation of the SIPP, matching 84% of the observations to Social Security Administration data, they focus more on identification of misreporting individuals and less on the distribution of the measurement errors.

Bricker and Engelhardt (2008) match the employer's report to the U.S. government on earnings in each job (W-2 administrative records) to the Health and Retirement Study (HRS), a U.S. longitudinal survey that served as a role model for the development of SHARE. The population of reference is therefore highly comparable with our own, and the results of that study will serve as a reference point for our analysis on gross household income. Mean measurement errors in male earnings are significantly positive in the 1991 wave of the Original Cohort of HRS, but are small (0.059 log points, or about $1500). The variance ratio is 0.322, and the distribution of measurement errors is non-normal and highly concentrated around zero. The authors confirm the general finding of a negative correlation between measurement error and the true level of earnings. However, they do not find any correlation between measurement error in earnings and other variables typically used in econometric analysis.

Measurement error in income-related variables has been most extensively studied, but there are also examples of measurement error studies of labour market status and schooling. Poterba and Summers (1986) underline the biasing effects of measurement errors in the reports of labour market status for the estimation of rates of labour market transitions and of labour market behavior in general. In particular, they show that when CPS respondents answer the same question about their labour market status a week after the official interview, 13% of those who reported being unemployed change their answer. Kane et al. (1999) focus on errors in self-reported education level, and develop a correction method to estimate returns to schooling. Notably, their method relies on the assumption that measurement errors in education are independent of earnings; we verify this assumption in our data.

While literature on measurement of education and labour market status simply underlines the existence of a measurement error problem, income and earnings validation studies depict non-classical measurement error and negative correlation between the error and the true underlying variable as stylized facts. In the remainder of our SHARE Denmark internal validation study we describe the extent and nature of measurement error for schooling, employment status and gross income.

3 Linkage and descriptive statistics

SHARE is a longitudinal survey that collects data across 19 European countries overall. By wave four, SHARE reports information from 150,000 interviews of 86,000 persons across all waves, and is one of the most extensive surveys on the elderly population worldwide (Börsch-Supan et al., 2013). We focus on the first wave of SHARE Denmark, which interviewed in 2004 a representative sample of residents of Denmark aged fifty or more (main respondents) and their spouses for a total of 1707 individual respondents. Our validation source is public administrative register data, which provide official information on demographics, socio-economic status and tax reports for the years 2003 and 2004.

To initiate the first wave of SHARE Denmark, a random sample of individuals aged 50 and above was drawn from the Central Person Register. This database contains vital statistics and current address for the population of residents of Denmark and each individual is indexed by a unique social security number (CPR). Sampled individuals in this register who agreed to be interviewed entered in the SHARE Denmark sample. As a consequence, CentERdata – SHARE's data-managing institution – is able to link each selected respondent with the associated Danish social security number. Data confidentiality requirements are such that only the data collection and survey agencies and Statistics Denmark observe social security numbers. We have access to unique individual identifiers in order to conduct our analysis but do not know the actual social security number ourselves.

SHARE surveys the sampled individuals and their spouse. However, the data collection agency only knows the CPR of sampled individuals, but not their spouse. We link spouses according to a cohabitor identification number (CNR) created by Statistics Denmark. This is generated for adults sharing the same address at time of interview who are married to each other or are in a registered partnership together. Non-registered cohabiting couples share a single CNR number only if they are of opposite gender, their age differential is less than fifteen years, and no other adult lives at the same address. Using the CNR we can obtain CPR of non-sampled spouses who were also interviewed. We retrieve administrative records for 1670 of the 1707 individual respondents, corresponding to 97% of the first wave of SHARE Denmark. Of the 37 observations we cannot match, 21 are spouses that have been interviewed in SHARE, but do not appear in the registers. In the remaining 16 observations (14 households) we cannot identify the main respondent.

SHARE collects a wide array of information, from health to employment status. In this paper we concentrate on education level, labour market status and income, as they are among the most commonly used variables in socio-economic analyses. In particular, we study gross household income in order to minimize reporting errors among different sources of income (e.g. financial and labor income).

Not all data collected in SHARE is first party reported. In most modules, if a respondent cannot answer, information is gathered though a proxy interview. In our analysis we do not distinguish among respondent and proxy interviews, as our aim is to assess general measurement error in SHARE Denmark. While proxy interviews could play a role in the magnitude of measurement errors, such a hypothesis is challenging to test convincingly, given that for the demographics and employment and pensions modules, first party response rate is above 90% for most countries. In the complete first wave data, first party response rate ranges from 90% in the Netherlands to 97.7% in Switzerland in the demographics module, and from 84.4% in Belgium to 96.6% in Austria. Excluding Israel, interviewed in 2005 and 2006, Denmark is the median country by aggregated first party response rate in both modules (96% in the demographic module, 93.4% in the employment and pensions module).

Income data may have yet another source of measurement error. Whenever the respondent cannot provide a precise assessment of income in the previous year, an unfolding sequence of bracket questions starts. Given this information, SHARE provides multiple imputations for the respondent's income (for details on the imputation procedure see Christelis, 2011). In the following analysis we aggregate multiple imputations by respondent, and we use their average as if it was a non-imputed response. Such a procedure gives unbiased estimates of mean outcomes, but underestimates the standard errors of the estimators using the imputed data. This issue is irrelevant for the remainder of this section, where we describe the data by a three-way comparison between aggregates from SHARE Denmark data, Register data using a linked sample, and aggregates from Register data using the whole Danish population aged fifty or more. However, the bias in standard errors is relevant in Section 4, where we assess measurement errors at the micro level and we test whether they can be considered classical. We continue the discussion on imputed data in Section 4.

The SHARE questionnaire asks for the highest level of education attained. As we show in the appendix, given the possible answers available to the respondent in the SHARE questionnaire, we use 1997 ISCED coding to collapse education into three categories indicating whether an individual has a low (0-2 ISCED scale), medium (3-4) or high (5-6) level of education.

Our source of validation for education are official registers used by the Danish administration. The central registration of education in Denmark began with the general population and housing census which was undertaken on November 9th, 1970. It was mandatory for all residents of Denmark to respond and use CPRs for identification. The census asked 13 questions about housing and 13 about persons, of which 3 were about schooling. These were under the heading "Education and vocational training status". The first question was about education or vocational training in progress. The second was about completed schooling. The third was about completed education or vocational training (Statistics Denmark, 1977). Full text of the questions is presented in an appendix. Five pages of instructions were followed for the coding of the education responses, with the objective of placing the written responses to each of the three education questions into a 3-digit coding frame.

After the census, information on education level has been recorded by a third party, typically the institution providing the educational qualification. Therefore, the information on education for individuals who obtained a qualification when the registers were in place was updated accordingly.

Most of the SHARE Denmark sample has a census-based administrative record of schooling. The proportion with a census-based record increases with age, respectively 22, 78 and 97% of 50, 60 and 70 years old in 2004.

Figure 1 shows aggregates for the three comparison samples, by gender and age cohort. All three samples exhibit both the same pattern of decreasing education level with age and a sharp educational differential between genders, especially at older ages. SHARE Denmark overestimates the proportion of individuals with a high-level education, and underestimates the proportion of individuals with a low level education. Together with the discrete nature of the education variable, this unbalance implies that the direction of the measurement error depends on the value observed in the register data. Namely, we observe more often individuals reporting a higher level of education than that reported in the registers than the other way around.

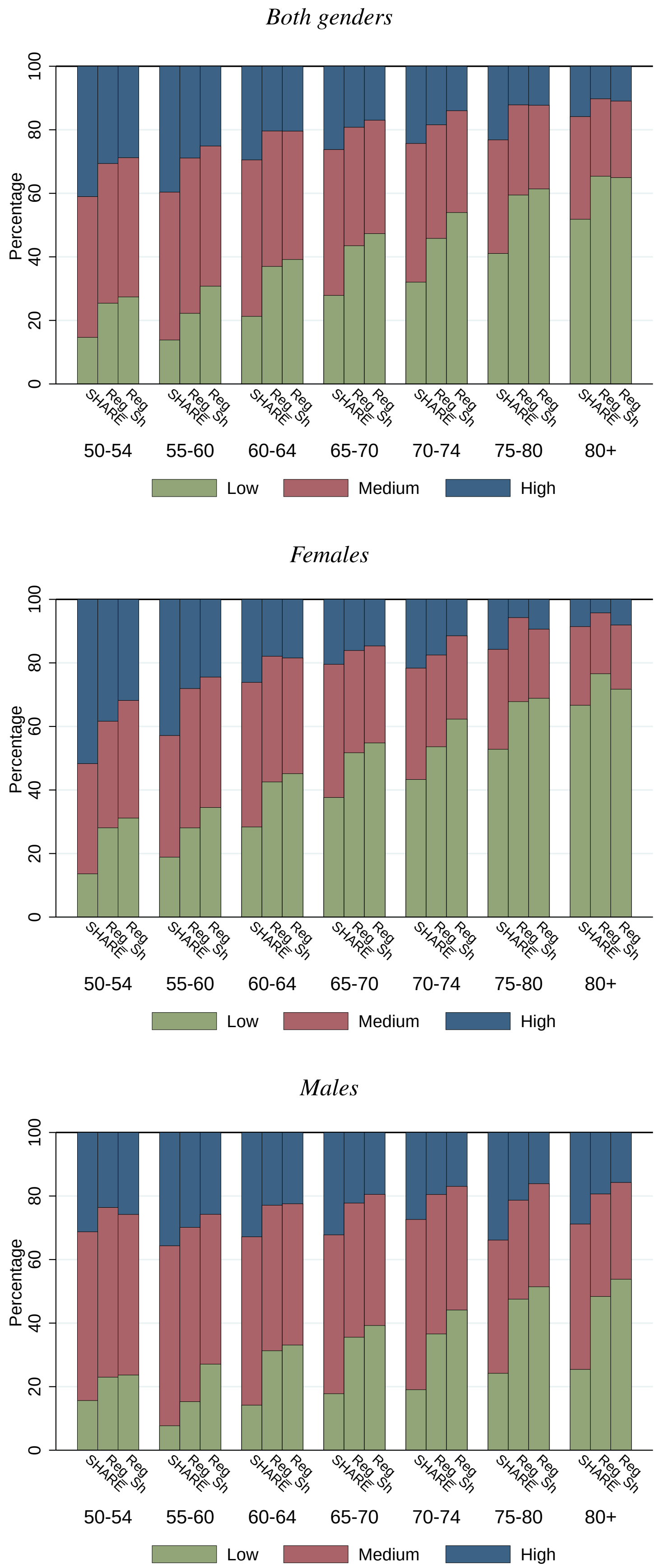
Because after the 1970 census any additional qualification received from a Danish institution had to be reported to the Ministry of Education by a third party, such discrepancy cannot be due to individuals acquiring further degrees over time, unless obtained abroad. However, only three individuals in our sample declare, while answering to the SHARE questionnaire, to have acquired any degree abroad. Therefore, while the linked SHARE Denmark sample does not differ much with respect to the whole Danish population, we expect a negative correlation between measurement error and true level of education in the SHARE Denmark sample.

