Australian Labour Force Data: How Representative is the 'Population Represented by the Matched Sample'?

Robert Dixon*

Abstract

This paper investigates two related matters. First, what proportion of the population is represent by the matched sample (i.e. by the gross flows data) in the Labour Force Survey, why is this proportion what it is and why does it vary over time? Second, given that slightly over 20% of the population are not represented in the matched sample, how representative are labour market indices derived from the matched sample data and, if biases are present, what is the source and what are the implications of the bias?

1. Introduction

Data on gross flows between various labour market states has been published on a monthly basis by the ABS since February 1980.[1] From time to time academic and non-academic researchers (Foster 1981, Foster and Gregory 1984, Fahrer and Heath 1992, Borland 1996b, Leeves 1997, Borland and Kennedy 1998, Debelle and Swann 1998, Leeves 1999, Leeves 2000) visit this data with a view to gaining extra insight into issues related to the determinants of movements in the level of unemployment over time and/or the equilibrium or natural rate of unemployment. Little attention however has been given to the implications of the survey methodology and related ABS statistical procedures for the representativeness of the data derived from the matched sample.

This paper aims to address two related sets of questions. The first set concerns the proportion of the population represented by the matched sample (and thus the gross flows data). What proportion of the total population is in fact represented by the matched sample, why is this proportion what it is and how does it vary over time? Second, given that slightly over 20% of the population are not represented in the matched sample, it is sensible to ask how representative are labour market indices derived from the matched sample data and, if bias is present, what can we say about the source of the bias? The structure of this paper is as follows. In section 2 we explain in some detail the way in which the Labour Force Survey is undertaken and the method by which gross flows data is derived from successive surveys. Section 3 examines the behaviour of the proportion of the population represented by the matched sample over time. Section 4 compares the time series properties of the

* Department of Economics, University of Melbourne, Victoria, 3010. I am grateful to Jim Thomson for helpful comments on an earlier draft. The first part of the paper relies heavily upon the description of the LFS and of the Gross Flows data given in various ABS publications.

[1] The Gross Flows data reported in 6202.0 and 6203.0 February 1980 between them covered the period from August 1979 through February 1980. Since February 1980 Gross Flows data has been published in 6203.0 only. Borland (1996a) provides a neat discussion of the sources and uses of the flows data.

unemployment rate[2] for those persons represented in the matched sample as against the rate for the population as a whole. Section 5 looks at the behaviour over time of the matched sample's unemployment rate and the unemployment rate for the groups not represented in the matched sample. Section 6 presents a framework which enables us to decompose the bias in the matched sample into its constituent parts and to evaluate their numerical importance. The final section considers the representativeness of the matched sample in capturing flows and transition rates and concludes.

2. The LFS and the Matched Sample

The Labour Force Survey (LFS) has been undertaken on a monthly basis[3] since February 1978. It is a component of the Monthly Population Survey, which is based on a multi-stage area sample of private dwellings (currently about 30,000 houses, flats, etc.) and a (much) smaller number of non-private dwellings (hotels, motels, etc.). Since 1992 it has covered approximately one-half of one percent of the population of Australia.[4]

LFS information is obtained from the occupants of selected dwellings by specially trained interviewers. Interviews are generally conducted during the two weeks beginning on the Monday between the 6th and the 12th of each month. The information obtained relates to the week before the interview—the 'reference' week. Households selected for the LFS are interviewed each month for eight months, with one-eighth of the sample being replaced each month. Prior to August 1996, all interviews were conducted face-to-face at the homes of respondents. Over the period August 1996 to February 1997, the ABS introduced the use of telephone interviewing to collect LFS data.[5] In this system the first interview is conducted face-to-face and subsequent interviews are conducted by telephone (if this is acceptable to the respondent). In both face-to-face and telephone interviews, interviewers attempt to collect all information about each household member from one adult. In the interviews an attempt is made (inter alia) to establish whether each person is in or out

[2] We could look at other series (e.g. the participation rate) but given the space constraints and the uses to which the gross flows data is usually put, we will focus here on its ability to mimic the unemployment rate for the whole population and for the groups not included in the matched sample. Also, to keep the paper of manageable length, all data in this paper refers to persons - there is no dis-aggregation into males and females.

