

## **Consistency Between Household-level Consumption Data from Registers and Surveys**

**Kim Abildgren**

(Danmarks Nationalbank)

**Andreas Kuchler**

(Danmarks Nationalbank)

**America Solange Lohmann Rasmussen**

(Statistics Denmark)

**Henrik Sejerbo Sørensen**

(Statistics Denmark)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 7D-2: **Well-being and Household Issues**

Time: Friday, August 24, 2018 [14:00-17:30]

Paper presented at the  
35th IARIW General Conference, Copenhagen, August 20-25, 2018

## **Consistency between household-level consumption data from registers and surveys**

by

Kim Abildgren, Danmarks Nationalbank (kpa@nationalbanken.dk)  
Andreas Kuchler, Danmarks Nationalbank (aku@nationalbanken.dk)  
America Solange Lohmann Rasmussen, Statistics Denmark (slr@dst.dk)  
Henrik Sejerbo Sørensen, Statistics Denmark (hss@dst.dk)

### **Abstract**

We explore the consistency at a 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 micro data. 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 in simple consumption functions 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 registry-based consumption figures aggregated over all households.

**This version:** 6 July 2018

**Key words:** Household micro data; consumption measures; consumption function.

**JEL Classification:** C81; D12; E21.

## **Introduction**

High-quality household-level data on savings and consumption with detailed household characteristics plays an important role in a wide range of empirical research. Traditionally, such micro data 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 (Meyer et al., 2015; Pistaferri, 2015). Back in the days – in 1931 – the participants in the Danish household budget survey had to complete detailed diary accounts of expenditures for each day of the entire year (Abildgren, 2018). Today, this period is two weeks in most European countries (Eurostat, 2015). One way forward might be to have more focus on exploiting low-cost data from administrative registers covering the entire population. Register data is usually considered to have a higher quality than survey data which might be affected by self-reporting bias.

The paper at hand has two main aims. The first is to explore the consistency at a 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 particular high quality. We then compare the figures with consumption data for the same households in the interview-based Danish 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 examine the results of the analysis with an aim to identify possible suggestions for future improvements of the information content in 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.

The second 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.

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 a household level is rather wide. The differences tend to be somewhat larger for high-wealth households. This might indicate that part of the difference between the two measures of consumption at a household level is related to the fairly crude adjustment for asset price fluctuations in our consumption data imputed from administrative registers. It would therefore be useful if the information content in the administrative registers in the future could be expanded to cover realised and unrealised capital gains and losses to the extent that such information is available in the Tax Registry 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 improvements in the household budget surveys in the 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 rents within social housing).

Furthermore, we find that the marginal propensity to consume out of income in simple consumption functions 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 registry-based consumption figures aggregated over all households.

## **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 imputed consumption data 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 interview-based Danish 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. Broadly 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 (Kojien *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. In the paper at hand we fill this gap in the literature.

The second strand of related research is the empirical literature on consumption functions based on household-level micro data. 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.

### **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 administrative panel data registers covering the entire population of Danish households over the period 2002-15. These registry-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 participate 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 minor gift from Statistics Denmark (e.g. 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 follows basically the standard national accounts definition (Eurostat, 2013). The implied definition used in the registry-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 household budget surveys and the registry-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 registry-based consumption data, a household is a family consisting of either one or two adults plus any number of children living at home.<sup>1</sup> 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 households involved in real-estate transactions in the transaction year. We have no registry-based consumption data for these types of households, cf. appendix A.

---

<sup>1</sup> 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 registry-based data, 2002-15

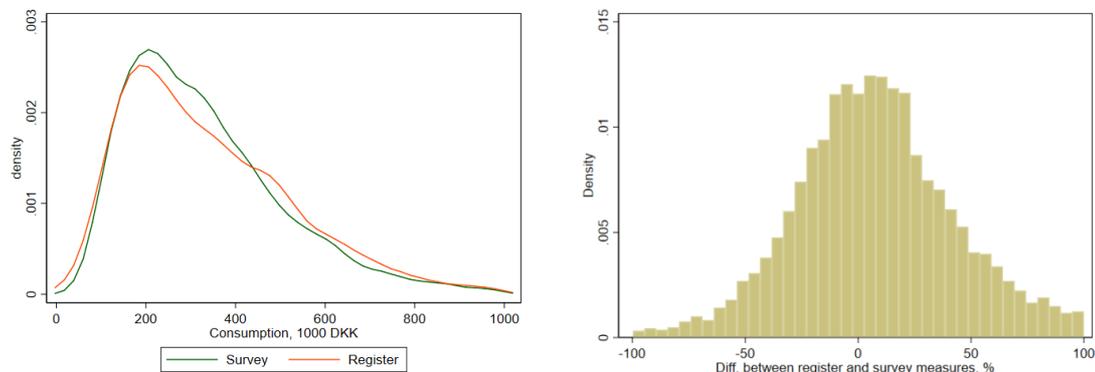
	Total number of household-year observations, all periods
Complete Household Budget Survey	11,117
- exclusion of cases with households involved in real-estate transactions	513
- exclusion of cases with household after-tax income below 25,000 DKK	41
- exclusion of cases where the two household definitions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

