

A PROTOTYPE MICRO-MACRO LINK FOR THE CANADIAN HOUSEHOLD SECTOR

BY HANS J. ADLER AND MICHAEL WOLFSON

Statistics Canada

This paper seeks to forge a link between Canadian macro and micro data relating to the household sector. The analysis is in three parts. The first part begins with National Accounts data on the personal sector. These data are adjusted to remove transactions relating to non-biological persons, so that the result is income and expenditure for the household sub-sector. The second part starts with the annual household survey used to collect income distribution data. These survey data are augmented in various ways to account for under-reporting and to add information from other micro data sets—particularly the periodic survey of household expenditure patterns and a sample of individual income tax returns. The result is a comprehensive, albeit partially synthetic, household micro data set. In the final part of the paper these two largely independent data sets are compared, and the general quality of the results is discussed.

INTRODUCTION

The development of an integrated system of National Accounts reached one of its high points with the creation of the 1968 UN Standard National Accounts (SNA). Also by the late 1960's, the confluence of economic events, new theoretical questions and a desire to make greater analytical use of computer technology led to some dissatisfaction with aggregate economic data and towards a desire for the development of micro data sets because of their analytical value. Most macro data or aggregates are summations of their constituent micro or individual data. The difference in the analytical use of micro and macro data is therefore neither in the overall facts or events which they portray, nor in the universe they cover. Rather, the difference lies in the analytical results which are desired, and in the theoretical variables emphasized and the questions posed and answered. Thus, for instance, the SNA macro accounts provide the basis for analysis of the total income, expenditure and savings of the four broad sectors of the accounts (business, persons, governments, and non-residents) and the economic behaviour of these large groups *vis à vis* each other. Such questions as which socio-economic or income size group within the personal sector saves and spends how much cannot be addressed by macro economic data.

Over the past two decades, significant advances in the creation, simulation and analysis of mainly personal micro data sets have been achieved for example by Orcutt, *et al.*, Pechman and others, and the potential of computer manipulations of micro data sets generally have been explored by, among others, Fellegi, Goldberg, Aukrust and Norbotten. During the same time, the Ruggleses built

Note: We are greatly indebted to our colleagues Brian Murphy, George Haydu and Martin Méthot for their invaluable contribution to the statistical part of this exercise.

The analysis presented in this paper is the responsibility of the authors and does not necessarily represent the views or policies of Statistics Canada.

some of the conceptual roads leading from SNA macro data in the personal sector towards the possibility of micro data manipulations and also carried out seminal work in the application of computer technology based analysis of the longitudinal establishment data files in the U.S. Department of Commerce. However, relatively little progress has so far been made in the integration of micro data analysis into an SNA framework.

There are several reasons for this seeming delay in the development and exploitation of micro data within an SNA framework. While computers and otherwise confidential micro data files have been available to government statisticians, the intellectual climate of economic analysis in the postwar years favored macro economics and the desire to quantify the large aggregates of Keynesian economics drove statistical development. However, interest in micro-analysis grew with the realization that there were limits to what could be explained with macro data. Techniques were developed which permitted data on individuals and/or households and their economic, social and demographic characteristics contained in either statistical or administrative files to be made available on a sample basis to the research public. Some of the micro data work in the personal field mentioned above has been a direct result of these developments.

It has however not been possible to link this work directly with the National Accounts because the so-called personal sector in the National Accounts encompasses a much wider range of actors than natural or biological persons; it includes non-profit institutions, personal associations, and unincorporated business. Furthermore it contains a significant number of imputations for non-market transactions, such as rent on owner-occupied housing; and last but not least in contradistinction to most other sectors, not all constituent micro data of the personal sector are derived from individual, family, or household files or accounts. Many of the data in the personal sector come from the records of businesses or governments e.g. income from government transfer payments in the personal sector is not a statistical aggregation from individual or household receipts, but comes from a record of the total out payments in the public accounts. A fair number of expenditures on consumer goods and services similarly are not the total of such expenditures as recorded by individuals, but the sales of such goods and services to individuals as reported in aggregate by business.

The purpose of this paper is to report on a prototype effort to construct, from existing data, a linked micro-macro account of the Canadian household sector. This effort encompasses two major processes. Essentially, we have started from the two existing and divergent types of data, and tried to meet somewhere in the middle. The next section of the paper reports on a series of adjustments that have been made to the personal sector in the SNA in order to extract a picture of the household sector from the conventional personal sector of which it is the major constituent. This involves trying to distinguish the economic activities of natural or biological persons from the activities of other institutions or groups (e.g. non-profit institutions) that have traditionally been put into the personal sector.

The following section then describes the construction of the Social Policy Simulation Database (SPSD). This is a microdata set that was originally constructed to serve as the basis for a policy oriented microsimulation model. However,

in order for the model to generate reliable and comprehensive estimates, the underlying database has had to be augmented and adjusted to cover the household sector as completely as possible. Thus, the SPSD can also be viewed (serendipitously) as an effort to construct a more complete micro foundation for the household sector.

Once the national accounts adjustments and the construction of the SPSD are described, we turn to an evaluation of their consistency. Starting from two largely independent sets of data, how close are we able to come? The reader will be the final judge, and the paper finishes with some concluding comments.

HOUSEHOLDS IN THE PERSONAL SECTOR OF THE NATIONAL ACCOUNTS

In order to avoid confusion, the term personal sector will henceforth be reserved to refer to the Personal and Unincorporated Business Sector of the Canadian SNA, while the term household sector will denote that aggregate which sums the transactions of natural persons and/or families and/or households. Aggregate data for the household sector will permit analysis of such macro economic concepts as income, expenditure, and saving from a truly natural person perspective. To elaborate on the latter term, what is really implied here is that one wants to create a sector which is the aggregate of decision-making units composed of people. The smallest meaningful unit is the individual. (Of course, it is typically important to be able to group individuals into the families or households of which they are members.) Income will thus be truly income as received by households; expenditures will exclude the outlays of all manner of collectives. As a result, the household savings rate can, like in the work by the Ruggleses for the U.S., hopefully become more economically meaningful than the current so called personal savings rate. The paper will then demonstrate the "linkage" of the macro-data to the 1984 SPSD, an actually constructed micro-data set, with which it is possible to examine the economic behaviour of significant and critical population subgroups such as the elderly, married couples with children, married couples without children and so on—where these groups can also be broken down by age, income, geography, household size, etc.

