

DANMARKS NATIONALBANK

3 DECEMBER 2018 — NO. 131

Consistency between household-level consumption data from registers and surveys

Kim Abildgren
kpa@nationalbanken.dk
DANMARKS NATIONALBANK

America Solange Lohmann Rasmussen
slr@dst.dk
STATISTICS DENMARK

Andreas Kuchler
aku@nationalbanken.dk
DANMARKS NATIONALBANK

Henrik Sejerbo Sørensen
hss@dst.dk
STATISTICS DENMARK

The Working Papers of Danmarks Nationalbank describe research and development, often still ongoing, as a contribution to the professional debate.

The viewpoints and conclusions stated in this paper are the responsibility of the individual contributors, and do not necessarily reflect the views of Danmarks Nationalbank or Statistics Denmark.

Consistency between household-level consumption data from registers and surveys

Abstract

We explore the consistency at household-level between register-imputed and survey-based consumption figures for Denmark over the period 2002-15. Furthermore, we analyse the differences between consumption functions estimated on the basis of the two types of microdata. We find that there is a close match regarding the median and mean level of consumption reported for the same households in the register-imputed and survey-based consumption data. Furthermore, we find that the marginal propensities to consume out of income estimated for the pre-crisis period (2005-07), the crisis period (2008-11) and the post-crisis period (2012-15) on the basis of register data are not significantly different to those estimated on the basis of survey data for the same households with income levels around the median. Finally, we find a relatively close match between total private consumption in the national-accounts statistics and the register-based consumption figures aggregated over all households. Overall, the findings confirm that low-cost consumption microdata derived from administrative registers covering the entire population of households can be an attractive alternative to survey data.

Resume

Vi undersøger konsistensen på husstands niveau mellem forbrug beregnet ud fra administrative registre og forbrug rapporteret i den interview-baserede forbrugsundersøgelse for Danmark i perioden 2002-15. Desuden analyserer vi forskellene mellem forbrugsfunktioner estimeret på basis af de to typer af mikrodata. Vi finder en tæt sammenhæng med hensyn til middel- og medianværdien af forbrugsniveauet for de samme husstande i de to typer af forbrugsdata. Endvidere finder vi, at de marginale tilbøjeligheder til at forbruge ud af indkomsten estimeret for årene før (2005-07), under (2008-11) og efter (2012-15) finanskrisen på grundlag af registerdata ikke er signifikant forskellige fra dem, som kan estimeres på grundlag af data fra forbrugsundersøgelsen for de samme husstande med indkomstniveauer omkring medianen. Endelig finder vi en forholdsvis tæt sammenhæng mellem det samlede privatforbrug i nationalregnskabsstatistikken og de register-baserede forbrugsdata aggregeret over alle husstande. Samlet set bekræfter resultaterne, at forbrugsmikrodata udledt fra administrative registre dækkende alle landets husstande kan være et attraktivt og billigt alternativ til interviewbaserede stikprøveundersøgelser.

Key words

Household microdata; consumption measures; consumption function.

JEL classification

C81; D12; E21.

Acknowledgements

We wish to thank participants for useful inputs at the ECB HFCN meeting in Bratislava in June 2018, the 35th IARIW General Conference in Copenhagen in August 2018 and the Deutsche Bundesbank Conference on "Heterogeneous households, firms and financial intermediaries" in Frankfurt in September 2018. We gratefully acknowledge comments by colleagues from Danmarks Nationalbank and Statistics Denmark. We also wish to thank Søren Leth-Petersen, University of Copenhagen, for useful discussions of the research project. The authors alone are responsible for any remaining errors.

1. Introduction

High-quality household-level data on savings and consumption with detailed household characteristics play an important role in a wide range of empirical research. Traditionally, such microdata were only available for rather limited sample sizes from interview-based surveys. However, surveys are expensive and it has been subject to discussion whether the data quality has declined over time (Meyer *et al.*, 2015; Pistaferri, 2015). Back in 1931, the participants in the Danish household budget survey had to complete detailed diary accounts of expenditures for every day in the entire year (Abildgren, 2018). Today, this period is two weeks in Denmark as well as in most other European countries (Eurostat, 2015). One way forward might be to focus more on exploiting low-cost data from administrative registers covering the entire population. Register data are usually considered to have a higher quality than survey data, which might be affected by self-reporting bias (Kleven *et al.*, 2011).

This paper has several aims. The first is to explore the consistency at household-level between register-imputed and survey-based consumption figures. As basis for the analysis, we impute measures of household-level consumption for the entire population of households in Denmark based on data from administrative registers over the period 2002-15. The Danish register data are primarily based on third-party reporting and therefore believed to be of particularly high quality. We then compare the figures with consumption data for the same households in the Danish interview-based household budget surveys over the same period and analyse the characteristics of those households where the largest differences occur. Furthermore, we explore whether the differences are related to macroeconomic factors such as the state of the business cycle or household-level events such as unemployment or major consumption events (e.g. car purchase). We also aim to identify possible suggestions for future improvements of the information content of the administrative registers and in particular the content and design of the interview-based household budget surveys. Furthermore, we discuss how various "Big Data sources" such as scanner and credit-card data might be used in relation to improvements of the Danish household budget surveys.

An additional aim of the paper is to analyse the consistency between consumption functions estimated on the basis of Danish household-level register data and similar consumption functions estimated on the basis of data from the Danish household budget surveys for the same households. The size of the estimated marginal propensity to consume out of income in a consumption function is of crucial importance for the consumption effects of an income shock and thereby for analyses of macroeconomic stability and the macroeconomic impact of various economic-policy measures. Finally, the paper aims to evaluate the correspondance between two separate measures of total private consumption, namely total private consumption in the national-accounts statistics and the register-based consumption figures aggregated over all households.

We find that there is a close match regarding the median and mean level of consumption reported for the same households in the consumption data imputed from administrative registers and in the interview-based household budget surveys. However, the distribution of the differences between the two measures of

consumption at household level is rather wide and the differences tend to be somewhat larger for high-wealth households. Some of the differences could in principle be related to the fairly crude adjustment for asset price fluctuations in the consumption data imputed from administrative registers. However, our robustness checks do not indicate that the adjustment for capital gains and losses is the main source of measurement error in the register-based measure of consumption. Nevertheless, it would be useful if the information content in the administrative registers in future could be expanded to cover realised and unrealised capital gains and losses to the extent that such information is available in the Tax Register or other registers.

We also perform a separate analysis of those households in the budget surveys who according to the car register purchased a car. We find that a relatively large share of these households did not report any car purchase in the budget survey. This is thus an area where there is scope for improvement of the household budget surveys in future. Perhaps it is better to make direct use of the value of car purchases that can be compiled on the basis of the car register than to collect the information from the households via interviews. The same could be the case within other areas where information on consumption at a household level is available from other sources (e.g. consumption of electricity or rent within social housing).

Furthermore, we find that the marginal propensities to consume out of income estimated for the pre-crisis period (2005-07), the crisis period (2008-11) and the post-crisis period (2012-15) on the basis of administrative data are not significantly different to those estimated on the basis of survey data for the same households with income levels around the median.

Finally, we find a relatively close match between total private consumption in the national-accounts statistics and the register-based consumption figures aggregated over all households. The compilation of private consumption in the Danish national accounts statistics is not based on register data on income and wealth - it is based on a commodity flow system that incorporates information from the development in the retail turnover index supplemented with information from the household budget surveys. The register-imputed consumption figures aggregated over all households and the figure for total private consumption from the national accounts can thus be seen as two independent estimates of the total private consumption in Denmark and the relatively close match between them supports the credibility of both sets of data.