There are also some differences in the definition of the "year" as the reference period in the the household budget surveys and the registry-based consumption data. The registry-based consumption data follows in principle 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, purchase of large durable goods (such as a car) and housings costs (such as rent, water and heating expenditures) is based on a personal interview about the households major expenditures over the past 12 months which is not necessary 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 the household level based on respectively registry data and survey data. 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 substantial. This reflects probably the well-known fact that the tails tend to be underrepresented in household budget surveys, cf. e.g. Pyatt (2003) 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.

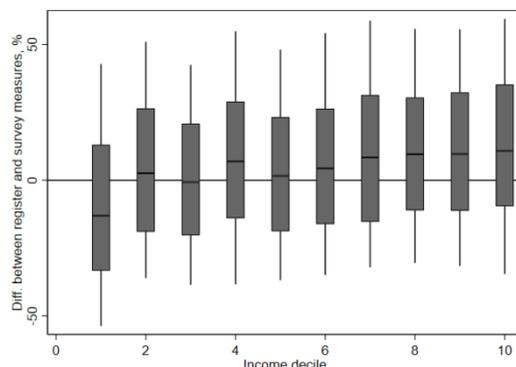
	Register measure of annual consumption level	Survey measure of annual consumption level	Difference between register and survey measure at a household level % of survey measure
	DKK	DKK	
Mean	347,677	332,455	7.41
Std. dev.	216,982	173,672	34.93
1st decile	127,848	143,255	-35.18
1st quartile	190,530	200,874	-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

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

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.

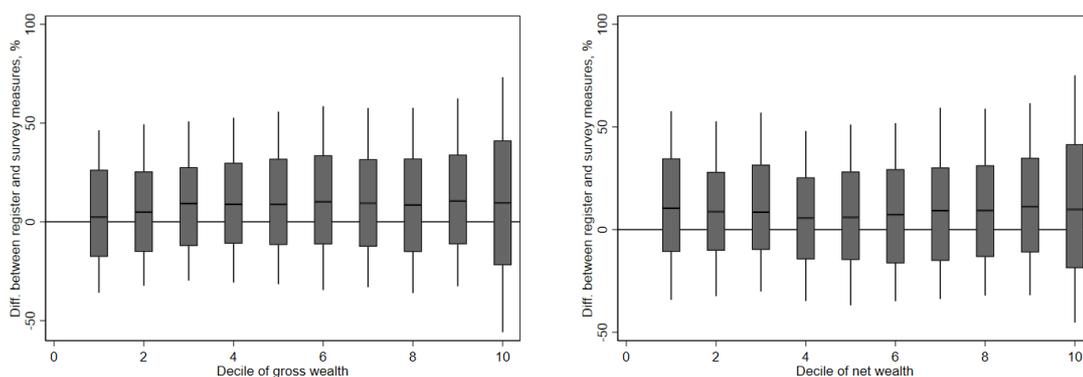
There are some indications that the differences between the registry- and survey-based consumption measures tend to be larger for low-income and high-wealth households, cf. Figure 2 and 3. The latter might indicate that part of the differences between the two measures of consumption at a household level are related to the fairly crude adjustment for asset price fluctuations in our consumption data imputed from administrative registers. It would therefore be useful if the information content in the administrative registers in the future could be expanded to cover realised and unrealised capital gains and losses to the extent that such information is available in the Tax Registry or other registers. Baker *et al.* (2018) have also

recently pointed at measurement errors in consumption measures imputed from administrative data for households with large equity portfolios.