The precise steps by which a household sector can be derived from the Canadian SNA personal sector depend generally on the statistical and analytical purposes for which such a household sector is to be used. A major statistical historical revision of the Canadian SNA has recently been completed, so that this paper does not propose a restructuring of the Canadian SNA.

Indeed, the sequence of adjustments outlined in this section are in the nature of supplementary and more detailed explanatory tables and can therefore be tailor-made. Thus it is not necessary to permanently decide what the exact boundaries of the household sector will be. As a matter of fact it can be argued that restructuring of such a well-developed system as the Canadian SNA should not be proposed until any suggested alternatives have demonstrated their analytical and statistical robustness in actual use and over some time. The more precise nature of these choices will become evident as the different analytical uses of the household sector are discussed. Also while the maintenance of a uniquely defined SNA structure in printform will probably be always necessary, it can be argued

that as both details and aggregate series become more and more computer accessible, analysts will find it progressively easier to satisfy their own needs by constructing alternate aggregates with officially supported micro-data sets, more detailed partial aggregations of macro-economic time series and computer programs.

There are fundamentally two reasons why the establishment of a household sector is analytically desirable. First, the behaviour of natural biological individuals is likely to be very different from the other types of agents included in the personal sector. Particular concern at the present time focuses on household saving. Second, in order to undertake economic analysis by the social, demographic and economic characteristics of households it is necessary to link to micro-data and this in turn requires a pure household sector. While the former reason may eventually lead to an "official," fully integrated SNA household sector, the latter analysis in the short term necessitates a household sector tailored to the particular definitions in specific micro-data sets—in the case of the present paper that of the Social Policy Simulation Database (SPSD) for 1984.

In this paper, the definitions adopted for the aggregate household sector are largely driven by the need to conform to the concepts embodied in the SPSPD. However, the discussion also points out some alternative concepts which in the longer term might be preferable.

The first and most obvious step from the personal to the household sector is, of course, the removal of all (or at least most) data which do not pertain to natural persons, i.e. the income and expenditure—and hence also savings—of all the non-profit institutions and associations of individuals now included in the personal account. In the Canadian SNA the largest single component among the non-profit institutions are the institutions of higher learning. All religious organizations fall into this category and other large groups include welfare and research organizations such as the Red Cross, the Canadian Cancer Society, the Heart Fund and many others. Trade unions and political parties as well as recreational, education and cultural organizations are others which are also in this category. Such a deconsolidation of the personal sector, however, also implies that fees and donations by households to these organizations, and transfers and gifts from these organizations to households must then be separately estimated and included in the household expenditure and income accounts, respectively.

It should be noted that the deconsolidation of the personal account only shows different data in the household saving account where institutions have been removed which have savings (or dissavings) and/or which receive or transmit funds from or to other sectors of the economy. For instance, in the case of institutions of higher learning the major part of their operations are financed by government and their removal from the personal account will lower the income side of the household account by the amount of this transfer income while the expenditure side will be lowered by the grand total of their expenditure. Since part of their expenditure is financed by households, the expenditure side will be lowered more in the first round of adjustment, and the savings remaining in the household account will be raised by the excess of expenditure over income of the institutions. However (assuming no net saving, endowment funds, or dissaving on the part of institutions) this will then be exactly offset in the next step when

fees paid by households to these institutions, which were omitted as intrasectoral transfers in the personal account, are included in household expenditure. Both the income and expenditure accounts in the household sector are then lower by the non-household financed expenses of these institutions. On the other hand, for example, the exclusion from the household account of a small soccer club which exactly matches its expenses for a coach, uniforms, and rent for the playing field by the fees it charges its members, will not result in a different aggregates between the personal and household accounts because the total expenses of the club removed from personal expenditure will be precisely matched by the fees charged by the club which will be added to household expenditure. This process is statistically speaking rather fortunate for it obviates the necessity of adjusting the personal sector for the numerous small, informal, *ad hoc* and quite frequently short-lived clubs and associations—more often than not without employees, formal records, investment income or capital formation. To attempt such an undertaking would indeed be a statistical nightmare and costly beyond any bounds of reasonableness.

The concepts of household income, expenditure and savings and their treatment poses a number of issues whose only long-term solution might be to supply the data users with all the options and permit them to “roll their own.” The main problems are treatment of private pension funds and to a lesser extent self-financed annuity and whole life insurance purchases. The Canadian personal income and expenditure account does not contain on the expenditure side the private employer/employee contributions to pension funds and/or annuities nor individually purchased whole life insurance and annuity contracts. Since the means from which these funds are financed are either explicitly or implicitly included on the income side, the savings of such institutions are part of total personal savings. The assets of such funds are in the personal income and expenditure account deemed to be personal assets and outpayments from the funds are thus not part of personal income but a distribution of assets, and the investment income earned on these assets is included in personal income. This, it may be noted, is in contradistinction to public service and other publicly funded pensions, contributions to and outpayment from which are counted as transfers and hence excluded from personal savings. The savings and assets generated by these latter funds are part of the government sector. Private pension funds contain both employer and employee contributions and it is unlikely that individuals are very knowledgeable about their employer’s contribution or their equity in the accumulated funds. (This is likely to change, however, as a result of major changes to the income tax rules whereby employees will be notified annually of the value of the employer’s contribution to the pension plan on their behalf; see Department of Finance, 1986.) Households have not exercised much control over these funds. (But again this also will probably change with recent pension reforms.) The Ruggleses have therefore proposed to remove employer contributions from the household account as well as remove the assets of or equities in such funds from household assets and to include pension benefits from these plans with household income.

