



## **Imputing Consumption from Income and Wealth Information**

Browning, Martin; Leth-Petersen, Søren

*Publication date:*  
2003

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Browning, M., & Leth-Petersen, S. (2003). *Imputing Consumption from Income and Wealth Information*.  
Copenhagen.



# **Imputing Consumption from Income and Wealth Information**

Martin Browning  
Søren Leth-Petersen

AKF Forlaget  
March 2003



# Preface

The study presented in this memo investigates if it is possible to derive a measure of total expenditure for individual households from information on income and wealth. The analysis gives indication that administrative register data on income, tax payments, and wealth can be used to construct a measure of total expenditure at the household level. This result is useful, in particular, for researchers studying consumption behaviour of households. Consumption analysis has up to now primarily been based on expenditure survey data where each household is interviewed only one time, or where repeated interviews for the same households has been undertaken but information only collected on a subset of consumption items. Using register data it is possible to construct a measure of total expenditure for individual households and to follow these households across time. This enables more precise modelling of intertemporal allocation behaviour. Moreover, register data have the advantage over survey data that information is available for the total population. This enables analyses that focus on particular subgroups. This is relevant for studying dynamics of poverty, for example. The results presented in this memo thus create a potential for dealing with a wide range of research questions within the consumption literature that has not been possible to address hitherto.

Martin Browning and Søren Leth-Petersen

March 2003



# Contents

|   |    |
|---|----|
| <b>Summary</b> .....  | 7  |
| <b>1 Introduction</b> .....   | 8  |
| <b>2 Deriving total expenditure from the income-tax registers</b> ..... | 12 |
| 2.1 The accounting identity .....                                       | 12 |
| 2.2 The consumer expenditure survey data .....                          | 14 |
| 2.3 Income and wealth information .....                                 | 14 |
| 2.4 The accounting imputation .....                                     | 17 |
| <b>3 An alternative imputation method</b> .....                         | 22 |
| 3.1 Sources of measurement error .....                                  | 22 |
| 3.2 An alternative imputation procedure .....                           | 23 |
| 3.3 The quality of the adjusted imputation .....                        | 25 |
| <b>4 Valid covariance</b> .....   | 28 |
| <b>5 Conclusion</b> .....   | 32 |
| <b>Appendix</b> .....   | 34 |
| <b>References</b> .....   | 39 |
| <b>Sammenfatning</b> .....  | 41 |



# Summary

We investigate the feasibility of deriving a measure of total expenditure at the household level from administrative micro-data on income and wealth. We use Danish administrative data that provide measures of disposable income and the holding of different assets at the end of the year. To check the quality of our imputation we exploit that the households in the 1994-1996 Danish Expenditure Survey can be linked to their administrative data for the years around the survey year. These matched data offer a unique possibility to construct a measure of total expenditure and to check directly on the reliability of the imputation. The results are promising: administrative register data on income, tax payments, and wealth can be used to construct estimates of total expenditure at the household level that have reasonable properties.



# 1 Introduction

A major impediment to a better understanding of household saving and consumption is the lack of good longitudinal data on expenditures and/or saving. Although various attempts have been made to overcome this problem, none of them are completely satisfactory. One widely used strategy is to use questions regarding sub-components of the total, such as the food expenditure questions in the Panel Study of Income Dynamics (PSID). One of the problems with this information, which is based on recall questions, is that it seems to be very noisy, albeit with no substantial bias; see Browning *et al.* (2002) for a discussion of the current state of play on asking recall expenditure questions. When we then first difference the noisy data we face a real signal extraction problem. It is also the case that the dynamics of food expenditures may be quite different from the dynamics of other non-durables and durables. An analysis based on food expenditure therefore tells us little directly about the dynamics of expenditure for other goods such as alcohol, tobacco and entertainment. Although attempts have been made to use the expenditure information in the PSID to impute total non-durable expenditure, see Skinner (1987) for the original attempt and Browning *et al.* (2002) for a discussion of imputing total expenditure from survey data, this strategy has not been widely used.

An alternative and widely used strategy to overcome the lack of genuine panel data is to use time series of cross-sections from expenditure surveys to construct quasi-panels (Browning *et al.* (1985)). Although this has proven to be useful in Euler equation estimation, such data cannot tell us a lot about the idiosyncratic dynamics of earnings, asset returns, consumption<sup>1</sup> and saving (see Moffitt 1993).

A third strategy is to impute total expenditure from income and wealth information. In this study we investigate the quality of such a measure of total expenditure derived from administrative income-tax-register data at the individual household level. To do this we use Danish register data, which give longitudinal wealth and income information on *all* individuals in Denmark from 1981 to 1996. These data ultimately derive from the fact that all Danes (or residents) have a personal number and this personal number is recorded in a great many transactions such as the tax return, bank accounts and interactions with the unemployment and welfare systems. These data are then collected centrally and are made available for research purposes (with, of course, due precautions to maintain confidentiality). In general, the use of administrative data for research has advantages and disadvantages. A primary advantage is that often the coverage of administrative data is wide so that we end up with large and representative samples. Moreover, many variables are better reported in administrative data than they are self-reported in surveys; income from government transfers is a classic example. The main disadvantage of administrative data is that they are not collected for research purposes and often critical information is missing. This then entails the use of imputation or latent variable techniques. An apposite example of information that is never recorded in administrative data is household expenditure information – the subject of this paper.

Our basic imputation method uses the formal version of the accounting identity that total expenditure is equal to income minus the change in wealth over the period. We can construct such measures because Denmark had a wealth tax from 1981 until 1996 which lead to the details of wealth holdings being automatically reported (by banks and other financial intermediaries) to the tax authorities for all Danish tax payers, even though the great majority of these were not liable to the wealth tax. Thus accurate reports of, for example, cash in the bank and bond holdings at the end of each year are available even for low earnings households who have no substantial asset income. Combined with information on income that is also reported automatically to the tax authorities from various sources during the calendar year, this allows us to construct imputed total expenditure dur-

ing the year. The high quality of this income and wealth data makes the register data well suited for this type of imputation.

The imputation requires access to data on disposable income and two or more years of data on wealth. This type of information is not special to Denmark and is available in other data sets such as the PSID and the Health and Retirement Study (HRS). For example, Hubbard *et al.* (1994) use two waves of the PSID with the wealth modules for 1984 and 1989 to calculate asset accumulation as a measure of savings. Ziliak (1998) attempts a similar approach to ours to imputing total consumption using the PSID data. One drawback of using the PSID for this purpose is that wealth is only recorded every fifth wave, so that annual information on wealth needs to be imputed. The imputation of savings is thus based on differencing a wealth measure that is itself imputed. The HRS also contains the necessary information for imputing total expenditure, and has the advantage over the PSID that households are interviewed about income and wealth every second year since 1992. It is also the case that the collection of wealth information in general surveys (often using unfolding brackets) is becoming more common; see, for example, the English Longitudinal Survey on Ageing (ELSA) and the Survey of Health, Ageing and Retirement in Europe (SHARE). Thus the techniques discussed here may have wider applicability. More ambitiously, it may be that the experience of using income and wealth data to impute consumption on currently available data can be used to inform future survey design.