We repeat a similar analysis in Figure 2, where we group labour market status into three macro-categories that, given the age of our sample, cover the main alternatives. Retirees and workers, which comprise both employed and self-employed respondents, constitute the great majority of our sample for both genders and for each age category. Therefore, we group all other labour market statuses recorded in SHARE data together, notably unemployed and homemakers.

As in Figure 1, the upper pane shows aggregates for genders, the middle pane for females only and the lower pane for males only. Again, the age pattern in all three samples is the same: the great majority of respondents aged between fifty and sixty are in the labour market, while soon after sixty-five almost everyone is retired. Most transitions from work to retirement take place as expected between sixty and sixty-five. The three samples are similar, and while SHARE Denmark slightly overestimates the "other" category, the figure does not indicate substantial bias.

As our linkage procedure relies on registered household composition, the setting of our analysis changes when we consider gross household income, shifting the focus from the individual to the household: according to the administrative registers, 117 of the respondents that SHARE records as singles have a co-habiting partner that is not indicated in the household grid in SHARE data. As the information we use for defining couples is constructed on the first of January, some of these missing observations might be caused by deaths before the time of the interview. Others can be due to imprecision in the CNR indicator. For example, if two individuals of different gender have been cohabiting for more than one year, but are not part of the same household, we would consider them anyway as a couple. While this does not create any problem for individual measures, it can bias the analysis conducted for household aggregates. Therefore, we define the sample as "enlarged" when we aggregate the unreported missing spouses and impute for them the SHARE gross household income that the respondent declared; as "small" (and consistent), where we drop the 117 respondents for whom SHARE does not report a spouse. We present some general results for both the enlarged and the restricted, consistent sample, but use the restricted sample for the main analysis on gross household income.

SHARE provides gross household income as the sum of different income sources (pension benefits, employment, capital etc., see the appendix for details) for the two interviewed members of the household. Accordingly, we aggregate the reported value of income in the tax registers for each



