

THE DATA: A BRIEF DESCRIPTION

JAN R. MAGNUS^a AND MARY S. MORGAN^{b*}

^a*CentER for Economic Research, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands*

^b*University of Amsterdam and London School of Economics, Houghton Street, London WC2A 2AE, UK*

SUMMARY

This paper provides a brief description of all the data we provided for use in our experiment in applied econometrics: aggregate time-series data for the USA and The Netherlands; cross-section sample-survey budget data for the USA; and cross-section household-level budget data for The Netherlands. We include brief descriptions of the data sets, definitions of the variables involved and documentation of original sources. We indicate where full information about the data, and the data themselves, can be found. © 1997 John Wiley & Sons, Ltd.

J. Appl. Econ., 12, 651–661 (1997)

No. of Figures: 0. No. of Tables: 0. No. of References: 3.

1. GENERAL INFORMATION

The data consist of both time-series and budget survey data from the USA and the Netherlands. We use 'BS' and 'TS' to denote budget survey and time series, respectively. Also, we use 'US' and 'NL' to denote data from the USA and the Netherlands respectively. The data were originally provided on a diskette containing nine files: five files for the USA and four for the Dutch data.

The US data consist of four sets of cross-section observations and one time-series set. Our aim has been to provide cross-section data on incomes and expenditures as close as possible to the type used by Tobin (1950); and to extend the data sets where possible to provide information on certain household characteristics. The four cross-section sets are from consumer expenditure interview surveys of samples of US households undertaken in 1941 (used by Tobin), 1950, 1960/1 and 1972/3. The data provided are for groups of urban households, extended to cover all households in the last survey. The survey methods used were roughly comparable in their approach. The four US budget surveys data were located in files bearing the following names: BS41US.WQ2, BS50US.WQ2, BS60US.WQ2 and BS72US.WQ2.

With the US time-series data, our aim has been to provide as consistent an updating as possible for Tobin's original data set. Where we have been unable to update Tobin's data, we have provided alternative series and appropriate scaling information to allow for closer comparison with household survey data. These revised and additional series cover the period 1912–89. The US time series data were located in the file: TS1289US.WQ2.

For the Netherlands, we have three budget surveys for the years 1965, 1980 and 1988. The 1965 survey gives data on 26 groups of households. The 1980 and 1988 surveys are at household level and are therefore much larger. They contain 2859 (in 1980) and 1950 (in 1988) households

* Correspondence to: Mary S. Morgan, Department of Economic History, London School of Economics, Houghton Street, London WC2A 2AE, UK

respectively. The data sets were provided (in ASCII format) under the following names: BS65.NL, BS80.NL and BS88.NL.

The Dutch time series run from 1948 to 1988 (41 years) and were located, also in ASCII format, in the file: TS4888.NL.

The data are defined, described and documented following this introduction: Sections 2–3 for the US data, Sections 4–6 for the Dutch data. Further technical notes, exact definitions for all variables, all the US data, Dutch data for the 1965 survey, summary statistics for the 1980 and 1988 survey data sets, the Dutch time-series data, and a full description of sources were all given in Magnus and Morgan (1995) and will also be reported in Magnus and Morgan (1998). The full data sets are available in the *Journal of Applied Econometrics* data archive <http://qed.econ.queensu.ca/jae>.

2. BUDGET SURVEYS FOR THE USA

The data are obtained from sample surveys of households carried out at four different periods covering expenditures and incomes. The reported observations are organized into matrices, vertically by income classes within different household sizes, beginning with one-person households and ending with all households taken together. Each observation represents a group of households in a particular income class and of a particular household size.

The budget surveys provided three different types of information, arranged horizontally in the data matrices with three different types of information: (i) information on sample size/population of households; (ii) characteristics of households in the group; and (iii) income/consumption data for households in the group.

2.1. Variable Definitions and Data Availability

Not all surveys have all the variables listed below. In addition, not all variables share exactly the same definition (differences are mostly minor). The definitions given here are the standard shared definitions for all four surveys.

