CAM

CAM

Centre for Applied
Microeconometrics

Institute of Economics
University of Copenhagen

http://www.econ.ku.dk/CAM/

Imputing Consumption from Income Tax Registers

Martin Browning
Søren Leth-Petersen

2002-07

The activities of CAM are financed by a grant from
The Danish National Research Foundation

Imputing Consumption

from Income Tax Registers

Martin Browning
CAM, Institute of Economics
University of Copenhagen
Studiestræde 6
1457 Copenhagen K – Denmark
Telephone: +45 3532 3070
E-mail: martin.browning@econ.ku.dk

Søren Leth-Petersen
Institute of Local Government Studies - Denmark
Nyropsgade 37
1602 Copenhagen V - Denmark
Telephone: +45 3311 0300
E-mail: slp@akf.dk

Abtract: In this study we investigate if it is possible to derive a measure of total expenditure from administrative income-tax-register data at the individual household level. We exploit that the households in the Danish Expenditure Survey 1995 can be linked to their administrative registers for the years around the survey year. These matched data offer a unique possibility to both construct measures of total expenditure and to check directly on the reliability of our imputations. The results are promising. It gives clear 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.

Acknowledgements: The research is sponsored by the Danish Social Science Research Council (SSF), grant 24-00-0282 and by the Danish National Research Foundation thorugh its grant to CAM.

1. Introduction

One of the biggest impediments to further development of empirical research on intertemporal allocation seems to be 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 the food expenditure questions in the 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, Crossley and Weber (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 nothing about the dynamics of expenditure for other goods such as alcohol, tobacco and entertainment. Also, attempts have been made to use the expenditure information in the PSID to impute total 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.

Another 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, Deaton and Irish (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[1] and saving (see Moffitt (1993)).

In this study we investigate if it is possible to derive a measure of total expenditure from administrative income-tax-register data at the individual household level. To do this we use

Danish register data, which give longitudinal information on *all* individuals in Denmark from 1981 to 1998. The basic method is to use the accounting identity relating total expenditure in a period to income and the change in wealth across the period. We can construct such measures because Denmark had a wealth tax from 1981 until 1996. This 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 from various sources during the calendar year, this allows us to construct imputed total expenditure during the year.

One immediate objection to the wealth differencing method is that it is very noisy. On this point, one advantage of the register data over survey data is that register data potentially allows us to construct samples covering the whole population, i.e. of infinite size. Another objection to the use of register data is that it is also likely to be subject to unknown biases, even if the wealth data are unusually reliable. To investigate these biases we use data drawn from the Danish Family Expenditure Survey (DES) for the year 1995. 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, so that we can directly check the validity of our imputation methods for this cross section. Although we can not check the time series

[1] In this paper we use the terms 'total expenditure' and 'consumption' interchangeably.

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. Issues of potential measurement errors and a remedy to compensate for such are discussed in section 3. 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, they provide another 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 section we define the identity that forms the basis for deriving total expenditure from income-tax registers at the household level. Section 2.2 presents the data. In section 2.3 the performance of the imputation is investigated by comparing the imputed measure with the level of total expenditure as stated by the households in the CES.

Total household expenditure, for the year we are interested in, is denoted c_t, where lower case t indicates the period of interest. To derive a measure of c_t start out with the following identity describing the total value of the stock of assets at the end of period t:

$$\sum_k p_{kt}A_{kt} \equiv \sum_k p_{kt-1}A_{kt-1} + \sum_k i_{kt}A_{kt-1} + \sum_k \left(p_{kt} - p_{kt-1}\right)A_{kt-1} + e_t - t_t - c_t$$ **Error! Unknown switch**

where p_{kt} is the price of asset k at the end of period t and A_{kt} is the stock of asset k at the end of period t, so that $\sum_k p_{kt}A_{kt}$ is the value of the total stock at the end of period t. This equals the value of the total stock at the end of the previous period, $\sum_k p_{kt-1}A_{kt-1}$, plus returns to these assets, $\sum_k i_{kt}A_{kt-1}$, where i_{kt} is the return, ie. interests or dividends, to asset k, plus capital gains, $\sum_k (p_{kt} - p_{kt-1})A_{kt-1}$, plus earnings, e_t, less taxes, t_t, and what is allocated for consumption, c_t. Rearranging (1.2) yields:

$$c_t \equiv \left(e_t + \sum_k i_{kt}A_{kt-1} - t_t\right) - \left(\sum_k p_{kt}A_{kt} - \sum_k p_{kt-1}A_{mk-1} - \sum_k \left(p_{kt} - p_{kt-1}\right)A_{kt-1}\right)$$ **Error! Unknown swi**

where the first set of brackets yields total disposable income and the second set of brackets yields the change in value of the total stock of assets less capital gains. We refer to this as the *accounting imputation* since it is based directly on the accounting identity. Later we shall present an alternative imputation.