For the present exercise, we have followed this treatment in order to bring household income, expenditure, savings and assets in consonance with that which

one obtains from current questionnaires directed at the household, and hence with the SPSPD. This makes the treatment parallel with that of public funds. On the other hand, it can be argued that the economic behaviour of households both at the micro and macro level is in no small way influenced by the fact that they have "forced" savings and future income security on account of the presence of such partially or wholly sponsored employer funds. An argument can therefore be made that the present treatment of these funds also has behavioural validity which cannot be ignored.¹

Since there is no compromise between these two positions, and since this whole exercise is concerned with adding to the analytical value of the accounts rather than bringing forth a new generally consistent structure of the Canadian SNA, both versions of household income, expenditure and savings should be made publicly available. It may further be noted on ground of consistency that if one wants to include private funds with the household sector because the knowledge that future assured retirement income exists influences present spending (and saving) patterns, a similar treatment should be accorded to public sector employment pension plans.

Somewhat analogous arguments to those made for pension funds can be adduced for individual purchases of whole life insurance and individual purchases of annuities. With respect to life insurance purchases, the Canadian treatment is that the administrative expenses of insurance companies are included in current expenditure on consumption goods and services and are deemed to represent the value of services rendered by life insurance companies. The accumulated funds of life insurance companies are deemed to be personal savings and the investment income on policy holders' funds is included with personal income. The Ruggleses make the case that whole life insurance has some similarity to savings accumulation in pension funds in that individuals have no exact knowledge of their equity, and have no control or access to their funds. Such investments nevertheless differ from equities in pension funds, at least to the extent that whole life insurance savings are accessible to individual policy holders by means of policy loans to the extent of the cash surrender value of the policy. Cash surrender values may be taken as reasonably equivalent to policy holders' equity and are probably not too difficult to obtain either via individual questionnaires or in aggregate from the record of insurance companies. As in the treatment of pension funds previously discussed, the exact aggregate one derives for the household account would depend on the intended analysis, and both the present treatment and the complete separation of households and the insurance sector may prove useful in analysis. In order to achieve this, one would have to include the total premium payments on the expenditure side and the claim payments on the income side of the household accounts. To accord with the SPSPD, all life insurance premiums and benefit payments have been added to the expenditure and income side respectively and investment income of life insurance companies has been removed from income.

¹Richard Ruggles, in private conversation with one author, has, however, pointed out that national accountants ignore many other influences and facts which lead to different saving patterns, e.g. the rational considerations which do or should determine the saving pattern of a university professor with tenure are quite different from those of a prominent boxer or football player!

Annuities present the third aspect of this particular problem. At the moment, purchases of annuities and return of their capital value are also considered intra-personal transfers, while the administration expenses and investment earnings are included with those of life insurance companies.

Again it can be argued that from a cash flow point of view, individuals considered payments into annuities as a purchase of (future) benefits and payments from annuities as an income. This latter version is the treatment in the SPSPD, and is the way in which annuities have been treated in this study. However for the reasons outlined in the discussion of pensions, the present treatment—even in the pure household account—is also analytically valid and hence again both versions might eventually be published. In the case of annuities there is however another argument which could be proffered in opposition to changing present methodology. The question whether annuities are in essence different from the purchase of any other financial asset—be it some form of a Registered Retirement Savings Plan (a tax-deferred personal savings vehicle), a bond or stock—cannot be answered unambiguously and depends in no small way on each person's perception of his own saving mechanism. The two version account therefore is probably still valid, although it must be admitted that in this instance one comes pretty close to counting true capital transactions as current.

The above mentioned RRSP's, as well as the now discontinued Registered Home Ownership Savings Plans (RHOSP), may constitute another problem, since the current interest income from these plans is included in the personal sector. It may be difficult to defend the exclusion of such income from household income conceptually since the distinction between RRSP's and other individually held investments such as bonds and stocks—the income from which should not be excluded from household income and savings—becomes quite esoteric. On the other hand the impact of both recently implemented and proposed tax changes relating to pension plans as well as pension reforms which will result in major flows of money from pension plans to RRSP's (as a consequence of early vesting and portability) may change both the nature and perception of these plans. The complete articulation of all such flows would permit alternate treatments and thus seems called for. Since purchase and redemption of RRSP's are not considered household expenditure or income in the SPSPD, no adjustment for RRSP's were necessary in this study. (Although at this time probably not yet statistically significant, the conversion of RRSP's into annuities would make the inclusion of that type of annuity income in the total annuities income flow, described in the preceding paragraph, somewhat ambiguous.)

On a more theoretical level, McCracken *et al.* have pointed out that specific treatment of the above described investments revolve around such questions as the degree to which such investments can serve as collateral for loans, the fungibility of the assets, the rigidity of the contractual obligations incurred or the possibility of discretionary changes. Whether decisions on these investments are motivated by personal or business reasons, are entirely voluntary or required by law or contract, are reached individually or collectively or result in funded or unfunded liabilities are other factors which may influence the specific treatment accorded to such investments.

The case for casualty insurance and other non-life insurance transactions is also perceptually more complicated at the individual level. In aggregate only the net premiums (defined as the total premiums paid less claims received) or the administrative cost of this insurance is recorded, because claim receipts can simply be viewed as intra sector transfers (although they are of course routed via another sector). In the linkup with micro data, the total premiums should be recorded as outlays, while any claims received may in many instances be viewed as other income. Since this individual view of the insurance transaction is gross, the net saving will remain unaffected, although income and expenditure may differ between micro records and macro aggregates. The present data do not permit a precise reconciliation for this item.

The next set of adjustments are those which from the point of view of obtaining aggregate household savings need really not be made, but which are essential from the point of view of reconciling the aggregate household income and expenditure flows with the existing SPSP micro data and which also would contribute to the conceptual tidiness of the household saving concept. We refer, of course, to imputed items which are aggregate attributions made on the basis of macro-economic desiderata and are not currently found in micro data. They are always included on both the income and expenditure sides and do not affect savings. The main items which have been excluded here in addition to private employer pension fund contributions previously dealt with are the value of farm products consumed in farm households, the imputed interest for banking services, net imputed house rent for owner-occupied housing, and board and lodging furnished to employees in lieu of wages as well as clothing issued to members of the armed forces. Of course, the same arguments that lead to the inclusion of these imputations in the macro accounts also apply to the micro data. Hagenars *et al.* present an initial effort at just such imputations at the micro level.