Education levels: low = {0,1,2}, medium = {3,4}, high = {5,6}, ISCED scale

Fig. 1: Education level by age group

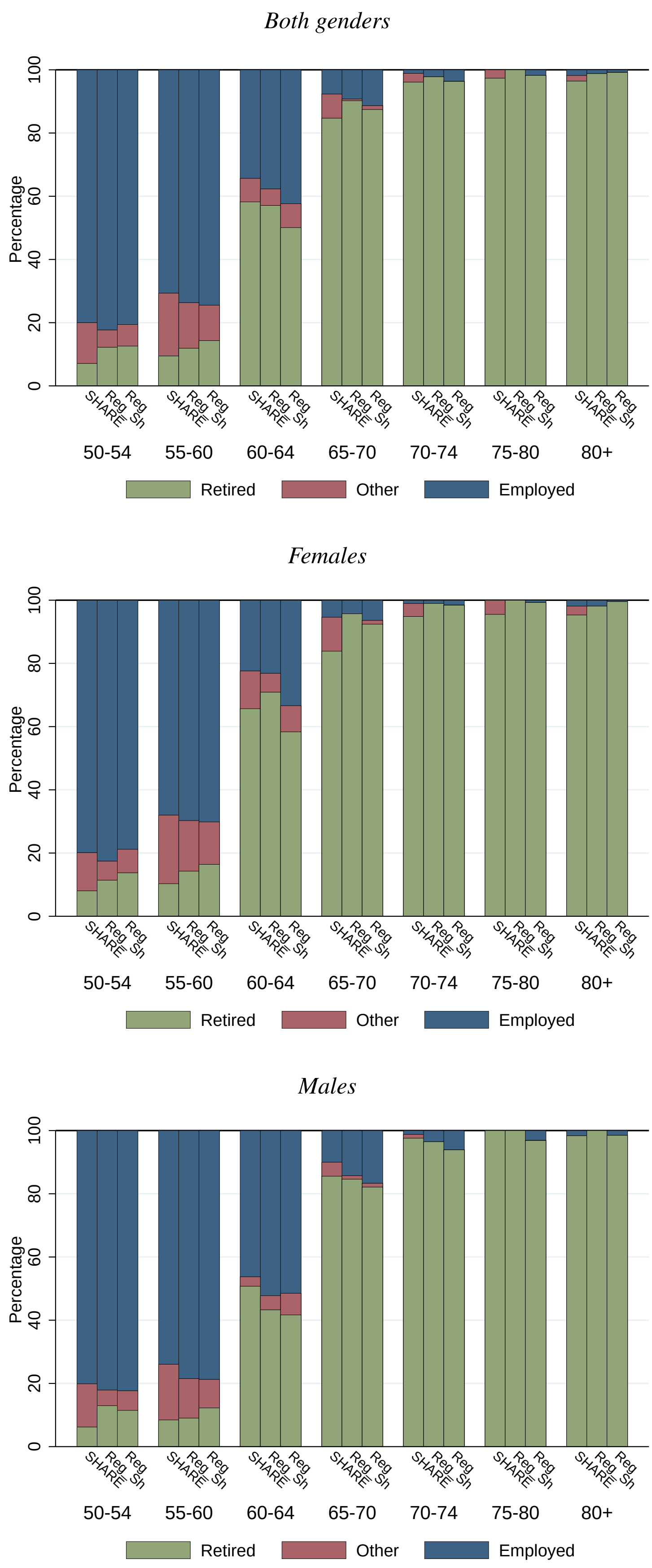


Fig. 2: Labour market composition by age group

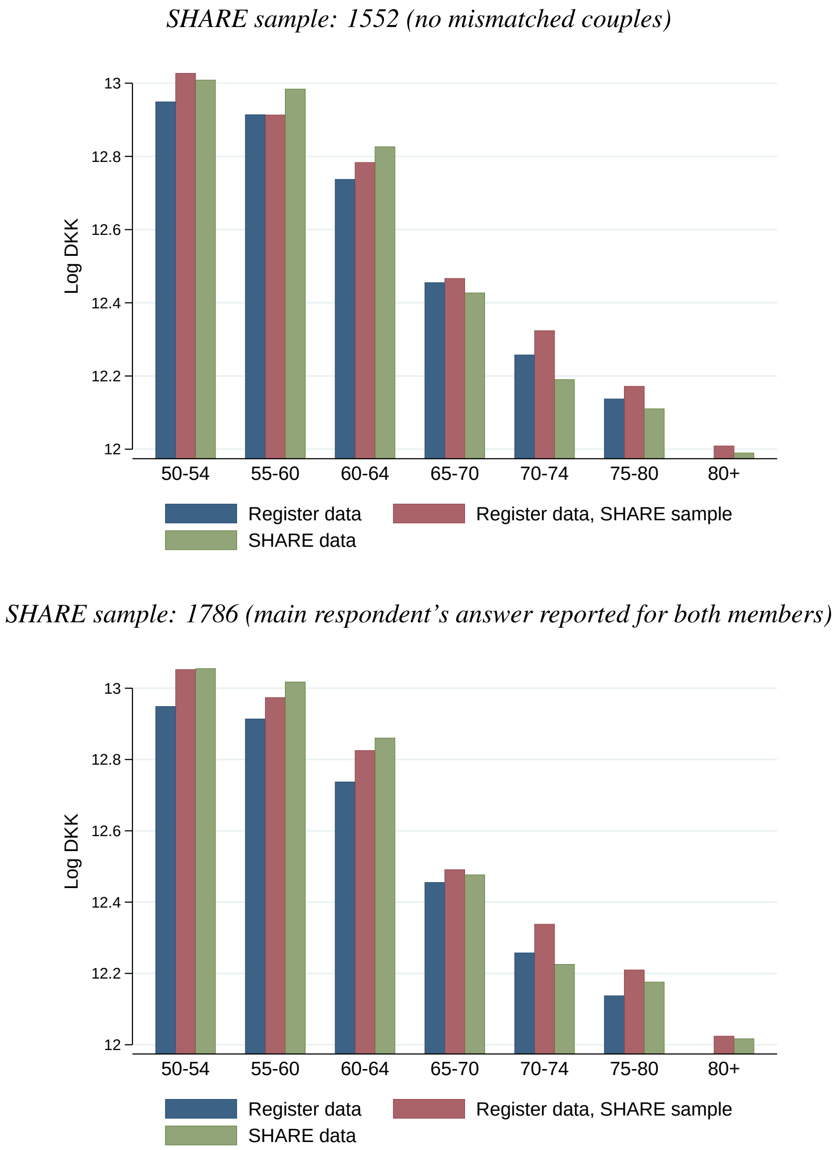


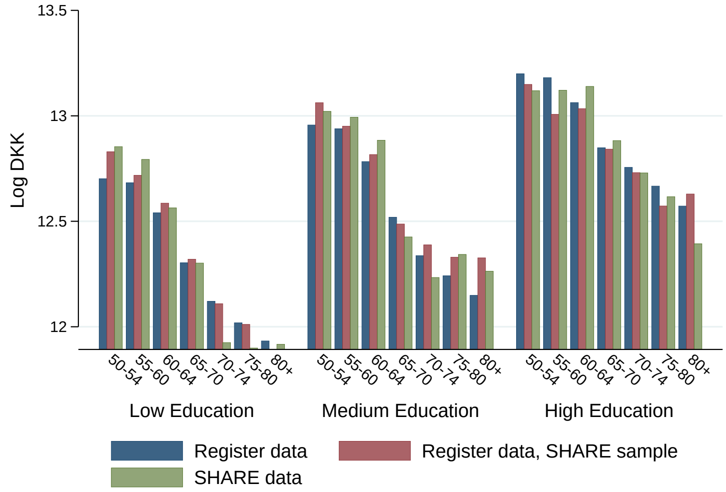
Fig. 3: Gross household income, averages by age group

member of the household. Figure 3 shows average logarithms of gross household income in SHARE Denmark and register data by age group. We use logarithms, following Bound and Krueger (1991), because reporting errors in income are likely to be heteroskedastic in levels. The upper pane shows averages for the small sample, for which SHARE interviews the co-habiting partner of the main respondent observed in the registers. The lower pane shows averages for the enlarged sample, where we include both partners as we observe them in the registers, and we impute the main respondent's answer about gross household income to both observations. By doing this operation, we implicitly assume the main respondent considered the income of the non-interviewed partner when reporting household income values. The figure does not show evident patterns in averages across ages. Previous studies showed, if any, a negative correlation between measurement error and income.

Figure 4 averages log-income by both age and education categories. As in Figure 3, we report averages for both small and enlarged samples.

The patterns that Figure 4 exhibits for average income on education level and age cohort correspond closely in SHARE Denmark and register data. Educational differentials are consistent in both

SHARE sample: 1552 (no mismatched couples)



SHARE sample: 1786 (main respondent's answer reported for both members)

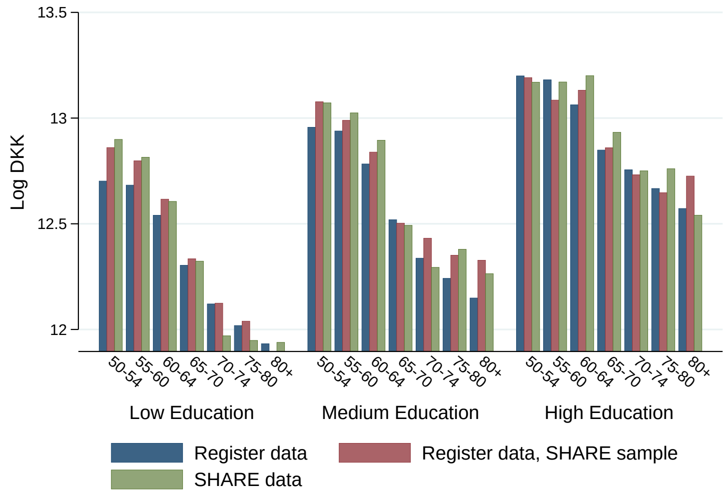
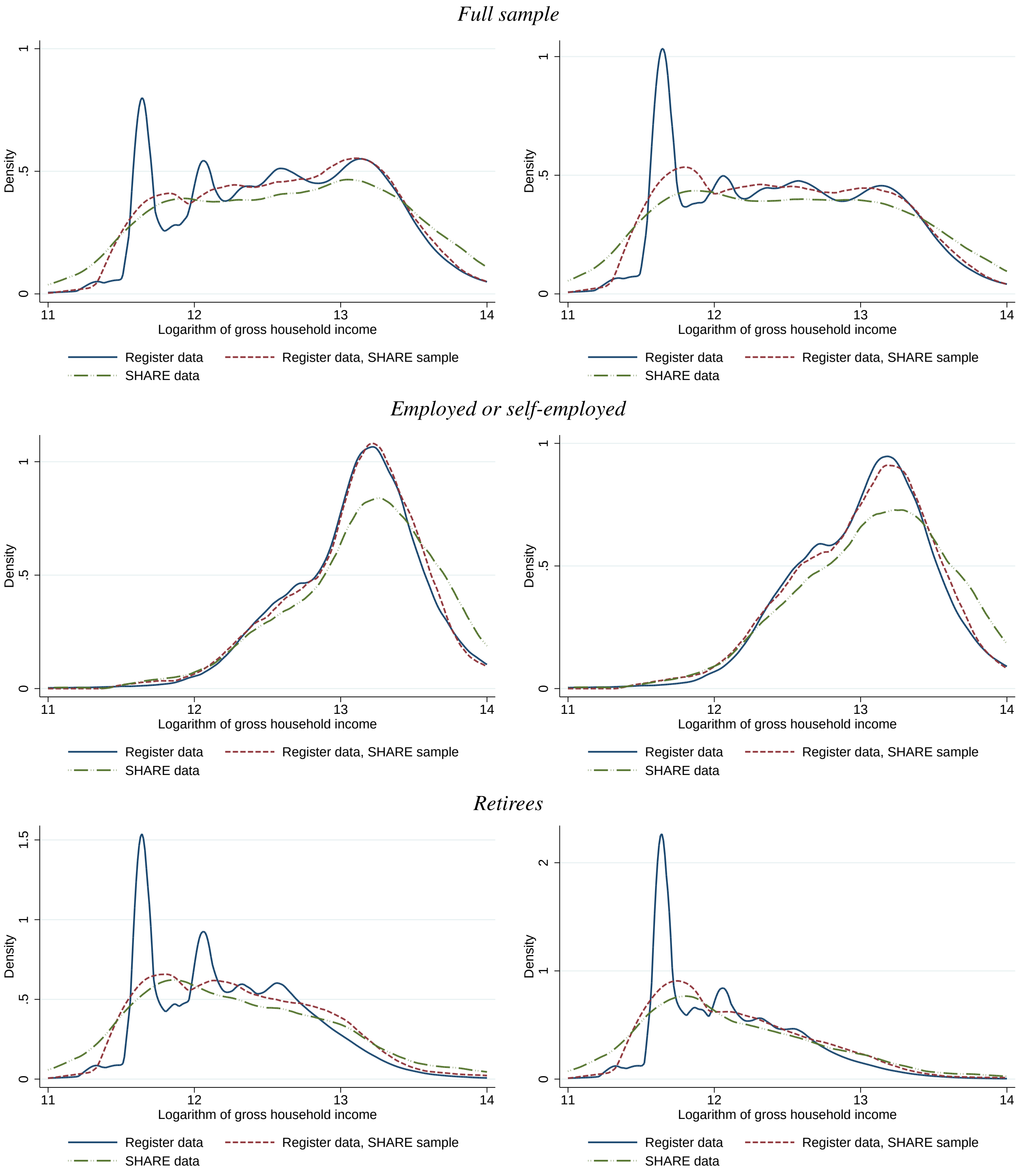