[3] The technical notes to ABS 6204.0 for 1978 contain a good description of the sample design and the procedure used to derive estimates from the sample data. ABS publication 6232.0 provides a copy of the questionnaire.

[4] Between population Censuses, the size of the sample grows in line with estimated population. Following each Census—that is, every five years—the sample is re-weighted and its size adjusted. Until late 1987 the survey covered approximately two-thirds of one percent of the population, between late 1987 and late 1992 it covered approximately three-fifths of one percent of the population and, as mentioned in the text, since 1992 it covered approximately one-half of one percent of the population.

[5] The ABS reports that "[d]uring the period of implementation [of telephone interviewing], the new method produced different estimates than would have been obtained under the old methodology. The effect peaked in November 1996, when it is estimated that the published estimate of total employment was about 36,000 to 48,000 lower than would have been measured under the old methodology. The effect dissipated over the final months of implementation and was no longer discernible from February 1997. Therefore, the estimates for February 1997 and onwards are directly comparable to estimates for periods prior to August 1996" (ABS, 6203.0 October 2000, p 36). See also the feature article 'The effect of telephone interviewing on Labour Force estimates' in the June 1997 issue of 6203.0.

of the labour force, if in whether employed or unemployed, and if employed whether the employment is full-time or part-time.

To derive labour force estimates for the entire population in the scope of the survey, expansion factors (weights) are applied to the sample responses. Weighting ensures that LFS estimates conform to the benchmark distribution of the population by age, gender and geographic area. A weight is allocated to each sample respondent according to his/her State/Territory of usual residence, region (capital city or other), age and gender. In essence, the weights are the inverse of the probabilities of selection, adjusted for any under-enumeration and non-response.

Data on gross flows between months are based only on the matched component of the sample—that is, persons surveyed in a given month whose responses in that month can be matched with responses in the previous month. The matched sample differs from the total sample for three reasons: the exclusion of respondents in non-private dwellings, sample rotation and 'non-response' by persons in the survey in the previous month.[6] We will deal with each in turn.

For the LFS, private dwellings (such as houses and flats) and non-private dwellings (such as hotels and motels, boarding houses and short-term caravan parks,[7] hospitals and nursing homes, educational colleges and aboriginal settlements) are separately identified and sampled. The sample of non-private dwellings is obtained by first compiling a list of all non-private dwellings in Australia. A sample is taken from this list in such a way that each region across Australia and each different type of dwelling is represented. For smaller non-private dwellings, each occupant is included in the survey; for larger dwellings, a sub-sample of occupants is taken. The "procedures used to select persons in non-private dwellings preclude the possibility of matching any of them who may be included in successive surveys" (ABS 6203.0, October 2000, p 42) and so matched sample data can only refer to persons in private dwellings.

The sample of private dwellings is obtained using a multi-stage approach. Australia is divided into about 70 geographical regions, which are then stratified according to population density and growth. A sample of census collection districts is then randomly selected to represent each region. Each collection district is divided into smaller areas called blocks. One block is selected randomly from each district to represent the others. In urban areas, a sub-sample of dwellings in the selected block is taken. Blocks in rural areas contain fewer dwellings and, as a result, all dwellings in a selected block are included in the survey.

As it is not reasonable to retain the same respondents in the survey for a long period of time, a proportion of the dwellings in the sample are replaced each month. This procedure is known as sample rotation. Since the monthly LFS commenced in 1978, dwellings have been retained in the survey for eight consecutive months so that about

[6] Obviously anyone for whom there is no response in the current month is not included in the current month's sample, matched or unmatched.