Other supplementary labour income, such as employer contributions to private health or disability insurance schemes have also been eliminated since such data are not available in the SPSP.

Related to imputations is the matter of current versus capital expenditures by households. While the treatment of expenditure on owner-occupied housing as capital rather than current expenditure—and hence its inclusion in savings—in the personal account poses no problem in going from the personal to the household sector, since both the requirements of economic analysis and personal perception of housing as capital are likely to be coincident, the treatment of hard major durable consumer items such as cars, washing machines etc. as current expenditure has been subject to questioning in the personal sector and hence *a fortiori* in the household sector. The data pertaining to the “capital” outlays are readily available. A few remarks about such a change of treatment may, however, be apropos. While it is quite true that from many points of view the acquisition of hard durables is a capital expenditure since consumption is spread out over several years, one may question what the point of such conceptual change would be. Particularly in the household sector, we surely would not want to add an imputation for the services rendered by such goods (which would be the logical although not essential consequence of such a treatment).² In this paper, the treatment of durable goods as part of consumer expenditure has been retained.

Charitable donations from corporations to the personal sector must also be removed from household income since all these donations are assumed to flow directly to non-profit institutions rather than to households. The problem of transfers composed of bad debt allowances from the business to the personal sector is somewhat different. From an aggregate household point of view these allowances should be left in the household sector since they constitute the amount by which goods reported as bought were not paid for and hence constitute an element of household savings and income. To reconcile with the micro data, however, these transfers have been removed from the aggregate since it is assumed that bad debts have not been reported as such by households, while the goods so acquired have been recorded as purchases.

Since the SPSD does not contain information on population in the Yukon and Northwest Territories, the household sector has been adjusted accordingly for this study.

Since the aggregate household expenditure on goods and services is based on aggregate sales data which include goods and services bought by non-residents, a special adjustment to exclude such purchases as well as to include goods and services bought by Canadians abroad is made to the aggregate data. The SNA aggregate will therefore correspond to the aggregate expenditure of SPSD. However, at present the detailed expenditure data may show a discrepancy on account of external transactions, particularly in such categories closely related to tourism.

One almost intractable problem remains. It is an issue which is present in all three data sets under discussion: the personal sector, the household sector and in the micro data. The concern is the inclusion in all three data of the net income of non-farm unincorporated business and the net income received by farm operators from farm production. What one would like to have in all three data sets is, of course, the withdrawals of proprietors rather than their total income, because the latter fudges the boundary between business and households. Such withdrawals are, at the moment, not obtainable from the national accounts data sources and the information is also not available in the SPSD. It is probably possible that one could eventually obtain these data. However, two factors mitigate against the active pursuit of this course. The first one is simply that it is felt that it would be more difficult to obtain proprietors' withdrawals instead of net income—and the latter is already probably one of the weakest components in both household accounts and surveys. This difficulty is also linked to the second aspect of the problem. The amounts withdrawn are often determined ex-post facto and revolve around lump sum or investment decisions on the part of a

²From another perspective one may also doubt whether these are truly capital goods of the same investment nature as housing. The consumer can, like houses, rent most of these types of goods *from* retail outlets which specialize in such services. However, there is one characteristic in which such goods differ significantly from housing as capital goods. Houses can be bought as an investment good, i.e. there exists a housing rental market for houses owned by individuals by means of which a houseowner can derive income from his investment. Such a market does not exist for the owner of hard durable goods i.e. while as stated above these goods can be rented from retail outlets, once they are owned by an individual, there exists no market on which the individual owner can rent out his possessions *to* others.

business person who may be motivated by both personal and/or business considerations. Thus an unincorporated proprietor may decide to withdraw less than usual from the business when a good business investment opportunity presents itself or he may increase his withdrawals to facilitate the purchase of a big ticket durable consumer item. Business and personal decisions and motivations can be inextricably intermingled in this sector and a strict separation of business and personal portions may not even be particularly desirable from an analytical point of view. For the time being, it has therefore been decided not to try to "purify" the household account for this item.

THE SOCIAL POLICY SIMULATION DATABASE (SPSD)

The SPSPD is part of a larger project in Statistics Canada to provide to the public a package of data and modelling software that will allow users to analyse proposals for various kinds of reforms to Canada's system of income and sales taxes and transfer programs. This analytical capacity occupies approximately 10 megabytes and can be used with any IBM personal computer or compatible. As of June 1987, the database and model was selectively released to users as a prototype for testing and evaluation. (Commercial release is scheduled for Fall 1988.)

The database contains 1984 data that have been combined from a variety of sources. The starting point is the Survey of Consumer Finance (SCF). This is an annual survey of households whose principal objective is to provide a series on income distribution. In 1984, there were about 100,000 individuals in about 40,000 households in the SCF. The sample is geographically stratified, but unfortunately from the viewpoint of its basic objective, it is not stratified by any strong correlate of income.

The strong points of the SCF are its almost complete coverage of the population, its detail on the structure of the familial relationships within the household, and its reasonably good income data. From a social policy perspective, it provides a very good basis for all the most commonly desired breakdowns for assessing the prospective impacts of tax and transfer policy changes—geography, family type, and income.

However, the SCF suffers from a number of major deficiencies from the viewpoint of social policy modelling. Thus, it has been substantially adjusted and augmented in order to produce the SPSPD. Fortunately, many of these adjustments and augmentations also serve to make the SPSPD a more accurate and comprehensive disaggregation, indeed micro foundation, for the household sector of the national accounts.

The creation of the SPSPD from the SCF is a complex multi-step process. The first main step was to adjust the weights for each individual in the database. This process was primarily motivated by the need to force individuals and the households to which they belonged to have identical weights. At the same time, the weight adjustments were also used to force the population into accord with a variety of externally derived totals. These included the population by age, sex and province from census estimates, the number of high income tax filers according to administrative data from tax returns, the volume of welfare benefits from

provincial government expenditure data, the number of unemployment insurance claimants from unemployment insurance administrative data, and distributions of families by size, labour market characteristics and province from data underlying the monthly Labour Force Survey sample frame. (The algorithm used is called raking or iterative proportional adjustment.)