Fig. 4: Gross household income, averages by age group and education level

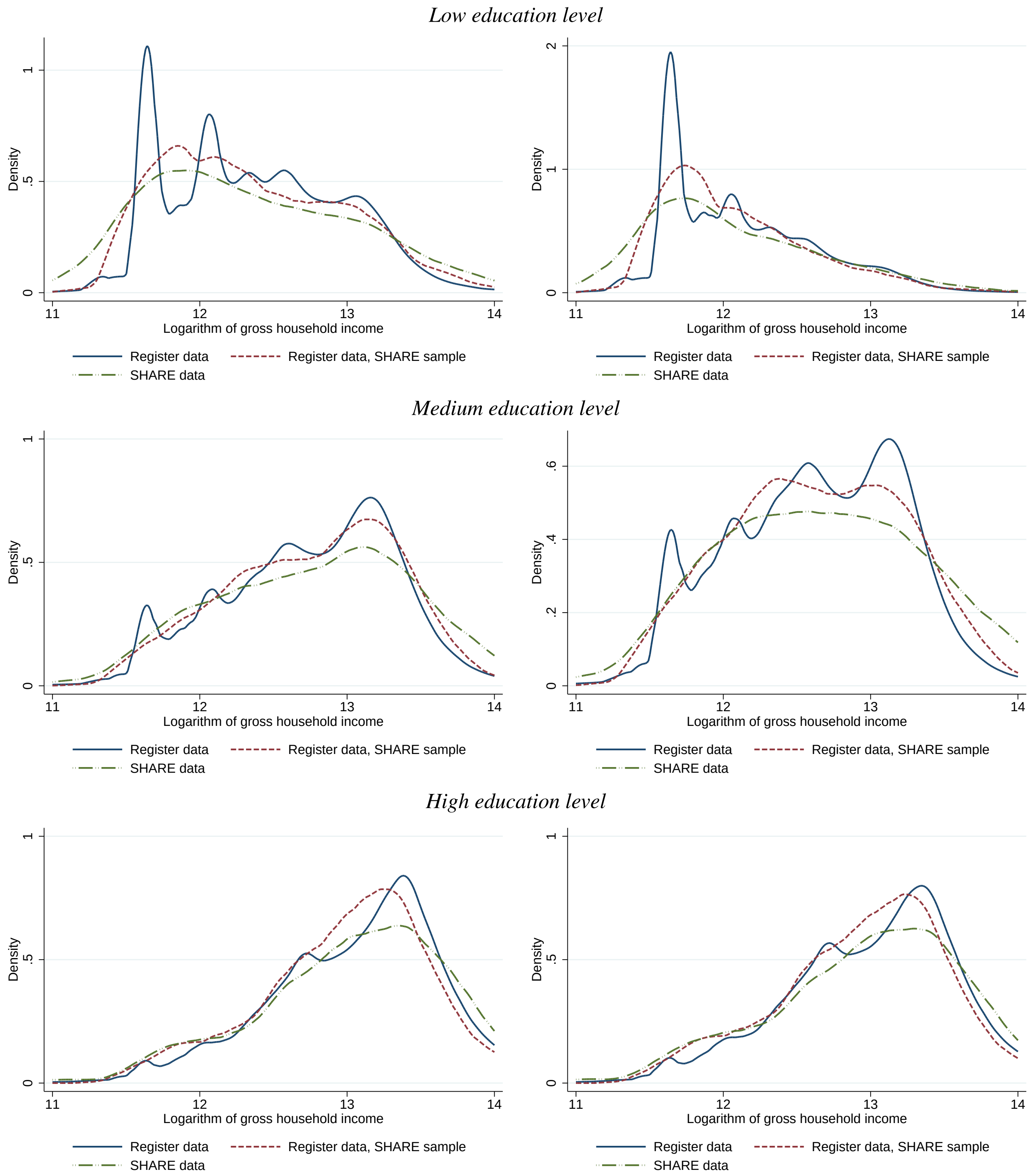


(a) Individuals

(b) Households†

† The labour market status considered for households is that of the household member contributing the most to household income.

Fig. 5: Gross household income, distribution by labour market status



(a) Individuals

(b) Households†

† The education level considered for households is the highest observed in the household.

Fig. 6: Gross household income, distribution by education level

data sources, and persist—or intensify—at older ages. Figure 4 shows no evident correlation between logs of household income and its associated measurement error.

Figures 5 and 6 plot kernel density estimates for the restricted and consistent samples. In both figures the left column treats data at the individual level, while the right column groups data at the household level. Spikes are apparent in the register population sample only because of the much higher number of observations, bandwidth being equal. These spikes correspond to levels of yearly benefits from old age pension (folkepension and ældrecheck), for singles and married couples (respectively 117,024 DKK, or 15,729 €, and 174,168 DKK, or 23,409 €). Otherwise, there are no substantial differences between the two columns.

Figure 5 classifies data by labour market status; Figure 6, by education level. For household grouped graphs, we use the labour market status of the member contributing the most to household income, and the highest education level observed among partners. Both figures show similar patterns in correlated categories. For retirees and individuals with a low level of education gross household income is more concentrated in the left part of the graphs, around minimal pension benefit and minimum wage levels. Bunching is more pronounced around pension benefits for the retired and for those with less schooling, who are more likely to be from older cohorts and retired. Generally, the distribution obtained using SHARE Denmark data is more spread than that obtained with register data, but data masses are consistently estimated. The difference in distributions is consistent with the hypothesis of additive measurement error, as the survey data distribution flattens out compared to the administrative data.

4 Comparison of micro-data and measurement error

Typically, we would like to assess both the magnitude of the measurement errors and whether they satisfy the usual assumptions on their distributions. However, assessing the magnitude of errors for categorical variables such as education and labour market status is challenging. Even if education, unlike labour market status, is treated as an ordered variable, a traditional error measure designed for continuous variables would be inaccurate and hardly comparable. Instead, following the norm in the measurement error literature, we define misclassification as whenever SHARE Denmark and register data report different values of a categorical variable for the same individual.

We focus on the identification and frequency of misclassification for education and labour market status. Following Pedace and Bates (2000), we study the probability of misclassification for both education level and labour market status with a probit model. In all the models where education, labour market status or gross household income enters as independent variables we use register data values. Moreover, we test for precision of individual reports of education level by fitting an ordered probit model on SHARE-defined education categories, with register data entering the model as independent variables. In this way, we are able to test whether the underlying continuous variable for education levels is uncorrelated with other variables, given the administrative record of education.

We study measurement errors in gross household income in a more standard way. Here we recall some standard results from the measurement error literature. Following the notation of Bound et al. (1994), we assume that while the true model is

$$Y = X\beta + \varepsilon, \tag{1}$$

income is measured with error according to

$$\widetilde{X} = X + u, \tag{2}$$

or

$$\widetilde{Y} = y + v \tag{3}$$

if income enters the model as a dependent or independent variable respectively. In other words, we assume an additive measurement error model.

As previously mentioned, the usual assumptions on measurement error (u or v) are independence of the true variable and zero correlation with other variables in the model (and with their specific measurement error). In general, even if these assumptions are satisfied, when the incorrectly measured variable is an independent variable, measurement error causes an attenuation bias. When the incorrectly measured variable is a dependent variable, there is a loss in efficiency.

Other forms of bias may occur when those assumptions are not satisfied. We define b as the coefficient from the model biased from measurement error, and we focus on the two possible biases arising from the correlation of the measurement error with the true value of the variable.

First, consider a model where income is the only independent variable, the dependent variable is measured without error and, for simplicity, $\beta \geq 0$. The independent variable bias is

$$\frac{\beta - b}{\beta} = \begin{cases} \frac{\sigma_u^2}{\sigma_X^2 + \sigma_u^2} & \text{if } Cov(X,u) = 0 \\ b_{u\widetilde{X}} & \text{if } Cov(X,u) \neq 0 \end{cases}, \tag{4}$$

where $b_{u\widetilde{X}}$ is the coefficient from the regression of u on the measured variable $\widetilde{X}$. If measurement errors are uncorrelated with the true variable X, then the proportional bias equals the ratio of variance due to measurement errors over the total variance. In other words, measurement error causes an underestimation of the true β of the model.

Second, if only the dependent variable is measured with error, a dependent variable bias in the estimation of the coefficient β occurs only if the measurement error v is correlated with the true variable Y, and it is equal to the coefficient from the regression of v on the true variable Y, b_{vY}.

For gross household income we therefore test for existence of both types of bias, measure b_{vY} and $b_{u\widetilde{X}}$, and study their magnitude. We then test for correlations between measurement errors and other variables typically used in empirical models. Finally, we provide empirical evidence of the consequences of measurement error bias with both simulated and real data. Because of the way we treat imputed income values (see Section 3), whenever we use SHARE Denmark income data, the standard errors associated with the estimators will be downward biased. As our goal is to test whether measurement errors are independently distributed, not accounting for imputed values works against us, in the sense that if we accept the hypothesis of independence under conservative standard errors, we would definitely accept it if we corrected confidence intervals for multiple imputations.

The remainder of this section determines the magnitude of measurement errors in SHARE Denmark data and studies their distributions. We conduct separate analyses for education, labour market status and gross household income in the next sub-sections.

4.1 Education

We divide both SHARE Denmark and administrative data in the three education categories defined in Section 3 (low, medium and high education). The first row of Figure 7 illustrates the frequency distribution of the education variables for the 1545 observations for which neither register nor SHARE Denmark data are missing. The sizes of the scatters represent the relative percentage of each cell, by SHARE Denmark (a) and register data (b).

Overall, 27.4% of the sample exhibits different values of education in the two datasets. Most of misclassified individuals (84.9%) report a higher education level in SHARE Denmark than the one we

(a) Percentages by SHARE Data (rows)   (b) Percentages by Register Data (columns)

Numbers indicate cell counts, circle size represents proportion of the sample

Fig. 7: Education level correspondence, by gender

observe in the registers. More than 50% of misclassifications have a low level of education according to registers, while less than 10% has a high level of education. The distribution of those variables suggests a negative relationship between measurement errors and education levels.

The second and third rows of Figure 7 divide observations by gender. Grouping individuals by register data categories shows that poorly educated males misreport their education level more than 50% of the time. However, even though the percentage of misclassified males is slightly higher, the gender differential is not significant according to a two-sample test of proportions (z-stat. = 0.83).

It is useful to identify the typical characteristics of individuals who are most likely to misreport levels of education. We therefore fit a probit model on misclassification to test for the presence of systematic correlations with the occurrence of measurement error. The first column of table 1 reports the estimated coefficients of this model.

