**GUIDE TO THE CALCULATION**

**OF THE**

**AUSTRALIAN CONSUMER PRICE INDEXES**

**1850 - 1914**

Stephen Laird

January 1993

Department of Economics, University of Adelaide

[Slightly re-formatted by Peter Lindert, September 2017]

Acknowledgments are due to Justin Myatt who wrote the majority of the programs, Ng Soo Nam who assisted with data entry, and also Stephen Woodland, Melissa Gibbs and John Souter for general computing assistance. This paper is part of a wider project (directed by Ian McLean) which has received financial support from the Australian Research Council. The author of this document, which was first issued in July 1991, takes no responsibility or liability for any loss associated with the use of this program. Section 8 of this paper was written by Stephen Woodland in 1992.

CONTENTS

Section Page Number (slightly altered in this re-formatted version)

1 Introduction 3

2 Raw Data Entry 5

2.1 Introduction 5

2.2 Format of Data Entry 6

3 Intermediate Processing of the Raw Data 14

4 Primary Data Analysis 16

5 Index Generation 17

5.1 Expenditure Weights 17

5.2 Base search 21

5.3 Calculation of Sub-indexes 23

5.4 Calculation of Aggregate Indexes 26

6 Programming Variations 30

7 Files Contained on Disk 31

8 W - Series Index Documentation

(compiled by Stephen Woodland). 35 a

9 Appendix 36

Section 1: Introduction

The aim of this paper is to explain the methods and procedures utilized in the construction of a consumer price index for Australia and the individual colonies/states over the period 1850-1914, with specific reference to the format of the data entry and the running and manipulation of the computer software used to generate it. The information contained in the paper will allow users to extend or modify the price index series, such as extending the data coverage, altering expenditure weights and estimate other price indexes. It is advisable to completely read this manual and the material mentioned in it before any running of the index programs takes place. There are two working papers by McLean and Woodland

, which provide background to this project and information on the data sources used.

This paper follows the order in which the indexes are generated as indicated in Figure 1. below.

As mentioned in McLean and Woodland (1992) there are three alternative forms of indexes which are currently calculated, these being the X, Y and Z series. All the necessary programs and data required to generate the indexes mentioned in these papers are contained on the accompanying disk. To run these programs you need the SAS Statistical Package (version 5.18) and a FORTRAN Compiler (version 5.6)

. A good knowledge of SAS is required and a small amount of FORTRAN knowledge. These programs are quite basic and no difficulties should be experienced when the files need to be adapted to accommodate desired variations in output. A brief description of each type of index follows.

X SERIES: This series is characterised by its fixed regimen of commodities, at the national level, over the entire period of index calculation. There occurs no change in the commodity coverage. Five variations of this series exist and these are named X1 to X5.

Y SERIES: This series is characterised by its changes in regimen in the following years, 1888, 1900, and 1913. The regimen is constant however within the subperiods, 1850-1887, 1888-1899, 1900-1912, and 1913-1914. There exists two variations of this series which are named Y1 and Y2.

Z SERIES: This series is somewhat unorthodox in the sense that it can have a varying regimen of commodities from one year to the next. An index in this series will include any commodity for which there exists both an expenditure weight and at least one item of price information during the base period for the specific index calculated. There are currently six variations of this series, these are named Z1 to Z6.

The variations in each of these index series are determined by which expenditure weight data set is used in the index calculation. This is explained in more detail in Section 5.1.

There has been a use of directories and subdirectories in the generation procedure. These have been set up such that similar levels of development and types of indexes have been aggregated. This procedure should be continued as it allows control in the development of a particular index

. Figure 2 shows the directory system in use.

A brief outline of the procedure follows. Primary data are entered into individual yearly data files contained in the directory DATA0. These files are then processed to standardise the data to a format that allows checks on the data to ensure that they are suitable for index generation. This occurs in the directories DATA1 and DATA2. For the Z series only, the next step is to choose the base period around which the index is calculated, with the aim of maximising the proportion of total household expenditure which is represented by the index at any point in time. This is performed in the subdirectory BASESEARCH. This procedure is not employed when generating any X or Y series index. Following this, sub-indexes can then be generated for each of the four categories of consumer expenditure, each of which has its own subdirectory as shown in Figure 2. This utilises the expenditure weights as contained in the WEIGHTS subdirectory, and the population data. The output of these files is then edited to form data files which are used to generate the aggregate price indexes in the FINAL subdirectory.

Section 2: Raw Data Entry

This section is concerned with the structure of the raw data files and the method used to enter the data into them.

Section 2.1: Introduction

The historical price data are primarily from the colonial and state *Statistical Registers*. Early in the design of the project a set of commodities was defined whose prices were to be collected. Each of these commodities was assigned a numeric code for each of the following: commodity type, the unit of measurement as indicated in the initial data sources, the location of the source of the price, and the price type e.g. retail, wholesale, or contract. A detailed description of these numeric codes is contained in the tables at the end of this section. The raw data files are organised by year so there are 65 individual data files for the period 1850-1914. A particular price quote and all the information relating to it are contained on a single line within a particular data file. Each of these files are named \*\*\*\*RAW.DAT,

eg. 1865RAW.DAT

These files are contained in the directory DATA0. The file protection status of these files should be set at all times to avoid the possibility of inadvertent deletion or editing of them. Each line within the file must contain a non negative number for each of the following:

:- year.