Another adjustment is used to create data representing the institutionalized elderly, who are the most important group excluded from the coverage of the SCF. Essentially, duplicates were made of the records of all unattached non-institutionalized elderly individuals. These records were then flagged as representing the institutionalized elderly, and their weights were adjusted according to hospital and other administrative data to correspond with the known counts of this population by province, age range, and sex.

As part of the weight adjustment process, the weights of high income individuals are increased to offset their under-representation in the SCF. This is based on comparisons with data drawn from a sample of 400,000 income tax returns. This tax return sample is stratified by income and contains about 25,000 records with incomes over \$80,000, the cut-off point used for defining high income. Analysis of the SCF shows that there are too few high income individuals, and also those SCF records that do have income over \$80,000 have lower average high incomes. Thus, it is important to correct for this under-reporting of high incomes, and further to obtain richer detail on the joint distribution of various income components subject to preferential tax treatment (e.g. dividends, capital gains) and deduction items (e.g. carrying charges) relevant for tax policy analysis. To achieve this, the incomes of high income individuals in the SCF are completely replaced with synthetic microdata from the corresponding high income population from the income tax return sample.

This replacement process is done in two steps. First, the sample of 25,000 high income tax returns is clustered into 5,000 groups each containing exactly five individuals. The clustering process grouped individuals who are similar with respect to income, composition by source, and the pattern of tax deductions. Then each group of five individuals is averaged to create a sample of 5,000 synthetic tax returns. This clustering and averaging process is used so that the resulting synthetic set of individuals will not reveal any confidential information if it is made public, which is a basic objective. The second main step is to match each of these 5,000 synthetic high income individuals with a high income SCF respondent. SCF records are cloned as often as needed to absorb all 5,000 tax return-derived synthetic records, with their weights correspondingly reduced. The two processes of first reweighting high income SCF records as part of the raking algorithm, and replacing high incomes as reported on the SCF with high incomes from tax return data results in an increase from \$11.3 to \$19.2 billion in income in the hands of the estimated 135,000 high income individuals.

Another major weakness of the SCF from the viewpoint of income tax policy analysis is the absence of data on tax deductions such as charitable donations and childcare expenses. These variables are imputed based on a complex hierarchical disaggregation of the full sample of 375,000 tax returns (i.e. excluding the high income returns already utilized), linked to a corresponding disaggregation of the SCF. The particular disaggregation varies with each of the 25 or so variables

that are being imputed. The finest level of disaggregation is designed to leave clusters of about 100 tax returns. Then within each of these clusters, a non-parametric density function is estimated for the variable in question. The process of imputation to the SCF then uses a random number generator for each SCF record and each deduction item to draw from the appropriate density function. The main relevance of this process from the viewpoint of the household sector account is that it adds information to the SCF on employee pension plan and individual RRSP contributions.

The last major adjustment to the SCF in creating the SPSPD of relevance to the household sector account is the imputation of data on expenditure patterns. These data are drawn from the Family Expenditure Survey or FAMEX. This is a geographically stratified sample of about 10,000 households from whom very detailed data on annual expenditure patterns, as well as income and family structure, are obtained. The raw FAMEX micro records are matched in a structured, but ultimately random way in a manner similar to that used for the synthetic high income tax returns. The FAMEX records are, however, synthetically matched at the household rather than at the individual level. Each FAMEX record has to be cloned on average about four times to bring the number of records up to the number of SCF households. Again a hierarchical disaggregation is first used to cluster both FAMEX and SCF records, then within each cluster corresponding sets of records are randomly matched.

In addition to the processes just described, the SPSPD includes synthetically matched records derived from a one percent sample of administrative data on unemployment insurance claim histories (30,000 records) designed to support policy oriented modelling of the unemployment insurance system. It also incorporates a certain amount of controlled randomization designed to add sufficient noise to the data to prevent any respondents to the SCF from being identifiable, while at the same time preserving the analytical utility of the database. Thus, the full SPSPD can be released as a public use microdata set.

This completes a sketch of the major steps in the construction of the SPSPD. Unlike the previous section which described a series of adjustments to the personal sector aggregates designed to isolate the household sub-sector, the primary driving principle in the construction of the SPSPD has been the pragmatic need to support microsimulation modelling of the tax and transfer systems. Of course, this implicitly means that the data should naturally come out closer to "true" household sector aggregates. However, the construction of the SPSPD was not particularly influenced by any national accounting style concerns about correct concepts, for example whether or not to include private pensions in the household sector.

Indeed, from the viewpoint of microsimulation modelling, this latter question is ill-posed. First, in general, more data is always preferred to less data. According to this principle, then, data on private pension plans should certainly be included if they are available. Second, and perhaps more fundamentally, the data will be included or ignored on pragmatic grounds depending on the particular question under consideration. Since microdata are necessarily children of the computer age, they are never used except with a computer and with powerful database, retrieval, and modelling software. In this kind of environment, there is no need

to agree on one correct definition. One can simply change an instruction in the software and roll up any aggregation or concept that one wants, given the available data.

JUXTAPOSITION OF THE HOUSEHOLD SECTOR AND SPSPD AGGREGATES

We turn now to a comparison of the results of these two largely independent data construction efforts. The basic preliminary results are shown in Table 1.

The first column of the table shows the aggregate amounts (in millions of dollars) that result from simply adding the corresponding item across all the individuals in the Social Policy Simulation Database (taking account of their respective sampling weights). These are the micro figures for this prototype link between a micro and macro view of Canada's household sector for 1984. The second column gives the corresponding macro aggregates derived from the National Accounts personal sector via the adjustments outlined earlier (the

TABLE 1A
SPSPD/NATIONAL ACCOUNTS RECONCILIATION
(All Figures in Millions of Dollars)