As expected, as the true level of education increases, the frequency of errors is significantly lower. Whether the cause of this negative correlation is the higher precision of more educated individuals or simply the attempt to conceal low levels of education is unclear. However, we know that gross household income is positively associated with the probability of misclassification, and that this positive relation is weaker for highly educated individuals. Therefore, the probability of misclassification is higher for high income, lower educated respondents, who more often report a higher level of education during SHARE personal interviews.

Column two of Table 1 supports this finding. Here we study the probability of reporting a higher level of education than that registered, and we therefore exclude highly educated individuals from the sample. Not surprisingly, high income and low education significantly affect the probability of positive misclassification as well. This finding suggests that the correlation between measurement error and the true level of the variable can be a source of measurement error bias.

However, there is no evidence that the probability of misclassification relates systematically to other variables commonly used in socio-economic analyses. Both labour market status and gender do not have any influence on the probability of misclassification. Moreover, precision in personal education assessment does not decrease with age. Misclassification is independent of the majority of the most common control variables. Moreover, to check for dependencies across measurement errors, we add to the model a dummy indicating whether the respondent does not report the same labor market status as the one we observe in the registers. We show that misclassification in education levels is independent of misclassification in labor market status, supporting the hypothesis of classical measurement errors, independently distributed across variables.

Finally, we assess precision of SHARE Denmark responses by estimating an ordered probit model on SHARE categories using as independent variables register data categories and the same controls we use in the probit estimations. Ideally, one would like to observe significant coefficients for the true value of education and insignificant coefficients for the other variables, in support of the hypothesis that if there are measurement errors in the underlying latent variable, they follow an independent random process.

Column three of Table 1 shows however that income correlates strongly with the underlying latent variable, and there is weak evidence for an interaction between levels of education and income. Nonetheless, neither age, gender nor labour market status influences measurement errors, indicating that, given income level, measurement errors are independent of many of the most common variables used in socio-economic studies.

Tab. 1: Education and labour market status misclassification

	Education			Labor Market Status
	Misclass.	Upward B.	O. Probit	Misclass.
Medium education	-0.548***	-0.714***	1.162***	-0.0706
	(0.0831)	(0.0854)	(0.0789)	(0.117)
High education	-0.999***		2.852***	-0.130
	(0.113)		(0.118)	(0.144)
Income	0.239***	0.228**	0.379***	-0.139
	(0.0869)	(0.0897)	(0.0790)	(0.122)
Income×Medium ed.	-0.203**	-0.0901	-0.224***	-0.135
	(0.0958)	(0.0972)	(0.0855)	(0.138)
Income×High ed.	-0.556***		-0.0894	-0.0378
	(0.124)		(0.117)	(0.149)
Employed	-0.00542	0.0792	0.135	-0.462***
	(0.111)	(0.122)	(0.100)	(0.135)
Other	-0.0696	0.0494	0.158	0.740***
	(0.180)	(0.203)	(0.161)	(0.170)
Age	0.000134	-0.00116	0.000781	-0.0395***
	(0.00534)	(0.00578)	(0.00488)	(0.00673)
Male	0.100	0.107	0.0832	0.0575
	(0.0722)	(0.0805)	(0.0654)	(0.0944)
Couple	0.0492	-0.0308	-0.135	-0.205*
	(0.0941)	(0.104)	(0.0851)	(0.123)
Misreported LMS	-0.0198	-0.0193	0.0173	
	(0.122)	(0.135)	(0.110)	
Misreported educ.				-0.0148
				(0.107)
Constant	-0.214	-0.110		1.419***
	(0.383)	(0.415)		(0.462)
Threshold 1			-0.0271	
			(0.351)	
Threshold 2			1.849***	
			(0.355)	
Observations	1542	1184	1542	1542
Pseudo-R^2	0.0732	0.0519	0.332	0.112

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; standard errors in parentheses

Estimated coefficients from binomial probit models in columns 1, 2 and 4, and from an ordered probit model in column 3

4.2 Labour market status

Labour market status is an unordered variable, and the correlation between the true variable and measurement errors is hard to quantify. Thus we will focus solely on identification of misclassification and to what extent it can be predicted.

We group respondents as outlined in Section 3. The connection between SHARE Denmark and register data is not strong: while administrative registers offer a detailed coding of the socio-economic status of residents of Denmark from income tax and transfer registers, SHARE respondents assess their own labour market status, choosing between employment, retirement, unemployment, disability or permanent sickness and homemaking.

Therefore, we would expect measurement errors to play an important role in individual assessments, especially because the boundaries between retirement and other states such as unemployment and homemaking may not be clear for individual respondents. Moreover, lack of precision in individual assessments might be correlated with demographics, such as education and age.

In fact, individuals assess their labour market status more precisely than their level of education. Figures 8 and 9 show the frequency distribution of labour market status, by SHARE Denmark (a) and register data (b), for the 1644 observations that responded to the labour market question in the SHARE questionnaire. Figure 8 groups respondents by gender; Figure 9, by education level. Over 90% of employees and self-employed (henceforth employed) and retired correctly assess their labour market status, and there are no differences in misclassification by gender. As expected employees are more precise than retirees, but the difference is small (less than 3 percentage points) and not significant at the 95% confidence level according to a two-sample test of proportions (z-stat. =1.92).

People with other labour market states (e.g. unemployed, homemakers, permanently sick or disabled) are generally less precise: 33.7% of them – probably the ones in transition between states or subject to mixed welfare programs – declare themselves to be either employed or retired. However, according to register data, only 5% of the linked SHARE Denmark sample is neither employed nor retired.

The distribution of responses appears to be independent of both gender and education. In order to broadly identify the characteristics of misclassified respondents we fit a probit model of a misclassification indicator, using labour market status and other demographics as independent variables. Similarly to the findings for education, there is no evidence of correlation between misclassification in education and labor market status.

The fourth column of Table 1 shows, as expected, that employed respondents are most precise, followed by the retired. Neither education nor gender has any influence on the probability of misclassification, but income affects it negatively. The distinction between retirement and other mixed states clarifies at older ages. The negative effect of age on misclassification is therefore not surprising.

4.3 Gross household income

As in previous sections, we assume that (income) registers provide the true value (of household income). We acknowledge that the tax register can be subject to error, for example regarding income from company ownership. However, such income sources are likely to be uncommon in the SHARE Denmark sample, and are most likely dropped from the analysis when we exclude outliers. While in the data used for this project we cannot discriminate between income sources, we observe that in November 2004 the register based labour force statistics dataset records 55 respondents in our sample as at least partially self-employed. Excluding these observations does not have any significant impact on our characterization of measurement error in income (see appendix).

(a) Percentages by SHARE Data (rows)

(b) Percentages by Register Data (columns)

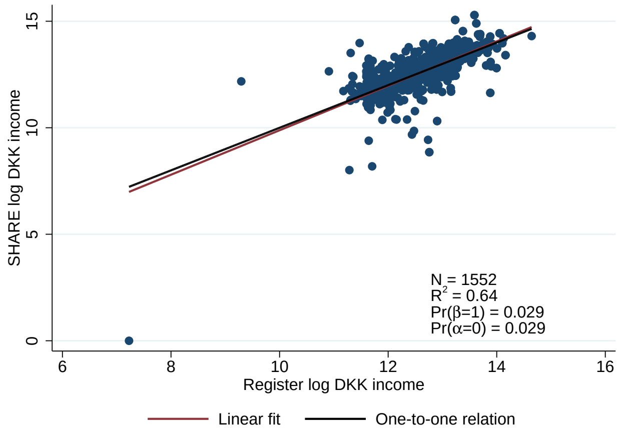
Numbers indicate cell counts, circle size represents proportion of the sample

Fig. 8: Labour market status correspondence†, by gender

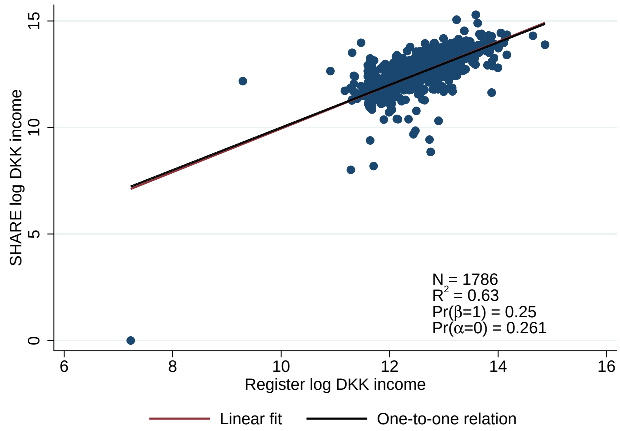
(a) Percentages by SHARE Data (rows) (b) Percentages by Register Data (columns)

Numbers indicate cell counts, circle size represents proportion of the sample

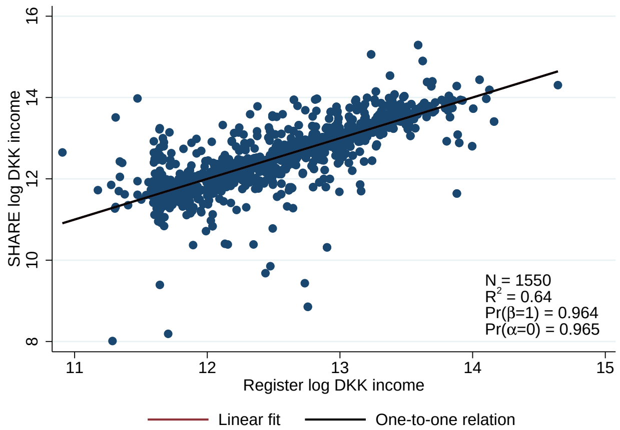
Fig. 9: Labour market status correspondence, by education



