Volume Title: Improving the Measurement of Consumer Expenditures

Volume Author/Editor: Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus, editors

Series: Studies in Income and Wealth, volume 74

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-12665-X, 978-0-226-12665-4

Volume URL: http://www.nber.org/books/carr11-1

Conference Date: December 2-3, 2011

Publication Date: May 2015

Chapter Title: Measuring the Accuracy of Survey Responses Using Administrative Register Data: Evidence from Denmark

Chapter Author(s): Claus Thustrup Kreiner, David Dreyer Lassen, Søren Leth-Petersen

Chapter URL: http://www.nber.org/chapters/c12663

Chapter pages in book: (p. 289 – 307)

10

Measuring the Accuracy of Survey Responses Using Administrative Register Data
Evidence from Denmark

Claus Thustrup Kreiner, David Dreyer Lassen, and Søren Leth-Petersen

10.1 Introduction

Danish administrative register data can readily be combined at the person level with survey data. This makes it possible to compare survey-based measures directly with corresponding measures based on information from administrative registers. Because register information is collected by third-party automatic reporting and completely independently from the survey collection, we believe this provides an inexpensive and powerful way to validate survey measures.

The objective of this chapter is to illustrate how Danish register and survey data may be combined at the person or household level and used for validating measures collected by survey, and we illustrate the potential of this methodology by two examples. In the first example we use administrative records about disposable income and wealth to validate the total expenditure measure collected in the Danish Family Expenditure Survey. In the second example we use third-party-reported information about gross personal income from the income tax register to validate a survey measure of gross personal income.

Claus Thustrup Kreiner is professor of economics at the University of Copenhagen. David Dreyer Lassen is professor of economics at the University of Copenhagen. Søren Leth-Petersen is professor of economics at the University of Copenhagen.

This chapter was prepared for the conference in Washington, DC, December 2–3, 2011, on Improving the Measurement of Consumer Expenditures, sponsored by the Conference on Research in Income and Wealth and the National Bureau of Economic Research, with support from the Centre for Microdata Methods and Practice. The authors would like to acknowledge that the Danish Social Science Research Council provided support for the work presented in this chapter. We are also thankful to Amalie Sofie Jensen and Gregers Nytoft Rasmussen for competent research assistance. For acknowledgments, sources of research support, and disclosure of the author's or authors' material financial relationships, if any, please see http://www.nber.org/chapters/c12663.ack.

Validating total expenditure requires assumptions as the register measure of total expenditure is itself ridden with error. The most important assumption is that the errors of the two measures are uncorrelated. This is not likely to be a restrictive assumption since the data are collected from completely independent sources. We find that total expenditure from the expenditure survey is mean unbiased, but noisy.

In the second example, where we validate survey information about gross income, the register measure of gross personal income is collected entirely from third-party automatically reported information. This is thought to be very close to the "truth" and the validation exercise therefore relies on few assumptions. We find that survey answers are noisy and mean biased. We also compare our results about the magnitude of income mismeasurement with the results from Bound et al. (1994). They compared survey responses with payroll data from a single US manufacturing company where workers are homogenous and received regular and well-defined payments. Consistent with our broader income measure and broader sample, we find larger errors than the study by Bound and colleagues.

The methodology presented in this chapter is simple but powerful. In the Danish context, it is possible to match survey and register data for any subsample of the population and it can be done at relatively low costs. In this way Denmark can be thought of as a "laboratory" for very detailed and focused validation studies to investigate the impact of survey methodology on the accuracy of survey responses so as to optimize the survey methodology across different groups and balancing this with survey costs.[1] It is possible for international researchers or statistical agencies to conduct new studies on Danish data through collaboration with researchers based in Denmark or directly with Statistics Denmark if necessary funding is available.

The next section outlines the Danish institutional setup facilitating the collection of administrative register data and the merging of register and survey records. Section 10.3 outlines the analytical framework that we use to asses the importance of measurement error in the survey data. Section 10.4 shows how income tax records with information about income, tax payments, and wealth have been used to impute a measure of total household expenditure that is then matched at the household level to data from the Danish expenditure survey in order to check how well the total expenditure measure in the survey matches the register-based imputation. The analysis presented in that section complements the analysis presented in Browning and Leth-Petersen (2003) and is based on the same data. In section 10.5 we combine income tax records with new survey data containing a measure of

1. Reducing measurement error is the primary mission of the Gemini project. The Gemini Project Vision Document (http://www.bls.gov/cex/ovrvwgeminivision.pdf), however, also emphasizes that the CEX budget is constant and that new initiatives to reduce measurement error should be balanced with the potential negative effects on response rates.

total gross personal income to directly validate the survey measure of gross income. Section 10.6 sums up and discusses the possibilities for future validation studies based on combining Danish register and survey data.

10.2 Matching Administrative Register Data with Survey Data

All persons in Denmark are assigned a unique personal identification number (CPR). This number is used by all government institutions to store person-specific information, including information relevant for taxation such as the information contained in tax returns, but also information about car ownership, contacts to the health care system, the educational system, and about family composition and place of residence, allowing for the construction of household units. Many administrative registers, including population registers and income tax registers, are collected by Statistics Denmark, which merges them and provides access to researchers working at authorized Danish research institutions. The data are confidential, are kept on servers at Statistics Denmark, and are accessed under comprehensive security precautions. The data must be kept at the servers and only aggregated numbers such as regression coefficients can be extracted.