	Receipts	SPSPD	National Accounts	Percent Difference
1	Wages & Salaries (Incl. Military Pay)	202,248	214,396	6.0
2	Self-Employment	16,082	17,992	11.9
3	Farm	2,864	2,981	4.1
4	Non-Farm	13,218	15,011	13.6
5	Investment	23,435	37,177	58.6
6	Dividends	4,742	7,610	60.5
7	Interest	13,147	24,474	86.2
8	Capital Gains	3,081		
9	Roomers & Boarders	125	118	-5.5
10	Other (Net Rental, Life Insurance Cash Dividends, Other)	2,341	4,975	112.5
11	Transfers	34,757	37,046	6.6
12	Family Allowance	2,291	2,393	4.5
13	Quebec Family Allowances	184	240	30.4
14	Old Age Security Benefits (OAS/GIS/SPA)	10,731	10,999	2.5
15	Canada/Quebec Pension Plan Benefits	4,869	5,599	15.0
16	Social Assistance	5,538	5,888	6.3
17	Unemployment Benefits	9,242	9,859	6.7
19	Other (Workers Compensation, Etc.)	1,902	2,068	8.7
20				
21	Miscellaneous	10,930	13,195	20.7
22	Pensions	7,751	8,810	13.7
23	Other Money Income (Alimony, Royalties, Strike Pay)	3,179	4,385	37.9
24	Other Money Receipts (Lottery, Lump Sum Settlements, Etc.)	5,692		
25	Total Receipts	293,144	319,806	9.1

TABLE 1B
SPSD/NATIONAL ACCOUNTS RECONCILIATION
(All Figures in Millions of Dollars)

	Disbursements	SPSD	National Accounts	Percent Difference
1	Expenditures	208,504	238,891	14.6
2	Food, Beverages and Tobacco	40,395	44,183	9.4
3	Clothing, Footwear and Accessories	13,193	15,473	17.3
4	Gross Rent, Fuel and Power	40,609	43,826	7.9
5	Furniture, Furnishings and Household Equipment and Operations	19,372	22,026	13.7
6	Medical Care and Health Services	5,012	9,361	86.8
7	Transportation and Communications	31,692	37,347	17.8
8	Recreation, Entertainment, Education and Cultural Services	16,788	21,391	27.4
9	Personal Goods and Services	28,441	41,071	44.4
10	Retirement Pensions	3,769	2,950	-21.7
11	Money and Other Gifts to Persons	6,370		
12	Other Expenditures N.E.C.	2,863		
13	Net Expenditures Abroad		1,263	
14	Taxes	52,490	55,352	5.5
15	CPP/QPP Contributions	2,600	2,787	7.2
16	Unemployment Insurance Contributions	3,426	3,164	-7.6
17	Personal Income Taxes	46,465	49,401	6.3
18	"Saving"			
19	Receipts minus Expenditures plus Taxes	32,150	25,563	-20.5
20	Net Change in Assets and Liabilities	17,024		
21	Account Balancing Difference	853		
22	Database Discrepancies	9,331		
23	RRSP Contributions	4,941	(5,636)	14.1
24	Total Disbursements	293,144	319,806	9.1

detailed data for the adjustments are outlined in Table 2). Finally, the third column shows the relative discrepancy between the two columns of figures.

Generally, the National Accounts-derived figures are higher. Still, for most items the differences are in the 5 to 15 percent range. Recall that these numbers almost always come from completely different sources. For example, wages and salaries in the National Accounts come from a Revenue Canada tabulation of wages and salaries submitted by employers, while in the SPSP they come from interviews with individuals.

The income and expenditure items have been arranged in two groups, receipts which are income plus some other lump sum receipts, and disbursements which correspond to expenditures plus taxes plus saving. There remain several conceptual differences that are apparent in the table. First, on the receipts side, capital gains are a source of income in household surveys and for income tax purposes, but are not included in the National Accounts. Second, the Family Expenditure Survey (FAMEX), in order to have a complete picture of receipts and disbursements to aid in the verification of interview responses, also asks for "other money receipts"—sources of funds which can be used to finance expenditures or saving but which are not usually thought of as income. Again, there is no corresponding concept in the National Accounts for either of two reasons. First, some of these

TABLE 2
THE ADJUSTMENT OF NIEA PERSONAL ACCOUNTS TRANSACTIONS TO SPSD CONCEPTS

Table 1 Category Number	NIEA Category/Item	Source ¹	1984 Data Millions of \$	Notes
<i>Receipts</i>				
1	Wages, Salaries & SLI	Q1.1	238,496	
	less			
	Food—other (imputed)	A62.2	498	
	Supplementary Labour Income (SLI)	Unpublished	22,595	Net of estimated SLI assumed to be in the same proportion as national SLI in W, S and SLI
	Yukon, NWT, and foreign countries	A38.11 and 12	889	
	Room and board—personal	Unpublished	118	Part of A62.6
	(imputed)			
	Total		214,396	
2	Accrued net income of farm operators	Q6.6	3,246	Excluding net accrued earnings of Canadian Wheat Board
	less			
	Farm product consumed on farm (imputed)	A62.1	197	Part of A62.6
	Farm fuel consumed on farm (imputed)	Unpublished	68	<i>Special note on line A62.6:</i>
	Total		2,981	The total amount of 517 m is divided as follows:
3	Net income of non-farm uninc. business,	Q6.7	24,715	Private room and board 118
	incl. rent			Fuel consumed on farm 68
	less			Institutional room and board 331
	Net rental	Q6.8	9,704	Cash and imputed, net of CCA
	Total		15,011	
6	Dividends paid to Canadian residents by Canadian corp.	Q15.14	7,202	The total of interest, dividend and misc. investment income received by persons from non-residents is 1,400 m divided as follows:
	Dividends received from abroad		408	
	Total		7,610	
7	Interest dividend and miscellaneous investment income	Q6.9	49,931	Dividends 408
	less			Interest 175
	Dividends		7,610	Miscellaneous 817
	Imputation for financial services	A62.5	3,381	

TABLE 2 (cont.)