[7] In 1981 (all) caravan parks were moved from the private dwelling sample to the non-private dwelling sample. In 1991 long-stay caravan parks were moved from the non-private dwelling component to the private dwelling component of the sample.

one-eighth of the sample has been replaced each month.[8] Thus the LFS sample can be thought of as consisting of eight sub-samples (or rotation groups), with a new rotation group being introduced into the sample each month to replace an outgoing rotation group. Dwellings in the replacement sample generally come from the same geographic area as those in the outgoing sample.[9] Given this, it is unlikely that the mere fact that some members of the sample are replaced each month will itself introduce any systematic bias.

Thus the rotation procedure used by the ABS is such that only seven-eighths of the private dwelling sample from one month is retained for the next month's survey. Persons normally residing in these dwellings who respond in both months form a 'matched sample' for the later month. Some respondents in the current month will not have responded in the previous month, however. These 'non-respondents' include persons who have moved to the dwelling in the sample between months, holiday-makers and others who were temporarily absent and for whom information could not be obtained in the previous month, other persons who could not be contacted in the previous month, and persons who refused to answer in the previous month.[10] Each of these unmatched groups is likely to have labour force characteristics that differ from those of the general population.

For future reference we note that, whilst the gross flows estimates are adjusted to account for non-response in the second month, the published estimates for gross flows "have not been adjusted to account for the unmatched sample component" (ABS 6203.0, October 2000, p 43). This means that the sample gross flows data for any month are expanded up to a 'population figure' which is equivalent only to the proportion of the persons responding in the second month who: (i) are living in private dwellings (i.e those living in non-private dwellings are excluded) at the time of the survey; and (ii) were in the survey in the previous month (i.e those rotated in or out are not included); and (iii) responded with useable information in the previous month (so non-respondents are excluded).

In the next section we look at the size of the non-matched sample and consider why it varies over time.

3. The Proportion of the Population Represented by the Matched Sample

We have seen that the procedures used to select persons in non-private dwellings preclude the possibility of matching any of them who may be included in successive surveys. For various reasons, including the mobility of the population and non-response in the earlier month, a proportion of the persons in those private dwellings that are included in the sample in successive months cannot be matched. The component of the sample that is common from one month to the next is referred to as

[8] This 'rotation rate' has been constant throughout the whole of the period except for certain months when redesigned samples were being phased in (October 1982, September–December 1987, September–December 1992 and September 1997–April 1998). There is some variation in the size of rotation groups even when rotation rates are constant, however, due to population growth and random variations in household size and response rates.

[9] The word 'area' here has to be taken literally; indeed it can refer to an area as small as a city block.

[10] There will also be a small number of people who were living at the dwelling in the previous month and who turned 15 between months.

the 'matched sample'. "Normally, those who can be matched represent about 80% of all persons in the survey" (ABS 6203.0, October 2000, p 42).

Let *PRMS* to denote the size of the "Population Represented by the Matched Sample"[11] and *POP* to denote the civilian population aged 15 years and over. We begin by asking: in practice, what proportion of the population is represented by the matched sample? We can establish this proportion (*PRMS/POP*) by comparing the size of the population represented by the matched sample with the total civilian population over the age of 15 for the second month of any pair.[12] Figure 1 shows[13] the proportion of the population represented by the matched sample (i.e. *PRMS*/*POP*) for each month over the period 1979:08–2000:10 as reported by the ABS at the time.[14] Figure 2 shows original and seasonally adjusted[15] series for the proportion of persons in the population who are <u>not</u> "represented by the Matched Sample" (i.e. *1 - (PRMS/POP)*) over the period 1979:08 - 2000:10. The mean of this series is 21.9%. On average then, 22% of persons in the population were not 'in' the matched sample.

Both the original and the seasonally adjusted series for *(PRMS/POP)* and *[1 - (PRMS/POP)]* are very noisy. Figure 3 shows smoothed values for the *(PRMS/POP)* series (solid line) superimposed on the seasonally adjusted series (dotted line). The smoothed series was obtained by fitting a Hodrick-Prescott filter trend (with a smoothing parameter of 14400) to the seasonally adjusted *PRMS* series and the *POP* series separately and then forming a ratio of the two trends.[16] It would appear that the proportion of the population represented by the matched sample has risen over the period, especially since the introduction of telephone interviewing in late 1996.