(i) Sample size/population shares:

%INPOPH: the per cent share of such (income/size) households in the population of households. It is 'population', because the sample has been grossed up using a weighting system

ESTPOPH: the estimated total number (the population) of such (income/size) households, measured in 000s

SAMPSIZE: the actual sample number of such households in the survey

(ii) Household group characteristics

AHSIZE: average household size

ANCHILD: average number of children per household

ANFTEARN: average number of full-time earners per household

ANOV65: average number of persons over 65 years per household

AGEHH: average age of head of household in years

EDUHH: average years of education of household head

EDUHH1: per cent of heads of household with 1–9 years education

<i>EDUHH2</i> :	per cent of heads of household with 9–12 years of education
<i>EDUHH3</i> :	per cent of heads of household with over 12 years of education
<i>NONWHITE</i> :	per cent of households which are non-white
<i>HOMEOWN</i> :	per cent of households which are homeowners
<i>RENTER</i> :	per cent of households which are renters
<i>AUTOOWN</i> :	per cent of households owing one or more automobiles

(iii) Income and consumption

<i>HINC</i> :	average household disposable income in \$ (all incomes after taxes, excluding net asset changes)
<i>TOTCON</i> :	average household total current consumption expenditure in \$ (excludes personal insurance, gifts and contributions and net changes in assets)
<i>FOODCON</i> :	food and beverages consumption expenditure in \$ (including at home and away from home)
<i>NETASSCH</i> :	average household net asset change in \$
<i>ACCBAL</i> :	account balancing item in \$. ¹

2.2. 1941 Urban US Budget Survey: Tobin's Data Set

The data used by Tobin (his Table 2) for his cross-section analysis of expenditure on food by family size came primarily from the 1942 survey designed to collect information on income, expenditure and saving over the period 1941 and the first quarter of 1942. This was an interview field survey, covering 62 urban areas (urban areas started at 2500 population). The usable number of interviews in the 1941-period sample ended up at 'about 1300' (according to a footnote, p. 32, this was 1220 full-period units and 79 part-period units).²

Not all the collected information (for example, on household characteristics such as race and occupation) were reported in the published tables, and Tobin used what was reported very effectively. Wherever possible, we have added data available in the survey to his data set. For example, we have added information on total consumption expenditures (*TOTCON*) by family size, and data on the highest income groups, which Tobin avoided, possibly because of the poorer quality of the sample. We also corrected a couple of errors in the numbers originally reported by Tobin.

The survey used nine income classes (ranging from \$0 to over \$10,000 annual pre-tax money income) and six household size classes (1, 2, 3, 4, 5 and more persons, and all households). The following variables are available for each income/family size group: *ESTPOPH* (for US urban population), *SAMPSIZE*,³ *HINC*, *TOTCON*, and *FOODCON*. Non-reported data occur where the original sample size was too small to admit reporting.

¹ The account balancing item is calculated as: $ACCBAL = \text{Total receipts} - \text{total disbursements}$ (including personal insurance, gifts and contributions) – net increase in assets. (Note that this is not the same as $HINC - TOTCON - NETASSCH$ because *TOTCON* is not equal to total disbursements.) *ACCBAL* is either reported in the tables, or calculated from them. According to the 1960–61 survey report, it is a 'net reporting discrepancy' and so it is given here as one indicator of the accuracy of the reported information. When positive, it indicates that reported income exceeds reported disbursements, when negative that reported spending plus saving exceeds income.

² The design appears to have been rather successful with a high response rate (the main difficulty being a rather high 'refusal' rate, necessitating substitutions, of those in the highest income brackets) and remarkably complete recording. The interviewers seem to have been so effective at obtaining data that net asset change reported is equal to the net surplus/deficit on current income/disbursement flows. (Thus there is no reported *ACCBAL* variable.)

³ Although the basic sample sizes for each group are not given, we were able to 'reconstruct' (and send to participants on 19 August 1995) the sample sizes for each cell to provide the variable *SAMPSIZE*. (These are available with the rest of the data: see Section 1 above.)

There are a few additional pieces of information available, particularly for 'all households' taken together, which we have added to the set of data Tobin used in his paper. These are: *AHSIZE*, *HOMEOWN*, *RENTER*, *AUTOOWN* and *NETASSCH*. For variable definitions, see Section 2.1 above. The source for the data and discussion of the survey is *Family Spending and Saving in Wartime*, Bulletin No. 822 of the Bureau of Labor Statistics, US Department of Labor, Washington, DC, 1945.

