

IN MEMORIAM: RICHARD RUGGLES (1916–2001)

I am deeply honored to be able to pay tribute to Richard Ruggles. I have known him for thirty years. He was my teacher, my boss, my thesis advisor, and a long-standing friend and colleague. I would first like to give a few personal reflections about my relationship with Richard and then to say a few words about his scholarly contributions.

I first met Richard in 1971. At that time, I was a second-year graduate student in the economics program at Yale University. I was also a research assistant, working on the IBM 370 at the Yale Computer Center. Those were the days of punch cards, rubber bands, and batch processing, where you submitted a job with a stack of cards and waited three or four hours (if things went well) to get your output back. However, if you had a big tape job—especially a multiple tape job—the job was run overnight, so that you received your output only the following day. I remember the Ruggles' house in those days piled with stacks of computer output, computer cards, and computer tapes.

Richard hired me as programmer in 1971 for his new project, the Measurement of Economic and Social Performance, at the National Bureau of Economic Research (NBER). The MESP project was my introduction to the world of microdata. Richard had, of course, been one of the founding fathers of national accounts in the United States. Indeed, in 1947, he had written the seminal textbook (and reference book) in this field, *An Introduction to National Income and Income Analysis*. He later followed up this book with *The Design of Economic Accounts* in 1970, written with Nancy, which then became the standard in this field.

However, in the early 1970s, Richard and Nancy had a vision about the future of national accounts. Microdata were just becoming available—the early Current Population Survey (CPS) files and the Public Use Samples from the decennial censuses provided by the US Bureau of the Census. Richard and Nancy saw the future of accounts as being derived from microdata sources, rather than the traditional means of aggregating up account totals from various administrative record sources. They saw the new accounts as a set of representative microdata samples representing each of the major sectors of the economy (households, businesses, governments, non-profits, and so on) linked together through interconnecting portals (workers employed in particular enterprises, for example). They also saw the main issue in national accounts as the reconciliation of microdata and the traditional aggregate totals. They realized that one data source could not fully represent the activities of any sector (for example, a full set of demographics, labor force activities, wealth profiles, and consumption behavior) and so this began the process of matching, merging, and even marrying different microdata sources.

At this time, Richard also introduced me to the IARIW. At that time, Nancy was Executive Secretary (a position she took over from Phyllis Deane of Cambridge University) and Richard was the managing editor of the *Review of Income and Wealth*. This was (and, I hope, still is) a great association—I am now Managing Editor and my wife Jane Forman is Executive Secretary—the acorn does not fall far from the tree. The first meeting of the Association he invited me to was in Rio de Janeiro, in January 1973, while I was still a graduate student at Yale (I think this spoiled me for future conferences). We would attend sessions in the morning and discuss weighty matters on the beach at Copacabana in the afternoon. In fact, I had my NYU job interview with Ned Nadiri in Rio (and somehow landed the job).

The Rio meeting was a “special” conference of the Association. My first general meeting was in Aulanko, Finland, in August 1975 (the IARIW has traditionally chosen rather esoteric places for their meetings). It was another fine meeting, with huge smorgasbords for lunch and dinner and saunas and (ice cold) lake swimming in the evening. I remember in the closing ceremony the head of the Finnish Statistical Office saying how terrific it was to be part of the “big Ruggles family.” That was what the association was—a big family, made possible by the warmth and commitment of the Ruggleses. I have attended every meeting since then (the meetings are held biannually). It still has the feel of a family reunion. It is not quite as much fun for me and Jane these days, since we now have to help organize and run it—but it is still the high point of the summer in those years.

Richard also provided a haven for the disaffected graduate students of the day. The early 1970s still reverberated from the political and social events of the 1960s—the protests against the War in Vietnam and the emergence of the so-called “New Left.” The Yale graduate students, at least in my class, were largely of this generation. URPE (the Union of Radical Political Economists) was strong on the campus. Richard himself also harbored a healthy skepticism for mainstream economics, and I think that was the basis of his bond with the “outsiders” in the graduate student community. I did my thesis under Richard’s direction—a structuralist analysis of Smith, Ricardo, and Quesnay.