One immediate objection to the wealth differencing method is that it is very noisy and the noise may not have zero mean. In this respect, the data we use are probably a good deal more reliable than the data collected in surveys. Moreover, we do not have any sample selection problems. Thus our analysis should be thought of a »best case« scenario: if we cannot reliably impute consumption using the high quality data we have then there is little hope for noisier data. Nonetheless, as we document below, significant biases may persist. To investigate the effect of these biases we use data drawn from the Danish Family Expenditure Survey (DES) for the years 1994-1996. These data give diary and interview-based information on expenditures on all goods and services, which can then be aggregated to give total expenditure in a sub-period within the calendar year. In common

with most expenditure surveys, these data are thought to be of high quality. What is particularly fortunate for us is that the households in the DES can be linked to their administrative records (including their asset and income information) for the years around their survey year. That is, we can make an exact match between the DES and the register information, and directly check the validity of our imputation methods for this cross section. Although we cannot check the time series properties of our imputation, these data offer a unique possibility to both construct measures of total expenditure and to check directly on the reliability of our imputations.

In the next section the basic approach to imputing a measure of total expenditure from the income-tax registers is outlined. The imputation is implemented and compared with total expenditure data from the DES. In section 3 we address the issues of measurement error and outline a method that is designed to minimise the impact of such errors on the imputation. In section 4 the basic and the corrected measures are compared. Comparison involves a check of the performance of the two measures in terms of correlations with household demographics. Since these have not been used in the imputations, this provides an independent check of our imputation methods. Section 5 concludes the paper.

## 2 Deriving total expenditure from the income-tax registers

### 2.1 The accounting identity

The simplest approach to deriving an expression for total household consumption from the income-tax register is based on a simple accounting identity in which total expenditure is calculated by subtracting savings components from disposable income for the household. The calculation of total disposable income from income-tax registers is, in principle, straight forward, while savings components are identified by calculating changes in wealth from the end of one tax year to the end of the next. In this subsection we define the identity that forms the basis for deriving total expenditure from income-tax registers at the household level. In sections 2.2 and 2.3 we describe the DES and the register information, respectively. In section 2.4 we assess the quality of the consumption imputation by comparing the imputed measure with the level of total expenditure as stated by the households in the DES.

The household begins year  $t$  with a portfolio (vector) of assets  $\{A_{kt-1}\}$  where  $A_{kt-1}$  is the level of asset  $k$  at the end of period  $t-1$ . These assets are held throughout the year and earn a return of  $i_{kt}$  for asset  $k$ . During the year the household also receives earnings (including transfers) of  $e_t$  and pays taxes of  $\tau_t$ . Total expenditure throughout the year is given by  $c_t$ . At the end of the year the household sells the assets  $\{A_{kt-1}\}$  at prices  $p_{kt}$  and buys a new portfolio  $\{A_{kt}\}$  at the same prices. The identity of revenue and purchases gives:

$$\begin{aligned}
c_t + \sum_k p_{kt} A_{kt} &\equiv \left( e_t + \sum_k i_{kt} A_{kt-1} - \tau_t \right) + \sum_k p_{kt} A_{kt-1} \\
&= y_t + \sum_k p_{kt} A_{kt-1}
\end{aligned} \tag{1}$$

so that consumption equals disposable income,  $y_t$ , if the agent leaves the end-of-period- $t$  value of the portfolio unchanged. If we could observe disposable income and all assets and asset prices then we could use this equation to construct a measure of consumption,  $c_t$ . In our data, however, we do not observe the stock of each asset (except for housing), but only the values of each at the beginning and end of the year:  $W_{ks} = p_{ks} A_{ks}$  for  $s = t-1, t$ . To deal with this, we re-arrange equation (1) to give:

$$\begin{aligned}
c_t &= y_t + \sum_k p_{kt} A_{kt-1} - \sum_k p_{kt} A_{kt} + \sum_k p_{kt-1} A_{kt-1} - \sum_k p_{kt-1} A_{kt-1} \\
&= y_t - \Delta W_t + \sum_k (p_{kt} - p_{kt-1}) A_{kt-1}
\end{aligned} \tag{2}$$

where  $W_t = \sum_k W_{kt}$  and  $\Delta$  is the first difference operator. The final term on the right hand side is the capital gains on the portfolio held at the beginning of the year. The price-change term is not observed. As we shall see below, most of our asset groups are composed of quite diverse assets which have very different returns; for example, our first asset group includes both cash-in-hand and interest-bearing bonds. Consequently, we do not attempt to construct a measure of the final term, except for housing. Below we shall return to the effect of ignoring capital gains. Thus we use the following equation for imputation<sup>2</sup>:

$$c_t^* = y_t - \Delta W_t \tag{3}$$

We refer to this as the *accounting imputation* since it is based directly on the accounting identity. In section 3 we shall present an alternative imputation method.

## 2.2 The consumer expenditure survey data

The sample used consists of the 3,866 households in the Danish Family Expenditure Survey (DES) for 1994-1996.<sup>3</sup> These households are surveyed at different times of the year so that observations are distributed across the calendar year. Each household keeps a diary for two weeks, in which they record a detailed account of all expenditures in the household. The non-durable components of this information are scaled (by 365/14) to obtain a measure of annual non-durable consumption. Each household also participates in an interview in which they state their holdings of durables and the purchase of durables within the 12 months before the interview date. All values are measured in current prices. The information from the DES is used to form a measure of annual total expenditure for each individual household. It is this measure that is used to evaluate the performance of the imputation based on equation (2). To control for seasonality, we regress total expenditure from the DES on a series of zero-sum monthly dummies, so that the mean is preserved, before we compare with the imputation. A more detailed description of the total expenditure measure is given in the appendix.

Besides information about total expenditure the DES collects information by interview on payments to pension schemes and some demographic information to characterise the household. Information about the characteristics of the dwelling occupied by the household is obtained from a public administrative register of buildings and merged with the DES. The DES does *not* collect interview-based information about income, saving (apart from pension contributions) and wealth.

## 2.3 Income and wealth information

The accounting imputation is based on the income and wealth information available in the income-tax registers. Since the content of the income-tax-register data is linked to the nature of the tax system, we present a short description of the tax system before turning to the description of the register data and the accounting imputation. Income taxation is on a strictly personal basis with all persons filling out a tax form early in the year after the

relevant tax year. The tax form contains sections that summarise labour gross earnings and capital gross income. For employed persons, gross earnings are reported directly from the employer to the tax authorities, and most components of capital income are reported directly by banks and other financial intermediaries. This information is pre-printed on the tax form that each person receives.