Overall, the findings in the paper confirm that low-cost consumption microdata derived from administrative registers covering the entire population of households can be an attractive alternative to consumption data from household budget surveys.

2. Brief review of related literature

Our paper relates most closely to two strands of earlier research. The first is the part of the literature that focuses on imputing household-level consumption panel data with detailed household characteristics based on data from administrative registers. Browning and Leth-Petersen (2003) and Kreiner *et al.* (2015) imputed household-level consumption figures for 1994-96 via information from the Danish administrative registers

on income and changes in financial wealth. They found that the consumption data imputed from administrative registers were rather noisy but nevertheless performed reasonably well in terms of matching the median and mean level of consumption reported for the same households in the Danish interview-based household budget surveys 1994-96. The worst match was found for households with significant holdings of bonds or equities due to lack of information that allowed for a proper treatment of capital gains and losses. Similar studies have been carried out for Norway over the period 1994-2014 (Fagereng and Halvorsen, 2017; Eika *et al.*, 2017) and Sweden 1999-2007 (Koijen *et al.*, 2015; Kolsrud *et al.*, 2017). There have not been any studies on the consistency between register-imputed and survey-based consumption figures in Denmark based on data from the last couple of decades. One of the aims of our paper is to fill this gap in the literature.

The second strand of related research is the empirical literature on consumption functions based on household-level microdata. This is a field with many contributions, including several recent papers on the estimation of consumption functions on Danish household-level data (Browning *et al.*, 2013; Andersen *et al.*, 2016; Jensen and Johannesen, 2017; Hviid and Kuchler, 2017; Crawley and Kuchler, 2018). Recent contributions based on data for other countries include Carroll *et al.* (2017). The paper at hand is the first to study whether the use of household-level survey data or register data is of any significance for the estimated parameters in consumption functions.

Finally, our paper is the first study to compare the consistency between total private consumption in national-accounts statistics and aggregated register-based consumption figures.

3. Data sources, definitions and sample selection

Our dataset consists of two parts. The first is household-level consumption figures with detailed household characteristics compiled on the basis of panel data on income and wealth from administrative registers covering the entire population of Danish households over the period 2002-15. These register-based consumption figures are basically estimated as the difference between after-tax income and savings, where the latter is imputed by net wealth differencing (adjusted for estimated capital gains and losses) and contributions to privately administered pension schemes, cf. Appendix A for compilation details. This corresponds in broad terms to the approach followed by Browning and Leth-Petersen (2003).

The second part of our dataset consists of the household-level responses to the Danish household budget surveys over the period 2002-15. For a given year, the survey is based on interviews with around 1,000 households on major expenditures over the past 12 months combined with the same households' completion of detailed accounting booklets for a two-week period, cf. Statistics Denmark (1999, 2017). Each household only participates in the survey for one year – i.e. the survey data are cross sectional without any panel dimension. Households participating in the survey receive a small gift from Statistics Denmark (e.g. a coffee maker or a vase) in order to motivate participation in the survey. The survey results in household-level consumption figures at a very detailed level (8 digits of the so-called COICOP level in the international

standard "Classification of Individual Consumption According to Purpose", cf. United Nations (2000)). Due to the Danish system with a unique civil registration number (CPR number) for all citizens, Statistics Denmark is able to merge the survey results at a household level with the household-level consumption figures imputed from register data.

The definition of consumption expenditures in the household budget surveys basically follows the standard national accounts definition (Eurostat, 2013). The implied definition used in the register-based consumption figures is somewhat broader. In order to make meaningful comparisons, we therefore construct a similar broad consumption measure based on household budget surveys, cf. Table 1. It is this broad consumption measure from the household budget surveys that we will use for the analyses in this paper. The broad consumption figure is around 10 per cent larger than the standard consumption figure.

Table 1: Derivation of a broad consumption measure from the household budget surveys

Consumption in the household budget surveys (standard definition)
+ Value of home improvements (extension and rebuilding <i>etc.</i>)
+ Payment for private life insurance <i>etc.</i>
+ Fines
+ Gifts, charity
+ Membership fees
= Consumption in the household budget surveys (broad definition)

There are some minor differences between the household definitions applied in the data from the household budget surveys and the register-based consumption data. In the household budget surveys, a household is defined as comprising one or several persons who live at the same address and have a high degree of shared economy (share their meals and have joint income and expenditures, *etc.*). In the register-based consumption data, a household is a family consisting of either one or two adults plus any number of children living at home.¹ However, in most cases the two definitions give the same result. In order to ensure comparability between the two datasets, we only consider cases where the two household definitions result in the same number of persons in a given household, cf. the overview of our sample selection in Table 2. It should also be noted that we exclude household-year observations in the transaction year for households involved in real-estate transactions. We have no register-based consumption data for these types of households in the transaction year, cf. Appendix A.

¹ Two adults are regarded as members of the same family, if they are living together and a) are married to each other or have entered into a registered partnership, b) have at least one common child registered in the Civil Registration System or, c) are of opposite sex and have an age difference of 15 years or less, are not closely related and live in a household with no other adults. Adults living at the same address but not meeting one of the three criteria are regarded as separate families. Children living with their parents are regarded as members of their parents' family if they are under 25 years old, have never been married or entered into a registered partnership and do not themselves have children. A family meeting these criteria can consist of only two generations. If three or more generations live at the same address, the two younger generations are considered one family, while the members of the eldest generation constitute a separate family.

Table 2: Selection of data sample from the household budget survey to be compared with the register-based data, 2002-15

	Total number of household-year observations, all periods
Complete Household Budget Survey	11,117
- exclusion of cases with households involved in real-estate transactions	513
- exclusion of cases with household after-tax income below 25,000 DKK	41
- exclusion of cases where the two household definitions result in different number of persons in a given household	696
= Household budget survey data to be compared with register-based data	9,867

Source: See the main text.

There are also some differences in the definition of the "year" as the reference period in the the household budget surveys and the register-based consumption data. The register-based consumption data in principle follows the calendar year (January 1 to December 31). The issue is more complicated in the household budget surveys. The annual expenditures on a range of goods and services for a given household are based on a simple annualisation of the consumption expenditures reported for a two-week period, adjusted to take into account that underreporting is usually worse in the second than in the first week. However, no adjustments are made regarding seasonality. Furthermore, purchases of large durable goods (such as a car) and housing costs (such as rent, water and heating expenditures) are based on a personal interview about the households' major expenditures over the past 12 months which is not necessarily equal to a calendar year since the interviews with the households are spread evenly over the year. Some information in the household budget surveys will thus partly be about spending in the previous year which introduces some noise in the comparisons, cf. Statistics Denmark (2017).

4. Comparison of register- and survey-based consumption figures

The left part of Figure 1 shows a density plot of the distribution of annual consumption expenditures at household level based on register data and survey data respectively. Summary statistics of the moments of the two distributions are offered in Table 3. The mean and median of the two consumption distributions seem reasonably similar. This finding is consistent with the studies by Browning and Leth-Petersen (2003) and Kreiner *et al.* (2015) regarding household-level consumption in Denmark in the years 1994-96. Furthermore, the left part of Figure 1 also indicates that consumption inequality estimated by the register-based measure is somewhat larger than when estimated from the survey dataset, although the difference does not seem to be substantial. This probably reflects the well-known fact that the tails tend to be underrepresented in household surveys on consumption, income and wealth, cf. e.g. Pyatt (2003), Eckerstorfer *et al.* (2016) or Bricker *et al.* (2016). As shown in the right part of Figure 1, the distribution of the differences between the two measures of consumption at a household level is rather wide.