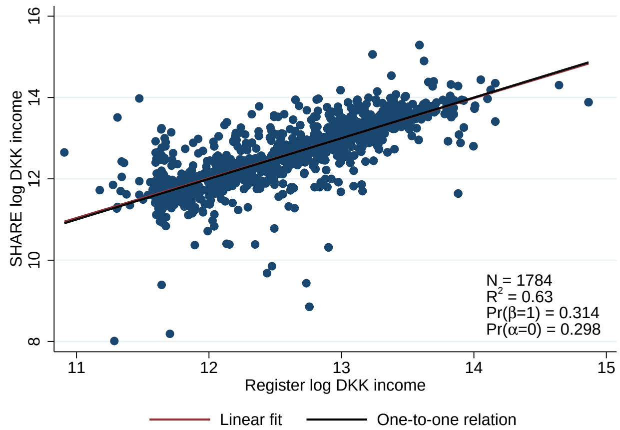
(a) Small, consistent sample: 1552 observations



(b) Enlarged Sample: 1786 observations



(c) Small, consistent and truncated sample: 1550 observations

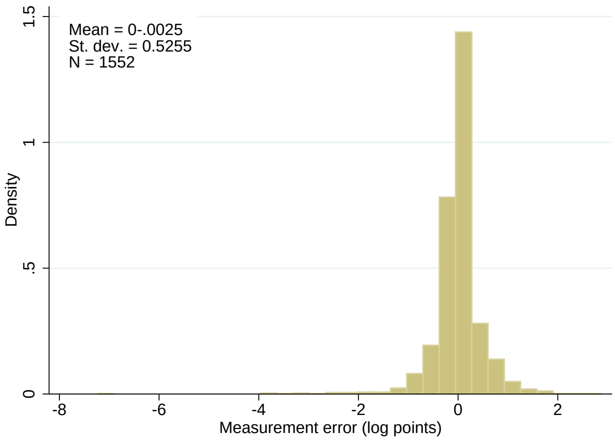


(d) Enlarged and truncated sample: 1784 observations

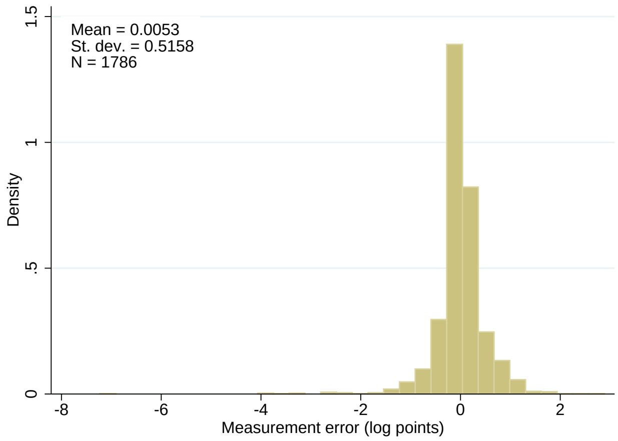
Fig. 10: Log household income scatterplot

The first row of Figure 10 shows the scatterplots of register versus SHARE Denmark measures of income. We show scatterplots for both the small consistent sample, for which we drop those 117 respondents whose spouse was not interviewed in SHARE Denmark, and the enlarged one, for which we add the not-interviewed spouse. While on average log household income measured in SHARE Denmark and that drawn from the tax reports approximate each other closely, the presence of measurement errors is evident. The figures plot a linear fit of register data on SHARE Denmark data and the 45 degree line of full agreement. If measurement error was completely independent of the value in the register data, the estimated coefficients for the linear relationship should be $[\alpha,\beta] = [0,1]$. The figures report the results from the single tests of hypothesis. A joint test of the two hypotheses is rejected at the 95% confidence level only in the top left pane in Figure 10.

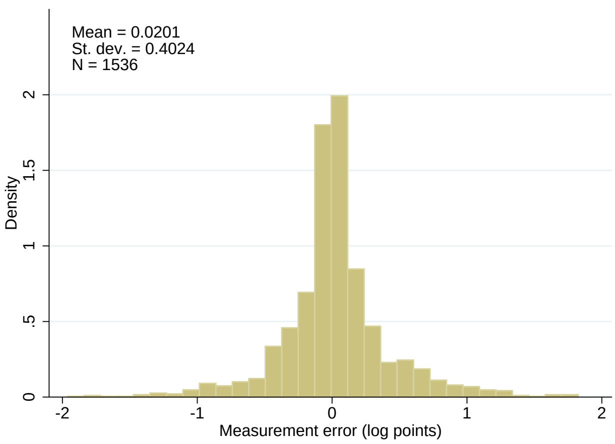
As most observations are concentrated at the right top of the graph, we drop the households with log yearly income lower than 10 (which corresponds to 22,000 DKK, roughly 3,000€) according to register data. This truncation eliminates two outliers, in particular one reporting zero income in SHARE Denmark data, and virtually eliminates the correlation between measurement error and income. Figure 11 shows that, net of a few outliers, the distribution of measurement errors is symmetric and highly concentrated around zero log points. The average measurement error is not significantly



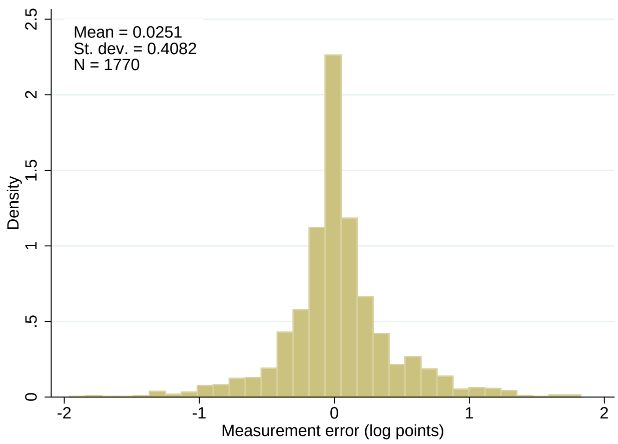
(a) Small sample, no truncation



(b) Enlarged Sample



(c) Small sample: truncation at ±2 log points



(d) Enlarged sample: truncation at ±2 log points

Fig. 11: Gross household income: measurement error distribution

Tab. 2: Income measurement errors; descriptives

	Individuals			Households		
	Register	SHARE	u	Register	SHARE	u
Mean	12.65	12.64	-0.00251	12.49	12.48	-0.00855
Std. Dev.	0.667	0.872	0.525	0.684	0.911	0.565
Observations	1552	1552	1552	1044	1044	1044

different from zero. Outliers are mostly in the left of the distribution where SHARE Denmark data underestimates the registered value of household income.

Table 2 presents averages and standard deviations of the logarithm of gross household income in SHARE Denmark and register data, and of the observed measurement error u. We replicate the table at the individual and at the household level. The aggregates do not differ much, and the average measurement error is not significantly different from zero. Essentially, Table 2 tests for the zero-mean hypothesis of the measurement error, which we cannot reject. However, the consequences for consistency of OLS estimators depend not only on the first moment of the measurement error distribution, but also on its variance. In particular, the bias due to measurement error depends crucially on the reliability ratio, i.e. the relative magnitude of measurement error variance with respect to the variance of the underlying variable.

Table 3 presents our estimates for measurement error biases according to the framework and the notation outlined at the beginning of this section. This exercise estimates the attenuation bias due to measurement error as the relative difference of OLS coefficients if we used SHARE Denmark data with respect to register-drawn, third party reported data. The first row shows the results from a simple OLS estimation. The sample proportion of total variance due to measurement errors represents the expected independent variable bias when measurement errors are independently distributed. Even under this assumption the variance ratio is not negligible, and we expect $\widehat{b} \simeq 0.61\beta$. The expected bias is slightly higher when we aggregate data to the household level.

Contrary to previous validation studies, we do not find a negative correlation between measurement errors and the registered value of the variable. As expected given the descriptives in Section 3 the correlation is positive, but $\widehat{b}_{vY}$ is marginally significant at the 95% confidence level only when we use data at the individual level. Such a correlation both introduces a dependent variable bias and increases the previously estimated independent variable bias.

However, this correlation is rather small, and increases the dependent variable bias measured by $\widehat{b}_{u\widetilde{X}}$ of only 1.6% (not significantly different from zero according to the estimated standard error for $\widehat{b}_{u\widetilde{X}}$). Moreover, the variance of errors is not independent of income (in particular, the sample variance decreases with income), and if we compute Huber-White heteroskedasticity robust standard errors, as we do in the second row of Table 3, $\widehat{b}_{vY}$ is not significantly different from zero.