:- number of price observations.

:- location code.

:- price type.

:- quantity code.

:- commodity code.

and :- price(s).

For the ease of entering additional raw data, a file named EXTRA.DAT is contained in this directory. It contains the basic format of the \*\*\*\*RAW.DAT files. The new data simply needs to be added to this file using a file editor and then when completed the new file needs to be appended to the relevant \*\*\*\*RAW.DAT file. This is particularly useful if large volumes of new information are obtained for a particular year as it allows the new data to be entered without having to directly access the existing data file.

Section 2.2: Format of Data Entry

The following contains the rules and format for entering the raw data into the files. Care is required to ensure that data are entered in the correct columns of the data file. Misplaced information will lead to incorrect output which might not in all cases be readily identifiable.

A brief overview of the contents of the raw data columns is provided in Table 1 below. This is followed by a detailed description of the numeric codes used, together with notes regarding variations or additions to them.

Table 1: Overview of Raw Data File Structure.

Column Number Data Type

1-4 Year in full.

6-7 Number of price quotations recorded.

9-10 Location of price quotation.

12 Price type.

14-15 Quantity unit of measurement.

17-19 Commodity code.

21-22 Price (Pounds only).

24-25 Price (Shillings only).

27-31 Price (Pence only).

Note: For a particular commodity in a particular location in a particular year there may be more than one price quotation, e.g. monthly or quarterly prices. In this case simply add those price quotations to the line of data above in the same format as given for columns 21 to 31, i.e. the next price quotation would be placed in columns 33 to 43 etc. Care needs to be taken in entering the data for the price. Note that the following conversion applies for pre-decimal prices:

1 Pound = 20 Shillings = 240 Pence

A detailed description of the numeric codes and columns follows.

Columns 1-4 Year in full eg. 1865, 1903.

Columns 6-7 Number of price quotations eg. 1 for annual prices and 4 for quarterly prices.

Columns 9-10 Location of the price quotation. Regions are coded as shown in Table 2 below.

Table 2: Codes for The Location of the price quotation.

Code Location

01 Australia.

11 New South Wales.

12 Victoria.

13 Queensland.

14 South Australia.

15 Western Australia.

16 Tasmania.

17 Northern Territory.

18 Australian Capital Territory.

21 Sydney.

22 Melbourne.

23 Brisbane.

24 Adelaide.

25 Perth.

26 Hobart.

27 Darwin.

28 Canberra.

Note: The existing data sets primarily contain price information relating to Sydney, Melbourne, Brisbane, Adelaide and Hobart.

Column 12 Type of price. The types of prices which have so far been encountered are:

1 = Retail.

2 = Wholesale.

3 = Contract.

4 = In bond.

5 = Duty paid.

Note: Retail prices and contract prices are effectively the same and are treated as such in any index calculation, so when running a retail price index, specify in the index programs mentioned in section 5.3 to include both of these price codes.

Columns 14-15 Code for the quantity unit with which the commodity is measured. Table 3 shows the codes which have been adopted. Also given is the conversion rates for various weights as well as examples of the commodity for which the code is used when the possibility of ambiguity arises.

Table 3: Unit Weight Codes

Code Unit/Description

1 Each (1 item)

2 Pound (lb)

3 Hundredweight = 1 cwt = 112 lbs

4 Pair (2)

5 Bushel = 60 lbs

6 Dozen (12)

7 2 lb loaf (bread)

8 Quart (1 quarter of a gallon)

9 Quarter or Hindquarter = 13.2 lbs (lamb)

10 Dozen Pounds (12 lbs)

11 Dozen Bunches (carrots)

12 100 Bunches (carrots)

13 Dozen Sticks (carrots)

14 Ton = 2000 lbs for wheat and 2240 lbs for all other commodities

15 Hogshead of beer = 52.5 gallons = 210 quarts = 420 pints

16 Gallon = 4 quarts = 8 pints

17 Head (lamb)

18 100 Pounds

19 Dozen quarts of beer

20 Case = 33 lbs (fruit)

21 4 lb loaf (bread)

22 Chest = 84 lbs (tea)

23 Stone = 14 lbs

24 Load (firewood)

25 Bar = 2 lbs (soap)

26 Quarter = 28 lbs (potatoes)

27 Tin (jam)

28 Bag (lettuce)

The unit measures of bushels (5) and cases (20) are volume measures. We have assumed that the unit weight of a bushel or case does not vary for different commodities, eg. a case of apples and pears would both be converted at 33 lbs per case.

In addition to the unit weight codes shown in Table 3 the following two unit-time codes are used for the housing data.

30 Week

31 Year

The raw data processing program which standardises the unit weights across time and location will need to be adjusted to incorporate any changes or additions to these unit codes such as a new unit of measurement or new conversion rates. For specific notes on these adaptations see Section 3 and the notes for the STANDARD.FOR program.

Columns 17-19 Commodity codes. These are split up into the following four categories of consumer expenditure, Food, Clothing, Other and Housing. Tables 4 to 7 show the codes for the commodities for which codes currently exist.

