THE DANISH
NATIONAL CENTRE
FOR SOCIAL RESEARCH

Paul Bingley
Alessandro Martinello

1:2014 WORKINGPAPER

MEASUREMENT ERROR IN INCOME AND SCHOOLING, AND THE BIAS OF LINEAR ESTIMATORS

SFI – THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH

MEASUREMENT ERROR IN INCOME AND SCHOOLING, AND THE BIAS OF LINEAR ESTIMATORS

Paul Bingley and Alessandro Martinello

THE DANISH NATIONAL CENTRE FOR SOCIAL RESEARCH, COPENHAGEN, DENMARK;

Working Paper 1:2014

Measurement error in income and schooling, and the bias of linear estimators

Paul Bingley, *SFI*
Alessandro Martinello, *University of Copenhagen and SFI*

January 9, 2014

The characteristics of measurement error determine the bias of linear estimators. We propose a method for validating economic survey data allowing for measurement error in the validation source, and we apply this method by validating Survey of Health, Ageing and Retirement in Europe (SHARE) data with Danish administrative registers. We find that measurement error in surveys is classical for annual gross income but non-classical for years of schooling, causing a 21% amplification bias in IV estimators of returns to schooling. Using a 1958 Danish schooling reform, we contextualize our result with an estimate of the income returns to schooling.

I. Introduction

Researchers have long known that measurement error in the data of interest can affect the consistency of parametric estimators of even the simplest linear model (Stefanski, 1985, 2000). In applied research, the implicit assumption is often that measurement error can be characterized as classical, i.e., an additive, independent

We are grateful for comments from James Banks, Martina Brandt, Arnaud Chevalier, Mette Ejrnæs, Anders Holm, Bertel Schjerning, Søren Leth-Petersen and Ian Walker. 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).

error term with constant variance (Fuller, 1987). Such an assumption is particularly convenient in the estimation of linear models, as instrumental variable (IV) estimators are robust to classical measurement error, while ordinary least squares (OLS) only suffers from proportional attenuation bias. However, while under this assumption the consequences of measurement error are trivial in linear models, all the seminal validation studies of Mellow and Sider (1983); Duncan and Hill (1985); Bound and Krueger (1991); Bound et al. (1994); Barron et al. (1997); Bollinger (1998), followed later by Bricker and Engelhardt (2008), suggest that measurement error in labor market-related outcomes is non-classical and negatively correlated with the quantity of interest. All these studies focus on the consequences of measurement error for OLS estimators, and maintain the assumption that the validation data source is measured without error.

This paper expands this line of research not only by examining the consequences of non-classical measurement error for IV estimators of linear models but also by challenging the notion that validation data is measured without error. By allowing for measurement error in our validation data, we present a novel methodological approach for the validation of economic survey data, an approach that incorporates the traditional validation analysis as a special case. We show that the negative correlation between annual gross income and measurement error estimated through a traditional validation study originates from moderate measurement error in the validation data. Moreover, we show that measurement error in bounded variables is by definition non-classical, as discussed in Kane et al. (1999): We estimate the measurement error properties of length of schooling and show that when we instrument an imperfectly measured discrete variable, we obtain inflated IV estimates[1]. Our paper bridges theoretical statistics and applied data analysis, offering researchers rules of thumb for quickly gauging the consequences for OLS and IV estimators once the properties of measurement error are known.

[1]Hyslop and Imbens (2001) show that such results apply to the Optimal Prediction Error (OPE) model, which can be interpreted as a special case of our model.

We proceed in four steps. We begin by expanding the measurement error model of Bound et al. (1994), then introduce an exclusion restriction that allows instrumental variables estimation. As we do not impose any distributional assumptions on the stochastic components, the resulting model is very general. We distinguish three cases in which measurement error can bias the estimation of linear models, according to whether we use the imperfectly measured variable as a dependent variable in OLS estimation, as an explanatory variable in OLS estimation, or as an explanatory variable in IV estimation. In each case the bias depends on different characteristics of the measurement error process; for example, the variance of measurement error affects only the OLS bias from an imperfectly measured independent variable.

However, our model shows that the variances of the measurement error and of the quantity of interest and their covariance identify the expected measurement error bias in any of the three cases arising in linear models. We provide simple rules for computing the expected measurement error bias in any given linear model once those sufficient statistics are known. While measurement error in general entails an efficiency loss, in this paper we focus on the consequences of measurement error for consistency. Doing so allows us to maintain a high degree of flexibility without imposing distributional assumptions.

As is typical for validation studies in labor economics, the second step in our approach is to match survey data with validation data from third party reports. In our case, we match survey measures of gross income and length of education with the corresponding administrative measures, drawn from tax reports and civil registries. Our approach is not limited to measurement error in surveys, and it can be applied to other data sources, especially as we allow some contamination of our validation data with measurement error. However, as surveys are exposed to more sources of measurement error than third-party reports and as they are widely used to gather information from a population of interest, the properties of measurement error in surveys are often more relevant to the researcher than the properties of measurement error in other data sources. Being first-party reports, surveys are exposed to non-classical measurement error arising from non-random recall error or a flawed interview process (Biemer et al., 2004). Non-classical measurement error of this type occurs whenever low-income individuals overstate their earnings or high-income individuals understate theirs (for example, when respondents do not report temporary shocks in their annual income flows, attempting to provide information about their "normal" level of income).

Validation studies of survey data are common in labor economics. Duncan and Hill (1985) and Bound et al. (1994) use employer-provided payroll data as a validation source for ad hoc surveys on labor market outcomes, replicating the questions asked in the Panel Study of Income Dynamics (PSID). Bound and Krueger (1991) link information on labor market earnings from the Current Population Survey (CPS) to Social Security Administration (SSA) records, a censored (at the top tax bracket threshold) record of earnings that the U.S. administration uses to determine unemployment insurance eligibility and Social Security benefits. Similarly, Bricker and Engelhardt (2008) link responses from the Health and Retirement Study (HRS) with W-2 earnings records, an uncensored administrative data source from the Internal Revenue Service. These studies assume that the validation data is measured without error, and consistently find evidence against classical measurement error, especially for the labor earnings of men. We show that for gross income (in logarithms)[2] even relatively reliable validation sources such as administrative reports are contaminated with measurement error, and that when we allow for imperfect validation data we can't reject the hypothesis of classical measurement error in our data.

We match the Danish portion of the Survey of Health, Ageing and Retirement in Europe (SHARE), a longitudinal survey that collects data across nineteen European countries on individuals aged 50 or more and their spouses, with administrative records provided by the Danish authorities. By wave four, SHARE reports infor-

[2]To avoid cumbersome repetition, in the rest of the paper we simply refer to gross income. Unless otherwise noted, it is understood that we refer to the log transformation of gross income.

mation from 150,000 interviews of 86,000 persons across all waves, and is one of the most extensive surveys of the elderly population worldwide. Moreover, as the Health and Retirement Study (HRS) in the U.S. served as a role model for the development of SHARE and other sister surveys such as the English Longitudinal Study of Ageing (ELSA) and the Japanese Study of Aging and Retirement (JSTAR), the data collection mechanism and the questions asked are similar across this family of surveys. Such similarities make our specific findings on measurement errors in SHARE particularly relevant for a larger research community than SHARE users alone.

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). As a consequence, CentERdata – SHARE's data-managing institution – is able to link each selected respondent with the associated Danish CPR. Statistics Denmark then constructed a database drawn from administrative tax reports and civil registries, to which we are able to link the corresponding SHARE responses. While data confidentiality requirements are such that only the data collection and management agencies and Statistics Denmark observe CPRs, we have access to encrypted unique individual identifiers in order to conduct our analysis. Because of this unique linkage, we are able to successfully match 97% of the SHARE sample with uncensored tax reports and administrative civil registries of high data quality (see Jensen and Rasmussen (2011) for schooling and Browning and Leth-Petersen (2003) for income data), thus creating an exceptional dataset for a validation study.

While few validation studies combine an almost complete matching with uncensored administrative data, which is often assumed to exactly measure the quantity of interest, we relax the assumption of perfect validation data and acknowledge that it can also be contaminated by measurement error. For example, tax reports cannot capture income from undisclosed second jobs that might appear in survey data. More generally, measurement error in the validation data can originate from differences in the definition of a flow variable such as income, as typically occurs whenever paydays are not precisely synchronized with calendar months, or whenever capital income matures in one calendar year and capitalized in the next. If such errors exists, then the properties of measurement error in the survey cannot be identified by the simple analysis of the difference between survey and register variables. Even if survey measurement error is classical, then this difference is negatively correlated with the validation variable.

Three recent papers relax the assumption of perfect validation data: Kapteyn and Ypma (2007), Kreiner et al. (2013) and Abowd and Stinson (2013). While Kapteyn and Ypma (2007) allow validation data to be imperfect only because of mismatching, Kreiner et al. (2013) (who also use Danish register as validation data) and Abowd and Stinson (2013) do not separately identify all the sufficient statistics one needs to calculate the expected measurement error bias in both OLS and IV estimations of linear models. Therefore, as a third step in our analysis, we build two systems of moment equations that identify the sufficient statistics characterizing measurement error in gross income and length of education. We then estimate the parameters of interest via GMM. In our empirical strategy, identification is provided by exclusion restrictions or assumptions on the nature of measurement error in the administrative reports. We exploit the notion that register data are third party reports, and are thus unlikely to suffer from non-classical measurement error due to non-random response error.

In comparison to previous studies, we provide less precise information on the general measurement error structure (and thus we say little about efficiency), whereas we provide more precise information on the expected measurement error bias in the general class of linear models. Where we estimate the properties of measurement error in gross income, our paper is closest to Kapteyn and Ypma (2007) for type of data and approach. As in that study, we also question the assumption that validation data are error-free, and match survey data with Scandinavian administrative registers.

However, in addition to our interest in both income and length of schooling variables, a few key differences allow us to answer similar questions from a different angle. First, Kapteyn and Ypma (2007) separately examine earnings and pension income, while we focus on gross income. Second, while they allow for mismatching between validation and survey observations, our register-based survey sampling frame allows us to assume that we correctly match our observations, and as a robustness check we repeat our analysis excluding those matches of which we are less certain. Third, they allow for a rich error structure and impose distributional assumptions on unobservables for identification and estimate by maximum likelihood. We estimate the parameters of our model only through first and second order moment equations, thus allowing for non-normal distributions of unobservables. Our identifying assumptions are with respect to the nature of measurement error in the validation data, and are a direct consequence of the properties of the administrative data collection process.

Because earnings, capital income, and pension income are third-party reported in Denmark (by the employer, bank, and the state or pension fund respectively), we assume that measurement error in the tax reports is not correlated with true gross income. This assumption is the same imposed by Kreiner et al. (2013) on the same data, and is justified if non-classical measurement error in continuous and unbounded variables originates from response error. We do not impose any additional distributional assumption other than finite and constant mean and variance for the error components. For identification we use exclusion restrictions provided by variables that we assume to be correlated with income and uncorrelated with its measurement error in the validation data.

However, the assumption of classical measurement error in the validation data is not justified for bounded variables such as length of schooling, because measurement error depends on the bounds, and thus on the true quantity of interest. Kane et al. (1999) point out that if schooling, which is an ordered categorical variable, is measured with error, then the measurement error must be negatively correlated with the true value of schooling because of the upper and lower bounds. Consider the extreme example of measurement error in a binomial variable. This will always be negatively correlated with the true quantity of interest: if the true value is one, measurement error can only be non-positive; if zero, non-negative. Kane et al. (1999) show that measurement error in length of schooling is non-classical in both survey responses in the National Longitudinal Study of the High School Class of 1972 (NLS- 72) and their validation source, a selected subsample of transcript data from the Post-secondary Education Transcript Survey (PETS).

Moreover, for us the original source of civil registry information are third-party institutional reports for only a quarter of our sample, whereas information for the remaining 75% of the sample is drawn from the last population census in 1970, which is ultimately a different survey conducted earlier. Therefore, both SHARE survey and census reports of length of schooling can be contaminated with non-classical measurement error, possibly with different variances. However, because length of education is a stock variable that seldom changes in adulthood, and because the variable definition is the same in both the survey and the administrative data, we assume that the institutional reports we have for 25% of our sample are error free.