The tax form also contains a section summarising wealth. Again, for most wealth components the amounts are reported directly by banks and financial intermediaries and, once again this information is pre-printed. A few items on the tax form rely on self reporting. This is the case for some components of capital income, such as interest from non-deposited bonds and mortgage deeds, and for some components of wealth, such as non-deposited bonds and mortgage deeds, and the value of cars and boats. If the household does not have any corrections or additions the tax form only needs to be signed and returned. Based on the information on the tax form, the final tax payments are calculated. It is the information from this tax form and the final tax payments that we gain access to in the administrative income-tax registers. Income from shares, capital payments from pension funds, and inheritance are taxed separately, and information on these components is included in our calculation of disposable income (see the appendix for details).

As a general rule all types of income are taxable, including transfers such as pensions, unemployment benefits and social assistance. Child benefit is paid to all families with children and is exempt from taxation. This information is, however, recorded in the registers and taken into account in our calculations of disposable income. Additionally, two types of income are exempt from taxes: contributions to pension schemes and realised capital gains from the sales of quoted shares below a specified amount. These latter are not accounted for in our calculations. Contributions to pension schemes are tax deductible, and are subsequently subject to tax payments when paid out again. Funds accumulated in pension schemes do not enter the wealth data because such funds are not liable to the wealth tax. Contributions to employer-organised pension schemes are withheld by the employer before paying out the salary, and do not appear on the tax form. Therefore, such contributions enter neither our income nor



our wealth data so that we do not have to take these into account in our imputation. Contributions to privately organised pension schemes are also tax deductible. These contributions enter the income data because they are paid into the scheme after the salary/income has been paid out. They do not, however, enter the wealth data because pension funds are not liable to the wealth tax. We are therefore not able to treat these contributions as savings in our imputation<sup>4</sup>.

The information on income and wealth is obtained from the income-tax registers for all persons in a household in the DES for the year of the DES. Furthermore, information about wealth is obtained for the data from the previous year, so that a change-in-wealth measure can be calculated. The income-tax registers contain individual records with the information appearing on the tax form, that is, total income, total tax payments and the value of assets and liabilities. The information is merged with the DES using the civil registration number (personal number) that is unique to each person and is recorded in both data sources. In this way we can make a perfect match between register information and information from the DES. The register information on income, tax payments and wealth is used to form a measure of disposable income, and change in wealth, so that a measure of total expenditure can be constructed according to equation (3). The calculation of disposable income is described in more detail in the appendix. For homeowners it is possible to trace if the household has moved or made extensions to the existing stock of housing assets. In this way it is possible to evaluate if any changes has taken place in the physical stock of housing assets. Consequently, we take into account the capital gains relating to housing assets.

Because of the treatment of capital gains for homeowners we check our imputation for renters and for house owners separately.<sup>5</sup> In this way we are able to focus on renters who have a relatively simple wealth position, before moving to the more complicated case for homeowners. Owner status is identified from the information obtained by interviews in the DES, from information in the public administrative building register, and information in the income-tax register. It is required that the information from these different sources is consistent for the households to be categorised as either renters, owners or living in co-operative housing. For example,

households recorded as owners in the DES are also required to have positive value of housing assets in the income-tax register. Households that are recorded as renters, but at the same time have housing assets, are left out. Moreover, households that are registered with housing assets in one year, but not the other, are also left out. Next, households with self-employed individuals are left out because such individuals have highly unstable income-tax conditions, and because own business wealth is not likely to be measured well. This leaves 3,433 from the original sample of 3,866 observations. Out of these 1,432 are categorised as renters.

## 2.4 The accounting imputation

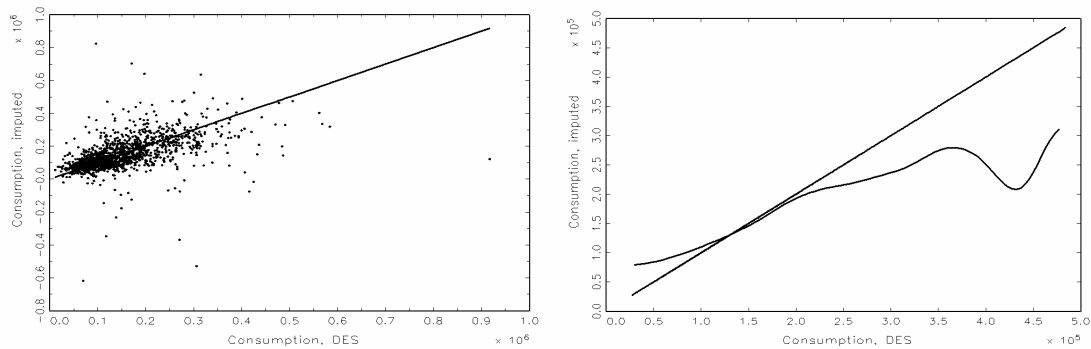
Measures of the distributions of total expenditure from the DES, the accounting distribution and individual differences between the two are given in table 2.1. Comparing the distribution of total expenditure from the DES with the accounting imputation for renters, cf. column 1 and 2 of table 2.1, it is first and foremost seen that total expenditure at the three quartiles is quite similar between the two measures. There is a tendency, though, that the accounting imputation has more spread than the DES measure. Also, the accounting imputation produces some very large negative values. In all, the accounting imputation generates 23 observations with negative values for renters.

Table 2.1 Quartiles of the distribution of total expenditure from the DES and of the accounting imputation and individual differences between the two measures for renters

|                       | DES<br>(1) | Accounting<br>(2) | Individual difference<br>(3) |
|-----------------------|------------|-------------------|------------------------------|
| # observations: 1,432 |            |                   |                              |
| Minimum               | 11,161     | -618,888          | -727,292                     |
| First quartile        | 90,324     | 86,493            | -30,529                      |
| Median                | 128,294    | 125,516           | 1,516                        |
| Third quartile        | 188,006    | 192,032           | 34,551                       |
| Maximum               | 916,956    | 823,286           | 835,557                      |

The pattern indicated from column 1 and 2 is potentially misleading since it is not the same households that are observed at the quartiles. One of the great advantages of our data sources is that we can compare the two measures of total expenditure on an individual basis. The third column of table 2.1 presents the distributions for individual differences between the two measures. The accounting-based imputation for renters is presented graphically in figure 2.1. The left panel gives the scatter plot of the DES measure of consumption against the imputed measure based on the accounting imputation, and the right panel presents the associated kernel regression. The comparison of the DES measure and the imputed measure at this individual level suggests that for the bulk of the data the accounting imputation matches the DES measure of total expenditure quite well, but for a few observations the imputation does a really poor job. Moreover, the nonparametric fit indicates that we overstate for low consumption households and underestimate for high consumption households. We conjecture that this is due to ignoring capital gains. For low consumption households the only asset is cash-in-hand which has a negative capital gain (since inflation is positive). Since we do not allow for this, consumption is overstated for these households (compare equations (2) and (3)). Conversely, high consumption households may experience significant positive capital gains. Allowing for this would increase the imputed value of total expenditure. In 1995 the rate of inflation was 2.1%, share prices rose by 11% and bond prices by 14.8% so that there is ample scope for corrections to allow for capital gains; this is left for future work.