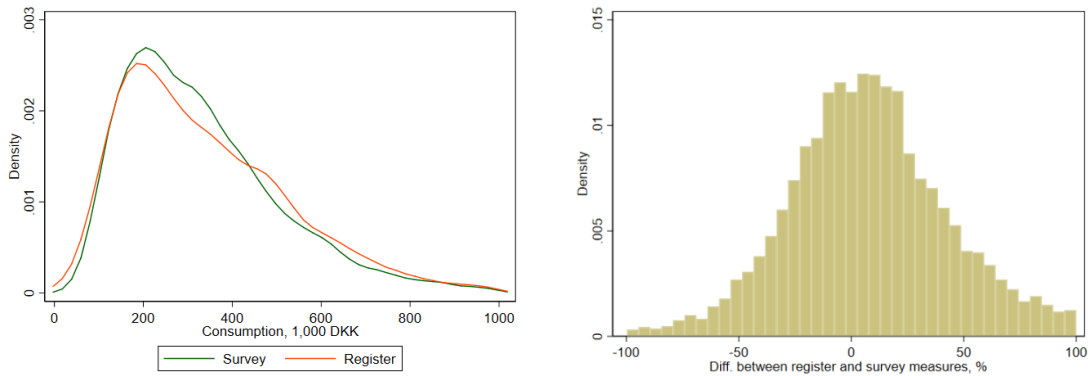


Figure 1. Distribution of annual consumption expenditures.

Notes: Sample period 2002-15. Left: Estimated by the Epanechnikov kernel density estimator. Outliers outside of the values of the x-axis are disregarded.

Table 3 Summary statistics for register- and survey-based household level consumption seen over the entire sample period 2002-15

	Register measure of annual consumption level	Survey measure of annual consumption level	Difference between register and survey measure at a household level
	DKK	DKK	% of survey measure
Mean	347,677	332,455	7.41
Standard deviation	216,982	173,672	34.93
1st decile	127,848	143,255	-35.18
1st quartile	190,530	200,874	-15.29
Median	303,027	298,374	6.37
3rd quartile	461,339	427,019	28.90
9th decile	623,035	576,596	54.21

Notes: In columns 1 and 2, outliers in the form of top and bottom 1% are excluded. In column 3, differences smaller than -100% and larger than 100% are excluded.

There are some indications that the differences between the register- and survey-based consumption measures tend to be larger for low- and high-income households and high-wealth households, cf. Figure 2 and 3. The latter could in principle indicate that part of the differences between the two measures of consumption at a household level are related to the fairly crude adjustment for asset price fluctuations in our consumption data imputed from administrative registers. Baker *et al.* (2018) have also recently pointed at measurement errors in consumption figures imputed from administrative data for households with large equity portfolios. In Appendix B we report results where all households in the sample that own more than 50,000 DKK worth of stocks and mutual funds shares have been excluded. The dispersion of the household level differences between the register- and survey-based measures are slightly smaller for households that do not own stocks and mutual funds shares. But the difference is so small that it does not seem to be the

adjustments for capital gains and losses that are the main source of measurement error in the register-based measure of consumption. Nevertheless, it would be useful if the information content in the administrative registers in future could be expanded to cover realised and unrealised capital gains and losses to the extent that such information is available in the Tax Register or other registers.

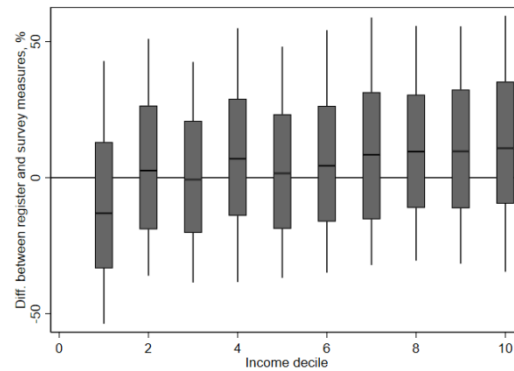


Figure 2. Differences between register- and survey-based consumption by income decile.

Notes: Sample period 2002-15. Thin lines indicate 9th and 1st decile, grey boxes indicate 3rd and 1st quartile and solid black lines indicate the median.

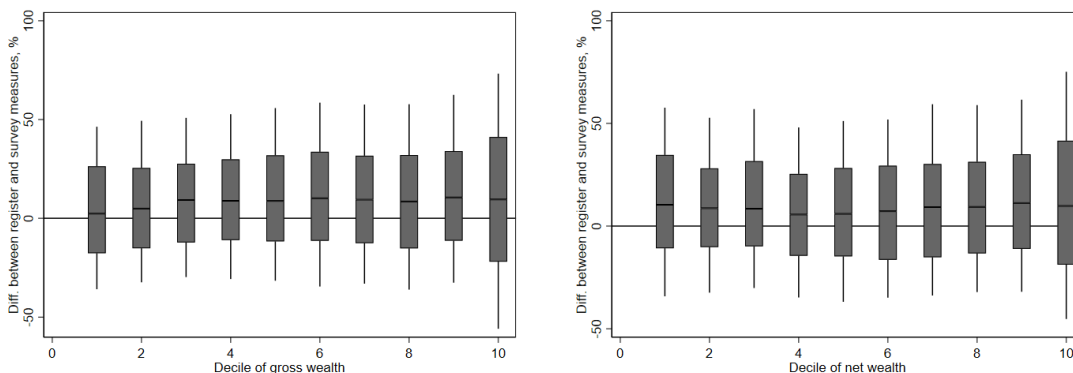


Figure 3. Differences between register- and survey-based consumption by gross and net wealth decile.

Notes: Sample period 2002-15. Thin lines indicate 9th and 1st decile, grey boxes indicate 3rd and 1st quartile and solid black lines indicate the median.

The differences between the register- and survey-based measures of consumption at a household level seem not to be clearly related to home-ownership, age, unemployment status or the business cycle, cf. Figure 4. However, for self-employed the distribution is wider. This might be expected since it is not possible to distinguish between assets and liabilities related to the "enterprise" part and the "home" part of the self-employed households in the register data.