We then estimate the properties of measurement error in length of education through the three-way comparison between SHARE survey data, institutional reports, and census responses. We split our sample into those whose administrative information we assume to be precisely measured (institutional reports) and those for whom validation data is of similar nature to the survey measure (census reports). Measurement error is characterized through a comparison of the differences between survey and validation data in the two samples. Intuitively, we perform an external validation study within an internal validation study. This is a novel approach to the validation of economic survey data. As for the analysis of gross income, we do not impose any structural assumptions on the measurement error components of the model, except a constant mean and variance.

We find evidence of mild measurement error in our validation data for income. This causes the difference between the survey and validation variable to correlate negatively with the validation variable in our sample. Once we account for such measurement error in the validation data, we find that measurement error in SHARE is classical for annual gross income but non-classical for years of schooling. If years of schooling enters the model as an explanatory variable, this causes a bias in both the OLS and IV estimators. More specifically, the bias for IV estimators is positive, leading to a 21% overestimation of the true returns to schooling.

In the fourth and final step, we contextualize our estimates through the application of both OLS and IV estimators to a simple model of gross income returns to schooling using Danish population data. In our IV estimation, to provide an instrument for length of schooling, we use a 1958 schooling reform that affected the cost of attending post-compulsory education. This reform was used in Arendt (2005, 2008) when studying the returns of schooling on hospitalization and other health outcomes. The results of our application support our findings on the effects of measurement error for OLS and IV estimators.

The remainder of the paper is organized as follows. Section II defines the general measurement error problem and identifies the sufficient statistics of measurement error that determine the bias for OLS and IV estimation of linear models. Section III presents our survey (SHARE Denmmark) and validation (Danish administrative registers) data, their similarities and differences, and justifies the assumptions we relax or impose on measurement error in our validation data. Section IV presents our empirical strategies for estimating the measurement error properties identified in section II for gross income and length of schooling, and shows how ours differs from standard validation study strategies. Section V first shows the results of our empirical analysis and estimates the biases for the estimation of linear models using survey measures of gross income and length of schooling. Second it provides an example of a regression of returns to schooling in Danish population data in order to contextualize our results. Section VI concludes.

II. Characterizing measurement error bias

The consequences of measurement error depend on the structure of the model of interest, on whether the variable measured with error is a dependent or an explanatory one, and on the properties of the error term (Hausman, 2001). We consider measurement error which can be expressed as an additive term to the true quantity of interest, and focus on the consequences of this additive measurement error for the consistency of ordinary least squares and instrumental variables linear estimators. Linear models are widely used in empirical microeconomics for their robustness and simplicity, and are the starting point of studies in the program evaluation literature. Common evaluation methods such as difference-in-differences or regression discontinuity designs ultimately require the computation of ordinary least squares or instrumental variable estimators, which suffer from measurement error bias according to the results in this paper.

We are interested in estimating the relationship between a dependent variable y and an explanatory variable x. We start from the simple univariate model of the type

$$y = \mu_y + (x - \mu_x)\,\beta + e_y \tag{1}$$

where $\mu_x = E[x]$, $\mu_y = E[y]$ and $Var(x) = \sigma_x^2$. Additionally, we assume that we can observe an instrument z for x such that

$$Cov(z, x) \neq 0 \ \wedge \ z \perp e_y, \tag{2}$$

thus satisfying the exclusion restriction for consistency of instrumental variable estimation of the parameter β. If $e_y \perp x$, we know that both OLS and IV estimators for equation (1) are consistent.

However, we do not observe one variable (or possibly both variables) in the model but instead we observe a measure (or measures)

$$
\begin{aligned}
m_s &= m + \underbrace{\kappa_s + \rho_s\left(m - \mu_m\right) + \varepsilon_s}_{\text{measurement error}} \\
&= \mu_m + \kappa_s + \left(1 + \rho_s\right)\left(m - \mu_m\right) + \varepsilon_s, \quad E\left[\varepsilon_s^2\right] = \sigma_s^2
\end{aligned}
\tag{3}
$$

where $m \in \{y, x\}$ and the subscript $\cdot_s$ indicates that we seek to validate survey information. Measurement error consists of three components: κ_s is a constant representing non-zero average measurement error; ε_s is an independent and identically distributed error term with mean zero and variance σ_s^2; and ρ_s represents the dependence between measurement error and the quantity of interest. Therefore, we allow for measurement error characterized by arbitrary mean, variance, and correlation with the true quantity of interest, and we explicitly separate the contribution of each of those characteristics.

The classical measurement error model is a special case of the model described in equation (3), for $\rho_s = \kappa_s = 0$. In this case, measurement error $m_s - m$ has mean zero and is independent of the quantity of interest. Our model also incorporates the specific error structure in Kapteyn and Ypma (2007), who provide a very precise, more detailed description of the measurement error structure, although under strict distributional assumptions. We do not model the dependence between other covariates and measurement error in m either. Under the assumption that measurement error $m_s - m$ is independent of other covariates, we can use the univariate model without loss of generality, because the omitted variable bias is independent of the measurement error bias.

A. Measurement error bias in linear estimation

According to the model in equation (3), measurement error is classical when $\rho_s = \kappa_s = 0$, such that m_s is equal to the sum of m and the i.i.d. component ε_s. The consequences of classical measurement error for the consistency of OLS and IV estimators are straightforward to predict (Stefanski, 2000, 1985). Because in a univariate model the OLS coefficient converges to the covariance between the dependent and the independent variable, normalized by the variance of the independent variable, then if only the independent variable is measured with classical measurement error, the OLS estimator converges to

$$\frac{Cov\left(x_s, y\right)}{Var\left(x_s\right)} = \beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_s^2} = \lambda_s \beta, \tag{4}$$

where λ_s is usually referred to as the reliability ratio (Fuller, 1987). As $\sigma_s^2 \geq 0$, classical measurement error in an independent variable attenuates towards zero the OLS estimator. This attenuation bias occurs solely because the variance of x_s is larger than the variance of x, thereby leading to an incorrect normalization.

However, because ε_s is i.i.d., classical measurement error does not affect the consistency of IV estimation of β by instrumenting x_s with z, because the exclusion restriction $z \perp y - \beta\left(x_s - \mu_x\right)$ holds. Neither does it affect the consistency of OLS or IV estimators when the mismeasured variable is y. The reason is that ε_s simply adds to the unobservable variation in the dependent variable and affects only the efficiency of the estimators. The fact that the IV estimator does not suffer from attenuation bias from classical measurement error, while the OLS estimator is attenuated, has been often cited as an explanation for IV estimates usually being larger than their OLS counterparts, even when we expect omitted variable bias to go in the opposite direction. Such findings are common in the labor economics literature, especially in studies estimating earnings returns to schooling (Card, 2001).

The properties of attenuation bias in OLS estimators and consistency of IV estimators for β break down when ρ_s is not equal to zero. When measurement error is non-classical the OLS estimator $\widehat{\beta}^{OLS}$ converges to

$$E\left(\widehat{\beta}^{OLS}\right) = \frac{Cov\left(x_s, y\right)}{Var\left(x_s\right)} = \beta \frac{\left(1+\rho_s\right)\sigma_x^2}{\left(1+\rho_s\right)^2 \sigma_m^2 + \sigma_s^2}. \tag{5}$$

The measurement error bias of the OLS estimator when the independent variable is contaminated with non-classical measurement error is still multiplicative, but it is not necessarily smaller than unity for negative ρ_s and small enough σ_s^2.

At the same time, the exclusion restriction for the instrument z, defined in equation (2) does not hold for non-zero ρ_s. If x_s is contaminated with non-classical measurement error, then the exclusion restriction $z \perp y - \beta(x_s - \mu_x)$ for consistency of IV estimators does not hold. Substituting for x_s in the exclusion restriction, we can write

$$\mu_y - \beta(\kappa_s + \varepsilon_s) - \beta\rho_s(x - \mu_x) + e_y \not\perp z \tag{6}$$

as $Cov(x, z) \neq 0$ from (2). In particular, the IV estimator of β, where z is an instrument for x_s, converges to

$$E\left(\widehat{\beta}^{IV}\right) = \beta\frac{1}{1+\rho_s}. \tag{7}$$

For negative ρ_s, equation (7) implies that the IV estimator on average overestimates the coefficient of interest β, even though z is a perfect instrument for x_s. The effect of non-classical measurement error on the consistency of IV estimators is often overlooked in the empirical microeconomics literature, and is especially relevant for bounded and discrete variables. Kane et al. (1999) stress that, as length of education is typically an ordered categorical variable, measurement error in schooling is almost by definition non-classical and negatively correlated with the true value of the variable, and thus IV estimates based on the contaminated variable on average return inflated estimates of the returns to schooling.

That bounded variables tend to be measured with non-classical error, negatively correlated with the true quantity of interest, is best illustrated by considering an example. We are interested in estimating the model in (1), where x follows a Bernoulli distribution with probability p. This situation is common in the program evaluation literature, where the interest often lies in the effect of a discrete treatment variable. However, assume that we observe x with probability π; with probability $1 - \pi$ the respondent misunderstands the question and gives the wrong answer $1 - x$. Thus, we can write the observed x_s as

$$\begin{aligned} x_s &= \pi x + (1-\pi)(1-x) \\ &= \underbrace{p}_{\mu_x} + \underbrace{(1-\pi)(1-2p)}_{\kappa_s} + \underbrace{(x-p)}_{(x-\mu_x)}\underbrace{(2\pi-1)}_{(1+\rho_s)} \end{aligned} \tag{8}$$

as p is equal to μ_x. Equation (8) rewrites x_s according to the notation in (3), and shows how measurement error in a discrete, bounded variable is non-classical for non-zero probability of error π. The coefficient ρ_s for discrete variables is negative and equal to $-2(1-\pi)$. Thus by applying equation (7) we know that even if we had the perfect instrument z for x_s, for an error probability of 10% the IV estimator $\widehat{\beta}^{IV}$ converges to an estimate of β inflated by 25%.

Similarly, non-classical measurement error biases OLS and IV estimators when the dependent variable is contaminated with error. Substituting y_s in (1), we see that the OLS and IV estimators of β when y_s is contaminated with non-classical measurement error converge to

$$E\left(\widehat{\beta}^{LHS}\right) = \beta\left(1+\rho_s\right) \quad (9)$$

and will therefore suffer from a bias proportional to the linear coefficient of a regression of the measurement error on the true quantity of interest. This result replicates that in Bound et al. (1994), who estimate this bias by regressing the difference between the survey and the validation variable on the validation variable, under the assumption that the validation data are measured without error. The bias from a dependent variable contaminated with non-classical measurement error is multiplicative, applying to all linear coefficients in a multivariate model. As the bias from an independent variable measured with error is also multiplicative if both the dependent and the independent variables are measured with error, the total bias is the product of the two biases.

Equations (5), (7) and (9) show that the biases of linear estimators depend on only three parameters defined in our measurement error model: ρ_s, $\sigma_x^2 = \sigma_m^2$ and σ_s^2. In practice, such parameters are defined by the second moments of the distributions of the quantity of interest and its measurement error. Measurement error bias in linear estimation does not depend on κ_s, which only affects the estimation of the constant term in a linear regression. Once ρ_s, σ_m^2 and σ_s^2 are known, computing the expected bias for any linear model is straightforward. The next step in our analysis is to estimate these three parameters in our data.

Non-linear models do not deliver such straightforward predictions. In discrete choice models, structural coefficients depend both on the reliability ratio and on the variance of the unobservervables in the model, as both of them are affected by measurement error in an explanatory variable. Consequently, the attenuation bias in the structural coefficient due to classical measurement error in an explanatory variable is stronger than for OLS. However, given the non-linearity of partial effects, one cannot a priori sign the direction of the bias when computing partial effects, even if the measurement error is classical.

In the last three decades econometric models have been developed for drawing correct inference in discrete probability models, as in Carroll et al. (1984), and in more general classes of non-linear models (Chen et al., 2005). At the same time these models reduce the data requirements or the strength of assumption needed for obtaining unbiased estimates (Hu and Schennach, 2008). All these methods require at least some knowledge about the measurement error generation process, either through validation studies or distributional assumptions. This leads to the need for exploring and investigating the properties of general measurement error in surveys.

III. Data

We study measurement error in total income and years of schooling as recorded in the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a longitudinal survey that collects data across 19 European countries. 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. We focus on the first wave of SHARE Denmark, which, in 2004 interviewed a representative sample of residents of Denmark aged 50 and above (main respondents) and their spouses, for a total of 1707 individual respondents. Our validation source is public administrative register data, which provides official demographic information for 1 January 2004 and tax reports for the year 2003.