Table 4: Codes for Food Commodities

Code Item

1 Wheaten Flour

2 Wheat

3 Wheaten Bread/Bread

4 Bacon

5 Milk

6 Butter (fresh)

7 Colonial Cheese

8 Mutton

9 Coffee (roasted)

10 Tea

11 Sugar

12 Salt

15 Potatoes

16 Aerated Bread

17 Eggs

18 Ham

19 Honey

20 Apples (cooking)

21 Lemons

22 Oranges

23 Pears

24 Carrots

25 Cabbages

26 Lettuce

27 Onions

28 Rice

29 Beef

30 Veal

31 Lamb

32 Jam

33 Biscuits

Table 5: Codes for Clothing Commodities

Code Item

51 Moleskin Jackets

52 Moleskin Coats

53 Waistcoats

54 Moleskin Trousers

55 Flushing Trousers

56 Coloured Shirts

57 Strong Boots

58 Strong Shoes

59 Shepherds' Coats

60 Straw Hats

61 Print Dresses

62 Merino Dresses

63 Flannel Petticoats

64 Calico Petticoats

65 Shoes

66 Shawls

67 Check Aprons

68 Flannel

69 Calico

70 Socks

71 Handkerchiefs (cotton)

72 Hats (felt)

73 Shirts

74 Suits (tweed/cloth)

75 Stockings

76 Shifts/Chemise

77 Stays

Table 6: Codes for Other Commodities

Code Item

13 Beer

14 Tobacco

34 Brandy

35 Rum

36 Whiskey

86 Soap

87 Candles (tallow)

88 Kerosene

89 Paraffin

90 Firewood

91 Coal (various types)

92 Blankets

93 Mattresses

94 Sheeting

95 Rugs

Table 7: Codes for Housing

Code Item

101 McLean's Rental Prices

102 Butlin's Rental Prices

For some commodities such as tea and coal there exists varying qualities for the particular item at both the state and national level. We have tried to maintain consistency when qualities vary. To aid this, records have been kept noting the qualities used, and the method by which unit prices are calculated. This information is contained in the working paper by McLean and Woodland (1992). This paper should be consulted prior to the extension of any of the price series for any commodity to ensure consistency is maintained.

Any price information for which there are uncertainties regarding its exact nature or there are no current weight conversions available should be included in the file UNSURE.DAT which is contained in the same directory as the other raw data files. This file is not included in any further processing.

Section 3: Intermediate Processing of the Raw Data

The raw data must be processed in two stages before it is in a suitable form to generate a price index. Both stages use Fortran programs.

The first stage program is called PENCE.FOR (see Appendix pp.36-37) and it is located with the raw data files in the DATA0 directory. This program converts all price observations from pounds / shillings / pence to a price expressed only in pence. It also produces a simple average price for each line of raw data for which more than one price observation exists.

To run the program, type the following at the computer prompt:

RUN PENCE

It will then ask you for the first and last years that you want the conversions to take place, and you must enter these at the prompt. They are usually 1850 and 1914. On entering this information it will then ask you which price types you want to process. Type in one of the following:

9 Processes all price types.

1 Processes retail prices only.

2 Processes wholesale prices only.

3 Processes contract prices only.

The output consists of a data file for each year that the program was asked to consider Each file is named \*\*\*\*STAGE1.DAT, where \*\*\*\* is the file year (eg. 1865STAGE1.DAT). Another data file is created, DATA.DAT, which is a combination of all of the individual yearly stage 1 files appended together. All of these files (\*\*\*\*STAGE1.DAT and DATA.DAT) are placed automatically in the directory DATA1.

The second-stage processing program is called STANDARD.FOR (see Appendix pp.37-41) and is contained in the directory DATA1. It converts the various quantity measurements of commodities into a standard measurement for a particular commodity for all locations during the time period specified (to the lowest common denominator), while simultaneously adjusting the unit price accordingly. The conversions between quantity codes that standard currently performs are shown in Table 3, any other additional conversions that may be required in the future will therefore require adjustments to the program STANDARD.FOR. This is a relatively simple procedure and is shown in Section 6. To run the standard program, type the following after the prompt:

RUN STANDARD

You will be asked for the first and last years that you want the conversions to take place (usually 1850 and 1914). On entering this information it will then ask you in what locations you want the conversions to take place. If you type N it will process all locations and if you want to choose a specific location then type in the specific location code as indicated in Section 2.2 for columns 9 and 10.

The output from this procedure is placed automatically in the directory DATA2. There should be 65 processed data files named \*\*\*\*STAGE2.DAT, where \*\*\*\* is the file year, (eg 1865STAGE2.DAT) as well as a data file which is all of the yearly data files in this section appended together. This file, named DATA.DAT, will be placed in the DATA2 sub-directory and is the data file which is read into the index generation programs once the primary data analysis has been performed. Another output file that the program STANDARD.FOR generates is the file EXCLUDED.DAT. This file contains the lines of price information for which unit weight conversions are currently not available.