Figure 2.1 Accounting-based imputation for renters. Left panel plots imputed consumption against the DES measure, together with the diagonal. The right panel depicts the kernel regression together with the diagonal



Note: Kernel regression is estimated during the Gaussian kernel on a sample that is trimmed by 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation<sup>6</sup> to be 23,208.

The accounting-based imputation is more complicated for house owners because of the particular design of the Danish mortgage system. The main problem is households re-financing their mortgage which impacts on the wealth at the end of the year; details are given in the appendix. The distribution of the total expenditure from the DES, the accounting imputation and the individual differences between the two former for house owners and people living in co-operative housing are given in table 2.2. The quartiles of the two measures (columns (1) and (2)) are quite close to each other. As before, the accounting imputation exhibits more spread than the DES measure and generates some very large negative values. For owners the accounting imputation generates 58 observations with negative values. The accounting imputation for homeowners, however, also generates some large positive values. Considering in column (3) the distribution of the individual differences between the DES measure and the accounting imputation it appears that the deviations are bigger at the lower and upper quartiles for owners than for renters. This is not so surprising, since the link between change in wealth and consumption is less direct for house owners than for renters. When the accounting imputation overstates consumption it

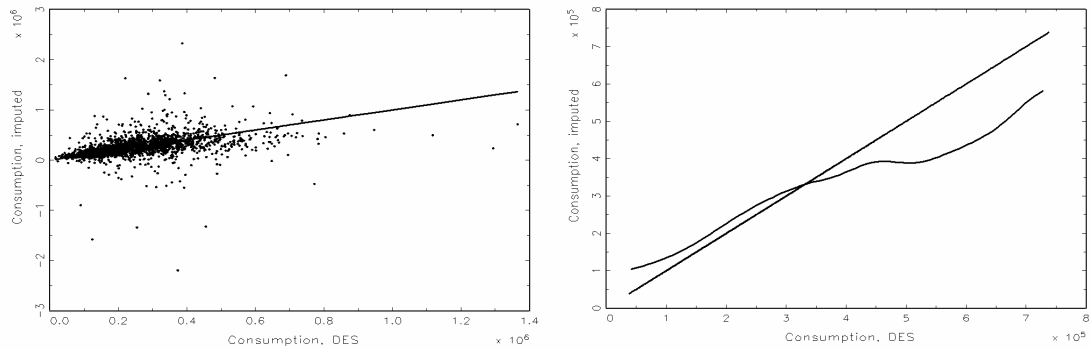
can, for example, be a consequence of re-mortgaging, and when it understates it can be a consequence of unobserved capital gains on other assets.

Table 2.2 Quartiles of the distributions of total expenditure from the DES, the accounting imputation and the individual difference between the two measures for house owners and people in co-operative housing

| # observations: 2,001 | DES<br>(1) | Accounting<br>(2) | Individual difference<br>(3) |
|-----------------------|------------|-------------------|------------------------------|
| Minimum               | 16,861     | -2,197,534        | -1,939,120                   |
| First quartile        | 155,794    | 137,704           | -69,596                      |
| Median                | 239,511    | 239,221           | -5,470                       |
| Third quartile        | 333,582    | 356,213           | 62,395                       |
| Maximum               | 1,365,365  | 2,325,625         | 2,570,891                    |

The accounting imputation for house owners and people living in co-operative housing is presented graphically in figure 2.2. The scatter plot and the nonparametric regression indicate that the imputed measure does reasonably well for the major part of the data. Once again, for low (respectively, high) levels of DES consumption the imputation seems to overstate (respectively, understate) the reported level of consumption.

Figure 2.2 Accounting-based imputations for house owners and people in co-operative housing. Left panel scatter-plots consumption in the DES against the accounting-based imputation together with the diagonal. The right panel depicts the kernel regression together with the diagonal



Note: The kernel regression is estimated using the Gaussian kernel on a sample that is trimmed 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation to be 38,713.

In general, the accounting measure appears to do reasonably well when compared to the level of consumption recorded by households in the DES. There are, however, some problems. In the next section we examine an alternative imputation procedure that exploits the fact that we have matched DES and administrative data which can be used in the construction of the imputation method as well as in evaluation.

## 3 An alternative imputation method

### 3.1 Sources of measurement error

The accounting identity-based imputation of total expenditure made in the previous section is characterised by performing reasonably well in terms of matching the level of consumption stated in the DES. As mentioned the imputed measure is rather noisy, and there is a tendency for the accounting imputation to underestimate total expenditure at high levels of reported total expenditure. A number of reasons exist as to why the imputation will not match exactly with the interview-based measure of total expenditure. First, some noise is introduced because of the particular timing of the variables entering the imputation and the DES. The imputation is strictly for the whole of the calendar year. However, the DES measure is a composite of a 14 day measure of expenditure on non-durables and a one year measure of durable expenditures. Thus for a household surveyed in the first two weeks of June 1995, the durables measure refers to June 1994 to May 1995. Although controlling for seasonality may moderate the effect for non-durables, it will not help for durables. Additionally, although the DES information is thought to be of high quality, it is not free from measurement error and infrequency problems, even for non-durables.

A mismatch between the accounting imputation and the DES measure can also be attributed to problems with the administrative data. First, the reporting of asset values is subject to error, particularly for those assets that are self-reported (such as cars). Second, high value items such as boats and cars enter differently in the two data sources. In the DES meas-

ure they are considered consumption whereas in the accounting imputation the purchase of a car may imply a decrease in, say, cash holdings, and an increase in »other wealth«, because the value of cars and boats is liable to the wealth tax. Furthermore, we have not controlled for capital gains. Finally, as already mentioned, the particular design of the Danish mortgage system can create a mismatch between the imputed measure and the actual level of consumption for one particular year. There is thus reason to believe that both the DES measure and the accounting imputation based on the register information are noisy measures of total expenditure during a calendar year. In all that follows we shall ignore the DES errors (apart from allowing for seasonality) and concentrate on an alternative imputation method that uses information from the DES as well as the administrative data.