The median regression in the third row of the table suggests that the presence of outliers leads to an overestimation of both $\widehat{b}_{u\widetilde{X}}$ and $\widehat{b}_{vY}$. We therefore exclude from the sample the single outlier, clearly shown in Figure 10, reporting zero income in the SHARE Denmark dataset. The last row of Table 3 shows the results from the OLS estimation on the 1551 observations sample. Dropping the single outlier reduces the proportion of total variance due to measurement error , at both the individual and the household level, and is sufficient to eliminate the positive correlation between measurement

Tab. 3: Income measurement errors; bias estimation

	Individuals			Households		
	$\frac{\sigma_u^2}{\sigma_X^2+\sigma_u^2}$	$\hat{b}_{u\tilde{X}}$	$\hat{b}_{vY}$	$\frac{\sigma_u^2}{\sigma_X^2+\sigma_u^2}$	$\hat{b}_{u\tilde{X}}$	$\hat{b}_{vY}$
Standard	0.383	0.389***	0.0436**	0.405	0.410***	0.0451*
		(0.0117)	(0.0200)		(0.0144)	(0.0255)
Robust	0.383	0.389***	0.0436	0.405	0.410***	0.0451
		(0.0319)	(0.0604)		(0.0379)	(0.0816)
Median	0.383	0.182***	0.0115	0.405	0.183***	0.00677
		(0.00878)	(0.00717)		(0.0117)	(0.00910)
No Outlier	0.363	0.360***	-0.0138	0.379	0.373***	-0.0348
		(0.0124)	(0.0192)		(0.0157)	(0.0242)
Observations		1552			1044	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; standard errors in parentheses

errors and the true value of income.

In Table 4 we test the hypothesis of zero correlation between measurement error and other variables in the model. We collapse the dataset at the household level, and regress the measurement error on a set of characteristics of both the financial respondent and, if observed, of the partner. As Table 4 shows, at the household level measurement errors have zero mean and are not affected by any of the observed characteristics of both the financial respondent and of the partner. Only conditioning on partner's characteristics, thus ignoring information from singles, we find mild correlation between measurement error in income and labor market status. We accept the null hypothesis of joint equality to zero of estimated coefficients in all models.

Moreover, we do not find evidence for cross-correlations between measurement errors in different variables: the misclassification dummies, indicating whether SHARE Denmark data on labor market status or education levels disagrees with the Danish registers, do not have a significant relationship with measurement error in income.

That measurement errors are uncorrelated with other variables greatly simplifies bias analysis and management. We have shown that, in a standard setting, measurement error biases the results of a linear regression only if income enters the model as an independent variable. While this bias is inevitable in the presence of measurement error, it is also easily tractable once the variance of the measurement error, which Table 3 provides for the Danish sample, is known (Stefanski, 2000).

Table 5 provides two examples of the independent variable bias occurring for gross household income. In the second column we compute the unreability ratio for log gross household income, providing the theoretical benchmark for the observed bias. In the third and fourth column we compute the 95% confidence intervals for the proportional biases allowing for some correlation between u and X, as in equation (4). Finally, in the last column we compute the observed measurement error bias, as the proportion between the coefficient calculated on measured data and that calculated on real data.

The first rows show the results from a simulated dataset, where the dependent variable is a linear function of the income as reported in the administrative registers, but we estimate a linear regression

Tab. 4: Income measurement errors: dependence analysis
Household Grouped Estimations

	Measurement Error			
	Financial respondent characteristics			
Age	-0.00265	-0.00203	-0.00236	
	(0.00246)	(0.00267)	(0.00296)	
Medium education	-0.0208	-0.0212	-0.0217	
	(0.0433)	(0.0444)	(0.0440)	
High education	0.0627	0.0660	0.0640	
	(0.0503)	(0.0519)	(0.0511)	
Male	0.0142	0.0202	0.0215	
	(0.0326)	(0.0350)	(0.0351)	
Employed	0.0620	0.0579	0.0585	
	(0.0535)	(0.0537)	(0.0532)	
Misreported educ.	0.0238	0.0266	0.0273	
	(0.0366)	(0.0364)	(0.0366)	
Misreported LMS	0.00206	-0.00274	-0.00402	
	(0.0709)	(0.0696)	(0.0708)	
	Partner characteristics			
Age		-0.000597	0.000646	-0.00137
		(0.00126)	(0.00339)	(0.00127)
Medium education		0.0236	0.0241	0.0368
		(0.0436)	(0.0439)	(0.0443)
High education		-0.00252	-0.000764	0.0331
		(0.0564)	(0.0568)	(0.0548)
Missing spouse ed.		0.0357	0.0104	-0.00602
		(0.0952)	(0.0952)	(0.0948)
Employed		0.0352	0.0479	0.0832*
		(0.0527)	(0.0647)	(0.0452)
Misreported educ.		0.0701	0.0698	0.0763
		(0.0484)	(0.0486)	(0.0490)
Misreported LMS		0.115	0.122	0.137*
		(0.0752)	(0.0794)	(0.0727)
Couple			-0.110	
			(0.257)	
Constant	0.123	0.0551	0.103	-0.000640
	(0.184)	(0.207)	(0.244)	(0.0972)
Observations	946	946	946	946
Adj. R^2	0.006	0.004	0.003	0.003
χ^2 F-Stat	2.855	2.173	2.052	3.251

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; standard errors in parentheses
OLS coefficient estimates

Tab. 5: Income measurement error: biases

	N	$\frac{\sigma_u^2}{\sigma_{\tilde{X}}^2+\sigma_u^2}$	$\hat{b}_{u\tilde{X}}$ 95% Conf. Int		$\frac{\hat{\beta}-\hat{b}}{\hat{\beta}}$
Simulation	1552	.383	.326	.451	.388
No outlier	1551	.363	.313	.408	.361
Assets	1537	.382	.326	.452	.436
No outlier	1536	.362	.312	.408	.357

model using SHARE Denmark data as independent variable. We repeat the simulation a thousand times, and we report the average relative bias in the fifth column. The unreliability ratio closely approximates the estimated average relative bias, even more so when the single outlier reporting zero income in SHARE Denmark data does not enter the sample. Removing the outlier also reduces the 95% bound that allows for correlation of the measurement error with the true value of the income variable, as that observation drives the correlation observed in the first row of Table 3.

The third and fourth rows replicate the exercise, using register data on assets held at the 31^{st} of December. In the full sample the observed relative bias falls within the predicted 95% confidence interval, although closely to its upper bound. Removing the single outlier not only shrinks the confidence bound, but also reduces by 18% the observed relative bias, which now falls closely to the unreliability ratio in the first column. This exercise suggests that, once outliers are removed, linear estimates of income effects from SHARE Denmark data underestimate the income coefficient by roughly one third.

5 Conclusion

We conducted a validity study of SHARE Denmark by linking 97% of respondents in Denmark to Public Administrative Register information on education level, labour market status and income. Measurement error in schooling, labour market status and income is found to be respectively modest, small and insignificant. SHARE Denmark respondents tend to overstate their level of schooling. This is driven by individuals with high (reported or registered) incomes reporting higher qualifications than those administratively registered. Labour market status is precisely reported, but the young are more likely to respond at odds with registers. Gross household income is on average not statistically different in SHARE Denmark and register data.

Unlike income validation studies for the US Panel Study of Income Dynamics and the Health and Retirement Study, we find that SHARE Denmark income measurement error is classical. As a consequence, econometric models are easily adjustable once the share of the variance due to measurement error (36.3% for household income in SHARE Denmark) is known. This finding may be a feature of the particular income measure we consider - gross household income. With respect to individual labour earnings, which is validated in the PSID-VS, gross household income is more likely to be correctly reported. However, the share of variance in gross household income due to measurement error is higher in SHARE Denmark than in HRS.

Our validation study is highly supportive of the quality of these three pieces of data collected in the first wave of SHARE for Denmark. We chose to validate the most common variables of interest in socio-economic analyses, for the first wave of data that we were able to link to register data. Other responses may not have this high quality, and neither may other Danish waves of the survey, or SHARE data collected in other countries. We leave for future work the validation of other variables and other waves of the Danish survey. Combining waves would allow analysis of repeated measures and transitions. Historical life-course information was collected in SHARE-life and linkage to administrative registers back to the 1960's could be used for validation there.

We conclude that, despite some inconsistencies, in particular for completed schooling, SHARE Denmark succeeds in providing researchers with high-quality information on education and retirement status and income, in which the measurement error bias can be accommodated using standard econometric techniques.

6 Acknowledgements

We are grateful for comments from James Banks, Martina Brandt, Axel Börsch-Supan, Søren Leth-Petersen and David Wise. Funding was provided by the Danish Strategic Research Council (DSF-09-065167) and Danish Social Sciences Research Council (FSE-09-063859). The usual disclaimer applies.

This paper uses data from SHARE release 2.5.0, as of May 24th 2011. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001- 00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT- 2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, R21 AG025169) as well as from various national sources is gratefully acknowledged (see `http://www.share-project.org` for a full list of funding institutions).

References

Barron, J. M., M. C. Berger, and D. A. Black (1997). How well do we measure training? *Journal of Labor Economics 15*(3), pp. 507–528.

Biemer, P. P., R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.) (2004). *Measurement errors in Surveys*. Wiley Series in Probability and Statistics. John Wiley & Sons.

Bingley, P. and A. Martinello (2014). Measurement error in income and schooling, and its bias on linear estimators. Working Paper 01:2014, SFI - The Danish National Centre for Social Research.

Bollinger, C. R. (1998). Measurement error in the current population survey: A nonparametric look. *Journal of Labor Economics 16*(3), pp. 576–594.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics 12*(3), pp. 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter Chapter 59, pp. 3705 – 3843. Elsevier.

Bound, J. and A. B. Krueger (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics 9*(1), pp. 1–24.