[Insert Figures 1, 2 and 3 near here]

[11] This is the term used by the ABS for the population equivalent of the number of those for whom information could be obtained in successive labour force surveys. This figure is provided in each issue of 6203.0. Note that the data given there are in population equivalents.

[12] The population in the *second* month of any pair is used because, "the expansion factors used in calculating the [gross flows] estimates [are] those applying to the second of each pair of months" ABS 6203.0, October 2000, p 43.

[13] The breaks in the data correspond to the periods when the size of the matched sample was abnormally low due to a new sample being rotated in (October 1982, September–December 1987, September–December 1992 and September–October 1997 - there seemed to be no disturbance past that date) and the period when telephone interviewing was being phased-in (August 1996 - January 1997).

[14] This series uses the (contemporary) population figures given in 6203.0 at the time of publication as the deflator. It is possible to use the present day (end of 2000) estimates of population size in past months instead. The two series give virtually identical results (the correlation coefficient between the two population series is 0.9996). Since it would have been the estimates of population at the time that were being used to form estimates of the population represented by the matched sample, it is the 'original' series which is the appropriate deflator for our purposes.

[15] An investigation of the seasonal component of the series indicate that there is a systematic tendency for the size of the unmatched proportion of the population to be relatively high in January and (especially) February of each year. Presumably this is because people are holidaying away from home in January and others are leaving home in January and February for various reasons (e.g. people moving to new locations to start new jobs).

[16] Henderson weighting schemes of varying lengths were tried but short-term cycles remained in the series. This is the reason the Hodrick-Prescott method was used.

Another useful way to discern any trends in the series (and, in this particular case, to possibly uncover the reasons for them) is to compute the means for the seasonally adjusted series[17] for the periods between the breaks when new samples were being rotated in and for the periods before and after the introduction of telephone interviewing. The means for the various sub-periods are set out in Table 1 below. The information in the Table reinforces the impression obtained by a scan of Figures 1 - 3, namely that the proportion of the total population covered by the matched sample rose in the early 90's and rose even further after the introduction of telephone interviewing at the end of 1996.

Table 1: Mean values of the proportion of the population represented (and not represented) by the matched sample in various sub-periods

	1979:08-1982:09	1982:11-1987:08	1988:01-1992:08	1993:01-1996:07	1997:02-1997:08	1997:11-2000:10
Represented	0.779	0.778	0.778	0.782	0.787	0.789
Not Repres.	0.221	0.222	0.222	0.218	0.213	0.211

Now, there are three reasons why the matched sample represents less than 100% of the population. One reason is that there is no attempt to match the 3% or so of the total population who reside in non-private dwellings.[18] Another reason is the practice of sample rotation, which has the effect that a maximum of 7/8 of the residents of private dwellings can potentially be matched across successive months. Finally, non-response by persons in the potentially matchable private dwellings reduces the size of the population represented by the matched sample below its potential maximum. The first two reasons just given explain why the potentially matchable population is less than the total population. The third element explains why the number in private dwellings who are actually matched is less than the number who potentially can be matched.

[Insert Figures 4 and 5 near here]

We commence our examination of these factors by looking at what has been happening to the proportion of persons in the whole population who are enumerated in non-private dwellings. Unfortunately, the data we need to make this calculation is only available from September 1984. Figure 4 shows the original values of this ratio over the period 1984:09–2000:10[19] and has the seasonally adjusted series (solid line) superimposed on the original series (dotted line). The proportion of the population enumerated in non-private dwellings to the total civilian population aged 15 years and over has a mean value of 3.25% with a maximum of 3.84% and a minimum of 2.38%.

[17] Averages are taken of the seasonally adjusted series because all of the series have clear seasonal components and the sub-periods are not identical in the months they contain.
[18] Strictly speaking I should write "the population enumerated in non-private dwellings".
[19] The seasonally adjusted and smoothed series were obtained using the same procedures as for *PRMS/POP*.