2.3. 1950 and 1960–61 Urban US Budget Surveys

The 1950 survey was an interview field survey carried out in early 1951 to collect information on all major items of expenditure and income for the whole year of 1950. It resulted in complete collection from 12,489 consumer units covering 91 urban areas. It was of the same type as the survey of 1941, but broader in coverage in every respect, in particular providing data on household characteristics. All data are based on 'complete reporters' only: those households where there was a certain consistency and completeness of all information about expenditures and lifestyles.

There are nine income classes (ranging from \$0 to over \$10,000 annual pre-tax money income) and seven household size classes (1, 2, 3, 4, 5, 6 and more persons, and all households). (The variables' definitions are given in Section 2.1 above.) Generally, the data are more reliable on expenditure than on income, for the middle-income groups than for the higher and lower groups, and for the major items of expenditure than for minor ones.

Variables available are: *%INPOPH*, *SAMPSIZE*, *AHSIZE*, *ANFTEARN*, *AGEHH*, *EDUHH*, *NONWHITE*, *HOMEOWN*, *HINC*, *TOTCON*, *FOODCON*, *NETASSCH* and *ACCBAL*. The source both for data and discussions is *Studies in Consumer Expenditure, Incomes and Savings*, tabulated by the Bureau of Labor Statistics, US Department of Labor for the Wharton School of Finance and Commerce, University of Pennsylvania, 1957; Vols III and XVIII.

The 1960–61 survey was similar to the 1950 interview survey (and also involved a diary survey, which provided no data in this case). It was somewhat broader in geographical scope (though it covered only 66 urban areas) and time period. Two samples of households were interviewed in 1961 and 1962, covering, respectively, the years 1960 and 1961. This resulted in a total sample of 9476 usable ('complete reporting') households.

There are ten income classes (ranging from \$0 to over \$15,000 annual *after-tax* money income) and seven household size classes (1, 2, 3, 4, 5, 6 and more persons, and all households). (The variables definitions are given in Section 2.1 above.) Particular care is needed with the extremes of the income distributions (as with all other survey data used here). The average household size variable (*SAMPSIZE*) is given only for each household size category (i.e. aggregated across incomes). The average household disposable income (*HINC*) data exclude value of *all* goods 'received without expenses' on the income side (compared with 1941 where they were included and 1950 where there was partial inclusion).

Variables available are: *%INPOPH*, *ESTPOPH*, *SAMPSIZE* (in part), *AHSIZE*, *ANCHILD*, *ANFTEARN*, *AGEHH*, *EDUHH*, *NONWHITE*, *HOMEOWN*, *AUTOOWN*, *HINC*, *TOTCON*, *FOODCON*, *NETASSCH* and *ACCBAL*. The sources for both data and discussion are *Consumer Expenditures and Income, Urban United States, 1960–61*, Survey of Consumer Expenditures, 1960–61, Bureau of Labor Statistics Report No. 237-38, April 1964, US Department of Labor; and Supplement 2—Part A, July 1964, *Cross-Classification of Family Characteristics*.

2.4. 1972–3 Urban and Rural US Budget Survey

The 1972–3 consumer expenditure survey covered all areas, not just urban ones. Like the 1960–1 survey, there were interview and diary elements, and, as before, we rely on the interview results. Unlike the previous surveys, this survey used a panel visited every quarter for five quarters from January 1972 to March 1973 and from January 1973 to March 1974. The data are averages of the two samples, involving 9914 units in the first period and 10,158 in the second. Expenditure information was collected in all periods, income information in periods 2 and 5, asset information in period 5.

There are 12 income classes (ranging from \$0 to over \$25,000 annual pre-tax money income) for the six individual household size classes (1, 2, 3, 4, 5, 6 and more persons) and 16 income classes (ranging to over \$50,000 income) for all households taken together. (The definitions of variables are given in Section 2.1 above.) Unfortunately, we do not have sample sizes for this set of data. For broad categories of expenditure and total expenditures, the results are thought to be rather accurate.⁴

Variables available are: *ESTPOPH*, *AHSIZE*, *ANCHILD*, *ANOV65*, *AGEHH*, *EDUHH1*, *EDUHH2*, *EDUHH3*, *NONWHITE*, *HOMEOWN*, *RENTER*, *AUTOOWN*, *HINC*, *TOTCON*, *FOODCON*, *NETASSCH* and *ACCBAL*. The sources for data and discussion are *Consumer Expenditure Survey Series: Interview Survey 1972–73*, Bureau of Labor Statistics Bulletin No. 1985 (August 1978), US Department of Labor; and December 1974 *US Monthly Labor Review*, Volume 97:2.