### 3.2 An alternative imputation procedure

To generate an alternative imputation method we adapt the methods of Lubotksy and Wittenberg (2001) to our context. Re-writing equation (3) with the change in wealth broken up into its components, we have that saving by household  $h$ ,  $s_h$ , is given by:

$$s_h = y_h - c_h = \sum \Delta W_{hk} \quad (4)$$

(where we have dropped the time subscripts, for convenience). We have measures of  $c_h$  and  $y_h$  from the DES and the administrative data, respectively. Using the matched data we can construct a measure of saving for each household in the DES. According to the equation above, this should be the sum of the changes in the values of the individual assets. This suggests using the regression:

$$s_h = \beta_0 + \sum_k \beta_k \Delta W_{hk} + \eta_h \quad (5)$$

for imputation. A naïve approach would be to take the OLS estimates of the  $\beta$  parameters and then to use these to predict savings for households in the administrative data that are not in the DES:



$$\hat{c}_h = y_h - \left( \hat{\beta}_o + \sum \hat{\beta}_k \Delta W_{kh} \right) \quad (6)$$

Unfortunately this does not take into account that the  $\Delta W_{kt}$  elements may be measured with quite disparate degrees of precision. Moreover, the measurement errors in these elements will give rise to attenuation biases in the parameter estimates. To control for this, Lubotksy and Wittenberg (2001) suggest the following procedure, which minimises the effect of attenuation bias on the estimate of savings. First calculate the covariance between the constructed saving measure and the individual changes in assets:

$$\hat{\rho}_k = \frac{\text{cov}(s, \Delta W_k)}{\text{cov}(s, \Delta W_1)} \quad (7)$$

(so that  $\hat{\rho}_1 = 1$ ). Then construct an imputed measure of consumption by:

$$\tilde{c}_h = y_h - \sum \hat{\beta}_k \hat{\rho}_k \Delta W_{hk} \quad (8)$$

Thus we scale each OLS parameter estimate by the relative covariance of the change in the associated asset and the constructed savings measure.

As before, we do everything separately for renters and owners. The regressions are presented in table 3.1. All the estimated parameters are significant and have the expected signs. An increase in asset holdings implies higher saving and an increase in liabilities is associated with lower saving. A one-unit increase in cash holdings corresponds to a 0.34 unit increase in the constructed savings measure for renters. A one-unit change in the change-in-other-assets implies 0.17 change in savings. This is a smaller effect than for the change-in-cash component. Recalling from equation (2) that capital gains are not considered savings, and that we have not been able to control for capital gains in the change-in-other-assets component this relative magnitude is expected. The coefficients for owners are uniformly closer to zero. These estimates are used to form an estimate of  $\tilde{c}_h$  according to equation (8).

Table 3.1 Regression of saving on wealth components

| Dependent variable                             | Renters |                 | Owners |                 |
|--|---------|-----------------|--------|-----------------|
|  | $\rho$  | $\beta$         | $\rho$ | $\beta$         |
| Change in cash holdings                        | 1       | 0.34<br>(0.08)  | 1      | 0.27<br>(0.04)  |
| Change in other assets                         | 1.20    | 0.17<br>(0.05)  | 0.26   | 0.11<br>(0.03)  |
| Change in liabilities                          | -0.76   | -0.19<br>(0.07) | -1.17  | -0.10<br>(0.02) |
| Constant                                       | -       | 77,596          | -      | 186,418         |
| $\left(\sum \hat{\beta}_k \hat{\rho}_k\right)$ |         | 0.689           |        | 0.423           |
| Number of observations                         |         | 1,432           |        | 2,001           |

Note: Robust standard errors in brackets.

### 3.3 The quality of the adjusted imputation

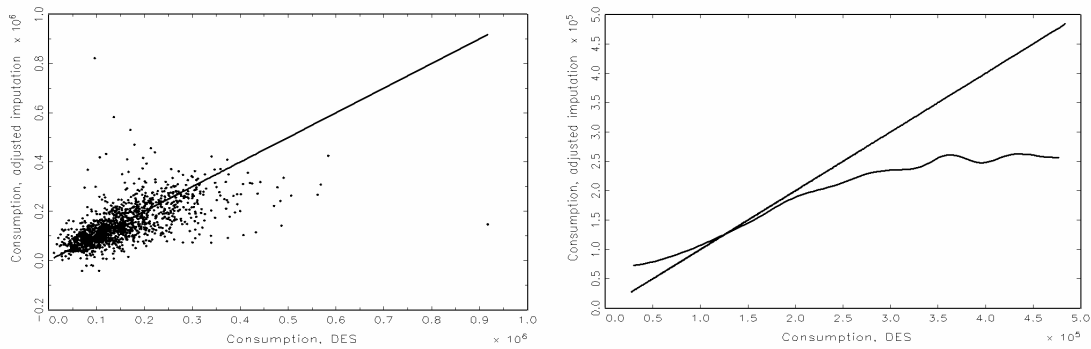
We now have two imputed measures of consumption: the accounting imputation,  $c^*$ , from equation (3) and the adjusted imputation measure,  $\tilde{c}$ , from equation (8). We first present the root mean squared error of the two measures as compared to the DES survey value:<sup>7</sup>

| RMSE    | Accounting, $c^*$ | Adjusted, $\tilde{c}$ |
|---------|-------------------|-----------------------|
| Renters | 89,287            | 71,247                |
| Owners  | 204,394           | 132,289               |

As can be seen the adjusted measure reduces the root mean squared error considerably for both renters and owners. The relative gain seems to be larger for owners. The gain is expected to be bigger for owners since the link between change-in-wealth and savings is less direct for this group due to the mortgaging system and for other reasons. Altogether this is taken as evidence that incorporating the survey information has improved on the imputed measure's ability to match the total expenditure measure from the DES.