2.2 The consumer expenditure survey data.

The sample used consists of the households that enter the DES for 1995.[2] For these households administrative register data on income, tax payments and wealth at the end of the year corresponding to the survey year and wealth information for the year before this is obtained and merged on to the DES. Merging these register data on to the DES facilitates a comparison of survey data of total expenditure with an imputed measure of total expenditure based on the information in the registers.

The Danish DES from 1995 contains information about 3,866 households. Information about the expenditures for these households has been collected over a period of three years, 1994, 1995 and 1996. The households in the survey have been contacted at different times of the year so that observations are distributed across the calendar year. Each household has participated in a comprehensive interview, where they have answered questions among other things about holdings of durables and about purchase of durables within the past 12 months from the interview date. Furthermore, each household has kept a diary for two weeks, where they have kept a detailed account of all expenditures in the household. This information is scaled to obtain an expression of annual consumption. The information from the DES is used to form a measure of total expenditure for each individual household that is used to evaluate the performance of the imputation based on (1.2). A more detailed description of the total expenditure measure is given in the appendix.

[2] This survey is the preferred survey to use for comparative studies of the kind undertaken here. The two previous surveys are 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.

2.3 The accounting imputation.

The register information on income, tax payments and wealth is used to form a measure of total expenditure as stated in (1.2). In order to implement this, it is necessary to have information about the physical stock of assets as well as the value of these assets. Only the total value of the stock of assets is observed. It is therefore in general not possible to control for capital gains. One exception, though, is housing assets. For this particular asset 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. This means that only the capital gains relating to housing assets can be taken in to account. We shall return to the missing capital gains information in the next section.

The income tax registers contains information about total taxable income, taxable wealth, and total final tax payments. Information in these registers is based on the tax form. Many entries on the tax form, both relating to the income, assets and liabilities, are reported directly from employers, banks and other credit institutions, and are therefore considered reliable. For the purpose of implementing (1.2) information about total taxable income, some non-taxable income components, final tax payments, and wealth are obtained for the year that the individual household enters the DES. Furthermore, wealth information is obtained for the previous year, so that a change-in-wealth measure can be calculated. One notable feature of the register data is that the data on asset holding can be divided in to three categories, cash holdings (including bonds), housing assets and other assets. While the content of the two first categories is well defined as cash holdings including bonds and the cash value of property as set by the tax authorities, the latter is more complex. The main content of this category is share holdings, but the category also contains self reported information about non-deposited bonds, a particular type of unquoted shares (in ships) as well as the value of investment objects and high value objects such as cars, boats. The quality of this information is low. Only relatively few observations, particularly among low-income households, are registered with values different from zero in this category. No information is held about accumulated pension funds. The bulk of wage earners are enrolled in employer organised pension schemes where pension contributions are deducted before the salary is paid out. As pension contributions are not taxable before they are paid out, pension funds do not appear on the tax form. One exception is if the scheme is privately organised in which case contributions are included in the total expenditure measure. The size of the liability stock is also available in the registers. This is because the wealth tax is paid of net wealth. Liabilities are generally registered for different categories such as mortgage and bank debt. A measure that is consistent across the observation period can, however, only be obtained for the total size of the liability stock. A more detailed description of the register data is given in the appendix.

The wealth situation is considerably more complicated for property owners than for renters. Therefore, we first check our imputation for renters and subsequently for house owners. In this way we are able to focus on households with relative simple wealth conditions before moving to the more complicated case. The sample is thus picked so as to focus on households that are either renters, owners or living in cooperative housing. Next, households with self-employed individuals are left out because such individuals have highly unstable income-tax conditions. Further, households that are registered with housing assets in one year but not the other as well as households that recorded as renters but at the same time have housing assets are left out. This leaves 3,434 from the original sample of 3,866 observations. Out of this 1,433 are categorised as renters.

Comparing the distribution of total expenditure from the DES with the accounting imputation or renters, cf. table 2.1, it is first and foremost seen that at the three quartiles are quite similar between the two measures. There is a tendency, though, that the lower quartile for the renters is smaller for the accounting imputation and the upper quartile is higher, indicating a tendency for more spread in the distribution of the accounting imputation. Generally the indication is that for the central values the imputation, so far, appears to do well. However, the accounting imputation produces some very negative values. For renters the accounting imputation generates 24 observations with negative values. The distribution of differences between the two measures confirms the previous considerations. For the bulk of the data the imputation matches the DES measure of total expenditure quite well, but a few observations exist for which the imputation do a really poor job.

[table 2.1 about here]

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.

[figure 2.1 about here]

The scatter plot suggests that the imputation does quite well. For some observations the imputed value of consumption is, however, quite far, and sometimes even negative, from the level of consumption stated in the DES. The kernel regression in the bottom panel shows that there is a tendency for the imputation to understate consumption for high levels of reported spending. For the bulk of the data it seems that the imputation does quite well, though.