As we mention in the introduction, the strength of this particular validation study is that the SHARE Denmark sample has been selected using our validation data; and thus a social security number (CPR) linking survey and registry data exists in principle for all the sampled respondents. We retrieve this linkage merging information from CentERdata and Statistics Denmark, which grants us access to administrative tax and civil registry information for our sample of interest. In this process only CentERdata and Statistics Denmark observe actual CPRs.

Once the sample is defined, SHARE surveys both the sampled individuals and their spouses, if relevant. However, while the data collection agency knows the CPR of the main respondent, it does not know the CPR of the spouse. We retrieve information on the spouses through a cohabitor identification number (CNR) created by Statistics Denmark. This number is generated for adults who have the same street address at the time of the interview and who are married to each other or are in a registered partnership. Non-registered cohabiting couples share a single CNR only if they are of the opposite gender, if their age differential is less than fifteen years, and if no other adult lives at the same address. Using the CNR, we can obtain the CPR of interviewed, non-sampled spouses of the main respondents.

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 interviewed whose main respondent appears as single in the registers. In the remaining 16 observations (14 households), we cannot identify the main respondent. Of the 1670 successfully matched respondents, only 19 report a year of birth different than that in the register data, and 12 out of 19 report a year of birth within one year of that recorded in the registers. Excluding these 19 observations has negligible or no impact on our results. We are thus confident that mismatching is not an issue for our analysis and that we can ignore it as an error component.

SHARE collects a wide array of information, from health to employment status. In this paper we concentrate on years of schooling and total gross[3] income. We

[3]During the first wave of SHARE, respondents were asked about gross income. From the second wave onwards, respondents were asked about net income instead.

choose to study total gross income instead of earnings for three reasons. First, the age composition in our sample is such that a large fraction of the respondents has either zero earnings or zero pension income. Thus focusing on total income helps us increase our sample size. Second, almost all Danish residents receive earnings from both employment and pensions in the same way, i.e. both are automatically transferred to the resident's primary bank account. Therefore no reason exists for suspecting that response error patterns should vary by income source. Third, SHARE collects information on the various income sources, which are then summed together. The same happens for our validation variable, drawn from Danish tax reports. Summing across different income components helps identify income classification discrepancies, if survey and validation data classify sources of income in different ways. For example, Danish tax authorities separately record bonuses, professional fees and employment earnings, while some respondents might consider them all as employment earnings.

Our validation data for income is drawn from 2003 official tax records from SKAT, the Danish tax authority. The 2003 tax year corresponds to the period that the respondents were asked to recall during the SHARE Denmark interviews in March 2004. In Denmark, employment earnings, pensions and other forms of social assistance are third-party reported, either by the employer or the state. Capital income from stocks, bonds, or mutual funds owned through a Danish institution are also third-party reported, thus making the Danish tax register a reliable validation source (Browning and Leth-Petersen, 2003). Tax returns were posted in April, to be returned with corrections by the end of the month. Thus their timing would not affect survey recall. Tax evasion motives can only affect reports of self-employment income or income from undisclosed second jobs (Kleven et al., 2011).

Not all data collected in SHARE is first-party reported. In most interview modules, if a respondent cannot answer, information is gathered though a proxy interview, where the infromation is collected from a designated third party. In our analysis we do not distinguish between respondent and proxy interviews, as our aim is to estimate sufficient statistics characterizing measurement error bias, not to give a detailed description of the measurement error process. In terms of incidence of proxy interviews, the Danish portion of SHARE is representative of the overall SHARE sample. 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 SHARE country by aggregated first-party response rate in both modules (96% in the demographic module and 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 bracketed response categories starts. Given this information, SHARE provides multiple imputations for each source of the respondent's income, if unknown (for details on the imputation procedure, see Christelis, 2011). Because of this imputation method and of the way the income variable is constructed, often only a small portion of a respondent's total income is imputed. While 35% of the matched observations have at least some imputed income, only a quarter of their income is imputed on average (the unconditional proportion of imputed income is roughly 9%).

One of the strengths in our study is the almost complete match of the survey respondents with the registers. Thus, to not introduce selection on unobservables and to maintain a sample size as large as possible, in the analysis on income measurement error we aggregate multiple imputations by respondent, and use their average as if it were a non-imputed response. Therefore, standard errors of our estimators will tend to be downward biased, and our tests more liberal. Alternatively, if we were to correct confidence intervals for multiple imputations, we would be more likely to accept the hypothesis that measurement error in income is independent of the true values. Furthermore, we show in the appendix the analysis on income for the selected sample of individuals for which less than 10% of gross income has been imputed. When we consider the selected sample, the results of our preferred specification and our main conclusions do not change.

The SHARE questionnaire asks for the highest level of education attained. SHARE provides the re-coded level of education according to 1997 ISCED coding and the imputed years of schooling, equal to the number of years enrollment that the education level normally requires. Information on the details of the education questions in SHARE Denmark and on how the length of education variable is constructed appears in the appendix.

Our validation source for education data are official registers used by the Danish governmant. These are based upon self-reports from a census and updates by institutional reports of qualifications. The central registration of education in Denmark began with the general population and housing census of November 9th, 1970, when all residents of Denmark had to respond using their CPR numbers. The census asked 13 housing questions and 13 people questions, three of which were about schooling. These were under the heading "Education and vocational training status". [4] Five pages of instructions were followed for the later 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 (Statistics Denmark, 1977).

After the census, information on education qualifications obtained was reported by a third party. For qualifications obtained in Denmark the institution providing the education and granting the qualification had to record it and report to the ministry of education. All such post-census information is updated monthly. In our sample, roughly a quarter of our data comes from institutional reports. These updates imply that any difference in means between the survey and the validation does not derive from people achieving higher levels of education after 1970.

Educational qualifications received abroad are not recorded in the administrative registers unless converted into an equivalent Danish degree. Only three of the

[4] The first question was about education or vocational training in progress; the second about completed schooling and the third, about completed education or vocational training. See appendix.

SHARE Denmark sample states they have a foreign qualification. For immigrants, Statistics Denmark conducted a schooling census in 1999, and has since surveyed new immigrants at two year intervals. There are 42 immigrants in the SHARE Denmark sample and we consider this source of information as self-reports similar to the 1970 census.

In sum, 75% of our sample have the original census record and 25% have an updated record. Thus the official record is mostly based on recall in 1970. We compare this with the 2004 SHARE response, 34 years later. The registers include a measure of years of schooling corresponding, as in SHARE, to the minimum number of years enrollment that the registered educational level requires. This is a minimum in the sense that it correspons to the shortest period of time required to obtain the qualification by the most direct route. We compare this measure of years of schooling with the one provided in SHARE Denmark.

IV. Identifying measurement error parameters

Our goal is to consistently estimate the characteristics of measurement error in our data for gross income and length of schooling. We observe two measures of a quantity m, m_s and m_r, from a survey and a validation source—register data—respectively. A standard assumption in validation studies is that m_r exactly measures the quantity m. Applying the notation defined in equation (3) to the register measure m_r, is equivalent to assuming $\kappa_r = \rho_r = \sigma_r^2 = 0$. If this assumption holds, then measurement error in the survey is precisely defined as the difference between the survey measure m_s and the validation measure $m_r = m$. We can then simply regress this difference on the validation measure to identify κ_r, ρ_r, σ_r^2 and σ_m^2, where the latter is the true variance of the quantity of interest and the rest are the parameters characterizing measurement error.

Most validation studies maintain the assumption that validation data are error-free. Table 1 lists some key results from three such studies, validating surveys collected over three decades. The first row reports the estimated average differences

Table 1

Studies assuming no error in the validation data source

Study	BK, 1991		BBDR, 1994		BE, 2008	
Survey	CPS		PSID-VS		HRS	
Val. source	SSA		Employer payrolls		W-2 earnings	
Year	1976	1977	1982	1986	1991	2003
$\widehat{\kappa}_s$	-	0.04**	0.007	0.003	0.059**	0.089**
λ_s	0.82	0.84	0.70	0.85	0.68	0.72
$\widehat{\rho}_s$	-0.194	-0.197	-0.172	-0.104	-0.304	-0.173
N	2924	2924	422	320	2670	635

NOTE—* $p < 0.1$, ** $p < 0.05$ for the hypothesis of $\widehat{\kappa}_s = 0$. All reported $\widehat{\rho}_s$ are significantly different than zero at the 5% confidence level. The symbols κ_s, λ_s and ρ_s refer to the notation introduced in equation (3).
SOURCES.—Bound and Krueger (1991) (BK, 1991), Bound et al. (1994) (BBDR, 1994) and Bricker and Engelhardt (2008) (BE, 2008).

between survey and validation measures for each of the validation studies, with the exception of 1976 earnings (not reported in Bound and Krueger, 1991). According to the notation in the measurement error model in (3), these differences identify κ_s under the assumption that the validation dataset is exactly measured. Except for the PSID-VS (Panel Study of Income Dynamics Validation Study) data, validated by Bound et al. (1994), surveys tend to overestimate average earnings. However, while a non-zero κ_s biases the constant term in a linear model, it does not affect estimators of relationships between variables and thus is of limited interest for the applied economist. Moreover, the reported estimates are hardly economically significant.

The second row of Table 1 reports the reliability ratios λ_s estimated in each of these studies. As Section II shows, under the assumption of classical measurement error, the reliability ratio provides an estimate of the attenuation bias caused by classical measurement error when the mismeasured variable is an independent variable. However, the third row of table 1 shows all these studies find that the cross-sectional difference between the survey and validation variable depends negatively on the validation variable. These researchers interpret this finding as evidence of non-classical, mean-reverting response error.

Bollinger (1998), using the same dataset as in Bound and Krueger (1991), finds that the negative correlation between the cross-sectional difference of survey and validation measures originates primarily from low income individuals (according to the Social Security Administration) who report higher values in the survey. A natural question is whether such low earnings reflect the true value the econometrician is interested in, or whether the validation dataset lacks information on certain types of unreported earnings, which are instead correctly reported in the survey data (Abowd and Stinson, 2013).

The presence of such errors in the validation data can produce evidence of mean-reverting response error if the validation data is incorrectly assumed to be the true values. Assuming a simple measurement error model such as

$$\begin{aligned} m'_s &= m + \varepsilon_s \\ m'_r &= m + \varepsilon_r \end{aligned} \tag{10}$$

where ρ_s is equal to zero, but $\sigma_r^2 = E\left[\varepsilon_r^2\right]$ is not, and calculating the measurement error as

$$m'_s - m'_r = \varepsilon_s - \varepsilon_r, \tag{11}$$

we find that measurement error is negatively correlated with m'_r because of the term ε_r appearing in both variables. The coefficient from a regression of the difference between the survey and the validation measure on the validation measure will converge in probability to the reliability ratio of the validation measure λ_r minus one.

While the studies in Table 1 acknowledge the possibility that evidence of mean-reverting error can derive from measurement error in the validation data, these researchers argue that this should not be the case in the data they examine (see Bound and Krueger (1991) for a discussion). Kapteyn and Ypma (2007) challenge these arguments by using an elaborate error structure that takes into account the possibility of mismatch between administrative and survey data but maintains the assumption that the validation data is exactly measured. They find no evidence of mean-reverting measurement error in labor earnings (and weak evidence in pension income) for a sample of Swedish respondents. In a model without covariates, their estimated linear relatioships ρ_s between measurement error and the variable of interest are equal to -0.013 and -0.131 for earnings and pension income respectively once mismatching is accounted for.

As we explain in Section III, the structure of our data is such that we can ignore the possibility of mismatching in our data. First, the SHARE sample was originally selected through the CPR numbers in our validation data. Second, removing the 19 observations that report a different year of birth than that reported in the civil registry does not affect our results. However, although the Danish civil and tax registries are at least as precise as the validation datasets previously used in other validation studies, we relax the assumption that our validation data is measured withour error, and we allow for errors in the registry reports of length of schooling and gross annual income. Without this assumption the parameters characterizing measurement error in the survey are not identified by the comparison of the two measures. We supply additional conditions for identification of those parameters separately for gross income and length of schooling, according to the structure and nature of our validation data described in Section III.