In those days and until Nancy Ruggles’ untimely death in 1987, Nancy and Richard were truly partners. They worked side by side in the Association and in almost all their scholarly work. Indeed, the central feature of the living room of their house on Prospect Street in New Haven was their partner desk, where they could almost always be found. Their house was always full of people, especially students, and Richard’s classes were held in the house. After Nancy died, Caridad came on the scene and filled the void in his life, also with tremendous gusto and warmth. The house became alive again and when I last visited them, just before Richard’s death, it was again full of life and vitality.

Richard was consistently upbeat, optimistic, and positive. Even during the darkest days of Richard Nixon, Ronald Reagan, and George Bush senior, he never despaired about the future. He had a wonderful personality—always warm and friendly. He drew people in and made them feel comfortable and accepted. I only hope that I can emulate half of what Richard has accomplished and the type of person he was.

THE SCHOLARLY CONTRIBUTIONS OF NANCY AND RICHARD RUGGLES

I am most familiar with the writings of Nancy and Richard Ruggles on the integration of micro and macro accounting data and the development of microdata. Four principal themes emerge in this body of work.

The first is the reconciliation of macrodata with microdata. The development of national accounts was based on double-entry bookkeeping such that any product entry is matched with a corresponding income entry. The principal aim is to maintain consistency between the product and income accounts. Thus, for example, the definition of Gross Domestic Product (GDP) on the product side consists of five components: household consumption, investment, government expenditures, exports, and imports (treated as a negative entry). However, GDP can also be defined on the income side as the sum of three components: employee compensation, corporate gross profits, and indirect business taxes. One of the principal goals in constructing the national income and product accounts is to maintain equality between the two ways of defining GDP.

The national accounting system was developed during the 1930s and 1940s through the pioneering work of Simon Kuznets, Richard Stone, and Nancy and Richard Ruggles, among others. The 1950s and 1960s saw the development of major microdata sources, first in the US and later in other OECD countries. Most of these were based on household surveys. In the case of the United States, this process began with the creation of samples from the U.S. decennial censuses and the production of monthly samples from the Current Population Survey (CPS). Later, samples (and censuses) of enterprises in the economy were developed in the US as well as censuses of governmental units.

In principle, the data contained in these new microdata sources should be consistent with the macrodata in the national accounts. For example, the sum of income entries in the March supplement of the CPS should match the corresponding entries of personal income in the national accounts; the sum of incomes and sales in the enterprise microdata should match the enterprise data in the national accounts; and the total for government expenditure in the government microdata should equal the corresponding entries in the national accounts. In practice, however, this was rarely the case and the two sources often produced rather disparate estimates. This is particularly true for the household sector for which estimates of total interest, dividends, transfer payments, and even wages and salaries derived from the two sources often differ.

Three of their papers address this issue in great detail: “The Role of Microdata in the National Economic Accounts,” “The Development of Integrated Data Bases for Social, Economic and Demographic Statistics,” and, particularly, “The Integration of Macro- and Micro Data for the Household Sector.” Several requirements are put forward by Nancy and Richard Ruggles to fully integrate the two sources. First, the definition of sectors should be the same in the macrodata and microdata. For example, while in the household microdata only households are included, the macro “household” accounts often include non-profit institutions. Second, definitions and imputations should be consistent between the two sources. For example, while employer-financed pension contributions are recorded as part of employee compensation in the national accounts, this component does not normally appear in household microdata as part of compensation.

Another example is that while the national accounts data include imputed rent to owner-occupied housing, this element is rarely present in household microdata. Third, alignment of macro and microdata should be a two-way street. Though national accounting data are internally consistent (between the product and income sides), this does not necessarily imply that macrodata are necessarily superior to the corresponding microdata. For example, in national accounting data, the interest received by the household sector as personal income is computed as a residual from that of other sectors, whereas in microdata the household provides a direct estimate of interest received.