The register data have many outstanding features, but the features most important in this context are that they cover the entire population and contain tax records with third-party-reported information about income and wealth. In this study we shall rely on register data from the income tax registers to validate survey information about spending and income. The income tax register is collected by the tax authorities in order to calculate the amount of taxes to be paid by all persons in Denmark by the end of each calendar year. The tax authorities collect information from many sources. Most important for this study are earnings and employers' pension contributions collected directly from employers, information about transfer income from government institutions, and information about interest payments/income and the value of assets and liabilities by the end of the year collected directly from banks. A recent study by Kleven et al. (2011) conducted a large-scale randomized tax auditing experiment in collaboration with the Danish tax authorities and documents that tax evasion in Denmark is very limited, in particular among wage earners. This means that the third-party-reported income information collected by the tax authorities is of very high quality.

The tax authorities use the information for different purposes. Information about earnings and capital income is preprinted on the tax return, whereas wealth information is used to cross-check if reported income is consistent with the level of asset accumulation from one year to the next. While the tax authorities collect this information at a high level of detail corresponding to individual entries at the tax return level for income and at the account level for wealth, this information is in some cases transferred to Statistics Denmark's research database as summary variables only; for example, we observe the sum of earnings from different employers, and for some capital income subcomponents, only net income is available. In addition to covering the entire population and being based on third-party-reported information, the income tax registers also have the attractive features that income and wealth information is not top coded and that longitudinal information can be retrieved as far back as 1980 for some variables.

A crucial feature for the present purpose is that it is possible to link to survey data via the CPR number. Matching surveys with register data is done at relatively low cost; for example, the survey used in the second part of this chapter consists of forty questions, was carried out as telephone interviews, and includes 6,000 completed interviews. The sample was randomized from the population based on register data covering the entire population, and the survey data was merged on to register data after collection. The total costs were about 200,000 USD.[2]

10.3 Analytical Framework

There are several ways of summarizing the accuracy of the survey data. In this chapter we focus on the magnitude of the attenuation bias in ordinary least squares (OLS) regressions of the register measure on the survey measure. The analytical setup is a generalization of the setup presented by Bound and Krueger (1991).

Consider

$$z^S = z^* + u^S \tag{1}$$

$$z^R = z^* + u^R, \tag{2}$$

where z^S is the observed survey-based measure, z^* is the true but unobserved measure, and u^S is the survey measurement error. Correspondingly, z^R is the observed register-based measure, and u^R is the register measurement error. All variables are measured in natural logarithms.[3] This amounts to assuming that the measurement error is multiplicative in levels. Subscripts identifying that each observation of (z^*, z^R, z^S, u^R, u^S) pertains to an individual are suppressed. In the case of gross income we believe that the register-based measure is very close to the truth, while this is obviously not the case in the other example where we compare total expenditure from survey data with imputed measures from the register data.

2. A number of survey agencies are specialized in conducting surveys and linking to administrative register data. Two of those are SFI survey (http://www.sfi.dk/Default.aspx?ID=2832) and Epinion (www.epinion.dk) who have collected the survey used in example 2. Also, Statistics Denmark (www.dst.dk) conducts surveys that can subsequently be merged on to register data.

3. The analytical framework, of course, does not require that the variables are measured in logarithms.

Assume

$$\text{(A.1)} \qquad \text{cov}(z^*, u^R) = 0$$

$$\text{(A.2)} \qquad \text{cov}(u^S, u^R) = 0.$$

Assumption (A.1) assumes that the error of the register measure is uncorrelated with the true level. This assumption is not testable with the data used in this chapter, and may in some cases be a reasonable assumption, while in others it may not. For example, it could be that people with a low level of true income have different errors than people with a high level of true income because they have different cognitive skills that influence the quality of their answer or have total income consisting of different subcomponents and different complexity, or because low-level-income people have different amounts of undeclared income. Similarly, in the case of total expenditure, consumers with a high level of expenditures are likely to have total expenditure consisting of different types of consumption than consumers with a low level of expenditures, and this may give rise to different measurement errors if subcomponents of total expenditure have different errors. Because we assume that the measurement error is multiplicative in levels, we do allow for the level of errors being larger at high levels than at low levels of income/total expenditure, but this is entirely determined by the logarithmic functional form that we employ in the applications. Assumption (A.2) assumes that the error of the survey measure is uncorrelated with the error of the register measure. This seems to be a reasonable assumption in both of the examples as will be discussed in connection with each example.

Consider a regression of z^R on the true but unobserved measure z^*:

$$\text{(3)} \qquad z^R = \delta_0 + \delta_1 z^* + u^R.$$

Now substitute in the survey measure for the true measure

$$\text{(4)} \qquad z^R = \delta_0 + \delta_1 z^S + u^R - \delta_1 u^S.$$

Using assumptions (A.1) and (A.2), the probability limit of the OLS estimator of δ_1 can be written

$$\text{(5)} \qquad p \lim \widehat{\delta_1} = \delta_1 \lambda,$$

where $\lambda = \text{cov}(z^R, z^S)/\text{var}(z^S)$ is just the OLS regression of the register measure on the survey measure. The bias due to the measurement error in the survey measure is then $(1 - \lambda)$.