A. Gross income

In the Danish tax system, earnings, capital income, and pensions are electronically third-party reported, respectively, from the employer, a financial institution, or the public administration. Moreover, income can be easily approximated as a continuous variable. Throughout the paper we assume that mean-reverting errors, and more generally correlations between measurement error and quantity of interest identified by ρ_s, are due to either response error or the nature of the data. Therefore, we assume that measurement error in the register data is not correlated with the true quantity of interest. However, we allow for additive independent errors in the validation measurement of income and non-zero average measurement error.

This structure of measurement error in the Danish tax registers is the same assumed by Kreiner et al., 2013. Because income is ultimately a flow variable, we interpret this error source not only as misreporting by the third parties but also as error originating from the difference between the period relevant for tax purposes and the period relevant for the individual decision process. For example, tax authorities report returns on investments maturing in the last months of 2003 but capitalized in 2004 as part of 2003 income. However, if we are interested in cash on hand, we consider them as 2004 income. Similar arguments apply to professional fees and, in general, income that matures and is received at different times.

For income, using the notation in equation (3) we can then write

$$\begin{aligned} m_s &= \mu_m + \kappa_s + (1+\rho_s)(m-\mu_m) + \varepsilon_s, \quad & E\left[\varepsilon_s^2\right] = \sigma_s^2 \\ m_r &= \mu_m + \kappa_r + (m-\mu_m) + \varepsilon_r & E\left[\varepsilon_r^2\right] = \sigma_r^2 \end{aligned} \tag{12}$$

where we assume $\rho_r = 0$. As shown in Section II, measurement error bias in linear models depends only on the parameters ρ_s, σ_s^2 and σ_m^2. Here we know that the covariance between m_r and m_s is equal to $\sigma_m^2(1+\rho_s)$. Thus we cannot identify the variance of m unless we know ρ_s. The OLS estimator for the relationship between m_s and m_r identifies

$$\frac{Cov(m_s, m_r)}{Var(m_r)} = (1+\rho)\frac{\sigma_x^2}{\sigma_x^2+\sigma_r^2} \tag{13}$$

because of the classical error component in m_r. Therefore, we need additional information to identify the parameters of the model. However, we know from Section II that IV estimation does not suffer from classical measurement error, which contaminates m_r. Thus, IV estimation of the linear relationship between m_s and m_r identifies $1-\rho_s$.

A suitable instrument z_m for this estimation needs only to satisfy the condition

$$Cov(z_m, m_r) \neq 0 \ \wedge \ z_m \perp \varepsilon_s, \tag{14}$$

because the unobservable component of m_s that is not explained by m is simply ε_s. In other words, any instrument that is correlated with m but not the measurement error in the survey measure except through m is a valid instrument. Because income is reported by a third party in our validation data, we consider any registered variable correlated with gross income as a valid instrument for the IV regression of survey gross income on register gross income. Under the assumption that ρ_r is equal to zero, this exclusion restriction identifies ρ_s.

Once ρ_s is identified, the second moments of m_s and m_r identify all other parameters of interest in the model, namely σ_s^2 and σ_m^2. To test for the presence of measurement error in our validation data, we also identify σ_r^2 from the variance of m_s. In other words, without imposing additional assumptions on the structure of the model or the distribution of the measurement errors, we can use generalized method of moments (GMM) to semiparametrically estimate the parameters of interest using

$$\begin{array}{ll}
1\text{: EXPV}_r & E\left[m_r - \widetilde{\mu}_r\right] = 0 \\
2\text{: EXPV}_s & E\left[m_s - \widetilde{\mu}_s\right] = 0 \\
3\text{: VAR}_r & E\left[\left(m_r - \widetilde{\mu}_r\right)^2 - \sigma_r^2 - \sigma_m^2\right] = 0 \\
4\text{: VAR}_s & E\left[\left(m_s - \widetilde{\mu}_s\right)^2 - \left(1+\rho_s\right)^2 \sigma_m^2 - \sigma_s^2\right] = 0 \\
5\text{: COV}_{sr} & E\left[\left(m_r - \widetilde{\mu}_r\right)\left(m_s - \widetilde{\mu}_s\right) - \left(1+\rho_s\right)\sigma_m^2\right] = 0 \\
6\text{: IV} & E\left[z_m\left(m_s - \left(1+\rho_s\right) m_r - \alpha\right)\right] = 0
\end{array} \tag{15}$$

as a system of moment restrictions to build the GMM criterion, where α is an auxiliary parameter that represents the constant term in the IV regression of m_r on m_s using z_m as an instrument. With one instrument for the sixth moment, the model is just-identified. However, adding more exclusion restrictions by using more than one instrument for the identification of ρ_s is straightforward.

Our method rests only on the assumptions that measurement error by third parties in our validation study is independent of the variable of interest, and that third-party measures of z_m are independent of first-party measurement error in the survey. In comparison to maximum likelihood estimation (Kapteyn and Ypma, 2007), using only second moments m_s and m_r for identification of the parameter of interest allows for flexibility in the measurement error structure and robustness to different distributions of measurement errors. Moreover, this method does not require particular data structures and repeated observations (Abowd and Stinson, 2013).

As κ_s affects only the constant term in a linear regression model, we do not separately identify the parameters κ_s and μ_s but rather their sum $\widetilde{\mu}_s$, which represents the expected value of m_s. The same holds for κ_r and μ_r. Therefore, while we allow for the average difference in measurement to be different from zero, we cannot say whether a non-zero average difference between m_s and m_r is due to average measurement error in the survey measure (κ_s), in the validation measure (κ_r), or both.

B. Length of schooling

When our variable of interest is years of schooling, we cannot use this strategy to characterize measurement error in the survey. For years of schooling, mean-reverting errors may arise not only from response error, but also from the bounded nature of the variable itself, if measurement error exists at all. Moreover, a large portion of our validation data is drawn from census self-reports, which can be contaminated with non-random response error. Thus we cannot argue that measurement error in our validation source is purely classical, or that an IV regression of m_s on m_r identies ρ_s.

However, educational qualifications received after 1970 are registered and reported directly to the Ministry of Education by the qualification granting institution. Because of this third-party centralized data collection method and because, unlike income, schooling is a categorical stock variable that seldom changes for seniors, we argue that years of schooling derived from institution reports are measured without error. We then use this information set to identify the parameters of interest, intuitively performing an internal validation study within the validation dataset.

Although we allow for differences in means, we assume that length of education has the same variance in the census sample and the institution report sample.

We define a variable c that indicates whether the source of our validation data is the 1970 census ($c = 1$) or institution reports ($c = 0$). We can then write the measurement error model for schooling in the survey and in the validation data according to the notation introduced in Section II, as

$$\begin{aligned} m_s &= (1-c)\left(\mu_{m0} + \kappa_s + (1+\rho_s)(m-\mu_{m0}) + \varepsilon_s\right) \\ &\quad + c\left(\mu_{m1} + \kappa_s + (1+\rho_s)(m-\mu_{m1}) + \varepsilon_s\right) \\ m_r &= (1-c)(m) \\ &\quad + c\left(\mu_{m1} + \kappa_r + (1+\rho_r)(m-\mu_{m1}) + \varepsilon_r\right) \end{aligned} \tag{16}$$

where $\mu_{mj} = E(m \mid c = j)$ for $j \in \{0,1\}$. We allow for dependence between c and m, and we indicate as σ^2_{m1} and σ^2_{m0} the variance of m in the census and the institution report sample respectively. The first and second moments of m_s and m_r can be rearranged as linear functions of c as

$$\begin{aligned} E[m_r] &= \mu_{m0} + c(\mu_{m1} - \mu_{m0} + \kappa_r) = \widetilde{\mu}_r \\ E[m_s] &= \mu_{m0} + \kappa_s + c(\mu_{m1} - \mu_{m0}) = \widetilde{\mu} \\ E\left[(m_r - \widetilde{\mu}_r)^2\right] &= \sigma^2_{m0} + c\left((1+\rho_r)^2\sigma^2_{m1} - \sigma^2_{m0} + \sigma^2_r\right) \\ E\left[(m_s - \widetilde{\mu}_s)^2\right] &= (1+\rho_s)^2\sigma^2_{m0} + \sigma^2_{s0} \\ &\quad + c\left((1+\rho_s)^2\left(\sigma^2_{m1} - \sigma^2_{m0}\right) + \sigma^2_{s1} - \sigma^2_{s0}\right) \\ E\left[(m_r - \widetilde{\mu}_r)(m_s - \widetilde{\mu}_s)\right] &= (1+\rho_s)\sigma^2_{m0} \\ &\quad + c(1+\rho_s)\left((1+\rho_r)\sigma^2_{m1} - \sigma^2_{m0}\right) \end{aligned} \tag{17}$$

where $\sigma^2_{sj} = E\left[\varepsilon^2_s \mid c = j\right]$ for $j \in \{0,1\}$.

Given that we observe c, this model has ten exclusion restrictions and eleven parameters, and is not identified if we allow such heterogeneity in the measurement error structure. Therefore, to estimate this model in our data, we impose at least one of the additional assumptions $\rho_r = \rho_s$ or $\sigma^2_{s0} = \sigma^2_{s1}$. The first assumption derives from the common nature of the survey and the census data, because for the census subsample we observe two responses to similar questions by the same individual at different times. That the non-classical measurement error component in the two surveys is the same is then plausible. The second assumption is implied by the stronger assumption $\varepsilon_s \perp c$, which loosely states that the source of measurement in the administrative registries is not related to the classical measurement error component in the survey. These assumption can be tested independently, not jointly.

Under at least one of these assumptions, the model is identified by the asymmetric error structure across the two samples in the two measures. Therefore, we interpret our identification strategy as an external validation study within an internal validation study. Specifically, as all the right hand sides of the equations in (17), the intercept of the variance of m_r point identifies σ^2_{m0}. Once σ^2_{m0} is known, the intercept of the covariance between m_r and m_s identifies ρ_s. Similarly, each of the coefficients in the five linear expressions of the first and second moments of m_r and m_s and their covariance identifies a parameter of the general model.

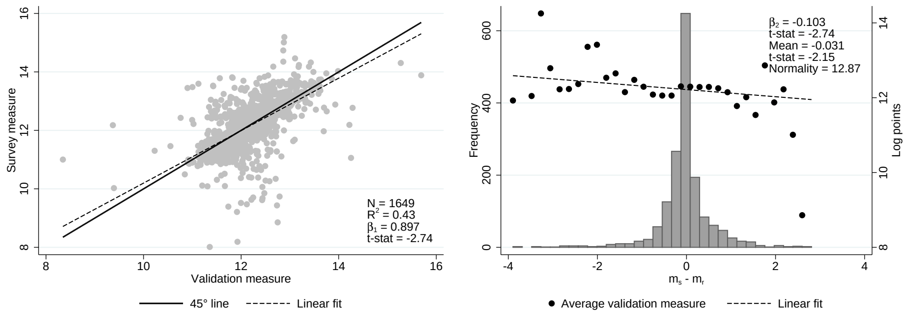
We estimate the parameters from this system of moments through GMM, which allows us to maintain a high degree of generality in that we do not impose additional distributional assumptions on the stochastic components of our model. A more structured model such as that developed by Kane et al. (1999), adapted to the specific data available, is likely to provide more precise estimates and allow correction not only for biases in the estimators of linear coefficients but also for efficiency losses. However, our framework is generally applicable, and provides sufficient information for us to apply the measurement error model presented in Section II.

V. Results

A. Gross Annual Income

We start by analyzing gross income and comparing the survey and register measures. As in the studies presented in Table 1, we initially assume that the validation data represents the true value of our variable of interest. We then construct our first measure of the error as the difference between the survey and the register values, $m_s - m_r$.

Both measurements are in Danish Kroner (DKK) in the original datasets. As is standard practice, we exclude 21 outliers reporting zero income in either the survey (19 observations) or the validation (2 observations) data. While including these observations does not change the broad conclusions of the paper, it greatly increases



NOTE.—In the left pane, which shows the scatterplot of survey versus register measure and their linear relationship, we report from top to bottom the sample size, the R^2 and β_1 for the linear regression and the t statistic from a t-test with H_0: $\beta_1 = 1$. The right pane shows the histogram of the difference between the two measures, the average register value for each histogram bin, and the linear regression line between $m_s - m_r$ and m_r. We report from top to bottom the associated β_2, the t statistic from a t-test with H_0: $\beta_2 = 0$, the average difference between measures, the t statistic from a t-test with H_0: $E[m_r] = 0$ and the z-statistic from a Shapiro-Wilk normality test.