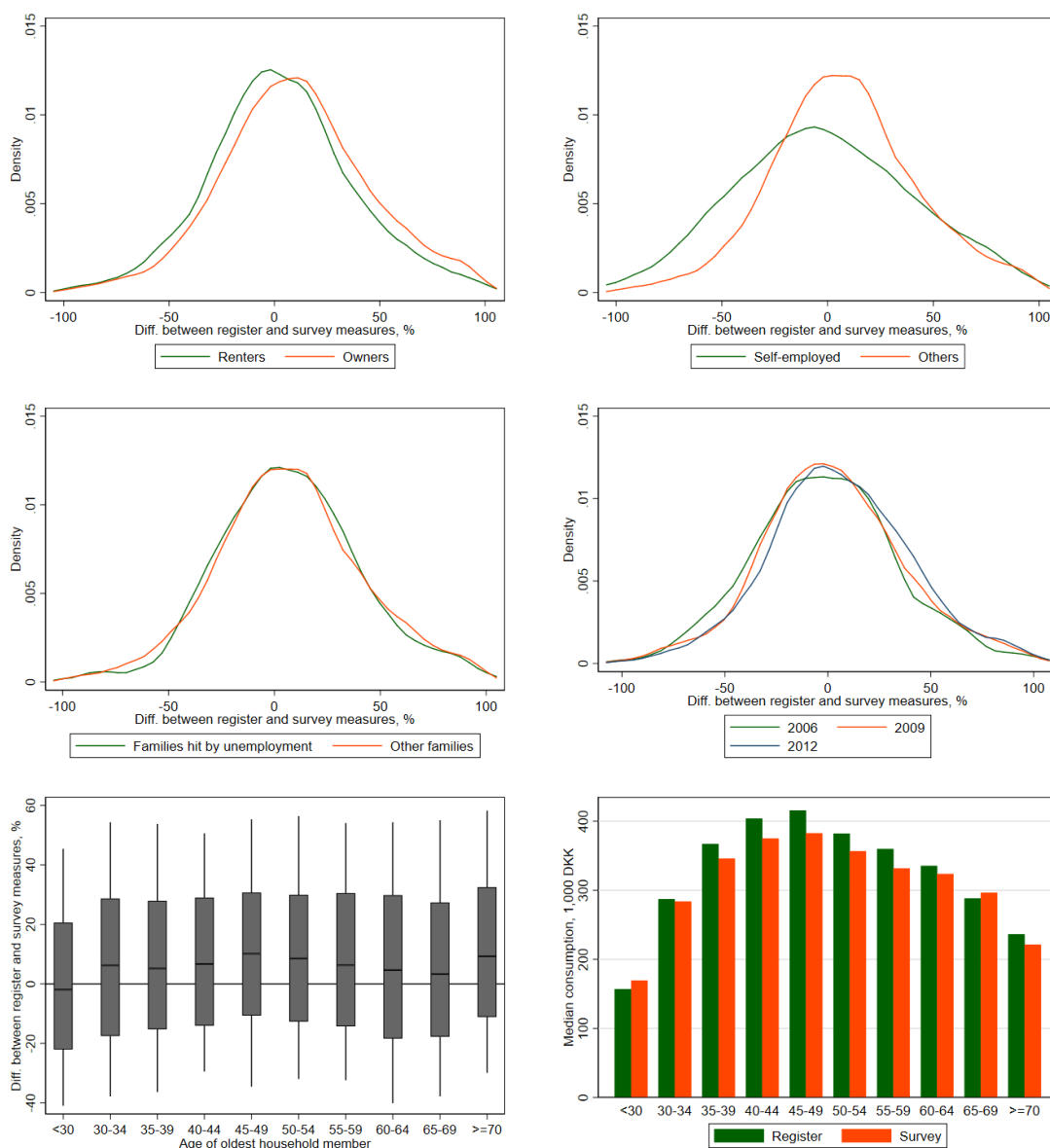


Figure 4. Differences between register- and survey-based consumption by various household characteristics.

Notes: Sample period 2002-15. Outliers outside of the values of the x-axis are disregarded. Families hit by unemployment are families, in which at least one member has experienced at least 6 months of unemployment over the past 2 years. Densities estimated by the Epanechnikov kernel density estimator. In box-plots: Thin lines indicate 9th and 1st decile, grey boxes indicate 3rd and 1st quartile and solid black lines indicate the median.

The different dimensions which are analysed in Figures 2-4 overlap. To explore more formally how the difference between the register- and survey-based consumption measures vary with household characteristics, we run a household-level regression of the absolute value of the percentage deviation between the two measures on a range of household characteristics. Results are reported in Table 4, and confirm the descriptive evidence. The differences between the survey and register measures are generally larger in the higher wealth deciles. This is in particular so for the 10th decile of gross wealth, where the

absolute value of the deviation between the two measures is 11 percentage points larger than in the 1st wealth decile when controlling for other household characteristics. Furthermore, the differences are larger in the lowest income decile. As expected, differences are also larger among self-employed. Otherwise, no clear patterns emerge regarding age, income and home-ownership status.

Table 4 Regression models. Dependent variable: Absolute deviation between register and survey measures in per cent of survey measure

Model	(1)	(2)
Age group		
30-39	0.040	1.366
40-49	-0.615	0.410
50-59	-0.958	0.158
60-64	-0.041	1.052
65+	-0.650	-0.227
Income	0.003*	
Income decile		
2		-5.522*
3		-11.244***
4		-7.734***
5		-10.070***
6		-9.756***
7		-9.525***
8		-10.557***
9		-9.579***
10		-8.272***
Total assets	0.001***	
Decile for total assets		
2		-0.274
3		0.062
4		0.801
5		1.827
6		3.185**
7		3.139**
8		3.315**
9		3.368**
10		11.007***
Tenants	0.302	0.337
Family size	-0.997***	-0.788**
Self-employed	2.712**	2.618**
Memo:		
Observations	6,210	6,210

Notes: Coefficient estimates (OLS). *** p<0.01, ** p<0.05, * p<0.1.

Underreporting is one of the main potential sources of error that can affect the quality of data from household budget surveys. With respect to car purchases, we can use register data to assess the degree of underreporting in the Danish household budget surveys. Figure 5 (left) reports the results of an analysis of those households in the budget surveys who according to the car register purchased a car. A large share of these households did not report any car purchase in the household budget survey. This finding is consistent with similar results in a recent study for Sweden (Koijen *et al.*, 2015) that found that almost 30 per cent of car transactions were not reported in the survey.

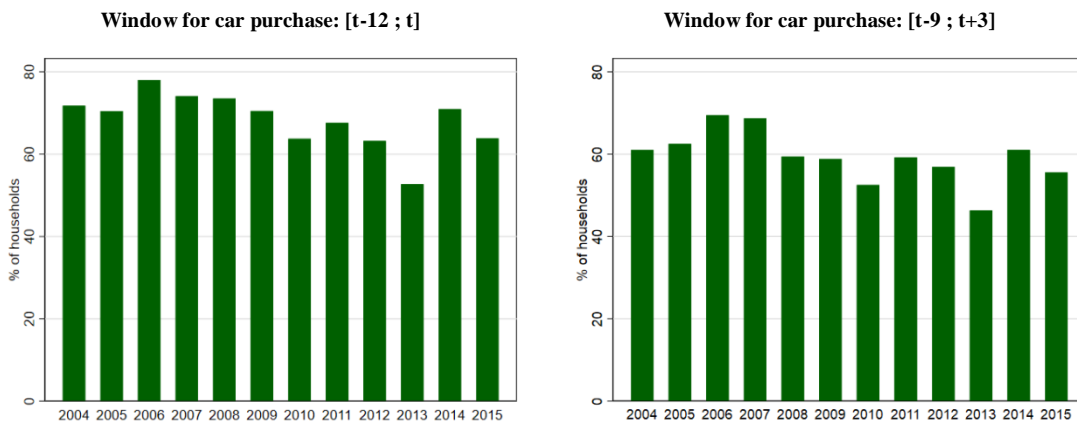


Figure 5. Car purchase according to register and survey data.

Notes: Share of households that buy a car according to the car register, who in the household budget survey have reported expenses for car purchase. Cars include motor cycles, campers etc. t = intended month of interview.

One potential limitation of our data is that the interviews for the budget survey in some cases are postponed, while only the originally intended interview dates are recorded in our data. To rule out that a large share of the apparent underreporting of car purchases is due to this issue, we also search in the registers for incidences of car purchase at a different window around the time of the interview, namely three months after the intended interview date, and one year back in time from that date. However, the results are very similar, cf. Figure 5 (right).

To see if this underreporting problem is larger for some groups of households than others, we estimate the probit model reported in Table 5. Middle-aged households and self-employed are somewhat more likely to underreport car purchases than other households whereas higher-than-median wealth households and tenants are less likely.

Table 5 Regression models: Underreporting of car purchases

Model	(1)	(2)
Age group		
30-39	0.032*	0.030**
40-49	0.040**	0.042***
50-59	0.051***	0.050***
60-64	0.016	0.018
65+	-0.003	0.007
Income	0.000*	
Income decile		
2		-0.016
3		-0.016
4		-0.005
5		-0.003
6		0.018
7		0.023
8		0.042*
9		0.023
10		0.032
Total assets	-0.000	
Decile for total assets		
2		-0.001
3		-0.000
4		0.003
5		-0.011
6		-0.027
7		-0.039**
8		-0.029*
9		-0.020
10		-0.016
Tenants	-0.034***	-0.034***
Family size	0.003	0.001
Self employed	0.022*	0.025*
Memo:		
Observations	7,402	7,402

Notes: Marginal effects (Probit). *** p<0.01, ** p<0.05, * p<0.1.