Section 4: Primary Data Analysis

Having generated the final data file DATA.DAT contained in the directory DATA2 we need to ensure that the data have been correctly entered before any index is generated. A number of testing procedures have been developed for this. In addition, there are programs which give a summary of the data collected. These procedures need not be run before any index is generated but it is wise to run them after large volumes of new data are entered, or to tailor them to perform checks on specific commodities. The programs available are contained in the directory DATA2. The programs available and their functions are:

COUNTER.SAS This program gives an overview of the price data collected. It shows the number of items for which there is a single line of price information for a particular price type in a particular location for a particular year. It also gives the total number of commodities for which there exists price information in a particular location in a given year.(See Appendix p.41)

RWCOMP.SAS This program plots retail prices against wholesale prices for a particular commodity in a particular location for the period over which those prices exist. It can also be adapted to include plots of other price types. We have found that visual checking of the relationship between retail and wholesale prices generated by this output is the simplest method available to ensure this relationship is correct. (See Appendix p.42)

LISTING.SAS This program lists all of the processed price data by commodity in each location. It is useful when used in conjunction with the output from the program RWCOMP.SAS. The program also prints the commodity unit weight as generated by the STANDARD.FOR program, so it must be used to check that these are consistent for a particular commodity in a particular location over the time period covered.(See Appendix pp.41-42)

Section 5: Index Generation

This section is concerned with the procedures used and the current variations available in generating both the sub-indexes and the aggregate indexes. It is divided into several sections and the order of index generation follows the order in which these sections appear below.

Section 5.1: Expenditure Weights

As shown in the flow chart, Figure 1, for a particular series, sub-indexes are initially calculated for each of the four categories of consumer expenditure, food, clothing, housing and other items. There are separate expenditure weights for each of these categories except for the housing sub-index which was calculated as a separate exercise ( see Section 5.3 ). The expenditure weights and the programs used to generate them are contained in the subdirectory DATA2.WEIGHTS.

For each category of expenditure there currently exists twelve different sets of expenditure weights, five for the X series, one for the Y series, and six for the Z series. This allows thirteen sub-index variations to be calculated for a particular category of expenditure in a particular location ( one of the variations Y2 utilizes the information generated by the series X1 to X4 ). Details of the construction and the sources of the various expenditure weights are given in the paper by McLean and Woodland

. Note that in each category the sum of the individual commodities expenditure weights sum to 100. Also, we do not currently have expenditure weights for all commodities for which price information exists. This effectively means that these commodities have a zero expenditure weight in any index calculated. What follows is a brief explanation by Index series of the different types of expenditure weights available, and how to generate some of them.

X SERIES: This series has five variations, accordingly it has five data sets of expenditure weights for a particular category of expenditure, thus giving a total of fifteen data sets. The data sets used to generate the sub-indexes X1 to X4 are:

W\*1861.DAT: based on the 1861 expenditure survey, for the index X1.

W\*1888.DAT: based on the 1888 expenditure survey, for the index X2.

W\*1900.DAT: based on the 1900 expenditure survey, for the index X3.

W\*1913.DAT: based on the 1913 expenditure survey, for the index X4.

These data sets exist for each category of expenditure so the \* in the above file names represents a C for clothing, a F for food, and a O for other items. For example WC1861.DAT is the expenditure weight data set used to calculate the sub-index X1 for clothing based on the 1861 consumer expenditure survey. This gives a total of 12 data sets of fixed regimen expenditure weights. These sets are simply data entered into files directly from the expenditure survey paper mentioned previously, and are contained on the disk.

The index X5, is based on a set of linearly interpolated fixed regimen weight sets. The expenditure weights used to generate this index vary over the time period of index calculation, so that the expenditure weight for a particular commodity is a function of the year. The expenditure weights for a particular commodity are non-zero in this weight set if that commodity is contained within the intersection of all of the previously mentioned fixed regimen weight sets for a particular category of expenditure. So if a commodity appears in the 1888, 1900, and 1913 expenditure surveys but not within the 1861 expenditure survey, then that commodity will be assigned a zero expenditure weight. Those commodities which appear in the intersection are then rescaled to sum to 100 within the particular category of expenditure. The expenditure weights for years in between the expenditure surveys are obtained by linear interpolation. Prior to 1861 the expenditure weights for these commodities are kept constant at their rescaled 1861 proportions, likewise the 1914 expenditure weights are the rescaled 1913 weights.

The X5 index's expenditure weights for a particular category of expenditure need to be generated using the program INTER.FOR (see Appendix pp.42-45). The data files that are read into this program, which are the weight sets used to generate the indexes X1 to X4, will need to be respecified for each category of expenditure, and the program re-executed as shown in Section 6 and the Appendix to this paper. The expenditure weight data sets which are generated, of which there are three, are named,

W\*INT.DAT

where \* represents the category of expenditure.

Y SERIES: There are currently two variations of this series available, Y1 and Y2. Of these, expenditure weight data sets only directly exist for the Y1 variant. The Y2 index is generated by splicing the X1 to X4 indexes for a particular category of expenditure, and thus do not require their own expenditure weights

. The expenditure weights for the Y1 indexes, of which there are three, are generated by the program INTAWGHT.FOR (see Appendix pp.45-47). This data set is similar to the X5 data sets except that non-zero expenditure weights are obtained for those commodities that exist within the intersection of adjoining expenditure surveys eg 1861-1888, 1888-1900 etc. This allows the regimen of commodities included in an index calculation based on these data sets to vary across time. Once again the user will have to specify which expenditure weight data sets are read into the program. These are the fixed regimen weight sets used to generate the indexes X1 to X4 above. The user will have to adjust these and re-execute the program to generate the expenditure weights for each category of consumer expenditure. The resulting output files are named,

W\*INTA.DAT

where \* represents a C for clothing etc.

Z SERIES: There are currently six variations available for this series, they are named Z1 to Z6. This gives a total of eighteen data files, six for each category of consumer expenditure. For the index variations Z1 to Z4, these files are.