Maintaining assumptions (A.1) and (A.2), this expression covers the case with classical measurement error where u^S are *iid* and $\lambda = \sigma^2_{u^S}/(\sigma^2_{z^*} + \sigma^2_{u^S})$ but is not limited to this special case. In particular, the present framework is more general since it allows for cases where the errors are not *iid*.

10.4 Example 1: Total Expenditure

Total expenditure is one of the most important variables collected in expenditure surveys and this variable is central to numerous studies of demand and intertemporal consumption allocation. However, there is little evidence on the quality of the information collected in expenditure surveys. In Denmark it is possible to link the household-level information from the Danish Family Expenditure Survey to administrative income tax records, including third-party-reported information about income and wealth that can be used to impute total expenditure.

10.4.1 Data

The sample used consists of the households entering the Danish Family Expenditure Survey (DES) 1994–1996. The households in this survey have been contacted at different times of the year so that observations are distributed across the calendar year. Each household has participated in a comprehensive interview, where they have answered questions about purchases of durables within the past twelve months from the interview date. Furthermore, each household has kept a diary for two weeks, where they have kept a detailed account of all expenditures in the household. This information is scaled to obtain an expression of annual consumption.

For the households entering the DES administrative register, data are collected on income, tax payments, and wealth at the end of the year (corresponding to the survey year) together with wealth information for the previous year, and this is merged with the DES data. Total expenditure is then imputed from the income and wealth information by simply calculating $\hat{c}_t = y_t - \Delta W_t$, where y_t is disposable income and W_t is net wealth measured at the end of period t. While simple in theory, there are many details involved in implementing this and we refer to Browning and Leth-Petersen (2003) for details. For the analysis we use the same sample selection criteria as Browning and Leth-Petersen (2003). This leaves us with a sample of 3,352 observations.

10.4.2 Results

We start out by presenting in figure 10.1 the distributions of the two measures of total expenditure and their individual-level difference. The left panel shows that the distributions have modal points very close to each other and the right panel shows that at the individual level the differences are centered at zero. It is, however, also evident that there are important differences in the spread of the distributions of the two measures, with the register-based measure exhibiting larger dispersion. The way the data are constructed implies that a fair amount of noise is expected. First, the interviews are distributed across the calendar year, and this means that recall questions about durable purchases, for example, do not necessarily pertain to the calendar year. Moreover,

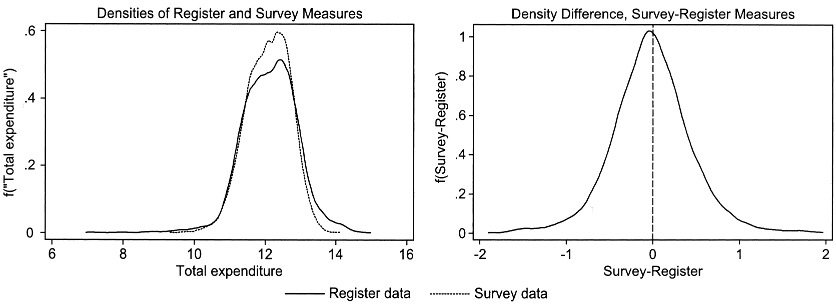

Fig. 10.1 Densities of the survey and register-based measures of total expenditure and of the individual differences

Note: The right panel includes only data in the interval –2;2. Thirty-two observations are selected away.

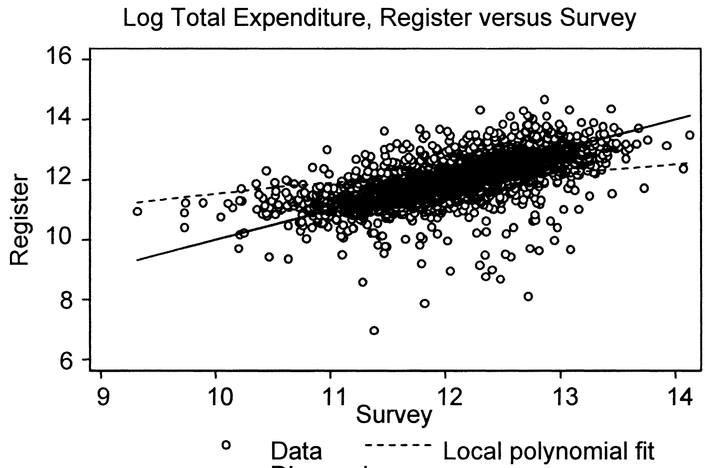

Fig. 10.2 Nonparametric regression of the register-based measure on the survey measure

Browning and Leth-Petersen (2003) show that the measurement error in the imputed measure is related to capital gains on wealth components used in the imputation.[4]

Figure 10.2 plots the data together with the diagonal and a nonparametric regression line. If the survey and the register measures coincided, all points

4. The Danish data hold information about the value of stocks and bonds at the household level, and this gives rise to measurement error. Without direct information about both the quantities and the prices of assets, it is not possible to distinguish active savings decisions from capital gains. Koijen, Van Nieuwerburgh, and Vestman (chapter 11, this volume) use Swedish register data with exact information about the holdings of stocks and bonds and are therefore able to address this issue.