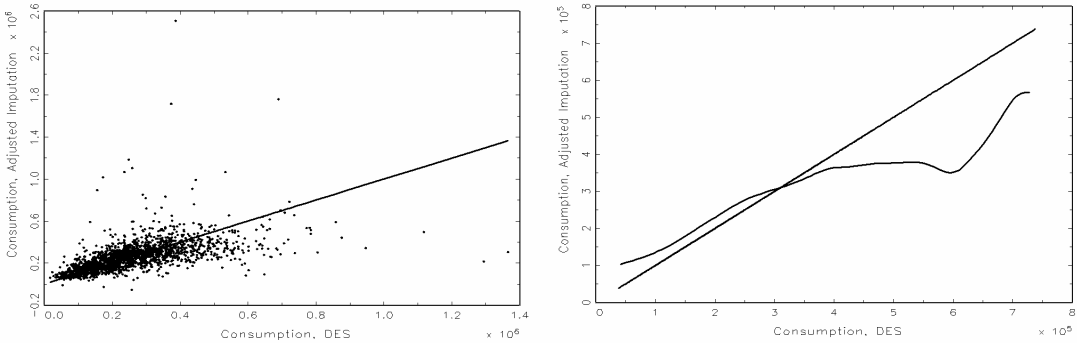
As before, we also present scatter plots and non-parametric regressions to gauge the fit (figures 3.1 and 3.2 for renters and owners, respectively). The scatter plots indicate the adjusted measure imputes fewer observations with extreme negative values than did the accounting imputation. However, considering the nonparametric fits in the right panels of figure 3.1 and 3.2 it becomes evident that the adjusted imputation also suffers from »undershooting« at higher levels of consumption. This is not, perhaps, too surprising since the adjusted imputation is designed to deal with the measurement error in the changes in asset values, but not the unobservability of capital gains.

Figure 3.1 Adjusted imputation against DES measure for renters. Left panel scatter-plot adjusted imputation against the DES measure together with the diagonal. The right panel depicts the kernel regression together with the diagonal



Note: The kernel regressions are estimated using the Gaussian kernel on a sample that is trimmed 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation to be 17,969.

Figure 3.2 Adjusted imputation against DES measure for owners. Left panel scatter-plot adjusted imputation against the DES measure together with the diagonal. The right panel depicts the kernel regression together with the diagonal



Note: The kernel regressions are estimated using the Gaussian kernel on a sample that is trimmed 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation to be 29,974.

## 4 Valid covariance

The ultimate purpose of imputing total expenditure is to obtain a measure of total expenditure that can be used, for example, in analysing the allocation of income between consumption and savings. So far, evidence has been presented that the pure accounting-based imputation provides a measure that does quite well for most of the data. Exploiting the information in the DES in creating an adjusted measure does seem to improve the performance in terms of the ability of the imputed measure to predict total expenditure from the DES. However, the validation tests carried out in section 2 and 3 focus only on the ability of an imputed measure to predict the total expenditure measure in the DES. This internal approach to validating the imputed measures completely neglects how the imputation approach affects the covariance of total expenditure with variables that are not used in the imputation, such as demographic variables. This is the valid covariance in terms of applying the imputation for analytical work.

The objective of this section is to investigate if the imputed measures have the same characteristics as the DES measure of total expenditure in terms of covariance with a number of important demographic variables describing family composition, age, and labour market participation. No »structural« interpretation should be given to these descriptive estimates since we do not condition on any measure of lifetime wealth. The parameter estimates obtained by regressing log-consumption from the DES on demographics are presented in column 1 of table 4.1. In column 2 and 3 the same exercise is made for the accounting and the adjusted imputations. All regressions are based on a sample in which observations with negative values of the accounting imputed consumption are dropped.

For the DES measure we find significant children, age and labour market participation effects; these are in line with effects found in studies on data from other countries. The presence of children is associated with increased consumption of 13-20%, and the single adult households are associated with a level of consumption that is about 44% lower than for households with two adults. Further, two dummies, one indicating age below 40 and another indicating more than 67 years of age, are included. The first dummy indicates that younger people tend to have a consumption level that is around 14% lower than households with the oldest person aged 40-67. Older households are found to be associated with a level of consumption that is about 18% lower. Finally, consumption is higher for higher labour supply.

Comparing the estimates in column 1 with the parameter estimates based on the accounting imputed measure, in column 2, it is seen that estimates resemble those of the DES based regression quite closely. All estimates are associated with a larger standard error, which is not surprising considering the results from sections 2 and 3. Differences do exist, though. Single adult households are now estimated to have a slightly lower level of consumption than what is indicated from the DES based regression. Estimated children effects are in most cases quite close to the corresponding estimates in column 1. The most important deviation seems to be that parameter estimate of school children is a bit lower than what is found when estimates are based on the DES measure of total expenditure. Estimates of age effects appear to match those presented in column 1 reasonably well. The dummy estimate for people aged more than 67 is a bit lower than the corresponding estimate in column 1, but the deviation is less than two standard deviations, though. Estimated labour supply effects match the estimates in column 1 quite well.

Table Regression of ln(Total expenditure) on demographics for three  
4.1 measures of expenditure

|                                     | 1.<br>DES           | 2.<br>Accounting    | 3.<br>Adjusted      |
|-------------------------------------|---------------------|---------------------|---------------------|
| Dependent variable, ln C            | $C = C^{DES}$       | $C = C^*$           | $C = \tilde{C}$     |
| Single adult                        | -0,4363**<br>0,0185 | -0,4870**<br>0,0232 | -0,4724**<br>0,0169 |
| # children, aged 0-7                | 0,1948**<br>0,0200  | 0,1838**<br>0,0265  | 0,1987**<br>0,0176  |
| # children, aged 8-14               | 0,1362**<br>0,0202  | 0,0772**<br>0,0276  | 0,0641**<br>0,0185  |
| # children, aged 15-20              | 0,1826**<br>0,0222  | 0,1687**<br>0,0324  | 0,1304**<br>0,0194  |
| # children, aged 21-30              | 0,1215**<br>0,0434  | 0,1983**<br>0,0704  | 0,1797**<br>0,0356  |
| Age of the oldest person <40, dummy | -0,1414**<br>0,0171 | -0,1573**<br>0,0229 | -0,2544**<br>0,0163 |
| Age of the oldest person >67, dummy | -0,1793**<br>0,0252 | -0,1268**<br>0,0345 | -0,1016**<br>0,0230 |
| Work time, man                      | 0,0074**<br>0,0005  | 0,0084**<br>0,0006  | 0,0092**<br>0,0005  |
| Work time, woman                    | 0,0066**<br>0,0005  | 0,0056**<br>0,0007  | 0,0072**<br>0,0005  |
| Constant                            | 12,0179**<br>0,0202 | 12,0331**<br>0,0274 | 12,0181**<br>0,0207 |
| Number of observations              | 3.352               | 3.352               | 3.352               |

Note: Robust standard errors in small numbers below parameter estimates. \*\* indicates significance at 5%.

In the regression based on the adjusted imputation, cf. column 3, an effect relating to single adult households is found that is quite similar to the one obtained in the regression based on the accounting imputation. The estimate is more than two standard deviations away from the estimate obtained in the regression based on the DES measure. Children effects are quite similar to the ones obtained in column 2. The parameter estimates associated with children age groups 8-14 and 15-20 are more than two standards errors from the estimates in column 1. Age effects are qualitatively similar to the age effects obtained in the regression based on the DES measure, but their magnitude is quite different. Labour supply effects appear to be slightly overestimated for the male part whereas the es-

timated female labour supply decision seems to match the estimate in column 1.