WZ\*1861.DAT: based on the 1861 expenditure survey, for the index Z1.

WZ\*1888.DAT: based on the 1888 expenditure survey, for the index Z2.

WZ\*1900.DAT: based on the 1900 expenditure survey, for the index Z3.

WZ\*1913.DAT: based on the 1913 expenditure survey, for the index Z4.

These data sets exist for each category of expenditure so the \* in the above files name denotes this. For example, the data file WZC1861.DAT is the expenditure weight data set used to generate the Z1 sub-index for clothing based on the 1861 expenditure survey and a varying regimen of commodities. This gives a total of twelve data sets of varying regimen expenditure weights. These sets are simply data entered into files directly from the expenditure survey paper mentioned previously, and are contained on the accompanying disk.

The expenditure weight data sets for the variants Z5 and Z6, are generated in the same way as for the variants X5 and Y1 respectively except that they are based on a varying regimen of commodities. The interpolated expenditure weights for the Z5 indexes are generated using the program INTER.FOR, and for the Z6 indexes by the program INTAWGHT.FOR. Once again the user must specify the expenditure weight data sets which are read into these programs. These data sets are those which are used to generate the Z1 to Z4 indexes above. The programs will therefore require the user to specify the data sets for each category of expenditure and re-execute the program. The output data set generated for the Z5 indexes is called WZ\*INT.DAT, and for the Z6 indexes is called WZ\*INTA.DAT. In both cases the \* represents the particular category of consumer expenditure. This gives a total of three expenditure weight data sets each for the Z5 and Z6 indexes.

Section 5.2: Basesearch

The aim of this process is to choose the five year baseperiod for the index calculations so as to maximise the proportion of consumer expenditure represented by the expenditure category indexes in each state and hence also for Australia. The basesearch procedure only applies to the Z series indexes (Z1 to Z5) as all other indexes are run with the commodity coverage being held constant (therefore, regardless of the baseperiod used, commodity coverage will always be maximised). This procedure utilizes the expenditure weights generated in the previous section for these indexes. All relevant programs and output are contained in the subdirectory DATA2. BASESEARCH.

This criteria cannot be utilized when generating the Z6 index. This is because the commodity coverage in the 1861-1888 period will be different to that in the 1888-1900 period. The later period is likely to include a greater number of commodities. Thus if we have prices for 88% of total consumer expenditure in both of the subperiods then the later is going to have a greater commodity coverage which is likely to include all of those commodities in the earlier period. This makes the maximisation criteria invalid so that when using these weight sets to generate an index the base period is chosen towards the last years of the data set when the commodity coverage is usually at its largest.

There are two SAS programs which are used to choose the initial base period for the indexes, these are:

BASE.SAS: This is used for the indexes which are based on one expenditure survey ie. Z1 to Z4. (See Appendix pp.47-48)

BASEINT.SAS: This is used to determine the Z5 indexes base periods. (See Appendix p.48)

The programs need to be adapted for each of the index variations being generated by ensuring that the correct expenditure weight data set is read into the program and that the output is correctly labelled. The output shows the proportion of total consumer expenditure which is represented in the particular category of expenditure within a particular state and in a particular year, as well as the number of commodities which this proportion represents. From the output generated choose the five-year base period for the specific sub-index on the basis of the above mentioned criteria.

Section 5.3: Calculation of Sub-indexes

There are twelve variations of the sub-indexes to be calculated for each of the expenditure categories defined. For each of these four expenditure categories there exists a subdirectory in which all of the possible index variations for that particular category are generated. The names of these subdirectories follows the name of the expenditure category, eg DATA2.FOOD.

Of the four categories of expenditure only those of food, clothing, and other are generated by the programs contained within the particular subdirectory. The housing sub-index was calculated as a separate exercise and then entered into the housing subdirectory as the data file RENT.DAT. However if more information relating to housing prices is obtained, the existing programs used for the remaining categories of consumer expenditure can be adapted to run a housing index, as explained below for the other three categories.

The method used to calculate the state and national level sub-indexes for food, clothing, and other items follows the same procedure. Here only the method of calculating the food sub-index will be illustrated for a particular series.

There are three SAS programs which are used to calculate a particular sub-index. These programs are used for each series. Two of the programs are subroutines common to every variation of sub-index generated, they are:

BASEPR.SAS: This subroutine calculates the average base-period prices for a commodity in a particular location over a five year period. (See Appendix p.49)

ANSWER5.SAS: This program is the main subroutine as it is this that calculates the actual index values.

(See Appendix pp.48-49)

The main program in which both of these subroutines are read into is denoted as:

SI#\*++++.SAS

Where: - # denotes the particular index series, X, Y, and Z.

\* denotes the category of expenditure eg: food (F), clothing (C), and other items (I).

++++ denotes the expenditure weight set used in the calculation of the sub-index, (1861,1888,1900,1913, INT, INTA ).

This program is basically the same for all variations except for a subtle difference when the varying expenditure weights are used (i.e. INT and INTA, see below). So for example the program SIXF1888.SAS (see Appendix, p.49) generates food sub-indexes at the state and national level which are based on the X series expenditure weights for the fixed 1888 consumer expenditure survey. What follows is an explanation of the program SIXF1888.SAS. Also indicated, is where changes are required within the program in order to run all other sub-index variations. In total there are thirty-six sub-index variations available as there are currently twelve variations of expenditure surveys for each of the three categories of consumer expenditure for which sub-indexes can be calculated.