The accounting based imputation is more complicated for house owners. This is because the particular design of the Danish mortgage system has implications for the imputation of total expenditures for house owners relative to renters. Therefore, before turning to the presentation of the imputed expenditure measure for house owners, it is useful to provide a description of the Danish mortgage system as it has important implications for the imputation of total expenditure from wealth data. In Denmark the financing of real property takes place via mortgage banks, so called mortgage credit institutions. Mortgage credit institutions offer loans where the borrower's real property is mortgaged 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 loans are funded by the issuing of callable mortgage credit bonds with fixed coupon rates and with a maturity of up to 30 years. 98% of all mortgage loans are issued as annuity loans. Because of the tax system, most mortgage loans are established as bond loans. The principal of the bond 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 makes the principal of the loan larger than the loan proceeds paid out. In this way the borrower will suffer a capital loss when establishing the loan, but the interest rate on the loan equals the coupon rate.[3]

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 capitalise on changes in the market rate and thereby 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 thereby making a gain. Major drops in interest rates most often trigger prepayment. When refinancing takes place the borrower is typically left with a lower monthly net pay, but with a higher principal, i.e. an increase in the outstanding debt.

[3] Conversely, the principal of a cash loan equals the loan proceeds paid out. In this type of loan the interest rate equals the yield to maturity of the underlying bond. The interest rate is thus higher than the coupon rate as the bonds are issued well below par.

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 is shown by the number of house owners that increased the level of liabilities from 1993 to 1994 being higher than for the following years.

In summary, changes in wealth for house owners does not necessarily directly imply adjustments in total expenditures, but can be simply a consequence of re-mortgaging. As a consequence, changes in liabilities for homeowners from one year to another may not be directly connected to changes in total expenditures in one particular year.

The distribution of the accounting imputation for house owners and people living in co-operative housing when compared to the distribution of total consumption from the DES has the same characteristics as for renters, cf. table 2.2. The quartiles are quite close to each other, and the accounting imputation generates some very negative values. For owners the accounting imputation generates 58 observations with negative values. The accounting imputation, however, also generates extremely large positive values. Considering the distribution of the 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 wealth situation is more complicated for owners and the link between change in wealth and consumption subsequently less direct.

[Table 2.2 around here]

The accounting imputation for house owners and people living in co-operative housing is presented graphically in figure 2 below. The scatter plot indicates that the imputed measure seem to do reasonably well for the major part of the data. For high levels of reported consumption the imputation seems to understate the reported level of consumption. The scatter plot shows that there are individual observations that are quite far away from the level of consumption stated in the DES, and that for some observations the imputed value of consumption is negative. The scatter plot also indicates that the spread of the imputed measure is larger than for renters. Considering the kernel regression in the right panel it is seen that there is a tendency for the imputed measure to understate at high levels of reported consumption. The general picture, however, is that the imputed measure is quite noisy, but does reasonably well for the vast part of the data.

[figure 2.2 around here]

In general, the accounting measure appears to do quite well for both house owners and renters when compared to the level of consumption stated by the households in the CES. There is a tendency to understate for higher levels of reported consumption. One characteristic of the accounting based measure is that it produces a few extreme deviations. Also, there is a tendency for it to be more noisy than the DES measure, and more so for house owners than for renters. Exploiting the information in the DES about total expenditure might be helpful in creating a better measure of total expenditure from the register data.

3. Measurement error

3.1 Implications 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 that the imputation underestimates total expenditures at high levels of reported total expenditures. 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 in part because of the particular timing of the variables entering the imputation. For example, cash savings are calculated as the difference in cash holding between the last day of year t and t-1. This timing does not line up with the actual consumption stated in the DES over a fourteen-day period for non-durables (and subsequently scaled up to 12 months of consumption). Also, consumption of durables in the DES is stated durable consumption over the past 12 months. Thus, for households interviewed in January the consumption of durables relates to the consumption in the previous year. Furthermore, we know *a priori* that for some of the assets holdings that are used for the imputation, it is not possible to control for capital gains. This implies that actual capital gains erroneously imply a lower level of imputed total expenditure relative to the true level of total expenditure. 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. These circumstances imply that it is reasonable to expect that a better imputation can be obtained by drawing on the interview-based information about total expenditure in constructing a measure of total expenditure. This is formalised in the next section.