FIG. 1.—Gross income measurement error, assuming $m_r = m$

the estimates of the measurement error variances and the standard errors of our estimates. This selection reduces our sample to 1649 observations (96.6% of the sample interviewed in the first wave of SHARE Denmark). Our survey variable has an average of 12.19 log points, and a standard deviation of 0.772; our register variable, an average of 12.22 log points and a standard deviation of 0.564. These standard deviations imply, under the hypothesis of $m_r = m$ and classical measurement error in m_s, a reliability ratio in the survey λ_s of 53.4%.

Figure 1 shows the relationship between the two measures and a first analysis of the measurement error, constructed according to our assumption that our validation dataset is error-free ($m_r = m$). The left pane of Figure 1 shows a scatterplot of the data, where the vertical and horizontal axes represent the survey measurement and the register measurement of income respectively. Despite a considerable amount of noise, the data are scattered around the 45°line, shown in solid black. However, the dashed linear prediction line shows that β_1, the estimated OLS coefficient for x_r, is equal to 0.897, which is statistically different from unity.

The scatterplot representation further clarifies the intuition behind the consequences of measurement error in the validation variable. As Section II shows, OLS estimators suffer from attenuation bias if the independent variable is measured with error. Thus a coefficient of 0.897 might result from a negative correlation between the survey measurement error and the true variable, a validation measurement error—with variance equal to 11.5% of the true variable variance—or a combination of the two. Most of the negative correlation between the difference in measures and the register measure of income is due to the tails of the register income distribution, confirming the findings in Bollinger (1998).

Maintaining the assumption of no measurement error in the validation dataset, the right pane in Figure 1 shows a histogram of the survey measurement error, defined as the difference between the survey and register income measures. The difference in measures has zero median and a negative mean of -0.03 log points, which, while being significantly different than zero (with a t-statistic of -2.15), is small in economic terms—amounting to 1977 DKK, or roughly $350 a year on average. We reject the hypothesis of a normal distribution through a Shapiro-Wilk normality test, with a z-statistic of over 12.

The black dots plot the average value of income as measured in the register for each histogram bin. We use these values, frequency weighted, to show with a dashed line the negative relationship between survey measurement error and register measurement. This regression of register income on the difference in measurements mirrors the OLS regression fitted in the left pane. The β_2 coefficient estimated in the right pane is equal to β_1, the coefficient estimated in the left pane, minus one.

The scatterplot suggests that a disproportionate amount of the negative correlation found in the data is due to outliers in the difference in measurements distribution. Such a finding is common in the measurement error literature. For example, Bound et al. (1994) find that when they omit a few outliers in terms of measurement error from the sample, the reliability ratio for 1986 earnings increases from 0.698 to 0.793, and the negative correlation between difference in measurements and validation measure of earnings decreases in magnitude from -0.17 to -0.04. Usually

Table 2

Income measurement error dependence, under the assumption $x_r = x$

	(1)	(2)	(3)	(4)
Register income	-0.103**	-0.119**	-0.387**	-0.125**
	(0.0375)	(0.0397)	(0.0506)	(0.0180)
Assets		0.00619	0.0128**	0.00372
		(0.00531)	(0.00488)	(0.00266)
Female		-0.0700	-0.0377	-0.0642**
		(0.0559)	(0.0529)	(0.0242)
Couple		-0.125	-0.0169	-0.0375
		(0.112)	(0.107)	(0.0507)
Female & couple		0.0228	-0.101	0.00936
		(0.0647)	(0.0638)	(0.0317)
Financial respondent			0.0527	0.0349*
			(0.0340)	(0.0203)
Age			-0.0104**	-0.00377**
			(0.00207)	(0.000911)
Labor income prop.			0.229**	0.118**
			(0.0593)	(0.0241)
Imputed proportion			-0.0394	-0.0165
			(0.0665)	(0.0337)
Imputed			0.0960**	0.0282
			(0.0368)	(0.0187)
Years of schooling			0.0240**	0.0103**
			(0.00471)	(0.00234)
Observations	1649	1649	1638	1638
Adjusted R^2	0.009	0.015	0.092	

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

these outliers are only detectable in terms of measurement error difference. Because they could not be detected as outliers without a validation source, such results are not self-contained.

We test whether the difference between measurements is correlated not only with the register value of education, but also with other financial and demographic characteristics. The results of this analysis appear in Table 2. The first three columns show results from OLS regressions of the difference between measurements and two sets of covariates. The first column reproduces the univariate regression graphically shown in Figure 1, and produces evidence of a mild correlation between the difference in measurements and income as measured by the administrative registers.

The second column introduces a first set of covariates observed in our validation data, including assets held on December 31, 2003 (in logarithms), gender, a couple indicator, and its interaction with gender. The couple indicator is defined as whether the respondent had a partner who had been interviewed in SHARE Denmark at the same time, and does not necessarily correspond to civil status. None of these register measured variables is significantly correlated with the difference in measurements.

The third column adds as additional covariates age, education, source of income, and survey-related measures capturing the imputation process and financial awareness of the respondent, all drawn from survey data. In the first wave of SHARE, a number of household level variables (such as food consumption and real estate value) were asked of only one household member, designated as financial respondent if the couple declared joint finances. The couple autonomously appointed the financial respondent. We use the financial respondent indicator as an indicator of financial awareness. This indicator does not appear to impact the average difference in measures. In other words, financially unaware respondents do not systematically overstate or understate their income level with respect to their partners.

In contrast, age and years of schooling, are both correlated with the difference in measurements[5]. The correlation of the difference between measurements and age is not explained solely by the source of income. We compute the proportion of income earned through labor with survey data, showing that individuals who earn most of their income through labor tend to overstate their income in the survey relative to retired individuals. A similar result holds for people whose income had been at least partially imputed. However, the proportion of imputed income does not impact the expected value of the difference in measurement.

The fourth column reports the results from a least absolute deviations (LAD) regression on the median and show that, as is often the case in the literature, the

[5]Correlation with other covariates is another, albeit relatively less studied, form on non-classical measurement error.

results on means are driven by a few outliers. Consistent with the literature, all the variables that affect the average difference in measurements have a much smaller effect on the median. The effect of register measured income on the average difference in measurements (third column) is roughly three times the size of the effect of register measured income on the average difference in measurements (fourth column).

The estimated relationships between the difference in measurements and the register measure of income are consistent with the effects found in the literature (see Table 1). Under the assumption of $m_r = m$, we find that measurement error in the survey is non-classical and negatively correlated with the true value of income. This may cause substantial biases for OLS and IV estimators, and we can easily compute the size of the bias given the results in Section II. According to the univariate analysis under the assumption $m_r = m$, we expect the OLS estimator to be 47.9% of the true parameter of interest (corresponding to a downward bias of 52.1%), and the IV estimator to be 1.115 times the true parameter of interest (corresponding to an upward bias of 11.5%) if m_s enters the model as a dependent variable. When m_s is used as an independent variable, both OLS and IV estimators are biased by an amount proportional to the linear relationship between measurement error and true value of income. According to the estimates in the univariate case, we expect a downward bias of 10.3% when estimating a linear model with m_s as a dependent variable.

We now drop the assumption of observing the true value of income in the validation dataset, and we turn to the model described in equation (12). In other words, we allow for non-classical measurement error in the survey data and classical measurement error in the validation data. We then estimate the model according to the strategy outlined in Section IV, using as instruments 2003 assets, a couple and a gender indicator and their interaction. These variables are the same as those in the second column of Table 2. Table 3 presents the results from the GMM estimation and highlights the differences in estimates with the model in which we assume

Table 3

GMM estimation of gross income measurement error model

	(1)	(2)	(3)	(4)
μ_r	12.22**	12.22**	12.22**	12.22**
	(0.0139)	(0.0139)	(0.0139)	(0.0139)
μ_s	12.19**	12.19**	12.18**	12.19**
	(0.0190)	(0.0190)	(0.0185)	(0.0188)
σ_x^2	0.318**	0.318**	0.288**	0.279**
	(0.0175)	(0.0175)	(0.0219)	(0.0220)
σ_s^2	0.277**	0.339**	0.300**	0.306**
	(0.0291)	(0.0262)	(0.0312)	(0.0314)
σ_r^2			0.0349*	0.0375**
			(0.0180)	(0.0181)
ρ		-0.103**	0.0177	0.0194
		(0.0375)	(0.0692)	(0.0710)
Observations	1649	1649	1649	1649
Hansen's J p-val.			0.0304	0.606
λ	0.534	0.483	0.490	0.477
OLS bias	0.534	0.479	0.490	0.477
IV bias		1.115	0.983	0.981
LHS bias		0.897	1.018	1.019

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

$m_r = m$.

The first two columns of Table 3 show the results for the models in which we impose a restrictive error structure as a benchmark. The first column shows estimates for a model assuming classical measurement error in the survey. In the second column we allow the measurement error in the survey to be correlated with the true value of income. In both columns we assume that the validation measure represents the true value of the income variable. The underlying model and assumptions in the second column of Table 3 are thus the same that produced the estimates in the first column of Table 2.

At the bottom of the table we use the point estimates of $\widehat{\sigma}_m^2$, $\widehat{\sigma}_s^2$ and $\widehat{\rho}_s$ to compute for each model the reliability ratio $\widehat{\lambda}_s$ and the expected OLS and IV bias when the survey measure of income enters a linear model as an independent variable (see equations (5) and (7)). Compared to the literature on earnings, in these constrained models, we estimate a much lower reliability ratio $\widehat{\lambda}_s$. This difference is likely due to the heterogenous composition of the income variable we are interested in. As capital income might be harder to correctly recall for a survey respondent, we expect the variance of the measurement error to be higher.

Columns 3 and 4 turn to the more general measurement error structure described in equation (12), in which we allow the validation data to be measured with error. The underlying models for the estimates produced in columns 1 and 2 are special cases of this general model. As discussed in Section IV, we need instruments that, while correlated with income, are otherwise independent on the measurement error component that does not depend linearly on income. In column 2 we present the results obtained with the covariates used in column 2 of Table 2 as instruments. In that model, those variables are uncorrelated with the difference in measures. The instruments used in the GMM estimation are thus gender, a couple indicator, their interaction, and the logarithm of assets at the end of 2003. All these variables are likely to be strongly correlated with income, and we assume they are independent of survey measurement error except through income. This assumption is particularly credible as we observe these variables in our validation data and independently of the measurement error process in the survey.

According to this identification strategy, we find no evidence of a correlation between measurement error in the survey and the true value of income, represented by $\widehat{\rho}_s$. At the same time, we reject at a 90% confidence level the hypothesis that the variance of the measurement error in the validation study is equal to zero, i.e., the assumption for which m_r is exactly equal to m. Thus we provide evidence that the negative correlation between $m_s - m_r$ and m_r has arisen because of the presence of mild measurement error in m_r.

As we have more instruments than parameters, we perform a Hansen test for overidentification in GMM models. With the instruments used for the estimations in column three, we reject the hypothesis of valid instruments at the 95% confidence level. If we reduce the set of instruments to only gender and the logarithm of assets, and re-estimate the model in column four, we accept the hypothesis of valid instruments. Although due to the acceptance of the test we consider the estimates in column four as the most robust, they closely resemble the estimates in column three. We find no evidence of correlation between survey measurement error and the true value of income, and we estimate a slightly higher, now significant at the 95% level, measurement error variance in the validation dataset.

When we allow for non-zero ρ_s and σ_r^2, our estimate of the reliability ratio of income in SHARE Denmark data (and thus the expected ratio between the corresponding OLS coefficient and the true parameter of the model) decreases to less than 50%. However, because we cannot reject the hypothesis of classical measurement error in the survey data, we do not expect any bias while using m_s as a dependent variable or as an independent variable in an IV estimation[6].

B. Length of schooling

For our analysis of measurement error in years of schooling we start, as with gross income, by constructing measurement error as the difference in measures $m_s - m_r$, under the assumption that our validation data exactly measures the true value of years of education. We cannot retrieve information about education level in our validation data for all respondents, especially for individuals born before 1920[7]. Therefore, we constrain our analysis to respondents born after 1920, and for whom we observe their education level in the administrative registers. We then exclude non-respondents to the SHARE Denmark questionnaire, and the single outlier reporting no education (i.e., zero years of schooling). Excluding this single observation does not have any effect on the results (the validation dataset also reports zero years of education for this particular individual, thus there is no measurement error), but it eases graphical representation. The selection process leaves us with a sample of 1538 validated observations.