In summary, the pure accounting-based imputation appears to do at least as well as the adjusted imputation when it is evaluated on its covariance with demographic effects. In some cases it seems to match the DES estimates even better than the adjusted measure. The most important difference between the two imputed measures appears to be that the accounting-based measure is somewhat noisier than the adjusted measure. The intuition behind this is that the adjustment for measurement error is targeted towards improving the ability of the change-in-wealth measure from the registers to predict the level of the DES measure. The adjustment is not concerned with improving the covariance structure of the register-based measure with variables not used in the imputation.



## 5 Conclusion

We have investigated the feasibility of constructing a reliable measure of household total expenditure from data on income and wealth. Our data have two distinct advantages over most other data. First, we have reasonably good measures of income and wealth for *all* households in Denmark. Second, we can exactly merge our administrative data to an expenditure survey (the DES) to directly check the quality of the imputation for those households that are in the DES. Our simplest imputation procedure is based on an accounting identity which ignores capital gains and measurement error in the asset values; that is we set imputed consumption equal to disposable income minus the change in (nominal) wealth. We find that this measure performs quite well in terms of matching individual households' self-reported total expenditure. The accounting measure also performs reasonably well in terms of its covariance with demographic variables. The latter is an important »external« check of validity since we do not use the demographic information in the imputation. The main problems with the accounting measure are that it is noisy (sometimes producing large negative values) and that it understates consumption for households with high levels of DES reported consumption.

In section 3 we investigated the performance of an alternative imputation measure that is designed to deal with the measurement error in the wealth information. Since this procedure requires information on consumption (or saving), disposable income and changes in individual components of wealth for the same household, the informational requirements needed to implement this imputation are significantly greater than for the simple accounting approach. This considerably reduces the scope of this applica-

tion; indeed, we are not aware of any other data in the world that could be used to implement this imputation procedure. The adjusted imputation does perform better in terms of the fit to the DES data, but it still exhibits the same bias as the accounting measure. On balance, it does not seem that the extra information needed for the adjusted procedure justifies the gains.

The results presented here are encouraging. The simple accounting procedure works reasonably well and »only« requires information on disposable income and wealth at two points in time; such data are now collected in several surveys such as the HRS and the PSID. Allowing for the noisiness of the measures of the values of different assets (the adjusted procedure) does not seem to bring any great gains in the quality of the imputation. The main reservation is in the treatment of capital gains (which we ignore); it is hoped that future work will provide some insight into how to deal with these and whether this reduces the »tilt« of the imputed measure as compared with survey data. Apart from this, our results indicate that for the bulk of the sample the simple accounting procedure does provide a reliable measure of total expenditure at the household level.

# Appendix

## Definitions

### Consumption in the expenditure survey

Consumption = Goods and services\*

- + Rental value of owner-occupied house
- + Contribution to unemployment insurance
- + Interest payments
- + Payments to alimony and child maintenance
- + Stamp duties, fees to authorities and fees in connection with house deal
- + Gifts, charity
- + Cost of extension/reconstruction of house
- + Contributions to privately organised pension schemes

\* Food, drink, clothes and shoes, housing, energy, furnishings, health, transport and communication, leisure and entertainment, cars, boats, other goods and services. Note, property taxes (ejendomsskatter) are included in housing expenditures.

### Disposable income from income-tax register information

The income-tax registers contain information about annual gross taxable income, tax and wealth as it is recorded in the income-tax register. A measure of disposable income is formed using total gross taxable income, as it appears on the yearly tax form, less payments of taxes. From this

rental value of the house for households in owner-occupied houses is deducted, as this is an income component introduced by the tax authorities for collecting taxes, and is hence not associated with any cash-flow. Furthermore, net payments from capital pensions, and income from shares, and inheritance are added, as these income components are taxed separately. Finally, as we are looking for a money-in-the-pocket measure, child benefits, which are tax exempt, are also added.

Disposable income = Gross taxable income according to the final tax form (including social welfare, unemployment insurance, alimony, public pensions, capital income other than income from shares)

- + Payments from capital pensions
- + Child benefits
- + Income from shares
- + Inheritance
- Total final taxes according to tax statement
- Taxes on payment from capital pension
- Rental value of house
- Tax on income from shares
- Inheritance tax

Note, that pension contributions are tax deductible. Therefore, pension contributions do not enter the income measure, except for contributions to privately organised schemes.

### **Wealth information in the income-tax registers**

In the income-tax registers a number of wealth variables exist. Definitions, however, vary from one year to the other, and wealth information can consistently be divided only into three categories: Cash holdings, cash value of the house, holdings of other assets. A substantial part of private savings is done in the form of pension contributions. Accumulated pension contributions are not taxable before they are paid out (and then enter as normal income), and do therefore not enter this wealth measure. The contents of the

wealth categories in terms of the variables that are available in the income-tax registers are listed in table A.1.

Table Wealth components

A.1

|  |
|--|
| <p><i>Assets:</i></p> <p><b>Cash holdings</b></p> <p>Cash holdings in bank, bonds, deposited mortgage deeds</p> <p><b>Cash value of house</b></p> <p><b>Other assets</b></p> <p>Value of shares</p> <p>Value of shares, main shareholder in company</p> <p>Own capital in domestic company</p> <p>Own capital in foreign company</p> <p>Other foreign wealth</p> <p>Other taxable assets (*)</p> <p><i>Liabilities:</i></p> <p><b>Sum of liabilities</b></p> |
|--|

(\*) This component is based on self-reporting. It is a »residual« wealth component, and it includes value on non-deposited mortgage deeds and bonds, cars, boats, value of co-operative apartments, premium bonds, unquoted shares in ships.

## The Danish mortgage system

In Denmark the financing of real property takes place via mortgage banks. Mortgage banks offer loans where the borrower's real property is used as collateral for the loan. It is possible to mortgage up to 80% of the property value, and the mortgage can be used to finance consumption of goods and services as well as housing. The most popular type of these loans is annuity loans that are funded by the issuing of callable mortgage credit bonds with fixed coupon rates and with a maturity of up to 30 years. The principal of the loan depends on the price of the underlying bond. When the bond price is below par a higher number of bonds must be sold to meet the funding requirements. This typically makes the principal of the loan larger than the loan proceeds paid out.

An important feature is that a borrower is entitled to redeem a callable bond *at par* at any time prior to maturity, for example by prepayment. This enables the borrower to exploit changes in the market rate of interest in order to reduce the costs of funding. If the interest rate falls, the borrower may prepay his loan, and raise a new loan at the lower coupon rate. This may lower his monthly net payment, but may also imply a larger principal of the new loan relative to the old loan if the price of the bond underlying the new loan is below par. The mechanics of re-mortgaging may be illustrated most clearly by an example. The example given below ignores many relevant aspects of a re-mortgaging, for example, the influence of the tax system.