Consider the accounting imputation for the n observations in the DES:

$$\hat{c}_{ACC} = y^* - \mathit{\Delta} W^* \qquad (2.3.1)$$

where $\hat{c}_{ACC}$ is a n×1 vector of imputed values, y^* is is a n×1 vector of disposable incomes and $\mathit{\Delta} W$ is an is a n×1 vector summarising change-in-wealth. This is an estimator of c, true total expenditure. Generally, y^* is considered to be measured quite well, since for wage earners practically all the relevant information is reported automatically to the tax authorities, and hence does not rely on self-reporting. However, for the reasons mentioned in the proceeding section it is expected that $\mathit{\Delta} W^*$ is not a perfect measure for savings. It is therefore assumed to be measured with error $\mathit{\Theta}$, so that $\mathit{\Delta} W^* = \mathit{\Delta} W + \mathit{\Theta}$. The mean squared error of the accounting based measure is then

$$\begin{aligned}\mathrm{MSE}[\hat{c}_{ACC}] &= E\left[(c-\hat{c}_{acc})(c-\hat{c}_{acc})'\right] \\ &= E\left[((y^*-\mathit{\Delta} W^*)-(y-\mathit{\Delta} W-\mathit{\Theta}))((y^*-\mathit{\Delta} W^*)-(y-\mathit{\Delta} W-\mathit{\Theta}))'\right] \quad (2.3.2) \\ &= E[\mathit{\Theta}\mathit{\Theta}']\end{aligned}$$

(2.3.2) states that any measurement error maps in to the imputed measure with full impact. Exploiting the self-reported information about total expenditure available in the DES may be helpful in constructing an alternative imputation where the measurement error problem is reduced. First, note that equation (2.1.2) says that disposable income is allocated between consumption and savings, and note that savings is measured by $\mathit{\Delta} W$ in the acconting based imputation. As emphasised above $\mathit{\Delta} W$ is expected to be measured with error whereas y is expected to be very precise measure of disposable income. The task is therefore to improve on the part of the imputation that relates to the savings component. To do this first construct a measure of savings, s_{DES}, that is based on the difference between dispoable income and selfreported total expenditure from the DES, $s_{DES} = y - c_{DES}$, and consider the regression equation

$$s_{DES} = \mathit{\Delta} W^* \beta + \varepsilon \qquad (2.3.3)$$

where s_{DES} is an n×1 vector, $\mathit{\Delta} W^*$ is the n×1 matrix of change-in-wealth, and β is a parameter to be estimated that relates change in wealth to s_{DES}, and ε is a vector of eror terms. Recall that $\mathit{\Delta} W^*$ is the unobserved true change-in-wealth measure, and that $\mathit{\Delta} W^* = \mathit{\Delta} W + \mathit{\Theta}$ with $\mathit{\Delta} W$ being the observed change-in-wealth and $\mathit{\Theta}$ the measurement error. Inserting $\mathit{\Delta} W$ in (2.3.3) above, and assuming mean-zero independent errors, gives rise to the classical measurement error problem with attenuation bias.

Lubotksy and Wittenberg (2001) suggest an estimator based on aggregating the parameters obtained from a multiple regression of s_{DES} on the vector of all the observed change-in-wealth components, $\mathit{\Delta} W$. The estimator is given by

$$\begin{aligned}\hat{\beta} &= \sum_{j=1}^{k} \left| \frac{\operatorname{cov}(s_{CES}, \Delta W_j)}{\operatorname{cov}(s_{CES}, \Delta W_1)} \right| \hat{b}_j \\ &= \rho' b\end{aligned} \qquad (2.3.4)$$

where $\operatorname{cov}(s_{DES}, \Delta W_1)$ is the pairwise covariance between s_{DES} and the first element in $\mathit{\Delta} W$, and $\operatorname{cov}(s_{DES}, \Delta W_j)$ is the pairwise covariance between the j'th element of $\mathit{\Delta} W$ and s_{DES}, and $\hat{b}_j$ is the parameter of $\mathit{\Delta} W_j$ in the multiple regression of s_{DES} on all the observed change-in-wealth components. The estimate $\hat{\beta}$ is now multiplied by $\mathit{\Delta} W$ to obtain an adjusted measure of savings,

$$\hat{s}_{ADJ} = \mathit{\Delta} W \hat{\beta} \qquad (2.3.5)$$

The imputed value of savings can now be used to construct an adjusted imputation of total expenditure

$$\hat{c}_{ADJ} = y - \hat{s}_{ADJ} \qquad (2.3.6)$$

Lubotski and Wittenberg (2001) show that the linear combination of the wealth components provided by (2.3.4) is the estimator that minimises the attenuation bias. The correction may thereby present an improvement over the pure accounting based measure in terms of reducing the MSE of the imputation.