**Figure 2.** Differences between registry- 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.



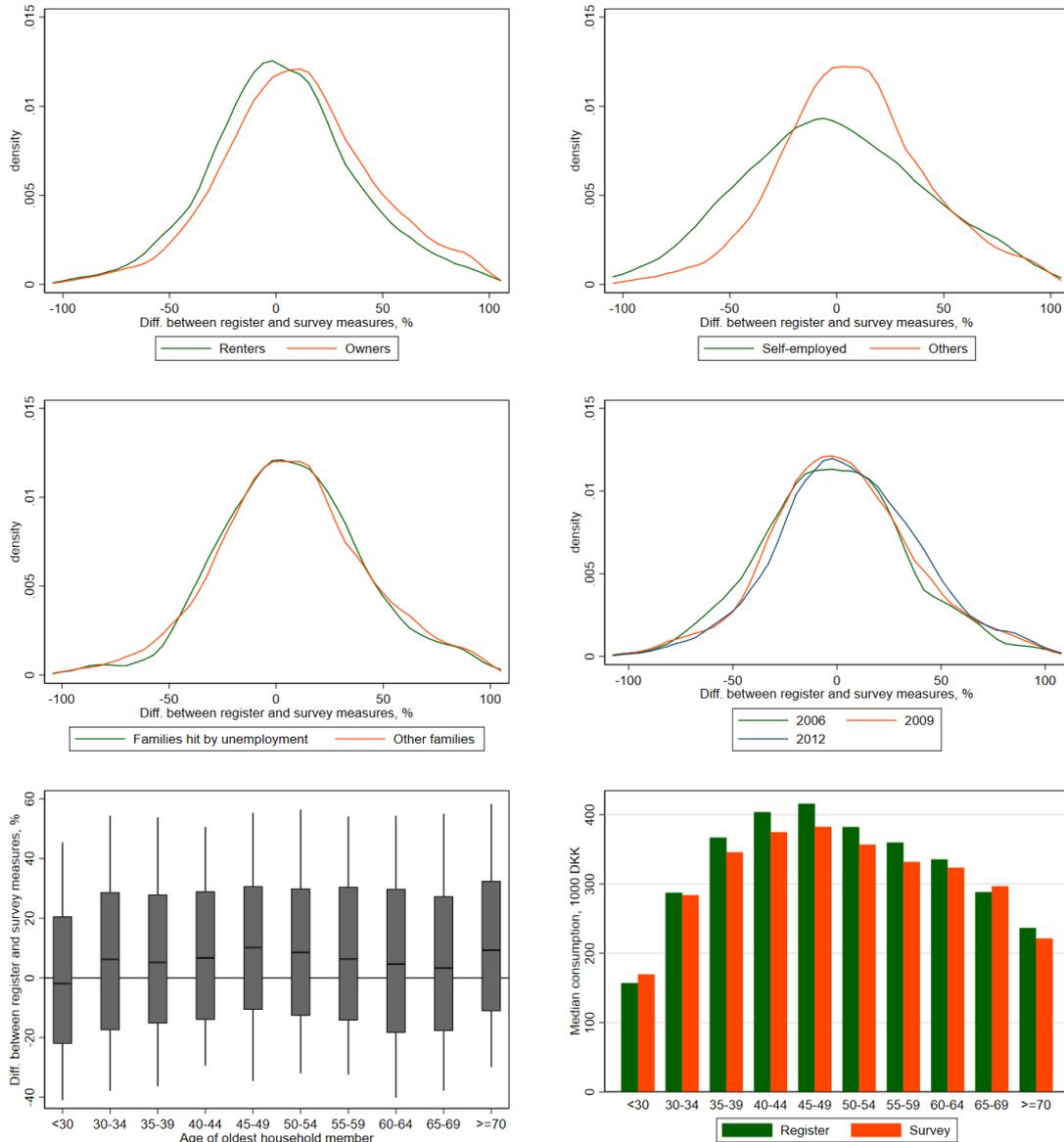
**Figure 3.** Differences between registry- and survey-based consumption by gross and net wealth decile.

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 two 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 registry data.

The different dimensions which are analysed in Figures 2-4 overlap. To explore more formally how the difference between the registry measure and the survey measure 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, 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.