3. TIME SERIES FOR THE USA

The US time series consist of three sets of data series:

(i) *Tobin's original data series* (1912–48, with 1942–4 missing; index numbers: 1935–9 = 100)

TOBPCFC: Tobin's per capita food consumption (his S)
TOBPCY: Tobin's per capita disposable income in \$ (his Y')
TOBFP: Tobin's food price index (his P)
TOBNFP: Tobin's non-food price index (his Q).

(ii) *Revised series*: Best estimates of Tobin's series, revised, updated, re-referenced (to 1967 = 100 where appropriate) and as complete as possible (1912–89)

PCFC: revised *TOBPCFC* for 1912–76
PCY: revised *TOBPCY* for 1912–89
FP: revised *TOBFP* for 1913–89
NFP: revised *TOBNFP* for 1913–89.

⁴ But the survey interviewers were less concerned than in the 1960–61 and earlier surveys to balance the household accounts. Thus there are larger gaps between total incomes and disbursements due to under-reporting of income, involving both forgetfulness and error. In addition, in the 1972–3 survey, large items of expenditure were placed fully in the year even if they were paid for over several years, whereas the 1960–61 survey managed to cope with lumpy expenditures rather more satisfactorily. Thus the relatively greater magnitude and variance of *ACCBAL* compared to earlier surveys — despite the omission of incomplete reporting units from the tables — suggests a greater unreliability in the overall totals in this survey than in earlier ones (see also footnote 1 above).

(iii) *Additional data series*: Alternative data for income and expenditures (1929–89), and population and household data for scaling purposes (1912–89)

<i>AGGEXP</i> :	aggregate personal expenditure in billion \$
<i>AGGEXPF</i> :	aggregate personal expenditure on food in billion \$
<i>AGGY</i> :	aggregate personal disposable income in billion \$
<i>POP</i> :	population in thousands
<i>NOH</i> :	number of households in thousands.

3.1. Data Description

TOBPCFC: Tobin's series of food consumption, an index from the Bureau of Agricultural Economics (US Department of Agriculture), was a price-weighted quantity index which measured the per capita 'disappearance' of the physical quantity of food to the civilian population. His index was 1935–9 = 100, with price weights for the same year, and years 1942–4 missing. Of all the data series, Tobin's data for food consumption is the most difficult to bring up to date.

PCFC: The Bureau provided a similar food consumption index for the years 1913–76 involving price base weights from 1947–9 for the years up to 1954 and weights from 1957–9 for years after that. This index has a reference year 1967 = 100. Because the price base weights have changed, this index does not replicate exactly the movements shown in Tobin's index. (It is very close up to 1941, but decidedly different from 1945.) We have provided this index.

AGGEXP and *AGGEXPF*: Since Tobin pointed out that his time-series quantity data were not strictly comparable to those used in budget studies, we have included two alternative series, namely time series of total personal consumption expenditures and of the sub-category for personal expenditure on food (including alcoholic beverages but excluding tobacco). The data are in billions of dollars, at current dollar prices.

These two data sets cover the period 1929 to 1989. They are consistent with each other and with the series for disposable personal income (*AGGY* and *PCY*), since they all form part of the national income and product accounts produced by the Bureau of Economic Analysis at the US Department of Commerce. (These appear to be of the same type as used by Stone in his reworking of the data to provide the Tobin–Stone estimates on page 142 of the discussion of Tobin (1950).) The income and expenditure totals were recently revised, but only for part of the period, and apparently not fully for the food data series. We therefore decided to end the data series at 1989, in order to provide a consistent set of series going back to 1929. Data for the period before 1929 were more than 5% different on the 1929 figure, were only partially available, and have been omitted.

TOBPCY: Tobin's per capita disposable income figures were computed from the currently available national income data divided by population figures. As elsewhere, 1942–4 were missing.