Unlike measurement error in income, we expect measurement error in years

[6] Testing for the IV bias to be equal to zero is equivalent to testing for $\rho = 0$, because the IV coefficient in the presence of non-classical measurement error is simply equal to the true coefficient multiplied by $(1 + \rho)^{-1}$.

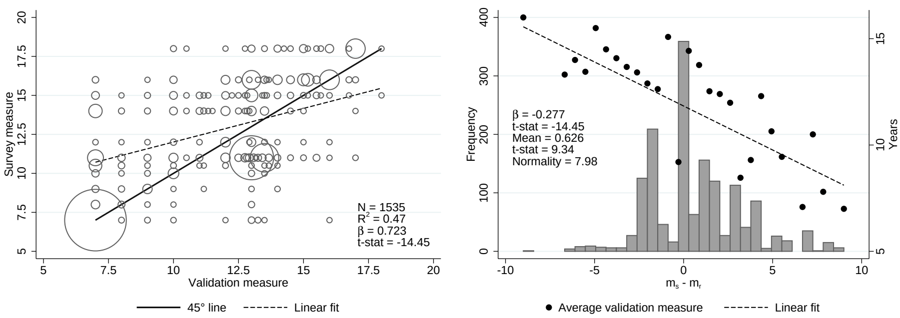
[7] The census also asked about schooling for those born 1910-19, but responses were not caded and included in the electronic record.

of schooling to be negatively correlated with its true value because of its bounded nature. People with few years of schooling can only err upwards, and vice versa. Therefore, by construction, people cannot have random response error if their true level of education is at the boundaries of the distribution. The bounded nature of the education data does not rule out a mean zero response error if respondents at both bounds of the distribution have the same likelihood of misreporting. However, this situation is clearly not the case in our data, where the average amount of years of schooling in our sample is 12.21 according to SHARE Denmark, and 11.57 according to the validation data.

Figure 2 shows the structure of our validated dataset and a first simple analysis of the difference in measurements, following the analytical structure adopted for measurement error in income. The left pane of the figure shows a scatterplot of our data, where the observations are organized with the survey measure on the vertical axis and the validation measure on the horizontal. Our validation source has more categories of education than the survey: in the validation data, length of schooling is recorded in months, allowing for finer measurement. The area of each circle in the plot is proportional to the number of observations sharing a particular combination of survey and register measurements. The largest cell corresponding to the minimum compulsory seven years of schooling in both measures has 225 observations (14.6% of the sample).

The scatterplot shows that large deviations from the validation variable are more common above the solid black 45° line, despite a large cluster of observations at approximately 13 years of schooling in the validation measure and at 11 years of schooling in the survey data. This cluster corresponds to a set of vocational degrees to which SHARE and Statistics Denmark attribute different years of schooling. That the distributions of large deviations from the one-to-one relationship are so different in the bottom and top parts of the pane suggest a strong negative correlation between the difference in measurements and the validation measure.

The linear relationship between validation and survey measures is far from one-



NOTE.—In the left pane, which shows the scatterplot of survey versus register measure and their linear relationship, we report from top to bottom the sample size, the R^2 and β_1 for the linear regression and the t statistic from a t-test with H_0: $\beta_1 = 1$. The right pane shows the histogram of the difference between the two measures, the average register value for each histogram bin, and the linear regression line between $m_s - m_r$ and m_r. We report from top to bottom the associated β_2, the t statistic from a t-test with H_0: $\beta_2 = 0$, the average difference between measures, the t statistic from a t-test with H_0: $E\left[m_r\right] = 0$ and the z-statistic from a Shapiro-Wilk normality test.

FIG. 2.—Years of schooling measurement error, assuming $m_r = m$

to-one. An OLS regression of the survey measure m_s on the validation measure m_r gives a coefficient of 0.724, which is statistically different from one. This negative correlation is more evident in the right pane of Figure 2, which shows a histogram of the difference in measures and a scatterplot of the average validation measure of years of schooling for each histogram bin. Unlike in the corresponding graph in Figure 1 for income, this negative relationship does not appear to be driven by outliers.

The histogram also shows that the average difference between the two measures is positive and large. Under the assumption that the validation dataset reflects the true value, we find that the survey overestimates the average length of schooling by 0.63 years. No such difference in median values exists. The median and modal difference between measures is zero. Interestingly, while the average length of education is higher in the survey than in the validation, the opposite is true for the medians. The median length of schooling is 13 years in the validation dataset and 11 years in the survey. Both lengths correspond to vocational educations. We cannot say whether the average difference between the measures of length of schooling results from an individual overestimation of one's own schooling or from an incorrect imputation of years of schooling after the survey responses were collected. Therefore, we cannot conclude whether it is SHARE Denmark that overestimates the average length of education, Statistics Denmark that underestimates it (either through imputation error or response error in the 1970 census), or (most likely) a combination of the two.

We now turn to the more general model described in equation (16), where both survey and validation measures can be contaminated with non-classical measurement error. As we explain in Section IV, we exploit the knowledge that part of our validation dataset comes from third-party reports, which we assume to be error-free because of the static, stock nature of length of schooling at these ages. This is the key assumption upon which the following analysis rests. While it is possible that the institution-reported information does not record education achievements obtained abroad unless converted into a Danish degree, only 3 individuals in our sample say they obtained an education abroad when intervewed in SHARE Denmark. Of those 3, only 1 has a much lower value in the register measure (13 years) than in the survey measure (18 years). Excluding those three observations from the analysis does not change our results.

Under the assumption that only institution reports exactly measure length of education, we can use the moments derived in equation (17) to estimate the properties of measurement error for the model in equation (16). We report the results from the GMM estimation in Table 4, where the first two columns serve as a benchmark. All symbols in the table refer to the notation defined in equation (16). In the first column we estimate a model assuming that measurement error in the survey is classical and that the validation data reflects the true value of education length. The second column, in which we allow for non-classical measurement error in the survey measure, reflects the analysis shown in Figure 2. As previously shown, we estimate a significant difference $\widehat{\kappa}_s$ of 0.63 years of schooling between the survey and the validation measures.

Table 4

GMM estimation of length of schooling measurement error model

	(1)	(2)	(3)	(4)	(5)
$\widehat{\mu}_m$	11.58**	11.58**			
	(0.0819)	(0.0819)			
$\widehat{\mu}_{m0}$			13.74**	13.74**	13.74**
			(0.0998)	(0.102)	(0.102)
$\widehat{\mu}_{m1}$			10.92**	10.92**	10.92**
			(0.144)	(0.144)	(0.144)
$\widehat{\kappa}_s$	0.629**	0.629**	0.566**	0.564**	0.564**
	(0.0669)	(0.0669)	(0.105)	(0.106)	(0.106)
$\widehat{\kappa}_r$			-0.0915	-0.0880	-0.0880
			(0.133)	(0.134)	(0.134)
$\widehat{\sigma}^2_m$	10.32**	10.32**			
	(0.236)	(0.236)			
$\widehat{\sigma}^2_{m0}$			4.146**	4.138**	4.138**
			(0.239)	(0.241)	(0.241)
$\widehat{\sigma}^2_{m1}$			9.974**	10.08**	9.945**
			(1.097)	(1.188)	(1.108)
$\widehat{\sigma}^2_s$	1.204**	6.107**	4.267**		4.321**
	(0.238)	(0.200)	(0.172)		(0.266)
$\widehat{\sigma}^2_{s0}$				4.321**	
				(0.266)	
$\widehat{\sigma}^2_{s1}$				4.229**	
				(0.224)	
$\widehat{\sigma}^2_r$			3.517**	3.509**	3.416**
			(0.231)	(0.233)	(0.451)
$\widehat{\rho}$		-0.276**	-0.176**	-0.180**	
		(0.0191)	(0.0423)	(0.0444)	
$\widehat{\rho}_r$					-0.169**
					(0.0522)
$\widehat{\rho}_s$					-0.180**
					(0.0444)
Observations	1538	1538	1538	1538	1538
λ	0.895	0.628			
λ, c=0			0.493	0.489	0.489
λ, c=1			0.700	0.705	0.697
OLS bias	0.895	0.649			
OLS bias, c=0			0.482	0.478	0.478
OLS bias, c=1			0.745	0.751	0.741
IV bias		1.380	1.214	1.219	1.219
LHS bias		0.724	0.824	0.820	0.820

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

At the bottom of the table we report the reliability ratio and the expected bias in linear models, given the point estimates in the top pane of the table and the results of Section II. We compute the expected bias for m_s entering a linear model as a dependent variable (LHS bias) or as an independent variable (OLS and IV bias). In the first column, because we impose ρ_s as equal to zero, the reliability ratio λ is equal to the expected OLS bias. We estimate a higher reliability ratio λ for years of schooling than we found for income. In the second colum however, under the assumption of precisely measured validation data, we estimate a stronger correlation between measurement error and true length of education.

Given the parameter estimates in the second column in Table 4, we expect an OLS estimated coefficient equal to 65% of the true parameter, or an attenuation bias of 35% (see equation (5)), and the IV estimated coefficient 1.38 times larger than the true parameter (see equation (7)). Moreover, the expected multiplicative OLS and IV bias for the coefficients of a linear model if m_s is a dependent variable equals 0.72, corresponding to an attenuation bias of 28%. Particularly relevant is the difference between the OLS and the IV measurement error bias. According to these estimates, measurement error alone can explain IV estimates of returns to schooling more than twice as large as their corresponding OLS counterparts, in the absence of omitted variable bias. However, only half of this difference is due to attenuation bias in the OLS estimates. The rest is because of amplification bias in the IV estimates due to non-classical measurement error.

In the third column, to identify parameters of interest, we estimate the complete measurement error model described in equation (16), imposing the additional assumptions

$$\begin{aligned} &\text{A)} \quad \sigma^2_{s0} = \sigma^2_{s1} = \sigma^2_s \\ &\text{B)} \quad \rho_s = \rho_r = \rho \end{aligned} \tag{18}$$

Once we allow for measurement error in the validation data, while we find significant measurement error in both the survey and the census reports, we find that the estimated $\widehat{\sigma}^2_s$ decreases in magnitude. The estimated non-classical component of measurement error in the survey decreases significantly, suggesting that measurement error in the validation data drives the strong correlation between the difference in measures and the validation measure shown in the second column of the table. However, we still find that measurement error in surveys is non-classical, causing an amplification bias of 21% for IV estimates of returns to years of schooling computed on survey data. This estimate of the IV bias is lower than that estimated by Kane et al. (1999) (approximately 34%) when using the National Educational Longitudinal Study (NELS)

Both the first and second moments of length of education depend on the source of measurement in our validation data. Institution reports are on average higher and more concentrated than survey reports. This dependence follows from the data collection method in the Danish education registers illustrated in Section III. Insitutions report education data only for qualifications obtained after 1970. Thus the insitution-reported subsample is younger and more educated than the subsample for which we observe census reports. This difference in underlying variances creates a significant difference in the two samples in terms of reliability ratio and OLS bias, as we estimate the same variance of the classical measurement error component in both subsamples. Because the relative variances of m and ε_s only affect OLS estimation of the effect of an imperfectly measured independent variable, the expected IV and LHS bias are the same in the two subsamples.

In terms of average measurement error, column three of Table 4 shows that, according to the our measurement error model, the survey overestimates average length of education by about seven months. At the same time, while census data appears to underestimate average length of education by about one month, this difference is not statistically significant. These small differences have no effect on the estimation of linear models, except for the consistency of the constant term.

We obtain these estimates by simultaneously imposing the additional assumptions stated in (18). While we cannot test these two assumptions jointly, we can relax one of them at a time and test the other independently. The fourth and fifth columns of the table, estimate our measurement error model while relaxing assumptions 18.A and 18.B, respectively. In the fourth column, we allow the variance of the classical component of survey measurement error to differ in the census and institution reports subsample. We find that while our estimate of $\widehat{\sigma}_s^2$ in the thrid column is a weighted average of the estimates of $\widehat{\sigma}_{s0}^2$ and $\widehat{\sigma}_{s1}^2$ in the fourth column, these latter estimates are not significantly different from one another. According to the results of the model under additional assumption 18.B only, the variance of the classical measurement error component in the survey is not dependent on c. None of our estimated parameters for the measurement error model change significantly when asssumption 18.A is relaxed.