Table 1 Category Number	NIEA Category/Item	Source ¹	1984 Data Millions of \$	Notes
	Investment income of:			
	Life insurance	Unpublished	5,456	
	Fraternal societies	Unpublished	144	
	Pension funds	Unpublished	6,972	
	Private non-profit organizations (PNPI), including Unions	Unpublished	651	
	Universities	Unpublished	159	
	Royalties	Unpublished	1,084	
	Total		24,474	
9	Room and Board—personal (imputed)	Unpublished	118	
10	Net rental income	Unpublished	945	
	Life insurance cash benefits	Unpublished	4,030	
	Total		4,975	
12	Family and youth allowances	A55.1	2,393	
13	Mothers and disabled allowances (provincial)	A55.26	240	
14 and 15	Old age security payments	A55.10	10,999	
16	Canada Pension Plan (CPP)	A55.38	4,045	
	Quebec Pension Plan (QPP)	A55.39	1,554	
	Total		5,599	
17	Direct relief (provincial)	A55.24	4,470	
	Old age, and blind pensions (provincial)	A55.25	705	
	Direct relief (local)	A55.34	713	
	Total		5,888	
18	Unemployment insurance benefits	A55.8	9,859	
20	Workmens' compensation benefits	A55.27	2,068	
22	Pensions to gov't employees	A55.9 and 28	2,594	
	Pensions from non-residents	Unpublished	428	
	Trusteed pension benefits	Unpublished	4,354	
	Life insurance annuity	Unpublished	1,434	
	Total		8,810	

Including fraternal society benefits. Original source is the Superintendent of Insurance Report. It shows death claims, annuity payments, surrender values and two other minor categories of benefits. This number is total, less dividends, and annuities.

TABLE 2 (cont.)

Table 1 Category Number	NIEA Category/Item	Source	1984 Data Millions of \$	Notes
24	Royalties	Unpublished	1,084	
	Strike pay	Unpublished	28	
	Adult occupational training	A55.13	103	
	Assistance to immigrants	A55.14	28	
	Charitable disbursements to individuals	Unpublished	440	
	Veteran's pensions and allowances	A55.2 and 3	1,092	
	RRSP benefits	Unpublished	1,610	
	Total		4,385	
<i>Disbursements</i>				
2	Food, beverage, and tobacco	A60.1	45,055	
	less			
	Food—other (imputed)		498	
	Farm product consumed on farm (imputed)		197	
	Territories	Unpublished	177	
	Total		44,183	
3	Clothing, footwear, and accessories	A60.5	15,518	
	less			
	Territories	Unpublished	45	
	Total		15,473	
4	Gross rent, fuel and powder	A60.9	57,159	
	less			
	Farm fuel consumed on farm (imputed)		68	
	Net imputed rent and CCA	A62.3	13,073	
	Territories	Unpublished	192	
	Total		43,826	
5	Furniture, etc.	A60.16	22,083	
	less			
	Territories	Unpublished	57	
	Total		22,026	

TABLE 2 (cont.)

Table 1 Category Number	NIEA Category/Item	Source ¹	1984 Data Millions of \$	Notes
6	Medical care and health services	A60.24	9,372	Small revisions required for purposes of this comparison
	less			
	Territories	Unpublished	11	
	Total		9,361	
7	Transportation and communications	A60.29	37,461	
	less			
	Territories	Unpublished	114	
	Total		37,347	
8	Recreation, entertainment, education and cultural services	A60.36	26,248	Small revisions required for purposes of this comparison
	Fees paid to universities	Unpublished	641	
	less			
	Expenditures of universities	Unpublished	5,436	
	Territories	Unpublished	62	
	Total		21,391	
9	Personal goods and services	A60.41	37,194	
	Premiums paid			
	—life insurance	Unpublished	8,846	
	—fraternal societies	Unpublished	273	
	Charitable contributions applicable to religion and welfare PNPI	Unpublished	2,322	Based on preliminary data
	Charitable dues	Unpublished	69	
	Union dues	Unpublished	536	
	Sub total		49,240	
	less			
	Expenses of religions and welfare PNPI	Unpublished	3,765	Based on preliminary data
	Trade unions	Unpublished	834	Contributions to political parties not available at this time.
	Political parties	Unpublished	103	
	Financial sector imputations	Unpublished	3,381	80% of actual number of \$108 m to allow for deductions.
	Territories	Unpublished	86	
	Total		41,071	

TABLE 2 (cont.)

Table 1 Category Number	NIEA Category/Item	Source ¹	1984 Data Millions of \$	Notes
10	Employee contributions to pension plans	Unpublished	2,950	
13	Net Expenditures abroad	A60.48	1,263	Not allocable by good or service category.
15	CPP	Q14.1	2,114	Lines Q14.1 and 14.2 are all employer <i>and</i> employee contributions. Adjustments have been made to retain all employee, and employer portion of self-employed contributions
	QPP	Q14.2	673	
			2,787	
16	Employee contributions to unemployment insurance	Unpublished	3,164	
17	Income taxes	Q6.18	49,555	
	less			
	Territories	G4.1 and 6	154	
	Total		49,401	
23	RRSP contributions		5,636	

Note: Source nomenclature. Q = *National Income and Expenditure Accounts*, Catalogue 13-001 Quarterly, table and line number. A = *National Income and Expenditure Accounts*, Catalogue 13-201 Annual, table and line number. G = *System of National Accounts, Provincial Economic Accounts*, Catalogue 13-213, Annual. Unpublished = Unpublished Statistics Canada data.

¹See Note at end of table for explanation of source nomenclature.

receipts are of a capital nature. Second, some reflect transactions entirely within the household sector. For example, these could be lump sum alimony settlements, bequests and inheritances. The corresponding disbursements by donors are shown under expenditures.

On the disbursements side of the account, the major conceptual difference relates to saving. For the National Accounts, saving is simply the difference between income and expenditure. In the FAMEX Survey, however, a series of questions are asked to determine saving directly. In part, this is done to provide redundancy in the questionnaire which can in turn be used for edit checks. This directly measured saving is referred to as the "Net Change in Assets and Liabilities." The residual error or discrepancy between income plus other money receipts on the one side and expenditures plus taxes plus saving on the other in the survey (household by household) is called the "Account Balancing Difference." In the FAMEX itself, this item serves to complete the identity between receipts and disbursements.

In Table 1, however, which has been derived from the SPSD, this is not sufficient. The reason is that not all variables have been added up from FAMEX data so that the underlying FAMEX identity need not hold. The market income variables (labour and investment income) have been taken mainly from the SCF, while payroll and income taxes and major transfers have been explicitly modelled. It should further be recalled that the SPSD also drew on income tax return data in order to adjust the incomes of high income individuals. As a result, a variable "Database Discrepancies" has been defined; it is simply the difference between receipts and disbursements as taken from the SPSD. It amounts to \$9.3 billion, and thus constitutes a substantial portion of household saving. This additional saving corresponds quite closely to the additional income imputed to high income individuals in the SCF based on tax return data.