To see this note that $\operatorname{plim} \rho' b = \beta \left(1 - \frac{1}{\sigma^2_{\Delta W} \rho' \Sigma_{\Theta\Theta} \rho + 1} \right)$ where $\Sigma_{\Theta\Theta}$ is the covariance matrix of the errors. The mean squared error of the adjusted imputation is thus

$$\begin{aligned}
\mathrm{MSE}\left[\hat{C}_{ADJ}\right] &= E\left[\left(C - \hat{C}_{ADJ}\right)\left(C - \hat{C}_{ADJ}\right)'\right] \\
&= E\left[\left(s_{CES} - \Delta W\hat{\beta}\right)\left(s_{CES} - \Delta W\hat{\beta}\right)'\right] \qquad (2.3.7) \\
&= E\left[\Delta W\left(\operatorname{var}(\hat{\beta}) + \operatorname{bias}\left(\hat{\beta}\right)^2\right)\Delta W'\right] \\
&= E\left[\Delta W\left(\operatorname{var}(\hat{\beta}) + \left(\frac{\beta}{\sigma^2_{\Delta W} \rho' \Sigma^{-1}_{\Theta\Theta} \rho + 1}\right)^2\right)\Delta W'\right]
\end{aligned}$$

If $\mathrm{MSE}\left[\hat{C}_{ADJ}\right] < \mathrm{MSE}\left[\hat{C}_{ACC}\right]$ then the adjusted measure may be preferred to the pure accounting based measure.

In the next section the regression-based weights are obtained.

3.2 Regression-Based Weights

The regression weights are obtained by regressing s_{DES} on the components of wealth change, i.e. change in cash holdings, other assets, and liabilities. The regressions are performed separately for renters and owners, because of the particular feature of the mortgaging system that makes the wealth situation more complicated for owners.

The regressions are presented in table 3.1. Parameter estimates for renters are presented in column 1 and parameter estimates for owners in column 2.

[table 3.1 about here. See the end of the paper]

All the estimated parameters for renters are significant and have the expected signs, cf. (2.1.1). An increase in asset holdings implies a lower level of consumption, and an increase in liabilities is associated with a higher level of consumption. A one-unit increase in cash holdings corresponds to a 0.37 unit increase in CES based 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 (2.1.2) 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 parameter estimates of the change-in-wealth variables for owners as well all take the expected sign, cf. column 2 of table 3.1. Noteworthy, in comparison with the parameter estimates for the renters, is that the parameter estimates are generally smaller. For example, the parameter of change-in-liabilities is about one half of the corresponding parameter estimate for renters. This is consistent with the fact that the mortgaging system can cause a change in liabilities from one period to the next that is not directly associated with consumption.

The estimates of b together with the estimates of ρ from table 3.1 are used to form an estimate of $\hat{\beta}$, cf. (2.3.4) that, in turn, is used in constructing the adjusted imputation. $\hat{\beta}$ is 0.41 for renters and 0.27 for owners. This means that the effect of a one unit change in the observed change-in-wealth corresponds to a 0.41 unit change in the CES based savings measure for renters. Correspondingly, a one unit change in the observed change-in-wealth for home owners is associated with a 0.27 unit change in the the CES based savings measure. In the next section the adjusted imputation is compared with the pure accounting based imputation of section 2.

3.3 Comparing the adjusted imputation with the accounting imputation

Assuming that c_{DES} represents the true value of total expenditure a sample-analogue of the mean squared error expressions of section 3.1 can be calculated. Scalar measures of (2.3.2) and (2.3.7) are given by

$$RMSE\left[\hat{C}_{ACC}^{\mathrm{Re}nter}\right]=\sqrt{\frac{1}{1,433}\sum\left(C_{DES}^{\mathrm{Re}nter}-\hat{C}_{ACC}^{\mathrm{Re}nter}\right)^2}=87,854$$

$$RMSE\left[\hat{C}_{ADJ}^{\mathrm{Re}nter}\right]=\sqrt{\frac{1}{1,433}\sum\left(C_{DES}^{\mathrm{Re}nter}-\hat{C}_{ADJ}^{\mathrm{Re}nter}\right)^2}=70,420$$

$$RMSE\left[\hat{C}_{ACC}^{Own}\right]=\sqrt{\frac{1}{2,001}\sum\left(C_{DES}^{Own}-\hat{C}_{ACC}^{Own}\right)^2}=205,975$$

$$RMSE\left[\hat{C}_{ADJ}^{Own}\right]=\sqrt{\frac{1}{2,001}\sum\left(C_{DES}^{Own}-\hat{C}_{ADJ}^{Own}\right)^2}=126,534$$

It is seen that 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-liabilities and is less direct for this group due to the mortgaging system. 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.

One of the main insights from section 2 is that the accounting based imputation is characterised by underestimating at higher levels of reported total expenditure. In figure 3.1 and figure 3.2 the adjusted measure is compared with the accounting based measure of section 2. The scatter plots indicate that the two measures are pretty close, especially for renters, and more so for renters with low levels of expenditure. This is hardly surprising, since for these households total expenditure is close to disposable income. Considering the non-parametric fits in the left panels it is seen that there is a tendency that the adjusted measure does give lower estimates of total expenditure for the highest levels of the accounting imputation. This points out how the adjusted measure try to correct for the accounting measures tendency to underestimate the CES measure total expenditure at low levels and overestimate at high levels.