Bricker, J. and G. V. Engelhardt (2008). Measurement error in earnings data in the health and retirement study. *Journal of Economic & Social Measurement 33*(1), pp. 39 – 61.

Browning, M. and S. Leth-Petersen (2003). Imputing consumption from income and wealth information*. *The Economic Journal 113*(488), F282–F301.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber (2013, Aug). Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *Int J Epidemiol 42*(4), pp. 992–1001.

Christelis, D. (2011). Imputation of missing data in waves 1 and 2 of share. SHARE Working Paper Series 01-2011, University of Mannheim.

Duncan, G. J. and D. H. Hill (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics 3*(4), pp. 508–532.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley series in probability and mathematical statistics. New York: John Wiley and Sons.

Greenberg, D. and H. Halsey (1983). Systematic misreporting and effects of income maintenance experiments on work effort: Evidence from the seattle-denver experiment. *Journal of Labor Economics 1*(4), pp. 380–407.

Jensen, V. M. and A. W. Rasmussen (2011). Danish education registers. *Scandinavian Journal of Public Health 39*(7 suppl), pp. 91–94.

Kane, T. J., C. E. Rouse, and D. Staiger (1999, July). Estimating returns to schooling when schooling is misreported. Working Paper 7235, National Bureau of Economic Research.

Keating, E., D. G. Paterson, and C. Stone (1950). Validity of work histories obtained by interview. *Journal of Applied Psychology 34*(1), pp. 6 – 11.

Leth-Sørensen, S. (1993). Ida: An integrated database for labour market research. Technical report, Statistics Denmark, Copenhagen.

Mellow, W. and H. Sider (1983). Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics 1*(4), pp. 331–344.

Miller, H. P. and L. R. Paley (1958, January). Income reported in the 1950 census and on income tax returns. In *An Appraisal of the 1950 Census Income Data*, NBER Chapters, pp. 177–204. National Bureau of Economic Research, Inc.

Pedace, R. and N. Bates (2000). Using administrative records to assess earnings reporting error in the survey of income and program participation. *Journal of Economic & Social Measurement 26*(3/4), pp. 173 – 192.

Poterba, J. M. and L. H. Summers (1986). Reporting errors and labor market dynamics. *Econometrica 54*(6), pp. 1319–1338.

Tab. 6: Education variable; wording of original questions

Question	Danish	English	ISCED
DN010_	Please look at card 2. What is the highest school leaving certificate or school degree that you have obtained?		
	7. klasse	7^{th} grade or lower	1
	8. klasse	8^{th} grade	2
	9. klasse	9^{th} grade	2
	10. klasse, realeksamen	10^{th} grade	2
	Studentereksamen eller HF	Gymnasium	3
	HH, HG, HHX, HTX	Technical secondary	3
DN012_	Please look at card 3. Which degrees of higher education or vocational training do you have?		
	Specialarbejderuddannelse	Vocational	3
	Lærlinge eller EFG-uddannelse	Vocational	3
	Anden faglig uddann. > 12 mdr.	Vocational > 12 months	3
	Kort videregående uddannelse	Higher education (<3y)	5
	Mellemlang videregående uddannelse	Higher education (3-4y)	5
	Lang videregående uddannelse	Higher education (>4y)	5

Statistics Denmark (1977). Folke- og boligtællingen 9. november 1970: C.4. uddannelse (housing and population census 9th november 1970: Section 4: Education). Technical report, Statistics Denmark.

Stefanski, L. A. (2000). Measurement error models. *Journal of the American Statistical Association 95*(452), pp. 1353–1358.

Appendix

A Survey questions

In this section we report the questions originally asked in the first wave of SHARE and how the variables for education, labor market status and gross household income were constructed. For further information and the exact Danish wording, we refer to the SHARE guideline and country-specific questionnaires available at `www.share-project.org`.

A.1 Education

The questions from which we draw information about education are those in module DV of SHARE wave 1, named DN010_ and DN012_ in the questionnaires. Table 6 shows the Danish wording of the options, the corresponding English translation, and the 1997 ISCED code that derives from the answers.

Tab. 7: Gross household income components

Variable	Question	Description
ydipv	ep205	Annual gross income from employment previous year
yindv	ep207	Annual gross income from self-employment previous year
ybaccv	as005	Interest income from bank accounts
ybondv	as009	Interest income from bonds
ystocv	as015	Dividends from stocks/shares
ymutfv	as058	Interest and dividend income from mutual funds
yrentv	ho030	Income from rent
yltcv	ep086	Monthly long-term care insurance previous year
pen1v	ep078_1	Monthly public old age pension
pen2v	ep078_3	Monthly public early or pre-retirement pension
pen3v	ep078_4	Monthly main public DI pension, or sickness benefits
pen4v	ep078_6	Monthly public unemployment benefit or insurance
pen5v	ep078_7	Monthly public survivor pension from partner
pen7v	ep078_9	Monthly war pension
pen8v	ep324_1	Monthly private (occupational) old age pension
pen9v	ep324_4	Monthly private (occupational) early retirement pension
pen10v	ep324_5	Monthly private (occupational) disability insurance
pen11v	ep324_6	Monthly private (occupational) survivor pension from partner's job
reg1v	ep094_1	Monthly life insurance payment received
reg2v	ep094_2	Monthly private annuity or private personal pension
reg4v	ep094_4	Monthly alimony received
reg5v	ep094_5	Monthly regular payments from charities received
yohmv	hh002	Annual other hhd members' net income
yohbv	hh011	Annual other hhd members' net income from other sources

A.2 Labor market status

We take the labor market status variable from the single question EP005_ in module EP of SHARE wave 1. The question is phrased as

```
   Please look at card 21.  In general, how would you describe your
current situation?

 (a) Retired

 (b) Employed or self-employed (including working for family business)

 (c) Unemployed

 (d) Permanently sick or disabled

 97. Other (specify)
```

A.3 Gross household income

Gross household income is the sum over all household members of a list of variables, each capturing a different portion of the income process of an household, each asked separately to the financial respondent. Table 7 shows the variables that form the gross household income variable and their source within the questionnaires. All questions refer to previous year income.

B Education questions in the 1970 Census

We hereby report the official English translation of the census questions regarding education level,

Section B. Education and vocational training status To be filled in for all persons who have turned 14, but not 70 years (i.e. born between November 9th, 1900 and November 8th, 1956)

6 Education or vocational training in progress
Persons who are **not** in process of education or vocational training, write: none For **school pupils** (i.e. up to and including secondary level) the class is to be listed, eg. 7th class, "2nd real", "1.g" **apprentices and trainees** should list this and the trade, eg, bricklayer's apprentice, cabinet maker's apprentice, traffic trainee, bank trainee For **students and others receiving an education**, the kind of education is to be listed as accurately as possible, eg.university student with language major or the like, correspondent - 3 languages, laboratory technician training, teacher's training, specialist teacher's training, agricultural school student.

7 Completed schooling
For persons who have left school, the highest examination passed is to be listed, e.g. "mellemskoleeks" (i.e. exam after 9 years of schooling), "realeks" (i.e. exam after 10 years of school), "nyspr. student" (i.e. exam after 12 years of school with language major), "HF" (i.e. exam after 11 years of school) or highest class in school which has been completed, e.g. 7th school year, 9th class, "2. real" (i.e. 10 years of school). For persons who have attended school abroad, the corresponding information is to be listed, the total number of years in school, and name of the country

8 Completed education or vocational training
This space is also to be filled in by persons who are economically inactive. The most important education or voctional training or further training is to be listed. For persons with an **exam or school leaving certificate** from university, higher school, or the like, the kind of education is to be listed as accurately as possible, e.g. university degree in languages or the like, university degree in engineering, degree from technical engineering school (college), chartered accountant, "HA" (i.e. degree from school of business and economics), school teacher, social worker. For persons with **apprentice's training or other vocational training**, the vocation is to be listed, e.g. electrician, trained office clerk, book seller's assistant, skilled baker, nurse, assistant nurse, technical assistant, laboratory worker, agricultural technician, catering officer. For persons whose vocational training is entirely practical this is to be listed and the nature of the work, e.g. practical office training, practical agricultural training. Persons without completed education or training including school pupils should write: none.

C Measurement error in income, excluding self-employed individuals

Tab. 8: Income measurement errors exclusing self-employed respondents; bias estimation

	Individuals			Households		
	$\frac{\sigma_u^2}{\sigma_{\tilde{X}}^2+\sigma_u^2}$	$\hat{b}_{u\tilde{X}}$	$\hat{b}_{vY}$	$\frac{\sigma_u^2}{\sigma_{\tilde{X}}^2+\sigma_u^2}$	$\hat{b}_{u\tilde{X}}$	$\hat{b}_{vY}$
Standard	0.380	0.386***	0.0463**	0.405	0.411***	0.0530**
		(0.0119)	(0.0202)		(0.0144)	(0.0257)
Robust	0.380	0.386***	0.0463	0.405	0.411***	0.0530
		(0.0335)	(0.0625)		(0.0381)	(0.0830)
Median	0.380	0.174***	0.00939	0.405	0.185***	0.00794
		(0.00860)	(0.00742)		(0.0117)	(0.00905)
No Outlier	0.358	0.356***	-0.0135	0.378	0.374***	-0.0284
		(0.0126)	(0.0193)		(0.0157)	(0.0243)
Observations		1552			1044	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; standard errors in parentheses