Figure 5 shows 13-term Henderson[20] and Hodrick-Prescott trends. Table 2 shows the means of the seasonally adjusted series for each of our sub-periods. The proportion of persons in the whole population who are enumerated in non-private dwellings appears to be trending downwards. Obviously, this is one reason why the ratio of matched to total population (*PRMS/POP*) has been rising.

Table 2: Mean values of the proportion of the population who are enumerated in non-private dwellings in various sub-periods

1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
0.035	0.034	0.031	0.034	0.030

Earlier we noted that the potentially matchable population will be 7/8 of the population resident in private dwellings.[21] Table 3 below shows how the potentially matchable population (*PMP*) as a proportion of the total population (*POP*) has changed over time.

Table 3: Mean values of the proportion of the population who are potentially matchable (*PMP/P*) in various sub-periods

1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
0.844	0.845	0.848	0.845	0.849

It is instructive to ask what proportion of the potentially matchable proportion of the population is in fact matched? To recover this information we simply divide the figures[22] given in the first row of Table 2 (this is *PRMS/POP*) by the figures given in Table 3 (this is *PMP/POP*) to get the ratio (*PRMS /PMP*). The result is reported in Table 4 below.

Table 4: Mean values of the proportion of the potentially matchable population who are in fact matched (i.e. *PRMS/PMP*) in various sub-periods

1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
0.922	0.921	0.922	0.931	0.929

[20] All Henderson trends computed in this paper are 13-term series calculated using the procedures described in ABS (1987).

[21] Strictly speaking it is dwellings that are rotated in or out, not persons. I am assuming that 7/8 of private dwellings is also 7/8 of persons in all the private dwellings in the sample.

[22] We also need a figure for *PRMS/POP* for the period 1984:09 - 1997:08, which is not given in Table 1. It is 0.778, which coincidentally is the same as that given in Table 1 for the whole of 1982:11 - 1987:08.

Earlier, (in Table 1) we saw that the proportion of the population represented by the matched sample has been tending to rise over time (e.g., it has risen from 0.778 in the period 1984:09 - 1987:08 to 0.789 in the period 1997:11 - 2000:10) and, correspondingly, the unmatched proportion has been tending to fall over time. We now see that there are essentially two reasons for this. First, the proportion of the population living in non-private dwellings included in the sample has fallen (Table 3). Second, the proportion of the potentially matchable population who reside in private dwellings and who are indeed matched has risen (Table 4). This rise, which may or may not be sustained, must in some sense reflect a rise in the response rate of persons in private dwellings included in the in the monthly LFS, following the introduction of telephone interviewing at the end of 1996.[23]

4. The Unemployment Rate for the Population Represented by the Matched Sample

Given that the matched sample represents less than 80% of the total survey (population), an obvious question to ask is - does the unemployment rate (say) in the matched sample accurately reflect that for the population as a whole? And if not, why not? Figure 6 shows (original) series for the unemployment rate in the whole population (UR_T) and for the matched sample (UR_{PM}) over the period[24] 1978:09 - 2000:10. The two series are highly correlated, the simple correlation coefficient being (r =) 0.968. However the mean and median values of the unemployment rate for the total population are 8.1% and 8.0% respectively and both are greater than the corresponding indices for the matched sample (7.8% and 7.7% respectively). Figure 7 shows seasonally adjusted and Figure 8 smoothed values for these two series. All three figures indicate clearly that the unemployment rate for matched sample has the same time series profile as the unemployment rate for whole population but that the level of the unemployment rate for the matched sample is consistently below that of the whole population.[25]

[Insert Figures 6, 7 and 8 near here]

As previously mentioned, another way to discern any trends in the series is to compute the means of the seasonally adjusted values for each sub-period. The relevant information is given in Table 5 below. We see that the means for the

[23] I write "in some sense" as the numbers in Table 4 reflect the matched responses of households in two successive surveys and so the figures must reflect: (a) the response rate in each of the two surveys and (b) the extent to which (non-) response is serially correlated. The interested reader might consult Dixon, Lim and Thomson (2001) for more details of this relationship.