As mentioned in section 3, there are some differences regarding the reference period ("the year") in the two sets of consumption data. Since households are interviewed about their major expenditures during the past year, one should in principle expect to find the smallest differences for households surveyed late in the calendar year. This is confirmed by Figure 6 where the mean of the differences between the register- and survey-based consumption figures narrows substantially for households interviewed in the second half of the

year. Koijen *et al.* (2015) also find smaller differences between register- and survey-based consumption figures in Sweden late in the calendar year (December). However, Figure 6 also indicates that there are seasonal issues related to the survey data since the closest match between the register- and survey-based figures are found in the first three months of the year. This suggests that one might consider including seasonal dummies (based on interview month) when using microdata from the household budget surveys in econometric works.

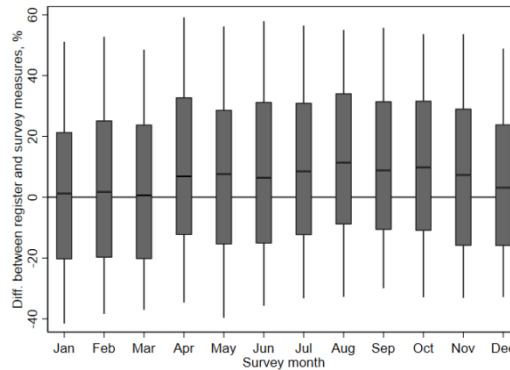


Figure 6. Differences between register- and survey-based consumption by location of survey in the calendar year.

Notes: Sample period 2002-15. Thin lines indicate 9th and 1st decile, grey boxes indicate 3rd and 1st quartile and solid black lines indicate the median.

5. Implications for estimation of consumption functions

The results presented so far indicate that the register-based and survey-based measures of household-level consumption are roughly equal on average, while the distribution of differences is quite wide. To assess the importance of these differences in analytically oriented work, we estimate and compare a number of simple consumption functions using the two measures of consumption for the same households. The size of the estimated marginal propensity to consume (MPC) out of income in a consumption function is of crucial importance for the consumption effects of an income shock and thereby for analyses of macroeconomic stability and simulations of the impact of various economic-policy measures.

First, to compare functional forms of consumption functions based on the two consumption measures, we estimate non-parametric regressions of consumption on income for different stages of the business cycle during the past 15 years or so, cf. Figure 7. For most periods, these regressions indicate that the shapes of consumption functions based on the two different consumption measures are relatively similar around the centre of the income distribution. However, for households in the lowest income groups, consumption functions based on the register-based consumption measure seem somewhat steeper than those estimated on the survey-based measure.

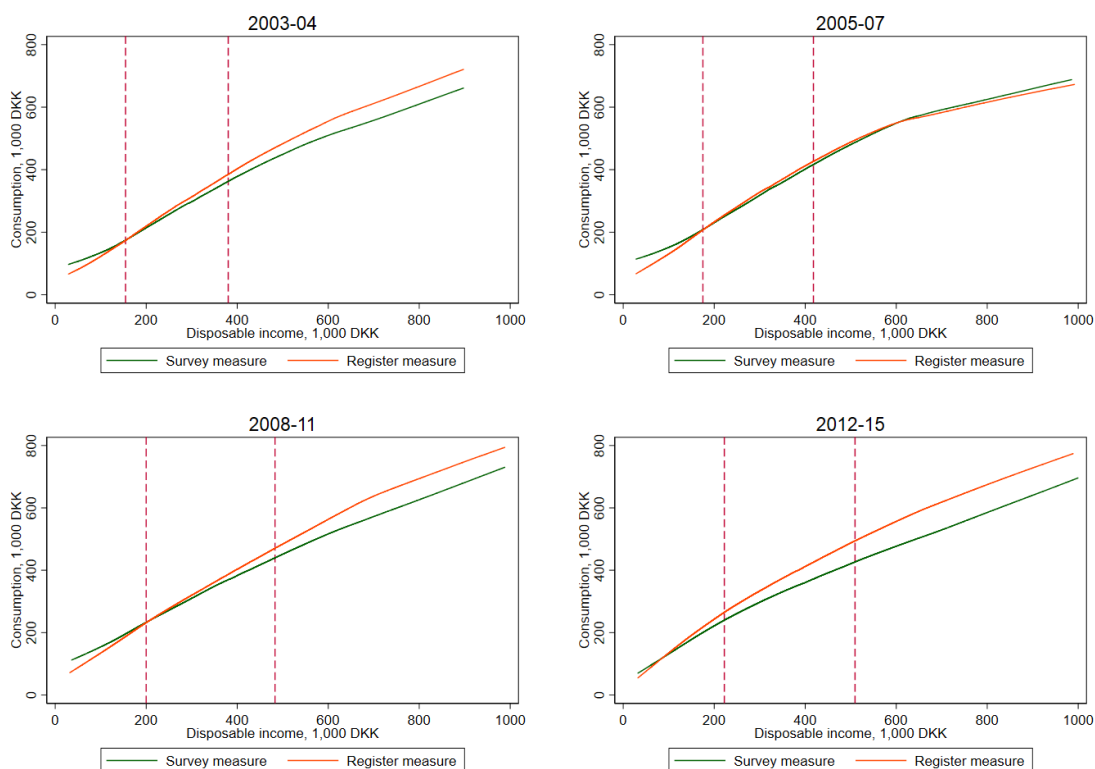


Figure 7. Consumption functions (locally weighted regression of consumption on income).
Notes: Vertical lines mark the 1st and 3rd quartiles of disposable income.

For households in the middle income quartiles (i.e. between the vertical lines in Figure 7), it is reasonable to assume that the consumption functions are linear in income. We therefore proceed by estimating a simple linear regression model of consumption on income for these households. We control for gross wealth and age group, as well as indicators for survey month following the discussion above. Consistent wealth data are only available from 2004, and we therefore only estimate the consumption functions for three periods around the most recent financial crisis: Pre-crisis period (2005-07), crisis period (2008-11) and post-crisis period (2012-15). Results are presented in Table 6.

Table 6. Consumption regressions (OLS regressions)

	2005-07		2008-11		2012-15	
	Survey	Register	Survey	Register	Survey	Register
Disposable income						
Coefficient (MPC)	0.670***	0.754***	0.715***	0.768***	0.705***	0.757***
Confidence interval	[0.536 ; 0.805]	[0.620 ; 0.887]	[0.616 ; 0.815]	[0.669 ; 0.867]	[0.623 ; 0.787]	[0.672 ; 0.841]
Z-test for equality (p-value)	0.283		0.364		0.331	
Memo:						
No. of observations	628	628	968	968	1,192	1,192
R-squared	0.238	0.262	0.232	0.264	0.272	0.328

Notes: The Table shows coefficient estimates and associated 95 per cent confidence intervals from a regression of consumption on income for households in the 2nd and 3rd income quartile. Survey month, age group and total assets (total financial and non-financial assets excluding pension savings) are included in the regression as control variables. In addition, the Table reports p-values from a Z-test for equality of coefficients. *, ** and *** denote rejection of the null hypothesis that the estimated coefficient or difference is zero at respectively a 10, 5 and 1-per-cent significance level.

The estimated MPCs out of income based on the register-based consumption measure are somewhat larger in magnitude than those based on survey data. However, the differences are not statistically significant different from zero in any of the three periods.