Initially the program reads in the data set DATA.DAT. In this part of the program you need to specify the price types for which you want the index to be generated. Both of the numbers 1 and 3 will produce a retail price index, and this holds for all the index series. The data are then sorted by commodity and location. The next stage is to read in the consumer expenditure weights from the WEIGHTS subdirectory, which in this case is the file WF1888.DAT. This needs to be re-specified for each index variation. The program then sorts and merges all data and then deletes all commodities for which price information exists but which has a zero expenditure weight.

Note: Do not use the same program for the fixed regimen expenditure sets (ie. 1861, 1888, 1900 and 1913 variations) and the varying regimen expenditure weights (ie. INT and INTA variations) as the data in the latter requires extra sorting of the entered data by year. A copy of each type of program is contained on the disk.

The next procedure specifically relates to the Z series indexes Z1 to Z5. It calculates the base-period prices for each location. The programmer must enter the base year high and low as determined by the BASESEARCH procedure for the particular Z series index into the program. Note that this section of the program also exists for all other sub-index variations in the X and Y series, and so a five-year period must also be entered when generating these sub-indexes. In these cases it is best to enter the five year period for which the final output indexes are centred around ie.1878-1882. Following this the subroutine ANSWER5.SAS is read into the program. This generates an index for each state and rescales the index from the base period entered to a common five year base centred around 1880. This enables the state indexes to be merged to form the national index at a later stage in the program. A price observation contributes to the calculation of the price index if the commodity has been allocated a non-zero expenditure weight and that at least one price observation exists for that commodity in the base period specified. This is always the case for the X and Y series indexes. For the Z series, in each year the sum of the expenditure weights which satisfy these conditions is calculated and in this process they are rescaled to sum to 100. The next step is to read in the state population data from the data file POP.DAT contained in the directory DATA2. The program then calculates the total Australian population as well as the proportion of the national population residing in each state. On the basis of this information the national sub-index for the category of expenditure is calculated by weighting the individual state sub-indexes by the proportion of total population in each state. If for a state there is no index value in a particular year then the proportions of the remaining states' populations are rescaled to sum to 100. The final stage of the program contains the instructions to print the indexes generated as well as the number of commodities which are included in that index value calculation, and the proportion that each state index contributes to the national index. In this section, the commands to put titles on the outputs will need to be adjusted for each variation of output (as shown on p.52 of the Appendix).

Therefore, so as to generate the variations of the price index programs contained on the accompanying disks (SIXF1861.SAS and SIXFINTA.SAS) the following changes need to be performed.

1. Check that the price type is correct.

2. Enter the appropriate expenditure weight data set name.

3. Specify the base years of the index calculation (five year average, particularly important for the Z Series variations where these years must be at the point of maximum commodity coverage).

4. Edit the output titles so that it is clear what index is being generated.

5. Rename the file according to the changes made ( ie. when creating an index using the 1900 food expenditure weights, the SAS file will be renamed SIXF1888.SAS).

In order to use the sub-index output to generate the aggregate expenditure indexes the former must be edited so that it is in the form of a data file. That is after editing, the file should only contain the years, the index values and the number of commodities that are used to generate the index value. The file created from this editing should be renamed as a data file (keeping the suffix of the title) and left in the subdirectory in which it was created.

Section 5.4: Calculation of the Aggregate Indexes

This is the final stage of the index generation process. All work involved in this section is done in the subdirectory DATA2.FINAL. As this is the final stage there are only twelve variations of the aggregate index to be generated. There are three SAS programs which are used. The first of these is a subroutine which is common to all variations of the indexes. The other two are the main programs, and they are the ones which will need adapting to take into account all of the variations possible. One variation is used to calculate the aggregate indexes based on a single expenditure survey and the other for the variations based on interpolated expenditure weights. These programs are:

COMBINE.SAS: This subroutine generates the aggregate index values for each state, and one of the national indexes (pp. 66-67 Appendix).

FI#\*\*\*\*.SAS: This is a program which organises the data sets and generates the output as well as generating the second form of the national index. In the above filename # represents the index series and \*\*\*\* represents the particular expenditure survey on which the index is based, e.g. 1861, 1888, 1900, 1913, INT and INTA. Note that the programs for the interpolated expenditure weight indexes are slightly different to those based on a single expenditure survey, a copy of each is contained on the disk (pp.57-66 Appendix).

Once again there are only slight variations in the form of the latter program, and these are explained below. The program FIX1888.SAS will generate state indexes and two types of national indexes which are based on all four categories of consumer expenditure using the fixed regimen 1888 expenditure survey. Initially the program reads in the share of total expenditure represented by each of the four categories of consumer expenditure in a specific survey. For the fixed regimen weight sets this is entered into the program before it is run. These weights are contained in the paper by McLean and Woodland (1992).

For the interpolated weight sets the procedure is a little more complicated due to the expenditure shares varying over time. A file showing these shares called INTWGHT.DAT has been developed as a separate exercise and is contained in the FINAL subdirectory. A copy is contained on the disk. As mentioned above the programs for the interpolated expenditure weight indexes is slightly different as this additional data set must be read into the program. The next stage is to read the edited output files generated in the previous section for a particular expenditure survey and index series. Each column of data within these four files has been given an alpha-numeric code for the index and the number of commodities which are used to calculate it. These must be entered into the program for each final index run. These codes are explained below.