*Example: Consider a household with a mortgage loan of 1,000,000 DKK. The loan is an annuity loan based on underlying bonds with a maturity of 30 years and a fixed coupon rate of 7%. The loan is assumed to have yearly instalments, and 27 instalments of 83,426 DKK are left before the loan matures. At this point the market rate of interest has decreased compared to when the loan was established so that a bond with a nominal value of 100 and a fixed coupon rate of 5% can now be sold at quote 95. The household decides to pay in the old loan at par. To finance this, a new loan based on bonds with a fixed coupon rate of 5% is established. Because the 5% bonds are quoted below par he needs to establish a new loan with a higher principal to get out proceeds to cover the principal of the old loan. The principal of the new loan becomes 1,052,632 DKK. However, because of the lower coupon rate the yearly instalments are now only 68,475 DKK, and the total payments until maturity are lower than the total payments on the old loan.*

Table A.2 Stylised example of re-mortgaging an annuity loan with a principal of 1,000,000 DKK

|  | Old loan  | New loan  |
|--|-----------|-----------|
| Fixed coupon rate, $r$                 | 7%        | 5%        |
| Maturity, $n$                          | 27        | 30        |
| Principal, $P$                         | 1,000,000 | 1,052,632 |
| Yearly payment, $y$                    | 83,426    | 68,475    |
| Total payments until maturity, $n * y$ | 2,252,495 | 2,054,256 |

Note: 
$$y = P * \frac{r}{1 - (1 + r)^{-n}}$$

The example illustrates that re-mortgaging can imply a decrease in net wealth from one particular year to the next in order to gain lower monthly pay for the remaining time of the loan. The change in wealth thus does not map exactly into a change in total expenditure for this particular period. In real-life the interest rate generally decreased over the period considered. However, a particularly pronounced drop occurred in 1993 and the re-mortgaging activity was particularly high in this period. In the data this shows by the number of house owners increasing the level of liabilities from 1993 to 1994 being higher than for subsequent years.

# References

Browning, M.; T. Crossley and G. Weber (2002): *Asking Consumption Questions in General Purpose Surveys*. CAM working paper 2002-05, Institute of Economics, Copenhagen.

Browning, M.; A. Deaton and M. Irish (1985): A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle. *Econometrica*, 53(3), 503-43.

Härdle, W. (1990): *Applied Nonparametric Regression*. Econometric Society Monographs, no. 19, Cambridge University Press.

Hubbard, R.G., J. Skinner and S.P. Zeldes (1994): Expanding the Life-Cycle Model: Precautionary Saving and Public Policy. *American Economic Review*, 84(2): 174-79.

Lubotksy, D. and M. Wittenberg (2001): *Interpretation of Regressions with Multiple Proxies*. Working paper #457, Princeton University, Industrial Relations Section.

Moffitt, R. (1993): Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross Sections. *Journal of Econometrics*, 59(1-2), 99-123.

Skinner, J. (1987): A Superior Measure of Consumption from the Panel Study of Income Dynamics. *Economics letters*, 23, 213-216.



Statistics Denmark (1999): *The Family Expenditure Survey, Documentation* (in Danish). Statistics Denmark, Copenhagen.

Ziliak, J.P. (1998): Does the Choice of Consumption Measure Matter? An Application to the Permanent-Income Hypothesis. *Journal of Monetary Economics*, 41(1), 201-16.

# Sammenfatning

I dette memo undersøges muligheden for at konstruere et mål for de samlede udgifter for individuelle husholdninger på grundlag af mikrodata for indkomst og formue. Vi bruger danske administrative registre, som indeholder informationer om disponibel indkomst og beholdningen af forskellige aktiver ved udgangen af året. For at undersøge kvaliteten af vores beregnede mål udnytter vi, at det for husholdningerne i den danske forbrugsundersøgelse for 1994-1996 er muligt at sammenkøre informationerne med de administrative registre for årene omkring undersøgelsesåret i forbrugsundersøgelsen. Disse sammenkørte data giver en enestående mulighed for at undersøge pålideligheden af vores beregnede mål for individuelle husholdningers samlede udgifter. Resultaterne er lovende: Administrative registerdata med information om indkomst og formue kan bruges til at beregne et estimat for de samlede udgifter for individuelle husholdninger, som har rimelige egenskaber.

## Notes

1. In this paper we use the terms »total expenditure« and »consumption« interchangeably.
2. With allowance for housing capital gains.
3. This survey is the preferred survey to use for comparative studies of the kind undertaken here. The only previous surveys, conducted in 1981 and 1987, were based on different collection methods, and were, furthermore, hampered by severely high non-response rates, cf. Statistics Denmark (1999). Later surveys do not facilitate merging of wealth data due to the abolition of the wealth tax after 1996.
4. As mentioned in the previous section on the DES, interview-based information about contributions to pension schemes is present in the DES. In order to form a measure of total expenditure from the DES that is comparable to the measure that we can construct from the register data we have included contributions to privately organised pension schemes in the DES measure of total expenditure. In the sample we are analysing 1,057 households (of which 236 are renters) report contributing to a privately organised pension scheme. The median contribution to a privately organised pension scheme among these households is 8,772 DKK. This is to be compared with a median level of total expenditure from the DES of 255,940 DKK in the same group.
5. In Denmark, living in a co-operative is quite common; such households are treated as homeowners since they have a share in the property value.
6. The generalised cross validation criterion picks the bandwidth  $b$  that minimises  $CV(b) = \sum_{k=1}^n (y_k - \hat{g}_b(x_k))^2 \Xi(w_{bk}(x_k))w(x_k)$ , see for example Härdle (1990). In this case  $y_k$  is total expenditure from the DES for household  $k$ ,  $x_k$  is the imputed measure for the same household, and  $\hat{g}_b(x_k)$  is the kernel estimate at bandwidth  $b$ .  $\Xi(w_{bk}(x_k)) = (1 - w_{bk}(x_k))^{-2}$  is the Generalised Cross-Validation function, where  $w_{bk}(x_k) = K(0) / \sum_{h=1}^n K\left(\frac{x_h - x_k}{h}\right)$ , and  $K(\cdot)$  is a kernel.  $w(x_k)$  is a weight function. The  $\Xi$ -function penalises bandwidths that are too small, and  $w(x_k)$  is introduced to reduce boundary effects.
7. For the accounting imputation this is  $\sqrt{N^{-1}(c^* - c^{DES})^2}$ . This is implicitly taking the (seasonally adjusted) DES measure,  $c^{DES}$ , as the »true« value. As we have discussed in the first subsection, the DES measure is also subject to measurement error so that this is not the »true« RMSE of the estimates.