Table 10.1 **Estimates of λ**

	$z^S - z^R$ unrestricted (1)	$-2 < z^S - z^R < 2$ (2)
λ	0.791***	0.816***
	(0.0148)	(0.0128)
Constant	2.519***	2.237***
	(0.1792)	(0.1546)
N	3,352	3,320
R^2	0.460	0.551

***Significant at the 1/10 of 1 percent level.
**Significant the 1 percent level.
*Significant at the 5 percent level.

would be located on the diagonal. The broken line is a nonparametric regression through the data cloud and comparing its slope to the diagonal shows the attenuation bias. One thing to notice is that the broken line is almost linear and it is also noticeable that the bias is apparent.

Table 10.1 presents the results from estimating the regression line by OLS. The estimate in column (1) shows that the bias is 0.21, suggesting that there is a fair amount of noise in the survey measure. Restricting the size of the errors does not change the estimate much, indicating that the bias is not caused by outliers. Of course, concluding that the survey measure is noisy relies on assumptions (1) and (2) being correct, in particular that the measurement errors of the two measures are uncorrelated. Since errors in the survey are related to the accurateness of the survey response and the register error is related to capital gains on the portfolio, this assumption does not appear restrictive. The assumption that the register error be uncorrelated with the true (but unobserved level) is not testable with our data and will, for example, be violated if respondents with a low level of true consumption overreport and people with high true levels of consumption underreport.

Using Swedish data, Koijen, Van Nieuwerburgh, and Vestman (chapter 11, this volume) run regressions similar to the ones presented in table 10.1 in order to quantify the amount of noise in their data, and it appears that there is more noise in the Swedish data than in our data. While there are differences between the two studies in terms of the imputation method and the timing of the surveys, it is not clear why this pattern emerges.

10.4.3 Summary, Example 1

In this example, the possibility to construct a register-based measure of total expenditure that can be compared with the survey measure is illustrated. While the validity of this exercise hinges on two important assumptions, we believe that the register approach provides an inexpensive way to get some insights on the precision of the survey measure that is difficult to obtain otherwise.

10.5 Example 2: Validating Survey Questions about Gross Income Using Third-Party-Reported Information from the Income Tax Registers

An income variable is included in almost any survey collected by social scientists, and surveying is often the only way to collect income jointly with other variables of interest. Danish register data on income are of very high quality because they are automatically third-party reported and are reported separately for different types of income. In this section we compare the responses to a one-shot recall question about gross personal income collected by telephone interview in January 2010 to the tax records of the respondents in order to assess the quality of the survey measure. As opposed to the previous example, the register information is now perceived to be close to the truth, and we therefore expect to be relying much less on assumptions (A.1) and (A.2)

10.5.1 Data

In January 2010 the authors of this chapter organized a telephone survey including 6,004 completed interviews. The purpose of the survey was to obtain information about their response to a stimulus policy implemented in 2009.[5] The sample is drawn randomly from the population of persons in employment at some point in the period 1998–2003, totaling 3.9 million persons or about 75 percent of the Danish population. As part of the survey, respondents were asked a one-shot recall question about their gross annual income in 2009. The question was:

"We are also interested in knowing about the development in your income before taxes. We are thinking about income such as earnings (including employers pension contribution), pension payments, payments from unemployment insurances, cash benefits, or other forms of transfer income. What was approximately your income before taxes in 2009?"

A total of 5,394 persons answered the question. Self-employed persons effectively self-report income to the tax authorities and we therefore do not have as much faith in the register information for this group as we have for wage earners and persons receiving transfer income. We therefore select away persons with own-business income. Finally, we deselect two observations with negative gross income[6] and are left with 4,793 observations. The survey data were subsequently merged at the person level with administrative register data about income from the income tax register covering the

5. The results are available in Kreiner, Lassen, and Leth-Petersen (2012), posted on our personal websites.

6. Negative gross income can occur because some components of capital income are available in our data set only as net measures and therefore adds negatively to gross income if the net-value is negative. This seems to be a small problem in the data set. For most people the major capital expenditure components are constituted by interest payments on bank debt and mortgages. Interest payments on bank and mortgage debt are observed, and when we take these components out, sixty-two cases are observed with negative capital income and half of these observations' negative capital income is less than 1,000 USD. We therefore conclude that this is a minor problem.