[figure 3.1 and figure 3.2 about here. See back of the paper]

In total, the adjusted imputation appears to be able to at least partially correct for some of the deficiencies that characterises the pure accounting based measure, namely that the pure accounting based imputation appeared to underestimate at low levels and to overestimate at higher levels of imputed total expenditure.

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 is 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. This type of analysis does not allow for any structural interpretation of consumption behaviour of households. It will only provide descriptive evidence about the patterns of covariance between imputed measures and some demographic variables on one hand and covariance between the DES measure and the same demographic variables on the other hand. Parameter estimates from regressions of consumption-income ratios are presented in table 4.1.

[table 4.1 about here. See the end of the paper]

The parameter estimates obtained by regressing the log-consumption based on the DES measure of total expenditure on demographics are presented in column 1. In column 2 and 3 the consumption-income ratio is based on the accounting measure and the adjusted measure of total expenditure respectively. Regressions are based on a sample where observations with negative values of accounting imputed consumption are deselected.

In the model based on the DES measure significant children, age and labour market participation effects are estimated. 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 42% lower than for households with two adults. Further, two dummies, one indicating age below 40 and another indicating more than 67 years. The first dummy indicates that younger people tend to have a consumption level that is around 15% 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, parameter estimates in column one indicate that, consumption is increased about 0.7% if the work time of the man is increased by 1%. The corresponding number for the female is slightly lower.

Comparing the estimates in column one 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 larger standard errors, which is not surprising considering the results from sections 2 and 3. Indications of differences do exist, though. Single adult households are now estimated to have a significantly lower level of consumption than what is indicated from the DES based regression. Estimated children effects are in most cases are quite close to the corresponding estimates in column 1. The most important deviation seems to be that effects of school children is estimated to be a bit lower than what is found when estimates are based on the DES measure of total expenditure. The differences are in no cases significantly different, though. Estimates of age effects appear to match those presented in column one quite closely as do labour supply effects.

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 two. The parameter estimate associated with children aged 8-20 is now more than two standards errors from the estimates in column one, but they are now in some cases more than two standard errors smaller than the corresponding estimates in column one. 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 estimated female labour supply decision seems to be 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 "true" estimates even better than when the adjusted measure is used. The most important difference between the two imputed measures seems to be that the accounting based measure is somewhat noisier than the adjusted measure.

5. Conclusion

The construction of a measure of total expenditure based on information about income and change in wealth that is available in the income-tax registers has been investigated. Evidence is found that a simple approach to imputing total expenditures based on an accounting identity provides a measure that performs quite well in terms of matching individual households subjective statements about total expenditure. Also, this measure performs reasonably well in terms of patterns of covariance with demographic variables that resembles those obtained from the DES quite closely.

The accounting measure appears to overstate at low levels of the imputed measure and to understate at high levels of the imputed measure. More precision in the ability of the imputed measure to predict total expenditure in the DES can be obtained by using weights formed by regressing total expenditure from the DES on all the income and wealth components. This approach comes at costs, though. This is in part because weights are completely sample-specific, which reduces the scope of application, and in part because the use of weights in some cases affects the covariance between total expenditure and demographic variables. Altogether this implies that the pure accounting based measure may be preferred in analytical work.

The results are promising. It gives clear 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.

References

Browning, Crossley and Weber (2002); Asking Consumption Questions in General Purpose Surveys; Draft

Browning, Martin and Deaton, Angus and Irish, Margaret (1985); A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle; Econometrica; 53(3), May 1985, pages 503-43.

Lubotksy, Darren and Wittenberg, Martin (2001); Interpretation of Regressions with Multiple Proxies; working paper #457, Princeton University, Industrial Relations Section

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

Skinner, Jonathan (1987); A Superior Measure of Consumption from the Panel Study of Income Dynamics; Economics letters, vol. 23, pp. 213-216

Statistics Denmark (1999); The Expenditure Survey, Documentation; (in Danish: Forbrugs-undersøgelsen, metodebeskrivelse); Statistics Denmark; Copenhagen

Definitions

Consumption in the consumer 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, other goods and services. Note, property taxes (ejendomsskatter) are included in housing expenditures.

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, two non-taxable income components, child benefits and rent support, are also added.

Disposable income = Gross taxable income according to the final tax form

+ Payments from capital pensions

+ Child benefits

+ Heat support

+ 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. Contributions to privately organised pension schemes are limited in extend.

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 in to 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 liabilities are divided in to mortgage and other debt. 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 below.

Table A.1 Wealth Components

Assets:
Cash holdings
Cash holdings in bank, bonds, deposited mortgage deeds
Cash value of house
Other assets
Value of shares
Value of shares, main share holder in company
Own capital in domestic company
Own capital in foreign company
Other Foreign wealth
Other taxable assets (*)
Liabilities:
Sum of liabilities

(*) This component is based on self-reporting. It is a "residual" wealth component, and it includes value on non-deposited mortgage deeds and bonds, car, boat, value of co-operative apartment, premium bonds, unquoted shares in ships.