The second theme is the synthesis of microdata from several sources. The problem addressed by Nancy and Richard Ruggles is that individual data sources, particularly household surveys, can ask the respondent only a limited number of questions. This is due both to practical limitations on the length of interviews and the cost of processing additional questions. As a result, different household surveys have concentrated on different kinds of household behavior. The CPS and the decennial census sources focus mainly on demographic characteristics of households and income receipts. The Consumer Expenditure Survey is very strong on consumption expenditure data but relatively weak on income and demographic details. The Federal Reserve Board's Survey of Consumer Finances concentrates primarily on household asset and liability data but offers very little information on consumption expenditures.

Another problem is that different microdata sources may focus on different parts of the distribution. This is particularly the case with household data on income. For example, both the decennial census and the CPS focus mainly on the broad middle classes but are relatively weak on the lower tail (the bottom 10 percent) and the upper tail (the top 5 percent) of the income distribution. In contrast, the Internal Revenue Service Tax Model, a sample of tax returns stratified by income, contains detailed income data on the upper tail on the distribution. However, it contains very limited information on the bottom tail on the income distribution, since most of these families do not file tax returns. It also has very sparse information on the demographic characteristics of the individuals in its sample.

The solution proposed by Nancy and Richard Ruggles is a statistical match of microdata sources. The idea of statistical matching is to combine microdata files which are complementary in terms of the variables they contain or the parts of the distribution that they sample. This approach is developed in two articles. The first, "A Strategy for Merging and Matching Microdata Sets," lays out one method for statistical matching. The basic procedure is to select common or overlapping variables found in the two datasets to be matched. The most important of these variables, called cohort variables, are chosen to be matched on an exact basis between the two datasets. These might include the type of household (married versus single), gender (in the case of singles), age, and race. The other overlapping variables, called "X variables" are matched on the basis of pre-assigned intervals rather than on exact values. For example, a typical X variable is total family income, which could be matched within hundred dollar or thousand dollar intervals. The two files are first sorted on the basis of the cohort variables and

then, within cohorts, on the basis of the X variables. The two files are then merged on the basis of their closest match.

The second article, “Merging Microdata: Rationale, Practice and Testing,” describes one such successful match carried out by the Ruggleses between the 1970 Census of Population Public Use Sample (PUS) and the 1969 Internal Revenue Service Tax Model (IRS). One problem was that the two files had different units of observation: the former was based on households and the latter on tax units. This problem was solved by assuming that all married couples filed joint returns and individuals, including unmarried adults in a household, filed single returns. The cohort variables (which were matched exactly) used in the statistical match were: (1) type of tax return; (2) sex of respondent (in the case of single returns); (3) race of head of household; (4) age of head of household; (5) number of children; and (6) owner-occupied home versus rental unit. The matching (or X) variables, for which statistical matching was done within intervals, were: (1) wage and salary income; (2) business earnings; (3) farm income; and (4) total income. Results of the match are documented in considerable detail in the article. Statistical tests were also applied which indicated that reliable synthetic datasets could be constructed from the sort-merge matching procedure developed by the Ruggleses.

The third theme is the creation of new longitudinal microdata sets. For the household sector there do exist “panel” datasets which re-interview the same households on a year-by-year basis. A cross-sectional sample of households is selected in the base year of the survey. This sample is interviewed in successive waves over time to create a longitudinal data file. One such example is the Panel Study of Income Dynamics (PSID), which was developed and is maintained by the Survey Research Center of the University of Michigan. This data source has led to a rich body of literature, on such topics as income mobility, poverty spells, and income variability over time.

There also exist administrative records kept by the federal government which could form the basis of the creation of longitudinal datasets. Though they are not specifically designed as panel data, they do contain identification numbers that would allow the records to be matched on a year-by-year basis.

Nancy and Richard Ruggles, in their paper, “The Analysis of Longitudinal Establishment Data,” demonstrate how this can be done in the case of establishment records kept for the US Census of Manufacturing. Every five years, the US Bureau of the Census does a census of all manufacturing plants in the US. It records data such as the number of employees, total sales, total profits, and total capital stock owned in the plant. It keeps records of these plants on the basis of an ID number. In non-census years, a sample of establishments is re-interviewed (in the Annual Survey of Manufacturing), and the same questions are asked. The combination of these two data sources can be used to develop a longitudinal dataset of manufacturing establishments.