In the fifth column of Table 4 we relax assumption 18.B, and maintain assumption 18.A. We thus allow for different non-classical measurement error components in the census and the survey data. Again, for column three, none of our parameter estimates changes significantly when assumption 18.B is relaxed; while the correlation between measurement error and the true value of years of schooling appears slightly lower in the census data than in the survey, this difference is insignificant. Columns four and five of Table 4 show that, when tested separately, the additional assumptions in (18) hold in our data. Our estimates of the consequences of measurement error bias in length of schooling for linear models are no different from the more parsimonious model estimated in column three.

Table 4 provides evidence that measurement error in years of schooling is non-classical, and, as expected, is negatively correlated with the true value of education length. This correlation is much smaller than results when we assume that the census data represent true education length. Using estimates from column four of Table 4, we expect a 21% amplification bias when m_s is used as an independent variable instead of m. This bias is approximately 56% of that estimated under the assumption that our validation data is precisely measured. According to the estimates shown in the third column of the table, measurement error alone explains a 63% increase between the OLS and IV estimates of returns to schooling. However, only about half of such an increase in estimates is due to attenuation bias in the OLS coefficient, the rest being due to an amplification bias in the IV estimates due to non-classical measurement error.

C. An application to the returns to schooling

We have shown that that the bias of OLS and IV estimators can be substantial when estimating returns to years of schooling. However, we have also shown that measurement error in gross income does not affect OLS or IV estimates if gross income is the dependent variable in a linear model. To contextualize our results, we consider a simple model, to identify the schooling returns on male income at older ages in Denmark. We consider both a standard OLS estimator, which gives the magnitude of the correlation between the two variables, and an IV estimator of the standard type used for assessing the direction of causality.

We use as an instrument a 1958 Danish schooling reform that affected the cost of accessing post-compulsory education in rural areas. This reform has been previously used by Arendt (2005, 2008) for studying the returns of schooling on hospitalization and, more generally, on health outcomes. Compulsory schooling lasted 7 years and from 1937 market towns (an administrative definition for towns of medium and large size) were required to offer 8th and 9th grade post-compulsory schooling. The 1958 reform required that rural areas also offer 8th and 9th grade schooling. Thus the cost of attending post compulsory school was reduced in rural areas for younger generations. In particular, the reform affected all individuals enrolled in the 7^{th} grade in the 1957/1958 school year and younger.

As we have individual-level information on the Danish population older than 45 in 2004 for month, year and place of birth, we construct our instrument as the double difference between older and younger cohorts born in urban and rural areas. We allow for differences in levels and in linear trends of schooling length between rural and urban areas, and between older and younger cohorts. This instrument is similar to those using college proximity to study returns to college education, such as Card (1993) and Kane and Rouse (1995), in that we assume that individuals respond to costs of schooling and that cost of schooling increases with distance to the institution providing education.

The reform does not have enough power to significantly affect education length in the SHARE Denmark sample. Therefore, for an example of the effect of measurement error in survey data, we compare the results from individuals for whom education data is drawn from the 1970 census with those from individuals with institution-reported education data. Because the populations of the two subgroups are different, when we compare the estimated effects of length of schooling on income as a senior we implicitly assume that the unobserved variation and the returns to schooling are homogeneous in both groups. We estimate our results on Danish males born between 1934 and 1954, i.e., ten years before and after the first cohort affected by the reform.

According to the results in Tables 3 and 4, we can form some hypotheses on the expected relationship between the coefficients in the two samples, and between OLS and IV estimators. First, as measurement error in gross income (our dependent variable here) is classical, if we assume that the data generation process is the same in the two subgroups and independent of the source of data given our observable controls, then there are only two reasons for which the OLS and IV coefficients in the two samples can differ. The first reason is omitted variable bias, i.e., endogenous selection into more schooling of higher ability individuals. We expect omitted variables to bias upward the estimates of return to schooling. Under our assumptions, omitted variable bias affects only OLS estimates of the returns of schooling in our data.

The second reason is measurement error, which affects only the census sample. The effect is different according to the type of estimator used. According to the results in Table 4 we expect an amplification bias for IV estimators and an attenuation bias for OLS estimators. Therefore, if the coefficient of interest β is the same in the two samples, we can use this information to draw three conslusion on the expected

Table 5

Returns of schooling on male gross income in Denmark, by data source

	OLS		IV	
	Register (1)	Census (2)	Register (3)	Census (4)
Years of schooling	0.0802**	0.0281**	0.0458	0.0706
	(0.00142)	(0.000274)	(0.0895)	(0.0536)
Born after reform	0.589**	0.513**	0.690**	0.408**
	(0.0233)	(0.00712)	(0.263)	(0.133)
Non-market towns	-0.0287	-0.0169**	-0.0355	0.0316
	(0.0223)	(0.00565)	(0.0245)	(0.0616)
Cohort trend	0.0561**	0.0500**	0.0583**	0.0502**
	(0.00208)	(0.000576)	(0.00600)	(0.000653)
Cohort trend, after ref.	-0.0517**	-0.0456**	-0.0599**	-0.0367**
	(0.00238)	(0.000591)	(0.0215)	(0.0112)
Cohort trend, non-market	0.00122	-0.000875*	0.00136	-0.00332
	(0.00129)	(0.000472)	(0.00103)	(0.00312)
Log assets	0.0633**	0.0572**	0.0665**	0.0518**
	(0.00132)	(0.000292)	(0.00831)	(0.00680)
Couple	0.184**	0.103**	0.191**	0.0734**
	(0.00515)	(0.00203)	(0.0180)	(0.0374)
Constant	9.993**	10.85**	10.45**	10.43**
	(0.0332)	(0.00730)	(1.179)	(0.529)
Observations	184025	467988	184025	467988
F-statistic			11.52	10.82

* $p < 0.1$, ** $p < 0.05$. Robust standard errors in parentheses. Municipality fixed effects are included in all regressions. The estimation sample includes all males in the 2004 Danish population born between 1934 and 1954.

coefficients. First, we expect the OLS estimates to be larger in the institutional reported sample than in the census sample, as measurement error bias attenuates only the latter. Second, we expect the IV estimates to be larger in the census-reported sample than in the institutional sample, as measurement error bias amplifies only the former. Third, we expect the OLS estimates to be larger than the IV estimates in the institution-reported sample, as only the former suffers from omitted variable bias and measurement error does not affect any of them. We test these propositions in Table 5.

The first two columns of Table 5 show the results from an OLS regression on the two samples. We allow for different levels and trends before and after the reform and in rural and urban areas. We additionally control for assets held at the end of 2003 (in log points), a couple indicator, and municipality fixed effects (defined as the municipality of residence in 2004). According to the results shown in Table 3, we expect an attenuation bias of the OLS coefficient. We find that an additional year of schooling is associated with an 8.7% increase in income in the census sample and with a 3.2% increase in income in the institutional-reported sample. These coefficients are significantly different, thus confirming our first proposition. If the data generation process is the same in the two samples, then the omitted variable bias is the same, and the only difference between coefficients must be due to attenuation bias in the OLS estimation in the census-reported sample.

The second two columns of Table 5 report the results for our IV regressions on the two samples. According to the F-statistics at the bottom of the table, the instrument is barely strong enough to provide reliable estimates of the effect of schooling on senior male income in both samples. Removing some controls from the model (chiefly the municipality fixed effects) improves the power of the instrument but does not change the results (see appendix). The standard errors of the estimators increase considerably, and our estimates are not statistically different from zero. Because of the imprecision in our estimates, we can neither reject nor confirm our second and third propositions.

Our results suggest that the relative magnitudes of the point estimates of the effect of schooling on male gross income in Denmark reverses in the IV regressions. Our IV point estimate of the effect of an additional year of schooling in the institutional reports sample is equal to 4.6%, while our point estimate for the census-based sample is 7.1%. Moreover, our IV point estimate in the institutional reports sample decreases compared to our OLS estimate, suggesting an omitted variable bias in the expected direction. However, none of these point estimate differences are significant.

Table 5 presents a clear, viable way of testing the implications of our results for the estimation of linear models under measurement error contamination. While we are able to only partially test our propositions because of the relative weakness of our instrument, we provide a structure that can easily be applied to different data sources and instruments. Overall, we are unable to find evidence that contradicts our results in the Danish population data.

VI. Conclusions

Measurement error is pervasive in both surveys and adminstrative datasets. Yet most validation studies assume no measurement error in their validation source. A standard treatment of measurement error assumes it to be classical for an explanatory variable, leading to attenuation bias in OLS estimates that can be corrected by IV. We show that if the measurement error in an independent variable is non-classical and is correlated with the true value of the quantity of interest, even if a perfect instrument exists for the quantity of interest, the IV estimator will be biased. This bias corresponds to an attenuation bias only if the correlation between measurement error and the quantity of interest is positive. But if correlation is negative, the IV estimator will tend to overestimate the magnitude of the coefficient of interest. Similar biases, but of opposite direction, exist for OLS and IV estimation of linear models when the dependent variable is contaminated with non-classical measurement error. The OLS and IV coefficient of linear models in this case are underestimated if the correlation between measurement error and quantity of interest is negative, and overestimated otherwise.

We show that ignoring errors in the validation data leads to incorrectly inferring non-classical measurement error in a validated variable. We build a framework that allows us to estimate the sufficient statistics determining measurement error bias in IV and OLS estimators of linear models through imperfectly measured validation data for length of schooling and gross income. Contrary to most validation studies, we find evidence of classical measurement error in gross income once we allow for imperfect validation measures. The substantial noise in the survey measure of gross income does not cause bias when income is the dependent variable or when using IV estimators. As income is usually the outcome of interest, we conclude that measurement error in a survey like SHARE only affects the efficiency of linear estimators.

We acknowledge that because years of schooling is a bounded variable, its measurement error is likely to be non-classical, independently of the response error generation process. As expected, we find that measurement error in years of schooling is negatively correlated with the true value of length of schooling. However, accounting for errors in our validation dataset reduces our estimates of the resulting IV amplification bias (from 38% to about 21%). We show that while measurement error alone can account for a 63% increase between an OLS and an IV estimate on the same sample in the absence of omitted variable bias, only half of this increase is due to the attenuation bias on the OLS estimates.

The general and flexible approach that we develop can be tailored according to the type of validation data available for assessing measurement error for other variables in other contexts. While our approach does not provide a sufficiently detailed description of the measurement error generation process to correct for efficiency losses, it allows for more precise identification of the characteristics of measurement error determining bias in OLS and IV estimation of linear models. Our approach extends the classical validation study techniques in the labor economics literature, because the traditional approach assuming that the validation data are exactly measured is a special case of our model. Moreover, our framework can test this assumption. A useful extension would be to consider the consequences of other types of non-classical measurement error, e.g. correlation of measurement error with other variables.

References

Abowd, J. M. and M. H. Stinson (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics 95*, no. 5:1451–1467.

Arendt, J. N. (2005). Does education cause better health? a panel data analysis using school reforms for identification. *Economics of Education Review 24*, no. 2:149 – 160.

Arendt, J. N. (2008). In sickness and in health—till education do us part: Education effects on hospitalization. *Economics of Education Review 27*, no. 2:161 – 172.

Barron, J. M., M. C. Berger, and D. A. Black (1997). How well do we measure training? *Journal of Labor Economics 15*, no. 3:507–528.

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.

Bollinger, C. R. (1998). Measurement error in the current population survey: A nonparametric look. *Journal of Labor Economics 16*, no. 3:576–594.

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*, no. 3:345–368.

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*, no. 1:1–24.

Bricker, J. and G. V. Engelhardt (2008). Measurement error in earnings data in the health and retirement study. *Journal of Economic & Social Measurement 33*, no. 1:39 – 61.

Browning, M. and S. Leth-Petersen (2003). Imputing consumption from income and wealth information*. *The Economic Journal 113*, no. 488:F282–F301.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Working Paper 4483, National Bureau of Economic Research.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*, no. 5:1127–1160.

Carroll, R. J., C. H. Spiegelman, K. K. G. Lan, K. T. Bailey, and R. D. Abbott (1984). On errors-in-variables for binary regression models. *Biometrika 71*, no. 1:19–25.

Chen, X., H. Hong, and E. Tamer (2005). Measurement error models with auxiliary data. *The Review of Economic Studies 72*, no. 2:343–366.

Christelis, D. (2011). Imputation of missing data in waves 1 and 2 of share. Working Paper Series 01-2011, SHARE - Survey of Health, Ageing and Retirement in Europe.

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*, no. 4:508–532.