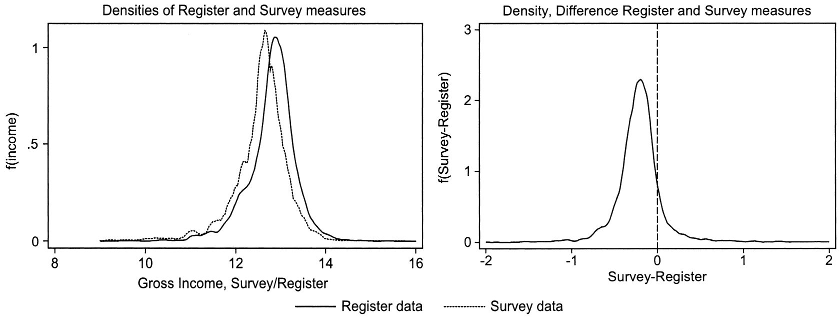

Fig. 10.3 Densities of the survey and register-based measures of gross income and of the individual differences

Note: The right panel includes only data in the interval –2;2. Sixty observations are selected away.

tax year 2009, that is, exactly the same period that the survey question was intended to cover.

10.5.2 Main Results

Figure 10.3 presents the densities of the survey and the register measure (left panel) and the density of the individual-level differences between the two measures (right panel). The left panel clearly reveals that the means of the two measures are not equal. It also suggests that the spread of the survey measure is larger, as would be expected if the survey measure carries an error and the register measure is accurate. The right panel confirms that the means are different when individual-level differences are considered, and also that individual-level errors have considerable spread, that is, that the survey measure is noisy.

Figure 10.4 graphs the register measure against the survey measure together with a smooth line through the data and a diagonal. The picture shows some very large outliers and also that the regression line has a smaller slope than the diagonal, indicating that the attenuation appears to be considerable.[7]

This is confirmed by a parametric regression reported in table 10.2. Regressing the register measure on the survey measure using the unrestricted sample yields an estimate of λ of 0.57 indicating important individual-level deviations

7. There is a graphically striking cluster of data points in the northeastern corner of the graph appearing to fall along a fairly tight regression line that is different from the mass of the data points. The apparent importance of this cluster is a visual deception because the cloud consists of only sixty-four observations. We have not been able to identify any significant differences between these observations and the rest of the data set apart from finding that they are, on average, four years younger than the rest of the sample. We also checked if interviewer effects could explain the pattern, but this was not the case.

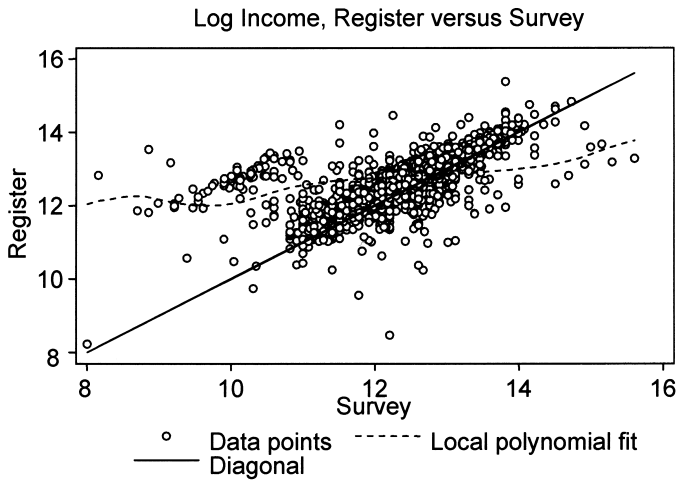

Fig. 10.4 Nonparametric regression of the register-based measure on the survey measure
Note: The graph includes only observations in the interval 8;16.

Table 10.2 Estimates of λ

	$z^S - z^R$ unrestricted (1)	$-2 < z^S - z^R < 2$ (2)
λ	0.570***	0.835***
	(0.0081)	(0.0072)
Constant	5.651***	2.283***
	(0.1024)	(0.0912)
N	4,793	4,707
R^2	0.505	0.739

***Significant at the 1/10 of 1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

between the survey and the register measure. In column (2) the errors are restricted to be within the –2;2 interval and this increases the estimate of λ to 0.84, suggesting that a limited number of outliers are responsible for a large part of the attenuation bias.

In table 10.3 the two measures of gross income and the individual-level errors are regressed on a set of "external" covariates. The idea is to see how the noise influences the covariance with other variables often used in empirical analyses. Comparing the numbers in columns (1) and (2) in table 10.3 suggests that the register and the survey measure have similar covariance with the set of external variables, but the parameter estimates obtained

Table 10.3 **Regressing on external covariates**

	z^R (1)	z^S (2)	$z^S - z^R$ (3)
Age	0.095***	0.114***	0.019***
	(0.0049)	(0.0066)	(0.0052)
Age^2	–0.001***	–0.001***	–0.000***
	(0.0001)	(0.0001)	(0.0001)
Woman	–0.145***	–0.232***	–0.087***
	(0.0127)	(0.0172)	(0.0136)
Single	–0.020	–0.067**	–0.047**
	(0.0162)	(0.0219)	(0.0173)
Number of children	0.012	–0.009	–0.021**
	(0.0072)	(0.0098)	(0.0077)
Education, short	0.113***	0.147***	0.034*
	(0.0160)	(0.0216)	(0.0171)
Education, medium	0.257***	0.288***	0.031
	(0.0188)	(0.0255)	(0.0202)
Education, long	0.362***	0.362***	–0.000
	(0.0241)	(0.0327)	(0.0258)
House owner	0.356***	0.245***	–0.111***
	(0.0147)	(0.0198)	(0.0157)
Constant	10.466***	9.954***	–0.512***
	(0.0980)	(0.1327)	(0.1049)
N	4,793	4,793	4,793
R^2	0.330	0.212	0.021

Note: Standard errors in parentheses.
***Significant at the 1/10 of 1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

using the survey measure do differ significantly from the parameter estimates obtained using the register measure for age, woman, number of children, single, and owner. Regressing the individual-level error on the same set of covariates suggests differences for the same variables. If one takes the register measure to be the truth, then the results of column (3) suggest that the measurement error associated with the survey measure is not classical.