Recent microdata-based literature has found a high degree of variation in the estimated size of the MPC out of income, cf. e.g. Jappelli and Pistaferri (2010, 2014) and Carroll *et al.* (2017). This reflects not only heterogeneity in consumption behaviour across household characteristics (age, wealth, employment status, *etc.*) but also variations in income definitions (permanent or transitory income), consumption definitions (durable, non-durable, total), state of the business cycle (recession or expansion) and data type (panel or cross-section data).

The total consumption functions in Table 6 are based on cross-sectional variation in levels. They are as mentioned based on data for middle income households only, which has to be taken into account when compared to other estimates. The estimated MPCs out of income in Table 6 are in the range of 0.67-0.77 which is roughly in line with a range of other estimates. For instance, Crawley and Kuchler (2018) find a MPC out of permanent income of around 0.8 based on Danish panel data whereas Gerlach-Kristena (2014) finds MPCs out of permanent income in the range of 0.7-1.2 in a cross-sectional setup based on Irish data for various household types and sample periods. Blundell *et al.* (2008) find a MPC out of permanent income of 0.65 based on US data. However, it should be emphasised that the purpose of the paper at hand has not been to estimate MPCs for carefully specified consumption functions with all possible bells and whistles but to study the effect of the choice of consumption data (register-imputed versus survey data).

6. Suggestions for improvements of the household budget survey

The household budget survey is carefully designed to deliver figures on consumption with a detailed breakdown by commodity for a few main household types for use in the national accounts statistics and as weights on the CPI statistics. The survey was not designed to deliver high-quality consumption figures at household level. However, we have identified a number of areas where changes to the design of the household budget survey might be improved which will also strengthen the value of the survey as a source of household-level consumption data.

Underreporting is as mentioned one of the main potential sources of error that can affect the quality of data from household budget surveys. Therefore, it seems obvious to make use of alternative household-level sources where such are available rather than to collect the information from the households via interviews. One might for instance obtain the value of car purchases (new and used) at household level directly from information on prices and quantities recorded in the nationwide car register.

The same could be the case in other areas where information on consumption at a household level already is known through registers or other sources. Tests are already ongoing or planned by Statistics Denmark regarding the consumption of electricity based on data from intelligent power meters and rents within social housing.

Potentially, these initiatives have two benefits, first to increase the precision of the consumption estimates and second to reduce the burden for the participating households. The latter might also encourage more households to participate in the survey.

Currently, scanner data (bar code data) from supermarket chains are used as a source of prices on certain commodities in the Danish CPI. It is part of a planned work at Statistics Denmark to compile retail turnover by zip code at a detailed commodity level based on scanner data. Such data might be useful as part of the validation of questionnaire results in the household budget survey, for instance whether there is a correspondence between the products bought in a given zip code and the statements in the diary accounts from the same zip code. Another area of planned work is to supply the respondents in the household budget survey with some guidance regarding the "typical consumption bundle" for similar household sizes and age groups based on the experience from earlier replies to the survey.

Looking somewhat further ahead, it would be interesting to combine scanner data with information from the credit and debit cards used in the payment transactions for the goods. Credit and debit card data may be linked to register data through the civil registration number. The result could therefore be (longitudinal) data for consumption of a long range of supermarket products at a household level. This might eliminate the need to collect information on daily groceries via a questionnaire in the household budget survey. It might also be possible to collect information on expenditures on hotels, restaurants, gas stations, *etc.* via credit and debit card data. The issue of linking scanner data with payment card data is one of the topics currently explored by

the Danish Research Data for the Social Sciences (DRDS), which is a collaboration between Copenhagen Business School, Statistics Denmark, Aalborg University, Aarhus University, University of Copenhagen, Roskilde University, University of Southern Denmark, National Research Centre for Welfare (VIVE), Danmarks Nationalbank, The Danish Economic Councils, and the ROCKWOOL Foundation Research Unit.

7. Register-imputed consumption compared to private consumption in the national accounts

It is a well-known fact that there might be large differences between total consumption compiled on the basis of household budget surveys and total private consumption expenditures as reported in the national accounts statistics (Garner *et al.*, 2006; Browning *et al.*, 2014; Aguiar and Bils, 2015).

As a general robustness check, it can therefore be of interest to compare the register-imputed consumption figures aggregated over all households with the total private consumption figures from the Danish national accounts statistics. The level of private consumption in the Danish national accounts statistics is not based on register data on income and wealth - it is mainly estimated on the basis of a commodity flow method and the development in the retail turnover index supplemented with information from the household budget survey, cf. Statistics Denmark (2016). The register-imputed consumption figures aggregated over all households and the figure for total private consumption from the national accounts can thus be seen as two independent estimates of the total private consumption in Denmark.

In order to ensure comparability, we compiled a broad national-accounts-based measure of private consumption that should be fairly comparable to the concept used in the register-based data, cf. Table 7. A few remarks should be given to this Table:

- The consumption of non-profit institutions serving households (NPISH) such as sports clubs, charitable organisations, trade unions, *etc.* is part of the total private consumption expenditures in the national accounts but is not covered by the register-based consumption data.
- Interest margin in financial intermediaries (FISIM) paid by households is part of the private consumption expenditures in the national accounts statistics. In register-based data it enters into the net income component and thus not in the consumption expenditures.
- Motor vehicle owner duty paid by households is part of the register-based consumption expenditures but is treated as a direct tax in the national accounts statistics.
- The value of major home improvements is treated as gross fixed capital formation in the national accounts but is part of the register-based consumption expenditures.

Table 7: Derivation of a broad consumption measure from the national accounts

Private consumption (standard national accounts definition)
– Final consumption expenditure of non-profit institutions serving households (NPISH)
= Final consumption expenditure of resident households
– Financial Intermediation Services Indirectly Measured (FISIM) paid by households
+ Motor vehicle owner duty paid by households
+ Gross fixed capital formation in housing, value of major home improvements
= Private consumption in the national accounts (broad definition)

In general, there seems to be a high degree of correspondence between the two datasets, cf. Figure 8. Over the period 2003-16, total private consumption was on average 862 billion kroner according to the national accounts statistics and 874 billion kroner according to the register data. If we adjust the register data for outliers, the average annual consumption figure was 859 billion kroner over the same period. There are, however, larger short-term fluctuations in the register-based consumption figures than in the national accounts figures.

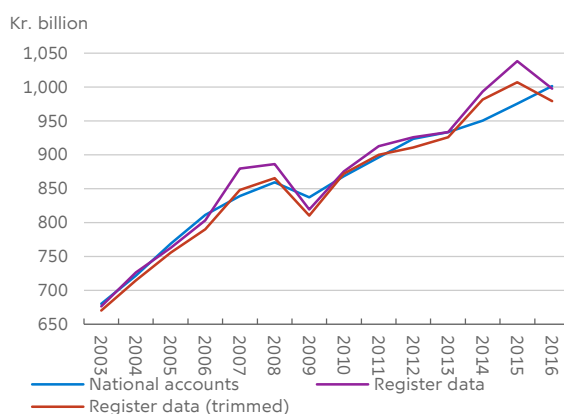


Figure 8. Total private consumption in national accounts compared to register-based consumption measure 2003-16.

Notes: The register-based consumption measure is compiled as the average consumption at household level (excluding households involved in housing transactions) multiplied by the total number of households (including households involved in housing transactions). In the trimmed consumption measure, the 1% top and bottom consumption observations have been excluded from the compilation of the average consumption at household level.