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

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *The Journal of Economic Perspectives 15*, no. 4:57–67.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*, no. 1:195–216.

Hyslop, D. R. and G. W. Imbens (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics 19*, no. 4:475–481.

Jensen, V. M. and A. W. Rasmussen (2011). Danish education registers. *Scandinavian Journal of Public Health 39*, no. 7 suppl:91–94.

Kane, T. J. and C. E. Rouse (1995). Labor-market returns to two- and four-year college. *The American Economic Review 85*, no. 3:600–614.

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

Kapteyn, A. and J. Y. Ypma (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics 25*, no. 3:513–551.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*, no. 3:651–692.

Kreiner, C. T., D. D. Lassen, and S. Leth-Petersen (2013). Measuring the accuracy of survey responses using administrative register data: Evidence from Denmark. In J. S. Christopher Carroll, Thomas Crossley (Ed.), *Improving the Measurement of Household Consumption Expenditures*, NBER Book Series Studies in Income and Wealth. NBER. Forthcoming.

Mellow, W. and H. Sider (1983). Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics 1*, no. 4:331–344.

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. (1985). The effects of measurement error on parameter estimation. *Biometrika 72*, no. 3:583–592.

Stefanski, L. A. (2000). Measurement error models. *Journal of the American Statistical Association 95*, no. 452:1353–1358.

Appendix

A. Survey and census questions

In this section we report the questions originally asked in the first wave of SHARE and how the variables for education and gross household income were constructed.

Table 6

SHARE education variable; wording of original questions

Question	Danish	English	ISCED	ISCEDY
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	7
	8. klasse	8^{th} grade	2	8
	9. klasse	9^{th} grade	2	9
	10. klasse, realeksamen	10^{th} grade	2	10
	Studentereksamen eller HF	Gymnasium	3	12
	HH, HG, HHX, HTX	Technical secondary	3	12
DN012_	Please look at card 3. Which degrees of higher education or vocational training do you have?			
	Specialarbejderuddannelse	Vocational	3	10.5
	Lærlinge eller EFG-uddannelse	Vocational	3	11/12*
	Anden faglig uddann. > 12 mdr.	Vocational > 12 months	3	14
	Kort videregående uddannelse	Higher education (<3y)	5	15
	Mellemlang videregående uddannelse	Higher education (3-4y)	5	16
	Lang videregående uddannelse	Higher education (>4y)	5	18

* The imputed years of education (ISCEDY) are 11 if the answer to the question DN010_ is Gymnasium or lower, 12 otherwise. Both question contemplate the option "None". Only one respondent reports "None" in DN010_, and ISCEDY is coded to 0. See `www.share-project.org` for more information.

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. Education in SHARE

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, the 1997 ISCED code that derives from the answers and the associated imputed standard years of schooling.

B. Gross income in SHARE

Gross income is the sum of a list of variables, each capturing a different portion of the income process of an individual, each asked separately to the financial respondent(s). Table 7 shows the variables that form gross income and their source within

Table 7

SHARE gross 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

See Christelis (2011) for more information

the questionnaires. All questions refer to previous year income.

C. 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 vocational 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.

B. Robustness checks

In this section we report some robustness checks for the results shown in Section V. In Table 8 we replicate the results of Table 3, excluding respondents who had more than 10% of income imputed. We find no evidence of correlation between survey measurement error and gross income.

Table 8

GMM estimation of gross income measurement error model, excluding respondent with more than 10% imputed income

	(1)	(2)	(3)	(4)
μ_r	12.23**	12.23**	12.23**	12.23**
	(0.0145)	(0.0145)	(0.0145)	(0.0145)
μ_s	12.19**	12.19**	12.18**	12.19**
	(0.0202)	(0.0202)	(0.0197)	(0.0200)
σ_x^2	0.298**	0.298**	0.287**	0.284**
	(0.0159)	(0.0159)	(0.0221)	(0.0224)
σ_s^2	0.281**	0.295**	0.265**	0.283**
	(0.0295)	(0.0273)	(0.0335)	(0.0342)
σ_r^2			0.0187	0.0138
			(0.0177)	(0.0181)
ρ		-0.0242	0.0400	0.0245
		(0.0379)	(0.0716)	(0.0726)
Observations	1412	1412	1412	1412
Hansen's J p-val.			0.0132	0.666
λ	0.515	0.502	0.520	0.501
OLS bias	0.515	0.502	0.519	0.501
IV bias		1.025	0.962	0.976
LHS bias		0.976	1.040	1.025

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

Similarly, in table 9 we replicate our results for measurement error in gross income for individuals whose earnings, as measured in the survey, exceed 50% of total gross income. We find that for those there is substantial measurement error in our validation data. Assuming that our validation data are exactly measured leads to estimating a large negative correlation between measurement error and quantity of interest. When we allow for classical measurement error in our validation data, we cannot reject the hypothesis of classical measurement error.

Table 9

GMM estimation of gross income measurement error model, more than 50% of income as earnings

	(1)	(2)	(3)	(4)
μ_r	12.60**	12.60**	12.61**	12.60**
	(0.0157)	(0.0157)	(0.0157)	(0.0157)
μ_s	12.62**	12.62**	12.61**	12.61**
	(0.0202)	(0.0202)	(0.0201)	(0.0201)
σ_x^2	0.157**	0.157**	0.120**	0.110**
	(0.0181)	(0.0181)	(0.0150)	(0.0153)
σ_s^2	0.103**	0.159**	0.118**	0.125**
	(0.0346)	(0.0295)	(0.0287)	(0.0287)
σ_r^2			0.0443**	0.0387**
			(0.0100)	(0.00998)
ρ		-0.200**	0.0898	0.102
		(0.0566)	(0.0931)	(0.0976)
Observations	634	634	634	634
Hansen's J p-val.			0.00258	0.00525
λ	0.604	0.496	0.503	0.467
OLS bias	0.604	0.483	0.501	0.468
IV bias		1.249	0.918	0.908
LHS bias		0.800	1.090	1.102

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

In Table 10 we replicate the results of the third column of Table 4, the most parsimonious model, for different samples. Here we address the concern about errors in the institution reported sample for individuals obtaining qualifications abroad. The first column of Table 10 reports the results shown in the third column of Table 4 for our preferred sample as a term of comparison. The second column adds to the sample the single observation reporting zero years of education, which we exclude for graphical presentation. The third and fourth columns exclude respondents stating having received a qualification abroad and born abroad respectively. One of the respondents receiving a qualification abroad was born in Denmark. In the fifth column, we apply both sample restrictions and include the respondent declaring zero years of education. The results do not change for any of these sample selections.

Table 11 replicates the results of Table 5, but excluding municipality fixed effects

Table 10

GMM estimation of length of education measurement error model (parsimonious), different samples

	(1)	(2)	(3)	(4)	(5)
$\widehat{\mu}_{m0}$	13.74**	13.74**	13.73**	13.70**	13.69**
	(0.0998)	(0.0998)	(0.100)	(0.101)	(0.101)
$\widehat{\mu}_{m1}$	10.92**	10.91**	10.93**	10.89**	10.88**
	(0.144)	(0.144)	(0.144)	(0.146)	(0.146)
$\widehat{\kappa}_s$	0.566**	0.566**	0.555**	0.534**	0.536**
	(0.105)	(0.105)	(0.106)	(0.106)	(0.106)
$\widehat{\kappa}_r$	-0.0915	-0.0909	-0.102	-0.133	-0.131
	(0.133)	(0.133)	(0.134)	(0.135)	(0.135)
$\widehat{\sigma}^2_{m0}$	4.146**	4.146**	4.135**	4.077**	4.075**
	(0.239)	(0.239)	(0.240)	(0.240)	(0.240)
$\widehat{\sigma}^2_{m1}$	9.974**	10.13**	9.990**	9.732**	9.870**
	(1.097)	(1.122)	(1.105)	(1.076)	(1.101)
$\widehat{\sigma}^2_s$	4.267**	4.269**	4.260**	4.227**	4.234**
	(0.172)	(0.172)	(0.171)	(0.171)	(0.171)
$\widehat{\sigma}^2_r$	3.517**	3.508**	3.516**	3.435**	3.426**
	(0.231)	(0.231)	(0.231)	(0.228)	(0.228)
$\widehat{\rho}$	-0.176**	-0.176**	-0.177**	-0.168**	-0.167**
	(0.0423)	(0.0423)	(0.0425)	(0.0428)	(0.0429)
Observations	1538	1539	1535	1489	1489
λ, c=0	0.493	0.493	0.493	0.491	0.490
λ, c=1	0.700	0.704	0.701	0.697	0.700
OLS bias, c=0	0.482	0.482	0.482	0.481	0.481
OLS bias, c=1	0.745	0.749	0.746	0.738	0.742
IV bias	1.214	1.214	1.215	1.201	1.201
LHS bias	0.824	0.824	0.823	0.832	0.833

NOTE—* $p < 0.1$, ** $p < 0.05$. Standard errors in parentheses.

from the set of controls. The instrument has more explanatory power, and the estimate of the returns to education on gross income is more precise. The conclusions drawn from the table do not change.

Similarly, Table 12 shows the estimated returns to schooling on gross income for a more parsimonius model, where we do not control for assets, marital status or the change in the post-cohort trends. In this model, changes in growth rate of length of schooling in the treated group factor in the estimated average treatment effect. We find a large, significant effect of an additional year of schooling in the IV estimate for the census sample, which we expect to be inflated. We do not find a significant effect of length of schooling in gross income for the insitution reported sample. The large standard error of this latter estimate means that the IV estimates in the two

Table 11

Returns to schooling estimation, no municipality fixed effects

	OLS		IV	
	Register (1)	Census (2)	Register (3)	Census (4)
Years of schooling	0.0865**	0.0315**	0.0472	0.0666
	(0.000825)	(0.000276)	(0.0907)	(0.0489)
Born after reform	101.3**	89.23**	120.9**	74.04**
	(3.710)	(1.167)	(45.40)	(21.25)
Non-market towns	-3.058	1.280	-3.613	5.726
	(1.966)	(0.875)	(2.465)	(6.285)
Cohort trend	0.0560**	0.0498**	0.0586**	0.0501**
	(0.00202)	(0.000555)	(0.00616)	(0.000727)
Cohort trend, after ref.	-0.0521**	-0.0459**	-0.0622**	-0.0381**
	(0.00191)	(0.000600)	(0.0233)	(0.0109)
Cohort trend, non-market	0.00157	-0.000669	0.00185	-0.00295
	(0.00101)	(0.000450)	(0.00126)	(0.00322)
Log assets	0.0645**	0.0575**	0.0682**	0.0535**
	(0.000729)	(0.000365)	(0.00861)	(0.00564)
Couple	0.184**	0.104**	0.189**	0.0815**
	(0.00367)	(0.00204)	(0.0105)	(0.0308)
Constant	-98.41**	-85.43**	-102.8**	-86.41**
	(3.930)	(1.078)	(10.79)	(1.765)
Observations	184025	467988	184025	467988
F-statistic			16.53	14.36

NOTE—* $p < 0.1$, ** $p < 0.05$. Robust standard errors in parentheses. The estimation sample includes all males in the 2004 Danish population born between 1934 and 1954.

samples do not significantly differ.

Table 12

Returns to schooling estimation by sample, no post-reform trends

	OLS		IV	
	Register (1)	Census (2)	Register (3)	Census (4)
Years of schooling	0.106**	0.0426**	0.0273	0.136**
	(0.000858)	(0.000283)	(0.109)	(0.0361)
Born after reform	0.179**	0.0176**	0.260**	-0.00236
	(0.00799)	(0.00338)	(0.111)	(0.00875)
Non-market towns	-5.132**	1.500	-7.122*	14.21**
	(2.084)	(0.929)	(3.682)	(5.053)
Cohort trend	0.00417**	0.0243**	-0.0104	0.0363**
	(0.00103)	(0.000519)	(0.0203)	(0.00471)
Cohort trend, non-market	0.00263**	-0.000781	0.00365*	-0.00729**
	(0.00107)	(0.000478)	(0.00188)	(0.00259)
Constant	2.917	-35.31**	32.47	-59.67**
	(2.011)	(1.008)	(41.00)	(9.556)
Observations	184025	467988	184025	467988
F-stat			13.30	32.71

NOTE—* $p < 0.1$, ** $p < 0.05$. Robust standard errors in parentheses. The estimation sample includes all males in the 2004 Danish population born between 1934 and 1954.