Table 2.1. Quartiles of the distribution of total expenditure from the DES and of the accounting imputation for renters.

# observations: 1,433	DES (1)	Accounting (2)	Difference (1)-(2)
Minimum	16,470	-619,221	-692,439
Lower quartile	92,988	86,535	-28,199
Median	129,240	125,570	991
Upper	188,449	191,943	34,120
Maximum	812,192	762,974	834,522

Table 2.2. Quartiles of the distribution of total expenditure from the DES and of the accounting imputation for house owners and people in co-operative housing.

# observations: 2,001	DES (1)	Accounting (2)	Difference (1)-(2)
Minimum	28,528	-2,197,534	-1,931,588
Lower quartile	158,684	137,788	-68,878
Median	239,440	238,946	-2,307
Upper	330,568	355,912	63,348
Maximum	1,456,764	2,325,566	2,574,712

Table 3.1. Regression of s_{DES} on wealth components

	Renters		Owners	
Dependent variable, s_{DES}	ρ	b	ρ	b
Change in cash holdings	1	0.3702** 0.0754	1	0.2858** 0.0402
Change in other assets	0.97	0.1680** 0.0489	0.25	0.1020** 0.0261
Change in liabilities	0.63	-0.1887** 0.0655	0.45	-0.0910** 0.0195
Constant	-	75,849.15** 2420.28	-	185,732.30** 3518.42
$\hat{\beta}$		0.4143		0.2749
Number of observations		1,433		2,001

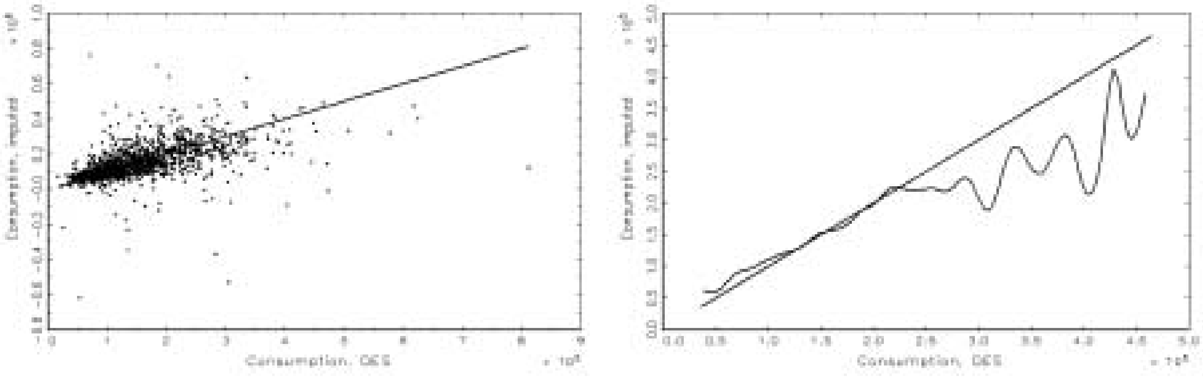
Note: Robust standard errors in small numbers below parameter estimates.** indicates significance at 5% level, *indicates significance at 10% level.

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

	1.	2.	3.
Dependent variable, ln C	$C=C_{ces}$	$C=C_{acc}$	$C=C_{adj}$
Single adult	-0,4291**	-0,4870**	-0,4748**
	0,0174	0,0232	0,01675
# children, aged 0-7	0,1986**	0,1834**	0,1887**
	0,01948	0,0264	0,0168
# children, aged 8-14	0,1289**	0,0777**	0,0608**
	0,0197	0,0274	0,0177
# children, aged 15-20	0,1874**	0,1692**	0,1420**
	0,0213	0,0324	0,0189
# children, aged 21-30	0,1278**	0,1996**	0,1809**
	0,0423	0,0704	0,0350
Age of the oldest person <40, dummy	-0,1442**	-0,1571**	-0,2229**
	0,0162	0,0228	0,0159
Age of the oldest person >67, dummy	-0,1831**	-0,1268**	-0,1091**
	0,0234	0,0347	0,0229
Work time, man	0,0072**	0,0084**	0,0093**
	0,0004	0,0006	0,0005
Work time, woman	0,0065**	0,0056**	0,0070**
	0,0004	0,0006	0,0005
Constant	12,0343**	12,0332**	12,0357**
	0,0190	0,0274	0,0201
Number of observations	3,352	3,352	3,352