10.5.3 Robustness

The survey question asks people to recall gross income including earnings, employers' pension contributions, transfer income, and capital income. Some of these are probably less salient to the respondent; employers' pension contribution is likely included in this category. This number does not appear separately on the paycheck, nor on the tax return, or the annual statement from the tax authorities since it is not liable to taxation before it is paid out. In a robustness check we subtract employers' pension contributions

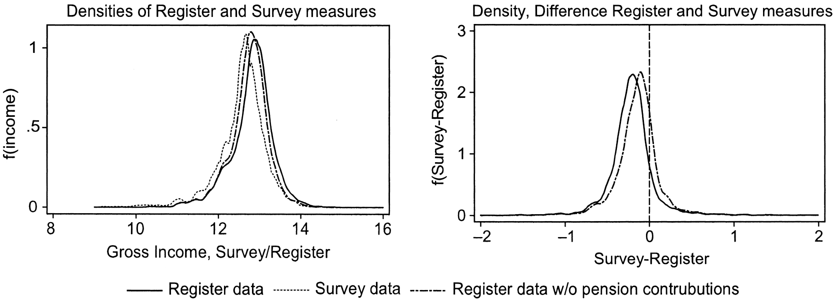

Fig. 10.5 Densities of the survey measure, the original register measure, and the register measure where employers' pension contributions are subtracted and of individual differences between the register measures and the survey measure

Table 10.4 Estimates of λ for register measure without employers' pension contributions

	$z^S - z^R$ unrestricted (1)	$-2 < z^S - z^R < 2$ (2)
λ	0.528***	0.773***
	(0.0077)	(0.0069)
Constant	6.093***	2.995***
	(0.0971)	(0.0872)
N	4,793	4,707
R^2	0.494	0.726

***Significant at the 1/10 of 1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

from the register measure and repeat the analysis of the previous section to check if the survey measure performs better when compared to the adjusted register measure. To do this we define an alternative gross income measure constructed from the registers where employers' pension contributions are deducted from the register measure used in the previous section. The idea is to investigate if respondents are more likely to have stated their income without employers' pension contributions even though it is clearly stated in the survey question that it should be included.

Figure 10.5 shows density graphs for the survey measure, the original register measure, and the register measure where employers' pension contributions are subtracted. The right panel shows densities of differences between the register measures and the survey measure. The figure shows that subtracting employers' pension contributions reduces the mean bias, but also that the spread is almost unaffected. Estimating λ by OLS reveals that this has not improved on the precision at the individual level. In fact, if anything, the estimates of λ in table 10.4 suggest that the attenuation bias has become more serious.

Table 10.5 **Regressing on external covariates**

	z^R (1)	z^S (2)	$z^S - z^R$ (3)
Age	0.086***	0.114***	0.028***
	(0.0046)	(0.0066)	(0.0053)
Age^2	–0.001***	–0.001***	–0.000***
	(0.0001)	(0.0001)	(0.0001)
Woman	–0.133***	–0.232***	–0.098***
	(0.0119)	(0.0172)	(0.0137)
Single	–0.007	–0.067**	–0.059***
	(0.0152)	(0.0219)	(0.0174)
Number of children	0.015*	–0.009	–0.024**
	(0.0068)	(0.0098)	(0.0078)
Education, short	0.102***	0.147***	0.045**
	(0.0150)	(0.0216)	(0.0172)
Education, medium	0.237***	0.288***	0.051*
	(0.0176)	(0.0255)	(0.0202)
Education, long	0.323***	0.362***	0.039
	(0.0226)	(0.0327)	(0.0259)
Owner	0.347***	0.245***	–0.102***
	(0.0137)	(0.0198)	(0.0157)
Constant	10.563***	9.954***	–0.608***
	(0.0918)	(0.1327)	(0.1053)
N	4,793	4,793	4,793
R^2	0.333	0.212	0.024

Note: Standard errors in parentheses.
***Significant at the 1/10 of 1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

Further, examining the correlation with the external covariates reveals even stronger correlations between the external covariates and the differences between the register and the survey measure. These results are presented in table 10.5. In the regression of individual-level differences between the two measures on external covariates, column (3), the parameters are now more significant, in particular in the case of education dummies. This suggests that the ability of the respondents to include employers' pension contributions when reporting their income varies across educational levels.