The development of such a longitudinal data file is not as straightforward as it might appear. First, there are both births (new entries) and deaths (plant closures) which complicate the process of keeping track of such plants over time. Second, some firms may merge while others may divest themselves of one or more plants by selling them to another firm. Since it is important to keep track of firm ownership information for purposes of statistical analysis, this problem further

complicates the development of a longitudinal dataset. Third, a plant may shift in terms of the products it produces, thus resulting in a change in industry classification.

The Ruggleses' pioneering work in this field produced a file which they called the LED file—Longitudinal Establishment Data file. The period covered was from 1974 to 1981. This file has now evolved into the Longitudinal Research Database (or LRD), which is maintained and updated annually by the Center of Economic Studies at the US Bureau of the Census. The database is now available to a wide range of researchers and has already given rise to numerous studies. One example is the 1996 book, *Job Creation and Destruction*, by Steven Davis, John Haltiwanger, and Scott Schuh.

The technique developed by the Ruggleses can today be applied to other governmental data sources. One example is the social security records on individual workers maintained by the Social Security Administration. Annual reports of wage and salary earnings, as well as social security contributions, are easily identified by social security number and could be linked over time to create a longitudinal dataset (this was done for a short period of time to create the so-called LEEDS file). Other possibilities are the personal income tax records and the corporate income tax returns of the Internal Revenue Service (IRS). Currently, the IRS provides researchers with a sample of annual tax returns for individuals. These could also be linked over time (again on the basis of the social security number) to create a longitudinal dataset.

The fourth theme in their work is the importance of institutional sectoring for the analysis of economic behavior. This problem is addressed in three papers: "Household and Enterprise Saving and Capital Formation in the United States, 1947–1991: Market Transactions View," "Accounting for Saving and Capital Formation in the United States, 1947–1991," and "The Integration of Macro- and Micro Data for the Household Sector." All three articles focus on the measurement of savings. Though most theories of savings, such as the permanent income hypothesis or the life-cycle model, implicitly assume that all savings is done by households, Nancy and Richard Ruggles argue that savings is done by different institutions within the US. In particular, besides households, the enterprise sector itself (as well as governments) engages directly in savings behavior.

In their accounting scheme, they develop separate current and capital accounts for the household sector, the enterprise sector, and the government sector. Their most provocative finding is that the household sector and the enterprise sector are each self-financing. In other words, almost all the financial savings done by households is used to pay for household capital formation—particularly, housing and consumer durables. On net, the household sector channels almost no financial savings to the enterprise sector. Conversely, almost all the capital formation done by enterprises is financed through enterprise savings—particularly, undistributed gross profits. These sets of results have wide-ranging implications for modern theories of savings and investment.

EDWARD N. WOLFF
New York University

REFERENCES

Davis, S. J., J. C. Haltiwanger, and S. Schuh, *Job Creation and Destruction*, MIT Press, Cambridge, MA, 1996.

- Ruggles, N. D., The Development of Integrated Data Bases for Social, Economic and Demographic Statistics, *Studies in Methods*, United Nations, Series F, No. 27, 1979.
- Ruggles, R., *An Introduction to National Income and Income Analysis*, McGraw Hill, New York, 1947.
- , Accounting for Saving and Capital Formation in the United States, 1947–1991, *Journal of Economic Perspectives*, 7(2), Spring 1993.
- Ruggles, R. and N. D. Ruggles, *The Design of Economic Accounts*, NBER, New York, 1970.
- , The Role of Microdata in the National Economic Accounts, *Review of Income and Wealth*, 21(2), 1975.
- , The Analysis of Longitudinal Establishment Data, Paper presented at the Bureau of the Census and NSF Conference on Longitudinal Establishment Data File and Diversification Study, Alexandria, VA, October 17–18, 1984.
- , The Integration of Macro- and Micro Data for the Household Sector, *Review of Income and Wealth*, 33(1), 1987.
- , Household and Enterprise Saving and Capital Formation in the United States, 1947–1991: Market Transactions View, *Review of Income and Wealth*, 38(2), June 1992.
- Ruggles, R., N. D. Ruggles, and E. N. Wolff, Merging Microdata: Rationale, Practice and Testing, *Annals of Economic and Social Measurement*, 6(4), 1977.