8. Final remarks and scope for further research

It should be underlined that it has not been the ambition of this paper to make any judgements regarding whether the register-imputed or survey-based consumption figures are closest to the "true" consumption. There are measurement issues in relation to both sets of consumption data and our main focus has been on the differences between the two datasets and the implication hereof for household-level consumption analyses. However, the findings in the paper confirm that low-cost consumption microdata derived from administrative registers covering the entire population of households can be an attractive alternative to consumption data from household budget surveys.

A combination of the two data sources may also be a potential way forward, for the purpose of both research and statistical production. Recently, Statistics Denmark has launched a new income and consumption register covering all 2.5 million households in Denmark (the so-called "Imputed Household Budget Survey"). The core of the register is based on administrative registers and contains information on income and taxes, car purchases, certain rent payments and a number of other consumption items for all households in Denmark. For households not covered by the household budget survey, the missing

consumption information is estimated by Statistics Denmark via a statistical mass-imputation procedure² based on household characteristics (including income) and the information from the household budget surveys. It could be of interest for future research to explore the consistency at a household-level between the register-imputed consumption figures and the consumption figures from the new household-level income and consumption register.

Finally, it is evident from the fat tails of the distribution of differences between register-based and survey-based consumption measures for all groups of households that one should carefully consider the treatment of outliers in any analytical application using the two datasets. It will often be necessary to exclude outliers and medians are often more informative than means. In many studies on household behaviour it might also be preferable to exclude self-employed (sole-proprietors) due to the difficulties of separating the "enterprise" part and the "home" part of household.

References

- Abildgren, Kim (2018), Household leverage and consumption during the Great Depression, *Journal of Financial Regulation and Compliance*, Vol. 26(2), pp. 287-299.
- Aguiar, Mark and Mark Bilal (2015), Has Consumption Inequality Mirrored Income Inequality?, *American Economic Review*, Vol. 105(9), pp. 2725-2756.
- Andersen, Asger Lau, Charlotte Duus and Thais Lærholm Jensen (2016), Household debt and spending during the financial crisis: Evidence from Danish micro data, *European Economic Review*, Vol. 89, pp. 96-115.
- Baker, Scott, Lorenz Kueng, Steen Meyerz and Michaela Pagel (2018), Measurement Error in Imputed Consumption, *NBER Working Paper*, No. 25078, September.
- Blundell, Richard, Luigi Pistaferri and Ian Preston (2008), Consumption Inequality and Partial Insurance, *American Economic Review*, Vol. 98(5), pp. 1887-1921.
- Bricker, Jesse, Alice Henriques, Jacob Krimmel and John Sabelhaus (2016), Measuring Income and Wealth at the Top Using Administrative and Survey Data, *Brookings Papers on Economic Activity*, pp. 261-312.
- Browning, Martin and Søren Leth-Petersen (2003), Imputing consumption from income and wealth information, *Economic Journal*, Vol. 113(488), pp. F282-F301.
- Browning, Martin, Thomas F. Crossley and Joachim Winter (2014), The Measurement of Household Consumption Expenditures, *Annual Review of Economics*, Vol. 6(1), pp. 475-501.
- Browning, Martin, Mette Gørtz and Søren Leth-Petersen (2013), Housing wealth and consumption: a micro panel study, *Economic Journal*, Vol. 123, pp. 401-428.
- Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka and Matthew N. White (2017), The distribution of wealth and the marginal propensity to consume, *Quantitative Economics*, Vol. 8(3), pp. 977-1020.
- Crawley, Edmund and Andreas Kuchler (2018), Consumption Heterogeneity: Micro Drivers and Macro Implications, *Danmarks Nationalbank Working Papers*, No. 129, November.
- Eckerstorfer, Paul, Johannes Halak, Jakob Kapeller, Bernhard Schütz, Florian Springholz and Rafael Wildauer (2016), Correcting for the Missing Rich: An Application to Wealth Survey Data, *Review of Income and Wealth*, Vol. 62(4), pp. 605-627.
- Eika, Lasse, Magne Mogstad and Ola L. Vestad (2017), *What can we learn about household consumption expenditure from data on income and assets?*, unpublished paper, March.
- Eurostat (2013), *European system of accounts – ESA 2010*, Luxembourg: Publications Office of the European Union.
- Eurostat (2015), Household Budget Survey. 2010 Wave. EU Quality report. Version 2, July, Doc. HBS/2015/01/EN Rev. 1.

² More specifically the Proc MassImputation procedure from the BANFF package in SAS, cf. Kozak (2005).

- Fagereng, Andreas and Elin Halvorsen (2017), Imputing consumption from Norwegian income and wealth registry data, *Journal of Economic and Social Measurement*, Vol. 42(1), pp. 67-100.
- Garner, Thesia I., George Janini, William Passero, Laura Paszkiewicz and Mark Vendemia (2006), The CE and the PCE: a comparison, *Monthly Labor Review*, September, pp. 20-46.
- Gerlach-Kristen, Petra (2014), Testing the Permanent Income Hypothesis for Irish Households, 1994 to 2005, *Economic and Social Review*, Vol. 45(4), pp. 511-535.
- Hviid, Simon Juul and Andreas Kuchler (2017), Consumption and savings in a low interest-rate environment, *Danmarks Nationalbank Working Papers*, No. 116, June.
- Jappelli, Tullio and Luigi Pistaferri (2010), The consumption response to income changes, *Annual Review of Economics*, Vol. 2, pp. 479-506.
- Jappelli, Tullio and Luigi Pistaferri (2014), Fiscal Policy and MPC Heterogeneity, *American Economic Journal: Macroeconomics*, Vol. 6(4), pp. 107-136.
- Jensen, Thais Lærkholm and Niels Johannesen (2017), The Consumption Effects of the 2007–2008 Financial Crisis: Evidence from Households in Denmark, *American Economic Review*, Vol. 107(11), pp. 3386-3414.
- Jørgensen, Casper Nordal and Andreas Kuchler (2017), Non-parametric estimation of marginal propensities to consume: the case of regression splines, *Danmarks Nationalbank Working Papers*, No. 112, March.
- Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen and Emmanuel Saez (2011), Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark, *Econometrica*, Vol. 79(3), pp. 651-692.
- Koijen, Ralph, Stijn Van Nieuwerburgh and Roine Vestmanz (2015), Judging the quality of survey data by comparison with 'truth' as measured by administrative records: evidence from Sweden. In: Christopher D. Carroll, Thomas F. Crossley and John Sabelhaus (eds.), *Improving the Measurement of Consumer Expenditures*, Chicago: University of Chicago Press, pp. 308-346.
- Kolsrud, Jonas, Camille Landais and Johannes Spinnewijn (2017), Studying consumption patterns using registry data: Lessons from Swedish administrative data, *CEPR Discussion Paper Series*, No. 12402, October.
- Kozak, Robert (2005), *The BANFF system for automated editing and imputation*, Proceedings of the Survey Methods Section at the SSC Annual Meeting, June.
- Kreiner, Claus Thustrup, David Dreyer Lassen and Søren Leth-Petersen (2015), Measuring the Accuracy of Survey Responses Using Administrative Register Data: Evidence from Denmark. In: Christopher D. Carroll, Thomas F. Crossley and John Sabelhaus (eds.), *Improving the Measurement of Consumer Expenditures*, Chicago: University of Chicago Press, pp. 289-307.
- Meyer, Bruce D., Wallace K. C. Mok and James X. Sullivan (2015), Household Surveys in Crisis, *Journal of Economic Perspectives*, Vol. 29(4), pp. 199-226.
- Pistaferri, Luigi (2015), Household consumption: Research questions, measurement issues, and data collection strategies, *Journal of Economic and Social Measurement*, Vol. 40(1-4), pp. 123-149.
- Pyatt, Graham (2003), Development and the distribution of living standards: a critique of the evolving data base, *Review of Income and Wealth*, Vol. 49(3), pp. 333-358.
- Statistics Denmark (1999), *Forbrugsundersøgelsen. Metodebeskrivelse. Fra dataindsamling til offentliggørelse*, Copenhagen: Statistics Denmark.
- Statistics Denmark (2016), *Danish GDP and GNI. Sources and methods 2012*, Copenhagen: Statistics Denmark.
- Statistics Denmark (2017), *Documentation of statistics for Household Budget Survey 2016*, Copenhagen: Statistics Denmark.
- United Nations (2000), Classifications of expenditure according to purpose, *UN Statistical Papers Series M*, No. 84.
- Vermeulen, Philip (2016), Estimating the Top Tail of the Wealth Distribution, *American Economic Review: Papers & Proceedings*, Vol. 106(5), pp. 646-650.
- Vermeulen, Philip (2018), How Fat is the Top Tail of the Wealth Distribution?, *Review of Income and Wealth*, Vol. 64(2), pp. 357-387.