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

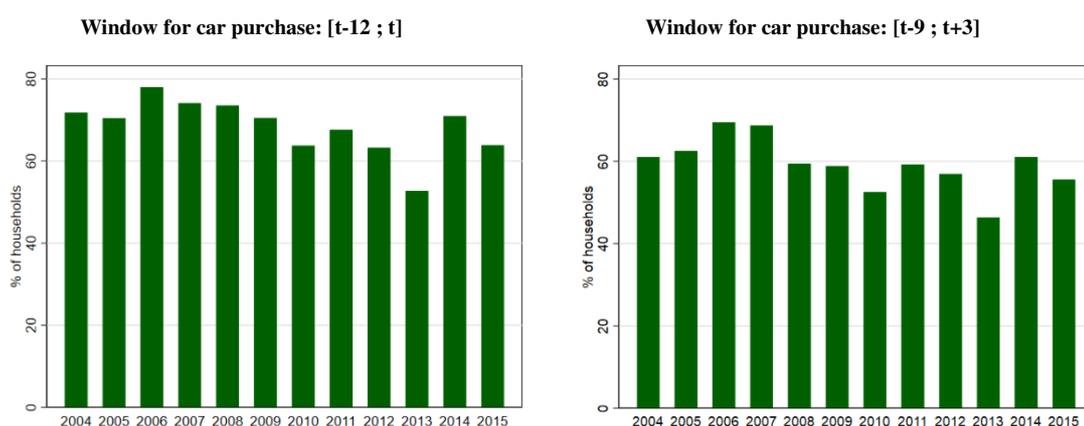
Notes: Sample period 2002-15. Outliers outside of the values of the x-axis are disregarded. Families hit by unemployment are families, in which at least one member have experienced at least 6 months of unemployment over the past 2 years. Densities estimated by the Epanechnikov kernel density estimator.

<b>Age group</b>		
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
<b>Income</b>	0.003*	
<b>Income decile</b>		
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***
<b>Total assets</b>	0.001***	
<b>Decile for total assets</b>		
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***
<b>Tenants</b>	0.302	0.337
<b>Family size</b>	-0.997***	-0.788**
<b>Self-employed</b>	2.712**	2.618**
<b>Observations</b>	6,210	6,210

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

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 survey. 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 (including motor cycles and campers). 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 (Kojien et al., 2015) that found that almost 30 percent of car transactions were not reported in the survey.



**Figure 5.** Car purchase according to registry 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 survey date is 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 estimated the probit regression 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. For the self-employed, the underreporting problem might reflect that it is not possible to distinguish clearly between the "enterprise" part and the "home" part of the self-employed households in the registry data.

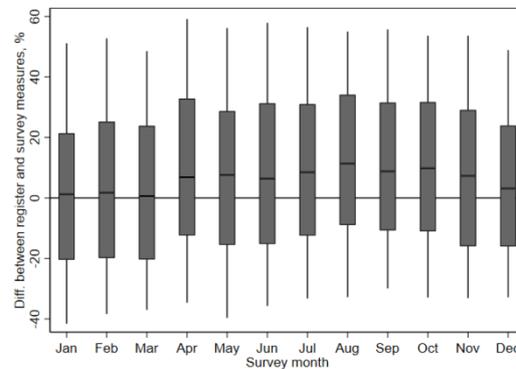
<b>Age group</b>		
30-39	0.032*	0.030**
40-49	0.040**	0.042***
50-59	0.051***	0.050***
60-64	0.016	0.018
65+	-0.003	0.007
<b>Income</b>	0.000*	
<b>Income decile</b>		
2		-0.016
3		-0.016
4		-0.005
5		-0.003
6		0.018
7		0.023
8		0.042*
9		0.023
10		0.032
<b>Total assets</b>	-0.000	
<b>Decile for total assets</b>		
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
<b>Tenants</b>	-0.034***	-0.034***
<b>Family size</b>	0.003	0.001
<b>Self employed</b>	0.022*	0.025*
Observations	7,402	7,402

**Table 5** Regression models: Underreporting of car purchases

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 registry- and survey-based consumption figures narrows substantially for households interviewed in the second half of the year. Kojen *et al.* (2015) finds also smaller differences between registry- 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 closed match between the registry- 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.