It is, of course, possible to construct many other concepts where other income components are subtracted. The most salient feature of income is arguably earnings and transfer income, which are received at regular intervals and where the recipient receives a letter stating the amount paid out. Capital income arrives in a less regular fashion and may therefore also be difficult to give an account for. We have experimented with subtracting capital income from the register measure, but this made little difference to the results and we therefore leave the results unreported. The calculations provided in this chapter are merely examples intended to illustrate the possibilities for identifying different subcomponents of income and how this may be used to identify what components of income respondents find it difficult to report in surveys.

10.5.4 Comparison to US Findings

The Consumer Expenditure Survey (CEX) and Panel Study of Income Dynamics (PSID) use annual recall questions about income but ask about different income components separately, for example, earned income, transfer income, and capital income. Bound et al. (1994) performed a validation study of the earnings question in the PSID by comparing answers to the PSID questions about earnings with company records for 418 workers from a single manufacturing company in 1983. This sample is called the PSID Validation Study. They find that the mean difference between the survey and the register measure is small, but that the standard deviation of the difference is substantial, amounting to 0.67 of the standard deviation of the company records. The corresponding measure in our data is 0.89. The slope coefficients from regressions of record on interview measure are similar between the two studies, 0.76 in the Bound et al. study and 0.84 in the trimmed version in our study.[8]

Both studies suggest substantial measurement error. One limitation of the Bound et al. study is that it is confined to validate the survey responses for a homogenous and small group. This could explain the smaller error. Our results suggest that the survey error is correlated with standard covariates and that the error is therefore not of the classical type. This leaves open the possibility that validation studies based on narrowly defined samples such as the PSID Validation Study do not give a complete picture of the size of the error in the main sample.

Another difference is that our study focuses on gross income including transfer income and capital income. This leaves open the possibility that our income measure is more noisy only because we include nonearned income. Gottschalk and Moffitt (2011) show evidence about the development of transitory family nonlabor income from the PSID, but to our knowledge the measures of transfer income and capital income in the PSID have not been validated.

Overall the results from the present study and the study by Bound et al. have implications for studies in many areas, but perhaps in particular for the interpretation of estimates from studies decomposing income variances into temporary and permanent components. The validation results suggest that

8. Bound, Brown, and Mathiowetz (2001) survey nine studies validating survey-based earnings measures from different US surveys against administrative records. Four of these studies report regression coefficients from a regression of the administrative record measure on the survey measure and three of these report regression coefficients in the vicinity of 0.75 using different data sets.

there is considerable noise in survey measures. This may explain why studies estimating income processes on US data collected in different ways find different results. Specifically, in a series of papers Gottschalk and Moffitt (1994) and Moffitt and Gottschalk (2002, 2011) use the PSID to decompose income into permanent and transitory variations and find that the transitory component is relatively big and increasing in the 1980s. For example, Moffitt and Gottschalk (2002) find that the variance of transitory log earnings for males is around 0.15–0.3. Kopczuk, Saez, and Song (2010) use Social Security Administration longitudinal earnings data for the period 1937–2004 and find that the transitory component is almost constant across time and relatively small, about 0.06–0.08 for the whole period and about 0.06 for the period 1980–, and that it cannot explain the increase in the variance of log earnings in the United States during the 1980s. While there are many other differences between these studies than the data collection mode, this does suggest the possibility that the size of the measurement error is important and not constant across time.

10.5.5 Summary, Example 2

The analysis of the quality of the recall question about annual gross income revealed that a one-shot recall question is inaccurate. Respondents tend to underreport their income level and the survey measure is noisy. Changing the definition of the register measure by excluding employers' pension contributions corrected for some of the mean bias, but did not reduce the spread much, and in particular did not reduce the attenuation bias in a regression of the register measure on the survey measure. The analysis also suggested that the individual-level differences between the survey and the register measure were correlated with observed characteristics of the respondents, suggesting that the errors associated with the survey measure are not of the classical type.

10.6 Summary and Suggestions for Future Work

This chapter has provided two examples illustrating how Danish third-party-reported register data can be matched at the individual- or household-level to survey records and used to validate the accuracy of responses to survey questions. The first example suggests that expenditure survey evidence on total expenditure is mean unbiased but noisy, and the second example suggests that a one-shot recall question about annual gross income is both mean biased and noisy.

The analyses presented in this chapter are possible because all persons living in Denmark are assigned a unique identification number to which all public authorities link up person-specific information and because surveys can be collected using the same person identifier. The potential of this validation methodology is big. In the Danish context, it is possible to match survey and register data for (potentially) the entire population, and it is also possible to match in the longitudinal dimension. In this way Denmark can be thought of as a laboratory where much more detailed and focused validation studies can be organized and where the impact of survey methodology on the accuracy of the survey responses are investigated so as to optimize the survey methodology across different groups and balancing this with survey costs.[9] For example, in the context of validating income questions the Danish setup allows researchers to merge survey records with tax records containing detailed information about different types of income and this provides a unique opportunity to test the ability of respondents to accurately report different types of income using different interviewing techniques and questions. Using the register data, it is also possible to consider individual as well as household units and to assess the extent to which it is important to ask all household members or just one in order to assess household income accurately. Finally, the Danish register data also contain very detailed information about car purchases with information about the exact type of car and the time of purchase. As in the Swedish case (see Koijen, Van Nieuwerburgh, and Vestman, chapter 11, this volume), this can be mapped directly to survey answers about purchases in order to test, for example, the impact of recall period length on precision of answers for that particular good.