Appendix A: Imputing register-based household-level consumption figures

Statistics Denmark – the Danish national institute of statistics – has a long tradition for collecting microdata from administrative registries and making them available to researchers in an anonymised form. The Tax Register contains information on household-level income, taxes, assets (bank deposits, stocks, bonds, houses and apartments owned) and debts, the Population and Family Registers deliver age, area of residence and family relations, the Health Register contains information on the number of days in hospital, the Property Register contains public valuation and sales prices on real estate, education information is available from the Education Register, employment status from the Labour Market Register, *etc.* Newer registries include data on mortgage loans with detailed loan-level characteristics, pension savings and car purchases. Information from all these registries can be combined due to the Danish system with a unique civil registration number (CPR number) for all citizens, and they can be aggregated to family level using information on family relations.

Our register-based consumption figure in year t for a given household (C_t) is basically estimated as the difference between the household's after-tax income (Y_t) and savings (S_t). The latter is imputed from changes in non-pension net financial assets after adjustments for capital gains and losses ($W_t - W_{t-1}$) and contributions to privately administered pension schemes ($PAPS_t$):

$$[A.1] C_t = Y_t - S_t = Y_t - [(W_t - W_{t-1}) + PAPS_t]$$

The household's after-tax income is gross income (wage income, dividends, net interest income, net government transfers, imputed rental value of houses and apartments owned, benefit payments from publicly or privately administered pension schemes, one-off payments from capital pensions, *etc.*) less all taxes as well as alimony payments. Contributions to employer-administered pension schemes paid directly by employers are not part of the household income and are therefore not included in the savings figures.

The non-pension net financial assets comprise gross financial assets (domestic and foreign bank deposits; market value of domestic and foreign bonds, mortgage deeds, stocks and mutual fund shares) less gross debt (domestic and foreign).

To adjust holdings of domestic stocks, *etc.* for price fluctuations, we rely on a fairly crude adjustment procedure where we assume that all households face value adjustments on their holdings of domestic stocks (including holdings of shares issued by Danish mutual funds) corresponding to the development in the general Danish stock price index (OMXC20, OMX Copenhagen 20 index). No value adjustments are made regarding holdings of domestic bonds or foreign securities.

Prior to 2009, we only have access to figures on housing loans from mortgage banks valued at the market prices of the underlying bonds financing the loans rather than the nominal value of the loans. However, for the post-2009 period we have access to information regarding nominal values as well as other loan characteristics. This enables us to calculate a more accurate measure of mortgage repayments than the

change in the market value. Interest rates and thereby the market value of debt have fluctuated more in the post-2009 period than prior to 2009. We therefore make use of the measure of actual repayments rather than the change in market value for this period.

Finally, it should be noted that our compilations exclude households involved in real-estate transactions in the relevant year. Variations in housing assets will therefore only reflect home improvements and capital gains and losses and we have not included housing assets in the net wealth figures used for derivation of our consumption figures. This implies that home improvements are part of our consumption figures whereas they from a national-accounts perspective are part of gross investments.

Appendix B: Results for households that do not own (more than a small amount of) stocks

Our calculation of the register-based measure of consumption involves a fairly crude adjustment for asset price fluctuations. In the following we report selected results based on a subsample where all households that own more than 50,000 DKK worth of stocks and mutual funds shares have been excluded. This corresponds to 16% of our sample (but only 10% of the full population).

The dispersion of the household-level differences between the register- and survey-based measures is slightly smaller for households that do not own stocks and mutual funds shares than in the full sample. But the difference is so small that it does not seem to be the adjustment for capital gains that is the main source of measurement error in the register-based measure of consumption.

Table B.1 Summary statistics for register- and survey-based household-level consumption seen over the entire sample period 2002-15 – only households that own less than 50,000 DKK worth of stocks.

	Register measure of annual consumption level	Survey measure of annual consumption level	Difference between register and survey measure at household level
	DKK	DKK	% of survey measure
Mean	327,144	318,621	6.79
Standard deviation	193,223	165,226	33.72
1st decile	124,923	136,892	-34.47
1st quartile	182,082	193,119	-15.16
Median	282,834	286,641	5.74
3rd quartile	438,954	410,273	27.41
9th decile	586,684	547,350	51.58

Notes: In columns 1 and 2, outliers in the form of top and bottom 1% are excluded. In column 3, differences smaller than -100% and larger than 100% are excluded.

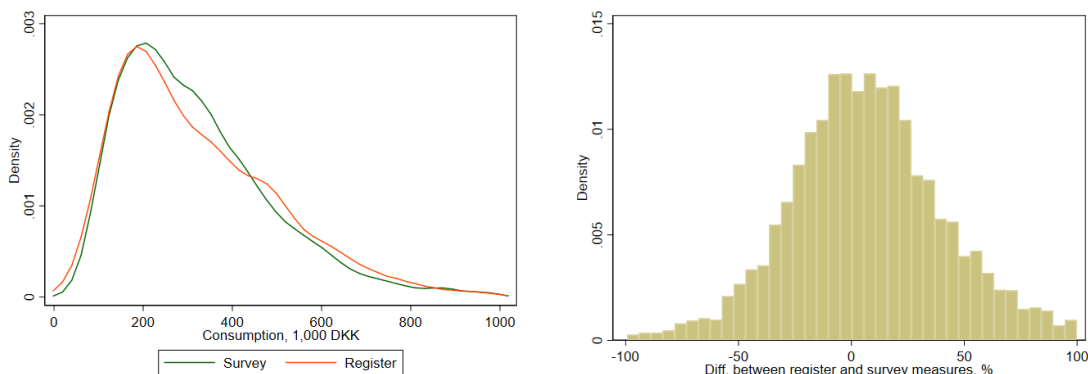


Figure B.1. Distribution of annual consumption expenditures – only households that own less than 50,000 DKK worth of stocks.

Notes: Sample period 2002-15. Left: Estimated by the Epanechnikov kernel density estimator. Outliers outside of the values of the x-axis are disregarded.

DANMARKS NATIONALBANK
HAVNEGADE 5
DK-1093 COPENHAGEN K
WWW.NATIONALBANKEN.DK



As a general rule, Working Papers are not translated, but are available in the original language used by the contributor.

Danmarks Nationalbank's Working Papers are published in PDF format at www.nationalbanken.dk. A free electronic subscription is also available at this Website. The subscriber receives an e-mail notification whenever a new Working Paper is published.

Text may be copied from this publication provided that Danmarks Nationalbank is specifically stated as the source. Changes to or misrepresentation of the content are not permitted.

Please direct any enquiries to Danmarks Nationalbank, Communications, kommunikation@nationalbanken.dk