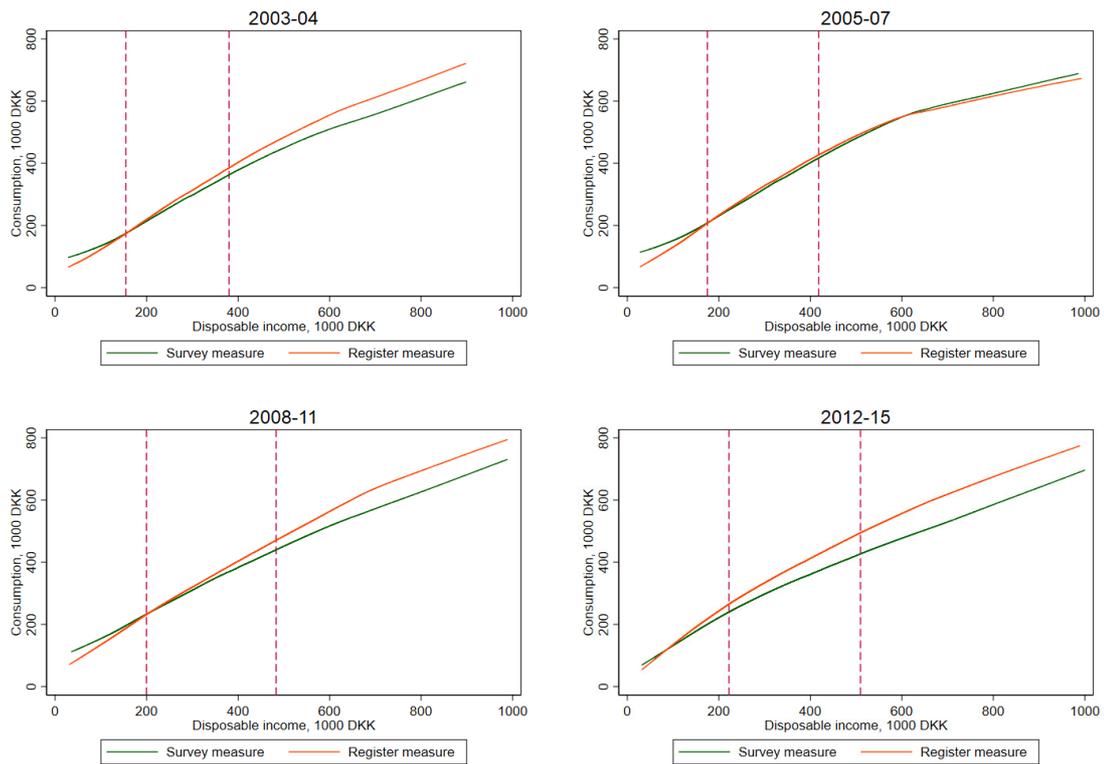
**Figure 6.** Differences between registry- and survey-based consumption by location of survey in the calendar year.

Notes: Sample period 2002-15.

## 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 very 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. 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.

	2005-07		2008-11		2012-15	
	Survey	Register	Survey	Register	Survey	Register
<b>Disposable income</b>						
Coefficient	0.674***	0.760***	0.730***	0.767***	0.714***	0.764***
Confidence interval	[0.539 ; 0.809]	[0.628 ; 0.893]	[0.631 ; 0.830]	[0.669 ; 0.865]	[0.632 ; 0.796]	[0.680 ; 0.847]
Z-test for equality (p-value)	0.269		0.523		0.348	
No. of observations	628	628	968	968	1,192	1,192
R-squared	0.207	0.253	0.208	0.261	0.256	0.322

**Table 6.** Consumption regressions (OLS regressions)

Notes: The Table shows coefficient estimates and associated 95% confidence intervals from a regression of consumption on income for households in the 2nd and 3rd income quartile. 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 \*\*\* denotes rejection of the null hypothesis that the estimated coefficient or difference is zero at respectively a 10, 5 and 1-per-cent significance level.

In the pre- and post-crisis periods, the estimated MPC out of income based on the register based consumption measure is somewhat larger in magnitude than that based on the survey measure. However, the differences are not statistically significant different from zero in any of the three periods.

Recent micro data 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 shocks), consumption definitions (durable, non-durable, total), state of the business cycle (recession or expansion) and data type (panel or cross-section data).

The identification of the total consumption functions in Table 6 is based on cross-sectional variation in levels. It is 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 certain other estimates. Crawley and Kuchler (2018) find for instance 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 break-down by commodity for a few main household types for use in the National Accounts statistics and as weights on the CPI statistics. The survey is not designed to deliver high-quality consumption figures at the household level. However, we have identified a number of areas where changes to the design of the household budget surveys might be improved which also will 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. It seems therefore 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 the household level directly from information on prices and quantities recorded in the nationwide car registers.

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.