PCY: There have been many revisions to the national income accounts and some also to the population figures. We have provided an updated and revised series for 'disposable personal income', per capita, in current dollars, using Tobin's data until 1918 and thereafter data drawn from the national income accounts data, and divided by the population series included in this data set. This series runs from 1912 to 1989.

AGGY: In addition, we have provided the aggregate income series used to calculate the above per capita series for the period 1919–89. This is total 'disposable personal income', in billions of

current dollars, drawn from the national income accounts. Disposable personal income includes all incomes (wages, rental incomes, benefits, etc.) of the personal (non-government, non-corporate) sector minus tax payments by the sector.

In addition, figures for resident population and for numbers of households are also given, to provide a choice of unit for scaling both total and food expenditures and disposable income. These figures come from the US Department of Commerce, Bureau of the Census.

POP: The population figures are for total resident population in thousands including armed forces abroad. They are derived from the decennial census as benchmarks using other data (such as on births, deaths and immigration) to estimate between census dates for 1 July each year.

NOH: The number of households, also in thousands, is based in part on census data and for later years in part on the Current Population Survey (CPS: a monthly sample survey used to estimate population characteristics).

TOBFP and *TOBNFP*: Tobin's price series were obtained from Bureau of Labor Statistics data series of Consumer Prices Indices for middle-income families in large cities. The series were for food, and for all items less food. They had reference years 1935–9 = 100, and years 1942–4 were missing.

FP and *NFP*: These price series have seen many weighting revisions and widening in coverage over the years since 1950 (and incidentally rely on the consumer expenditure surveys data to provide quantity weights). Despite all these changes it is possible to provide a continuous series from individual series which overlap in a satisfactory way beginning with Tobin's series for 1913–41. Both indices have been referenced to 1967 = 100.

3.2. Sources/References

All US time-series data have been collected from the following four sources:

Economic Report of the President for 1978, 1984, 1990, 1991, 1994 (US Government Printing Office, Washington, DC);

Historical Statistics of the United States, Colonial Times to 1970 (US Bureau of the Census, Washington, DC, 1975);

Statistical Abstract of the United States 1975, 1978 (US Department of Commerce, Bureau of the Census, Washington, DC); and

The National Income and Product Accounts of the United States 1929–1976 (Washington, DC, 1981).

4. BUDGET SURVEYS FOR THE NETHERLANDS

Since 1978, Statistics Netherlands (in Dutch: Centraal Bureau voor de Statistiek (CBS)) has conducted Budget Surveys (BSs) annually. Before 1978, Statistics Netherlands conducted BSs only occasionally. The aim of the BSs is to gather information on the expenditure behaviour of Dutch households in relation with household characteristics such as household composition, income and education level of all members of the household.

We have expenditure data for three years: 1965, 1980 and 1988. The 1965 survey gives data on 26 *groups* of households. The 1980 and 1988 surveys are at household level and are therefore much larger. They contain 2859 (in 1980) and 1950 (in 1988) households, respectively. We thank Statistics Netherlands for providing these data and allowing their usage in the experiment.

For all three surveys total household consumption ($V1$) is divided into six groups (the sum $V11$ to $V66$ is not exactly equal to $V1$ due to measurement error):

- $V1$ Total household consumption
- $V11$ Food
- $V22$ Housing (includes rental value for owner-occupied homes)
- $V33$ Clothing and footwear
- $V44$ Hygiene and medical care
- $V55$ Education, recreation and transport
- $V66$ Other consumption (includes banking and insurance services).

More detailed expenditure data is available for seven subgroups of food (again, the sum of $V110$ to $V119$ is not exactly equal to $V11$ due to measurement error):

- $V110$ Bread, pastry and flour products
- $V111$ Potatoes, vegetables and fruit
- $V113$ Products containing sugar and beverages
- $V115$ Oils and fats
- $V116$ Meat, meat products and fish
- $V118$ Dairy products
- $V119$ Other food products.

Further variables available for each survey are discussed below.