If over the entire period for which the index is calculated no index values exist in a particular location then this needs to be specified where the data file is read in by the code MISS for the index value and NMISS for the number of commodities used to calculate a particular index value. For the three expenditure categories for which sub-indexes have been generated a four letter code is used and this is structured as follows.

\*??+ where: \* denotes an I for index and an N for the number of commodities which were used to calculate it.

?? denotes the category of expenditure, ie. FD for food, CL for clothing, and OT for other.

+ denotes the location of the index or the number of commodities used in its calculation. The following numbers are currently used for particular locations.

1 for Sydney

2 for Melbourne

3 for Brisbane

4 for Adelaide

5 for Perth

6 for Hobart.

So for example the code ICL5 represents the clothing index values for Perth. For the housing sub-index the following codes are used.

IRTOB: Rental index for Australia derived by Butlin.

NRTOB: Number of items of price information contained in the calculation of the above index.

IRTOO: Rental index for Australia by McLean.

NRTOO: Number of items of price information contained in the calculation of the McLean index.

Specific rental indexes for particular locations are coded in the same way as for the other categories of expenditure sub-indexes given above with RT being the expenditure category notation. The Butlin series is only a national index. In this section of the program you will have to drop one of the two national rentals indexes from the data which are used to generate the final indexes, and accordingly make a note of this in the output titles. Note that the same sub-indexes generated for the other categories of expenditure are used if either housing series is used.

Next the program reads in the subroutine COMBINE.SAS, this generates the aggregate indexes at the state level by weighting the individual state category indexes by their shares of aggregate expenditure. It also generates one of the national indexes AUSIND, which is the weighted average by the four expenditure category shares of the various national expenditure category indexes. The program will rescale the expenditure shares if there are missing values for a category of expenditure. The program then calculates the second type of national index AUSIND2, which is the weighted average by state population shares of the aggregate state level indexes generated in the subroutine above. The final stage prepares the output which includes the aggregate expenditure indexes for each state and as the two national indexes, the number of commodities which are utilized to calculate that particular index value, as well as the shares of each sub-index's contribution to a particular location's aggregate index. The output titles which are generated at this stage of the program will need to be adjusted for each index variation and the resulting output file renamed.

Section 6: Programming Variations

The aim of this section is to provide details of the procedure to adapt the program STANDARD.FOR, and also how to execute a FORTRAN program.

The program STANDARD.FOR will need to be adjusted if any of the following arise:

:- there is a new unit weight conversion.

:- a new unit weight is encountered.

To do this you need to add to the program extra commands in the section which processes these unit weights. When entering new conversion weights check to see if the conversion needs to be commodity specific, for example in the case of wheaten flour a ton is equal to 2200 lbs but for all other commodities it is 2240 lbs, as shown in the example below. If this is the case then additional information is required in this section to also specify the commodity number.

e.g. ELSE IF (weight(linenum) .EQ. 14) THEN

IF (commodity(linenum) .EQ. 1) THEN

pence(linenum) = pence(linenum)/2000

weight(linenum) = 2

ELSE

pence(linenum) = pence(linenum)/2240

weight(linenum) = 2.

END IF.

After making changes to any of the FORTRAN programs listed in Section 7 below the programs will need to be re-executed to ensure that the adaptations incorporated are correct and functional. This procedure is as follows:

FOR 'filename'

LINK 'filename'

RUN 'filename'.

The output generated includes the file with the suffix EXE, if this file has been generated then the program has functioned properly. This file must be kept in the directory in which it is created if further versions are to be generated.

Section 7: Files Contained on the Disk

The following files have been put on the accompanying disk. There are 110 individual files. Note that not every data file and program mentioned in the paper have been included on the disk. Two programs are included to generate the interpolated expenditure weight data sets and as mentioned in the text, these will require adjustment and renaming to generate all of the variations. For the sub and aggregate indexes two programs for each of these levels of index generation are included on the disk. In each case there is one for the indexes based on a single expenditure survey and another for those based on the interpolated expenditure weights. These four programs will also require adjustment and renaming to produce the other sub and aggregate indexes.

A brief outline of each file's function, its directory placement, and any other relevant notes is given in the listing below.

Directory Filename Function/Notes

Data0

(1850-1914) Raw.dat 65 raw data files.

Pence.For Price standardising program.

Extra.Dat Data file format/appending file.

Unsure.Dat Ambiguous data.

Data1 Standard.For Unit of measurement standardising program.

Data2 Counter.Sas Data checking program.

Rwcomp.Sas Data checking program.

Listing.Sas Data summary program.

Pop.Dat State population data.

Data2.Weights

Wc(1861-1913).Dat 4 clothing expenditure weight data sets used to calculate the X1, X2, X3,X4, X5, Y1 and Y2 sub-indexes.

Directory Filename Function/Notes

Data2.Weights Wf(1861-1913).Dat 4 food expenditure weight data sets used to calculate the X1, X2, X3, X4, X5, Y1, and Y2 sub-indexes.

Wf(1861-1913).Dat 4 other expenditure weight data sets used to calculate the X1, X2, X3, X4, X5, Y1, and Y2 sub-indexes.