These initiatives have potentially two benefits, first increase the precision of the consumption estimates and second 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 planned work at Statistics Denmark to compile retail turnover by zip code at a detailed commodity level based on such 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 cards data can 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

As a more general robustness check, it can be of interest to compare the register-imputed consumption figures with the total private consumption figures from the national accounts statistics. In order to ensure comparability between the two datasets, we compiled a broad national-accounts-based measure of private consumption that should be fairly comparable to the concept used in the registry-based data, cf. Table 8. A few remarks should be given to this Table:

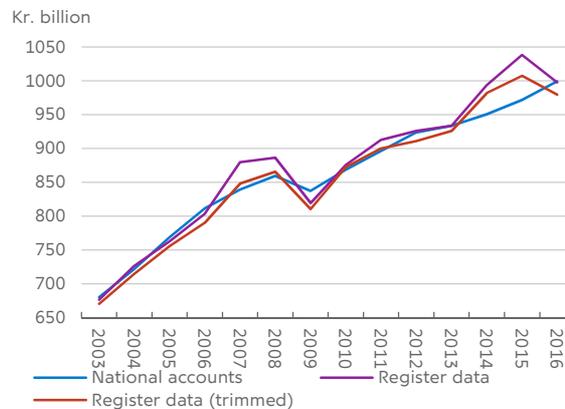
- The consumption of non-profit institutions serving households (NPISH) such as sports clubs, charitable organisations, trade unions *etc.* are part of the total private consumption expenditures in the national accounts but is not covered by the registry-based consumption data.
- Interest margin in financial intermediaries (FISIM) paid by households is part of the private consumption expenditures in the national accounts statistics. In registry-based data it enters into the net income component and thus not in the consumption expenditures.
- Motor vehicle weight duty paid by households are part of the registry-based consumption expenditures but are treated as direct taxes 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 registry-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 weight 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, Figure 8. Over the period 2003-16, total private consumption was on average 861 billion kroner according to the national accounts statistics and 874 billion kroner according to the registry 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 registry-based consumption figures than in the national accounts figures.



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

Notes: The registry-based consumption measure is compiled as the average consumption at a 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 a household level.

## 8. Final remarks and scope for further research

It should be underlined that it has not been the ambition with this paper to make any judgements whether the register-imputed or survey-based consumption figures are closest to the "true" consumption figures. There are measurement issues in relation to both sets of consumption data and our main focus has been on the differences between the two data sets and the implication hereof for household-level consumption analyses.

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 of the two data sets. It will often be necessary to exclude outliers and medians are often more informative than means.

Recently, Statistics Denmark has launched a new income and consumption register covering all 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. The consumption items included cover 8 per cent of total consumption. For those households covered by the Household Budget Survey, the register also cover the information on other consumption items collected via the Survey. For households not covered by the Household Budget Survey, the missing consumption

information is estimated via a statistical mass-imputation procedure<sup>2</sup> based on household characteristics (including income) and the information from the household budget surveys. It is part of our future research agenda 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.

### **Acknowledgements**

The authors wish to thank participants at the meeting of the ECB Household Finance and Consumption Network in Bratislava, June 2018, as well as colleagues from Danmarks Nationalbank and Statistics Denmark for useful comments on preliminary versions of the content presented in this paper. We also wish to thank Søren Leth-Petersen, University of Copenhagen, for useful discussions of the research project behind the paper. The viewpoints and conclusions stated do not necessarily reflect the views of Danmarks Nationalbank and Statistics Denmark. The authors alone are responsible for any remaining errors.

### **References**

- Abildgren, Kim (2018), Household leverage and consumption during the Great Depression, *Journal of Financial Regulation and Compliance*, Vol. 26(2), pp. 287-299.
- Andersen, Asger Lau, Charlotte Duus and Thais Lærkholm 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 Kuengy, Steen Meyerz and Michaela Pagel (2018), Measurement Error in Imputed Consumption, *SSRN Working Paper*, No. 3179117, May.
- 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, 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 Tokunaka 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), *Measuring Heterogeneity in Consumption Behavior: Micro Drivers and Macro Implications*, unpublished paper, June.
- Eika, Lasse, Magne Mogstad and Ola L. Vestad (2017), *What can we learn about household consumption expenditure from data on income and assets?*, unpublished working paper, March.
- Eurostat (2013), *European system of accounts – ESA 2010*, Luxembourg: 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.

---

<sup>2</sup> 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.
- 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 Lærholm 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.
- 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 (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.

## Appendix A: Imputing registry-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 Registry 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 Registry contains information on the number of days in the hospital, the Property Registry contains public valuation and sales prices on real estate, education information are 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 a family level using information on family relations.

Our registry-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 holding 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 have only 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 make therefore use 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.