4.1. The 1965 Dutch Survey

The 1965 budget study was conducted between June 1963 and May 1965. The information provided is scarce, mainly because the original micro data have been lost and all information has thus been gathered from the official Statistics Netherlands publication of 1966. The survey provides data on groups of households, not on individual households. Households have been categorized into groups according to three variables:

HINC: Gross (!) household income in 1965 Dutch guilders in two classes (i) $f6000-9000$ and (ii) $f9000-12,000$

HSIZE: Household size, the number of persons within the household in five classes (no single-person households)

URB: Degree of urbanization in three classes: (i) three largest cities: Amsterdam, Rotterdam and The Hague; (ii) other cities/urban areas; and (iii) villages/rural areas.

Altogether we have 26 groups of households (rather than 30, since four groups are empty). The 1965 survey is not as rich as the 1980/88 surveys. The degree of urbanization, however, is only available for this cross-section. Only households with more than two persons, where the head of household is a blue- or white-collar worker or an agricultural worker or a farmer, are selected. To make the data set representative for the Dutch population, weight variables were created based on profession (blue- or white-collar worker, agricultural worker or farmer) and number of persons in the household. The weights themselves are not available. The presented data are weighted, if applicable.

Thus, apart from *NUMH* (the number of households within the group), the data set also includes *WNUMH*. At the household level, each household and its consumption expenditures⁵ were weighted before aggregating over households within the 26 groups of households. The resulting variables are the weighted number of households (*WNUMH*) and the average weighted consumption expenditures (*V1* to *V66*). Thus *WNUMH* is simply the sum of the weights within one group.

Statistics Netherlands also provides information about the 'food equivalent' household sizes. (It is important to note that these equivalence scales only apply to food consumption.) A child eats less than a grown man and this is reflected in *ADULTEQ*: the weighted number of adult equivalents in the household.

In summary: apart from the average weighted expenditures per household group for total consumption and consumption categories listed above, the following variables are available for the 1965 survey: *HSIZE*, *NUMH*, *WNUMH*, *ADULTEQ*, *URB* and *HINC*.

4.2. The 1980 and 1988 Dutch Surveys

Information at the household level is available in the 1980 and 1988 budget surveys. Both surveys have been collected by Statistics Netherlands (see their 1992 publication). The survey data are provided by a combination of household record keeping on standard expenditures and consumptions for a given period (either a month in 1980 or a half-month in 1988 which is grossed up into an annual record) and additional year long records on larger expenditures and holidays. Further data are collected by interview. The 1980 survey contains 2859 households and the 1988 survey 1950 households.

The sample drawn takes account not only of geographical distribution but also of socio-economic characteristics to obtain representativeness. Not all households were willing to participate in a survey and in 1980 the population of employees was intentionally oversampled. Therefore, a weighting variable is constructed (*WEIGHT*) which can be used to make the sample representative for the Dutch population.⁶

Apart from the total and specific consumption expenditure variables noted above, the two surveys include the following:

- WEIGHT*: Weight variable per household
- HSIZE*: Household size, the number of persons within the household (five classes, including single-person households)
- HTYPE*: Type of household (three classes: single person, one-family household, and other types)
- AGEHH*: Age of head of household (five classes)
- EDUHH*: Education level of head of household (five classes)
- LFPHH*: Labour force participation status of head of household (three classes: full-time job, part-time job, and otherwise)

⁵ Consumption is the use of goods and services for non-productive purposes of which the creation is considered as production in the National Accounts. The consumption of a good or service is not necessarily equal to expenditure on it. For instance, for owner-occupied homes the rental value is noted as consumption, not the actual mortgage payments and interest. Purchase and sale of durable goods are included. This may result in negative expenditures, especially in transport (car sales).

⁶ The weight variable is based on the following characteristics: socio-economic category (self-employed, employee, non-labour force participant) combined with household income, household size, home owner or renter, gender and single person or not.

- LFPHP*: Labour force participation status of partner (two classes: same as head of household, not equal to head of household⁷)
- SOCGHH*: Socio-economic group of head of household (three classes: employee, self-employed, and unemployed/non-participation)
- HINC*: Household income,⁸ net of taxes and social premiums (Dutch guilders) (two classes, 0–*f*80,000, *f*97,300 if net household income if larger than *f*80,000⁹ (1980 only) or *f*100,800 if net household income if larger than *f*80,000 (1988 only).