Wzc(1861-1913).Dat 4 clothing expenditure weight data sets used to calculate the Z1, Z2, Z3, Z4, Z5, and Z6 sub-indexes.

Wzf(1861-1913).Dat 4 food expenditure weight datasets used to calculate the Z1, Z2, Z3, Z4, Z5, and Z6 sub-indexes.

Wzo(1861-1913).Dat 4 other expenditure weight data sets used to calculate the Z1, Z2, Z3 Z4, Z5, and Z6 sub-indexes.

Inter.For Program used to generate the expenditure weights for the X5 and Z5 sub-indexes.

Intawght.For Program used to generate the expenditure weights for the Y1 and Z6 sub-indexes.

Intwght.Dat Data set of the percentage shares of total expenditure for each of the four categories of expenditure by year.

Data2.Basesearch Base.Sas Program used to determine the initial baseperiod for the X1, X2, X3, and X4 sub-indexes.

Baseint.Sas Program used to determine the initial baseperiod for the X5 sub-indexes.

Data2.Food

Basepr.Sas Subroutine that calculates average prices for the index baseperiod.

Answer5.Sas Subroutine that calculates the index values.

Directory Filename Function/Notes

Data2.Food Sixf1861.Sas Program used to generate the food sub-index for the X1 series. It can be adapted to also generate the X2, X3, X4, Y2, Z1, Z2, Z3, and Z4 sub-indexes for this category of expenditure, as well as for the clothing and other categories.

Sixfinta.Sas Program used to generate the food sub-index for the X5 series. It can be adapted to also generate the Y1, Z5, and Z6 sub-indexes for this category of expenditure, as well as for the clothing and other categories.

Data2.Final Rent.Dat Previously generated indexes for housing.

Combine.Sas Subroutine that combines sub-indexes.

Fix1861.Sas Program used to generate the aggregate X1 index. It can also be adapted to generate the X2, X3, X4, Y2, Z1, Z2 , Z3 ,and Z4 indexes.

Fixinta.Sas Program used to generate the aggregate X6 index. It can also be adapted to generate the Y1, Z5, and Z6 indexes.

References

Ian W. McLean and Stephen J. Woodland (1991), "Trends in the Compostion of Consumer Expenditure: Austrailia 1854-1913", Working Paper No. 91-8, Department of Economics, University of Adelaide

Ian W. McLean and Stephen J. Woodland (1992), "Consumer Prices in Australia 1850-1914", Department of Economics, University of Adelaide.

Footnotes

Ian W. McLean and Stephen J. Woodland, "Trends in the Compostion of Consumer Expenditure: Austrailia 1854-1913", Working Paper No. 91-8, Department of Economics, University of Adelaide; and "Consumer Prices in Australia 1850-1914", Department of Economics, University of Adelaide,1992.

The version numbers are those which were used. It may be possible to use other versions to run these programs.

The existing programs have specific reference to these directories when they read data files and program subroutines. If you desire to change the subdirectory setup then these aspects of the program will also need changing. The file input specifications will probably need tailoring for your system and directory setup also.

The code number 29 currently has no unit weight attached to it.

The code numbers 13 and 14 are contained in the category of Other items in table 6.

For information on how these prices were obtained see McLean and Woodland "Consumer Prices in Australia 1850-1914".

McLean and Woodland "Trends in the Composition of Consumer Expenditure: Australia 1854-1913".

See McLean and Woodland (1992) for details of construction.

The programs are currently structured such that a copy of each of these files must be kept in each expenditure category subdirectory.

The columns in the data files are as explained in Table 1, Pg 4.

The columns of this file are set out in the following order, Year, New South Wales, Victoria, South Australia, Tasmania, Queensland, and Western Australia.

The columns in the data files are set out in the following order, Commodity No, Expenditure Weight.

The columns in this file are set out in the following order, Year, Foods Share, Clothings Share, Others Share, and Housings Share.

These four programs are also common to the "clothing" and "other" subdirectories. Remember to change the filename and the expenditure weight data set that is read into the program when transferring the index specific programs to these subdirectories.

The columns in this file are set out in the following order, Year, IRTOB, NRTOB, IRTOO, NRTOO, IRT1, NRT1, IRT6, NRT6, as given on Pg 20.

[Cal]culate that particular index value, as well as the shares of each sub-index's contribution to a particular location's aggregate index. The output titles which are generated\_

znc†zWWz

~ŸqqqqfZNB

tttttth]NC

zl\_\_\_\_\_THT:

oo`QB<

›ØØØØØØØ

›ØØØØØØØ

ghijklmnopqrstuv\_Ȓ

Times New Roman

Symbol

Courier

prestige

lettergothic

gothicPS

cubicPS

lineprinter

Helvetica

avantegarde

spartan

presentation

emperorPS

madaleine

zapf humanist

classic

roman f

roman g

roman h

timesroman

century

Palatino

souvenir

garamond

caledonia

bodini

university

BÿScript

scriptPS

script c

script d

commercial script

park avenue

coronet

script h

hebrew

roman s

russian

roman u

roman v

roman w

narrator

emphasis

zapf chancery

decor d

old english

decor f

decor g

cooper black

linedraw

bar3of9

pcline

tech h

Apple LaserWriter II NTX

pscript

Apple LaserWriter II NTX

:,COM2:,COM3:,FILE:,LPT1

INDGEN.DOC

stephen laird Simon Moll