This study focused on cross-sections of Danish households and persons. The Danish setup also allows asking the same people repeatedly and to match with panel data on income and wealth. Very little is known about the time series properties of measurement error in recall data. Bound et al. (1994) used panel data on earnings for the PSID validation study, but this is limited in size and only concerns a very narrowly defined group of people for two years. The Danish setup is much broader in scope since it potentially covers the entire Danish population with longitudinal information from the administrative registers. This provides a unique opportunity to learn about the time series properties of survey errors in the future. For example, it should be critical to understand if the size of the survey error is constant across time, if it always over/undershoots at the individual level, if the error is mean reverting but persistent, and so forth. The survey used in example 2 in this study has been repeated to cover questions concerning income in 2010 and will be repeated to cover 2011 through to 2013. When register data

9. For example, Olson, Smyth, and Wood (2012) explore if giving people their preferred survey mode increases the response rate. This possibility, however, potentially has a cost side to it by changing the level of precision for respondents who would have participated irrespective of the mode. Safir and Goldenberg (2008) attempts to measure this using natural data, but this approach ignores potentially important selection effects. In the Danish setting it would be possible to implement a randomized design that would be able to quantify the impact of self-selected mode choice on the precision of answers. As another example, it would be possible to asses the loss in precision by applying proxy reporting (http://www.bls.gov/cex/methwrkshp proxyrpting.pdf) by which survey responses are provided by a respondent about another member of the sampled unit or household.

have been released, we will be able to examine the time series properties of the survey errors.

The Danish setup allows matching new survey data with register data relatively easy and at relatively low costs. Matched survey and register data are kept at Statistics Denmark's servers and only researchers working at authorized Danish research institutions can get access to work with the matched data. However, researchers or statistical agencies with good research questions and appropriate funding wishing to start new research projects using combined survey and register data can do that in collaboration with Danish-based researchers. This can, for example, be done by contacting one of the authors of this chapter.

Appendix

Sample Statistics

Table 10A.1 Sample statistics, expenditure survey, and expenditure imputation from register data from example 1

	Register	Survey	Survey-register
N	3,352	3,352	3,352
Mean	12.077	12.085	0.008
Variance	0.518	0.381	0.296
Min.	6.951	9.302	–21.490
p1	10.106	10.550	–13.331
p50	12.105	12.121	–0.0210
p99	13.626	13.371	18.046
Max	14.660	14.127	46.236
Iqr.	0.988	0.875	0.551

Table 10A.2 Sample statistics, income survey, and income register data from example 2

	Register	Survey	Survey-register
N	4,793	4,793	4,793
Mean	12.804	12.561	–0.243
Variance	0.282	0.439	0.221
Min.	8.236	2.485	–8.934
p1	11.180	10.275	–0.254
p50	12.861	12.612	–0.214
p99	13.988	13.816	0.880
Max.	15.375	17.148	3.739
Iqr.	0.547	0.575	0.238

References

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. 1994. "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics* 12 (3): 345–68.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." *Handbook of Econometrics*, vol. 5, edited by J. Heckman and E. Leamer, 3705–830. Amsterdam: North Holland.

Bound, John, and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* 9 (1): 1–24.

Browning, Martin, and Søren Leth-Petersen. 2003. "Imputing Consumption from Income and Wealth Information." *Economic Journal* 113 (488): F282–301.

Gottschalk, Peter, and Robert Moffitt. 1994. "The Growth of Earnings Instability in the US Labor Market." *Brookings Papers on Economic Activity* 1994 (2): 217–72.

———. 2011. "The Rising Instability of US Earnings: Addendum." Unpublished manuscript, Johns Hopkins University, Boston College. http://www.econ2.jhu .edu/people/moffitt/jep%20addendum%20v1%202011–7–17.pdf.

Kleven, Henrik Jacobsen, Martin Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79:651–92.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics* 125 (1): 91–128.

Kreiner, Claus Thustrup, David Dreier Lassen, and Søren Leth-Petersen. 2012. "Consumption Responses to Fiscal Stimulus Policy and the Household Price of Liquidity." CEPR Discussion Paper no. 9161, Centre for Economic Policy Research.

Moffitt, Robert A., and Peter Gottschalk. 2002. "Trends in the Transitory Variance of Earnings in the United States." *Economic Journal* 112 (478): C68–73.

———. 2011. "Trends in the Transitory Variance of Male Earnings in the US, 1970–2004." NBER Working Paper no. 16833, Cambridge, MA.

Olson, Kristen, Jolene Smyth, and Heather Wood. 2012. "Does Giving People Their Preferred Survey Mode Actually Increase Participation Rates? An Experimental Examination." *Public Opinion Quarterly* 76:611–35.

Safir, Adam, and Karen Goldenberg. 2008. "Mode Effects in a Survey of Consumer Expenditures." Bureau of Labor Statistics. http://www.bls.gov/cex/cesrvymeths safir1.pdf.