5. TIME SERIES FOR THE NETHERLANDS

We have annual observations from 1948 to 1988 (41 years) at the aggregate level on population, income, and on expenditures and price levels for the various consumption categories:

- POP*: Number of inhabitants (thousands of persons)
- NOH*: Number of households (thousands of households)
- AGGY*: Disposable income (millions of guilders).

Consumption expenditure variables (in millions of guilders) are provided for total household consumption (*V1*), which is divided into six groups (*V11*, *V22*, ..., *V66*). The first group is food, which is further divided into seven subgroups. (The other groups are not subdivided.) Price indices are given for each commodity group (1951 = 100) (*P1* to *P66*). The categories of expenditure (and corresponding prices) are those listed in Section 4 above.

5.1. Data Description

POP (the number of inhabitants) and *NOH* (the number of households) are obtained from a 1989 publication of Statistics Netherlands. The data on *NOH* are available only from 1960 onwards.

The data on *AGGY* (aggregate disposable income) are given in various publications of the National Accounts (*Nationale Rekeningen*). Disposable income consists of consumption and household savings. These savings include contractual savings, e.g. savings through pension funds or life insurance companies. Savings through pension funds are mandatory in the Netherlands for most employees. It is important to note that in the budget surveys these contractual savings are not part of 'savings'. (In the budget surveys savings are 'private' savings only.)

Consumption expenditures are also taken from the National Accounts. In 1977 a revision of the National Accounts took place. This was not so much a change of definitions, but rather an improvement on the precision. The revised data are available from 1969 onwards. Before 1969,

⁷ It is important to note that in case the head of the household is working full-time, it is not possible to know whether the partner is working part-time or is not employed. The head of the household (throughout the surveys) is defined to be the person (man or woman) who owns or rents the house where the household resides.

⁸ Net household income is defined as income from employment, profit from enterprise, income from capital (1980 excluding own house, 1988 including own house), social security benefits and other income such as rent subsidies or employers' contribution to the state medical insurance scheme, minus pension premiums, social security contributions and wage/income tax.

⁹ Net household incomes above *f*80,000 are not recorded in this data set. The average household income for the group of households with net household income above *f*80,000 is recorded instead. For this group of households *f*97,300 is the average net household income in 1980; in 1988 the average equals *f*100,800.

only consumption based on the old classification is available. For disposable income, the revised definition is available only from 1977 onwards.

The corresponding price indices are Laspeyres index numbers. The base years are 1951, 1959, 1964, 1969, 1975, 1980 and 1985.

6. DUTCH DATA SOURCES

Centraal Bureau voor de Statistiek, 1966, *Nationaal Budgetonderzoek 1963/'65: verbruiksrekeningen van Hand-, Land- en Hoofdarbeiders en Boeren, Onderscheiden Naar Inkomensgroepen, Grootte van de Huishouding en Woonplaatsen*, Staatsuitgeverij, 's-Gravenhage.

Centraal Bureau voor de Statistiek, 1989, *1899–1989, Negentig Jaren Statistiek in Tijdreeksen*, Voorburg/Heerlen.

Centraal Bureau voor de Statistiek, 1991, *Private Consumption Expenditure and Price Index Numbers for The Netherlands, 1921–1939 and 1948–1988*, Voorburg/Heerlen.

Centraal Bureau voor de Statistiek, 1992, *Gebruikershandboek Budgetonderzoek 1988–1989*, Voorburg/Heerlen.

Centraal Bureau voor de Statistiek, published annually, *Nationale Rekeningen*, Voorburg/Heerlen.

ACKNOWLEDGEMENTS

We are grateful to Adriaan Kalwij at Tilburg University and Adam Sutton and Paolo Zaffaroni at LSE for expert help with the collection of the Dutch and US data respectively. This description of the data is a heavily cut and then slightly revised version of Chapters 6–10 of the *Experiment Information Pack*; see Magnus and Morgan (1995).

REFERENCES

- Magnus, J. R. and M. S. Morgan (1995), *The Experiment in Applied Econometrics: Information Pack*, CentER for Economic Research, Tilburg University.
- Magnus, J. R. and M. S. Morgan (eds) (1998), *The Experiment in Applied Econometrics*, John Wiley: Chichester and New York, to appear.
- Tobin, J. (1950), 'A statistical demand function for food in the U.S.A.', *Journal of the Royal Statistical Society, Series A*, 113, Part II, 113–41.