An estimate from the National Accounts is shown for the total amount of contributions to RRSP's. It is shown in parentheses because this does not contribute to the National Accounts estimate of household saving, since that figure is derived purely as a residual difference.

The \$2.9 billion of expenditures not elsewhere classified (N.E.C.—row 12) in the SPSD column are expenditure that should really be allocated to one of the explicit categories (i.e. rows 2 to 10). They were grouped separately for the SPSD because of their tax exempt status under the sales tax. Similarly, the \$1.3 billion of household expenditures abroad (row 13) in the National Accounts should also be allocated to specific areas of expenditure, but the necessary data do not exist.

Overall, we have been quite surprised at how well this juxtaposition of two almost entirely differently derived views of the household sector compare. The most serious discrepancy is the \$13.8 billion difference in dividend and interest income. From the SPSD side, there are two possible reasons these numbers could be low. The first and most likely is simply under-reporting on the SCF. Recall that for high income individuals, incomes were taken from income tax returns, so that this under-reporting would apply only to the (large) group of individuals with incomes below \$80,000. Most of these individuals have relatively small amounts of investment income (if any), and they were all eligible for an income

tax exemption on the first \$1,000 of investment income. The other less likely possibility is tax evasion on the part of high income individuals. On the National Accounts side, the investment income figures could be higher than they should be if the portfolio income of some non-biological entities in the personal sector was not removed in the derivation of the household sector; for example, no data were available to remove the investment income of estates and trusts.

CONCLUDING COMMENTS

In this paper, we have presented a prototype link between micro and macro data for the Canadian household sector. The link is not a direct one—the aggregate figures are not the sum of the micro figures, even though in principle they could and perhaps should be. The reason is that many items in the personal sector of the National Accounts are not based on data from individuals; rather they are derived from the business and government institutions that transact with individuals.

As a result, the process described in this paper has been one of starting from both ends and working toward the middle. One starting point was the personal sector of the National Accounts. A series of adjustments was made to these macro data in order to isolate the incomes and expenditures ascribable to the household sector—the subset of agents conventionally grouped in the personal sector that are living, breathing individuals.

The other starting point was a variety of household surveys and individual administrative data sets. The independent micro data sets were compiled in one synthetically matched and adjusted micro data set called the Social Policy Simulation Database (SPSD). The SPSPD has as its primary motivation to serve as an input to the Social Policy Simulation Model (SPSM), a new analytical tool developed by Statistics Canada for general public use, designed to support analysis of prospective changes in the areas of income and sales tax and transfer program policy.

The key result in this paper is a juxtaposition and comparison of the view of the Canadian household sector derived from these two very different starting points. In our view, as a first effort, the comparison looks quite good.

Further work is clearly possible. In future years, more conscious effort could be devoted to keeping track of key items in the personal sector of the National Accounts to allow more reliable deconsolidation of the household sector. Further work is also required to verify the detailed conceptual differences that remain in the items juxtaposed in the main table of the paper. Finally, future versions of the SPSPD could embody further adjustments and imputations to bring it more in line with the National Accounts derived aggregates.

BIBLIOGRAPHY

- Aukrust, O. and Nordbotten, S., Files of Individual Data and their Potential for Social Research, *The Review of Income and Wealth*, 19 (2).
Department of Finance, *A Better Pension System, Saving for Retirement*, Ottawa, Oct. 1986.

- Fellegi, I. P. and Goldberg, S. A., The Computer and Government Statistics, in *The Role of the Computer in Economic and Social Research in Latin America*, Ruggles, N. D. (ed.), National Bureau of Economic Research, New York, 1974.
- Gigantes, T., Fellegi, I. P., Goldberg, S. A., and Podoluk, J., Micro Data Sets, Simulation and Statistical Systems, Workshop in Micro Data Sets, Washington, Oct. 1970.
- Hagenaars, A., Hauser, R., Hedstrom, P., Saunders, P., Smeeding, T., and Torrey, B., Improving the LIS Income Measure: Microdata Estimates of the Size Distribution of Cash and Noncash Income in Seven Countries, Presented at the 20th General Conference of the IARIW, Rome, 1987.
- McCracken, M. C., Jacobson, P. M., and Henson, H., A Review of Issues Underlying the Treatment of the Persons and Unincorporated Business Sector in the Canadian System of National Accounts, Unpublished report completed under contract to Statistics Canada.
- Okner, B. A., Constructing a New Data Base from Existing Microdata Sets: the 1966 MERGE File, *Annals of Economic and Social Measurement*, 1, 325-342, 1972.
- Okner, B. A. and Pechman, J. A., Uses of Tax Files combined with Field Surveys, in *The Role of the Computer in Economic and Social Research in Latin America*, Ruggles, N. D. (ed.), National Bureau of Economic Research, New York, 1974.
- Orcutt, G., Caldwell, S., and Wertheimer, R., II, *Policy Exploration Through Microanalytic Simulation*, The Urban Institute, Washington D.C., 1976.
- Pechman, J. A. and Okner, B. A., *Who Bears the Tax Burden*, The Brookings Institution, Washington D.C., 1974.
- Ruggles, N. and Ruggles R., A Proposal for a System of Economic and Social Accounts, in *The Measurement of Economic and Social Performance*, Milton Moss (ed.), National Bureau of Economic Research, New York, 1973.
- Ruggles, R. and Ruggles N., The Role of Microdata in the National Economy and Social Accounts, *The Review of Income and Wealth*, 21 (2).
- Ruggles, R. and Ruggles, N., The Integration of Macro and Micro Data for the Household Sector, *The Review of Income and Wealth*, 32 (3).
- Statistics Canada, *Social Policy Simulation Database—Technical Reference Guide*, mimeo, Social and Economic Studies Division, Ottawa, 1987.
- Stone, R., The Disaggregation of the Household Sector in the National Accounts, World Bank SAM Conference, Cambridge, April 1978.