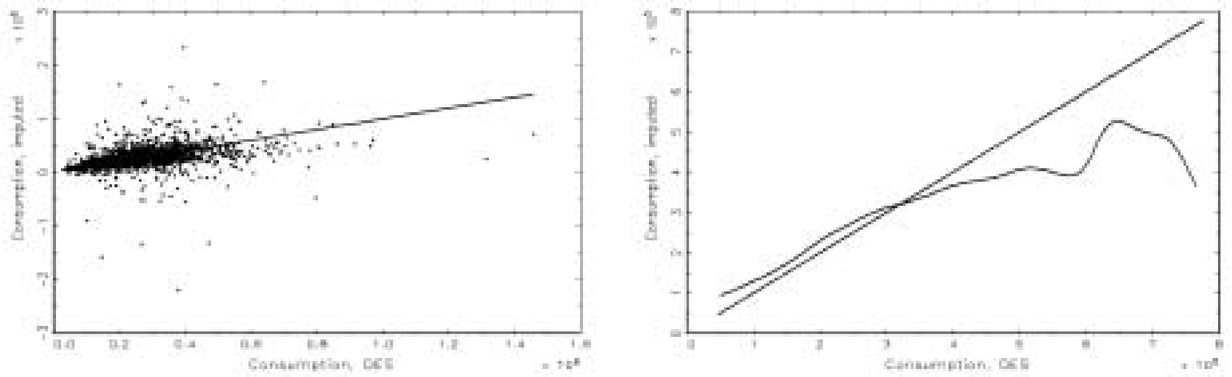
Note: Robust standard errors in small numbers below parameter estimates.** indicates significance at 5% level, *indicates significance at 10% level.

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



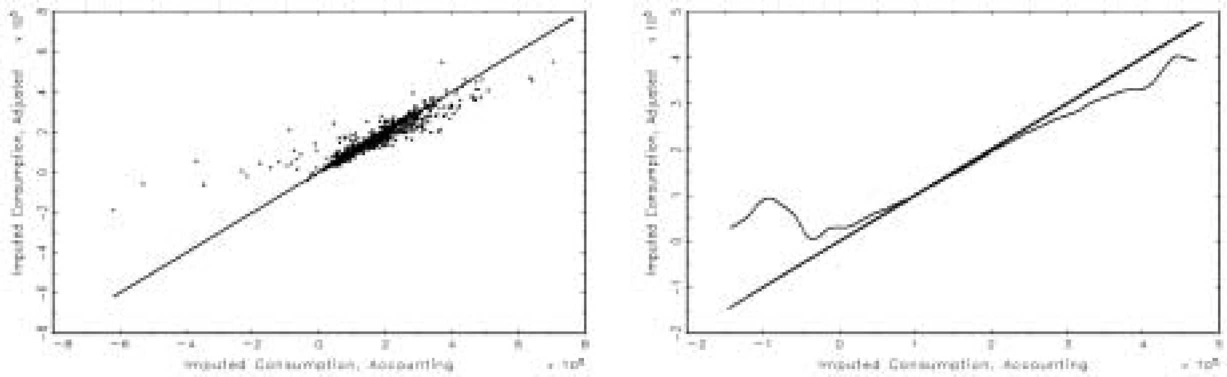
Note: Kernel regression are estimated using the Gaussian kernel on a sample that is trimmed by 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation to be 8,341.

Figure 2.2. Accounting based imputation 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.



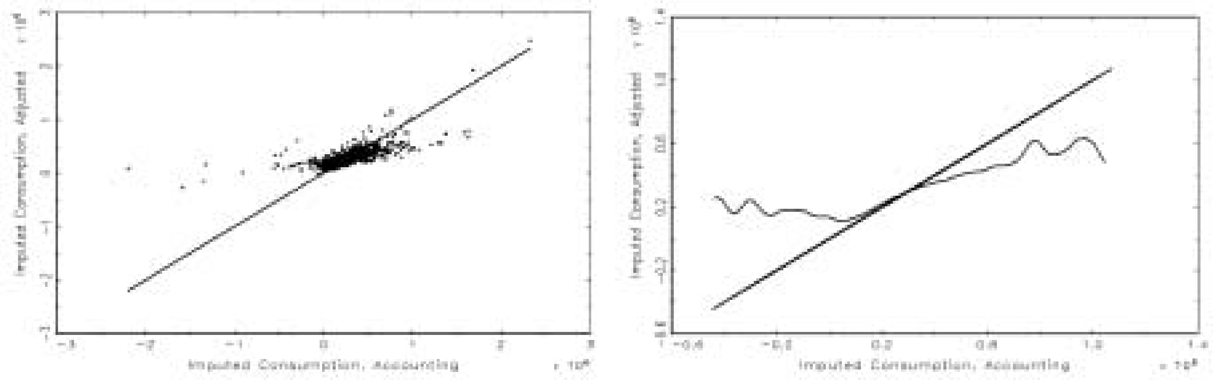
Note: The kernel regression is estimated using the Gaussian kernel on a sample that is trimmed 0.5%. Bandwidth for the kernel regression has been chosen by generalised cross validation to be 29,974.

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



Note: The kernel regressions 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 13,913.

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



Note: The kernel regressions 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 29,974.