[24] Note that we are again excluding the periods when the size of the matched sample was abnormally low due to a new sample being rotated in (October 1982, September–December 1987, September–December 1992 and September–October 1997) and the period when telephone interviewing was being introduced (August 1996 - January 1997). The ABS is of the view that during the phasing in of telephone interviewing survey estimates of employment and unemployment were biased. See n5 above.

[25] However, the difference between the two is not constant. Indeed, it may be worth noting that close examination of the relationship between the two suggests that the difference between the two is smallest 6-12 months before the unemployment rate begins to rise as the economy moves into recession. This can be seen in Figures 6 and 7. However, the amount of noise in the data would preclude using this as a predictive tool to forecast unemployment rate turning points.

unemployment rate for the population as a whole are above those for the population represented by the matched sample in every sub-period. It is also the case that the difference between the two is greatest when the economy is moving into or is in, recession.[26]

Table 5: Mean values of UR_T, UR_{PM} and (UR_T-UR_{PM}) (%) in various sub-periods

	1979:08 - 1982:09	1982:11 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
UR_T	6.2	8.8	8.0	9.5	8.7	7.4
UR_{PM}	5.8	8.4	7.6	9.3	8.4	7.2
(UR_T-UR_{PM})	0.4	0.4	0.4	0.2	0.3	0.2

We have seen that the unemployment rate for the matched sample is consistently below that for the population as a whole. To investigate the reasons why this is so we need to remind ourselves about the form which the Labour Force Survey (LFS) takes and the process by which certain individuals who are selected to be in the sample are matched, while others are not. We can then look at the labour market experience of the matched and unmatched groups in the population.

5. The Behaviour over Time of the Matched Sample's Unemployment Rate and the Unemployment Rate for Groups Not Represented in the Matched Sample

We know from the discussion of LFS methodology in Section 2 that the "gross flows estimates relate only to [those] persons in private dwellings" and, within this group, only to those persons "for whom information was obtained in successive surveys" (ABS 6203.0, October 2000, p 42f). It follows that the labour force characteristics of the matched and total population would differ if the characteristics of persons in private and non-private dwellings were to differ. They would also differ if the characteristics of persons in private dwellings who are represented in the matched sample differ from the characteristics of those persons in private dwellings who are not represented in the matched sample (this would be because of sample rotation or because of non-response). All of which is to say that any difference between the aggregate unemployment rate and the unemployment rate for the matched sample (UR_T - UR_{PM}) must reflect: (a) any difference between the unemployment rate for persons resident in non-private dwellings and that for persons resident in private dwellings and/or (b) any difference between the unemployment rate for persons resident in private dwellings who are represented in the matched sample and that for persons resident in private dwellings who are not represented in the matched sample.

We may therefore usefully think of the Australian population as being divided into those who are not enumerated/resident in private dwellings, and for this reason are not represented in the matched sample, (I will use the subscript *NP* to indicate this group) and those enumerated/resident in private dwellings (denoted by the subscript *P*). This

[26] Boehm and Summers (1999) date the onset of the recession in the early eighties as 1981:05 with the recovery commencing in 1985:11. They date the onset of the recession in the late eighties as 1989:11 with the recovery commencing in 1992:12.

second group may be further divided into those persons who are resident in private dwellings and who are represented in the matched sample (these persons will be denoted by the subscript *PM*); and, those who are residents of private dwellings but who - for one reason or another - are not represented in the matched sample (this group will be denoted by the subscript *PNM*).

[Insert Figures 9 - 16 near here]

Figures 9 through 16 show the behaviour over time of the unemployment rates for the various groups we have identified. Figure 9 shows original series for the unemployment rate in the whole population (UR_T) and for the matched sample (UR_{PM}) over the period 1984:09 - 2000:10. Figure 10 shows smoothed seasonally adjusted series for the unemployment rate in the whole population and for the matched sample over the same period. (These two figures are really conveying the same information as Figures 4 - 6 except for the shorter time span here). As we noted in section 2, the unemployment rate for matched sample has the same time series profile as the unemployment rate for the whole population but the level of the unemployment rate for the matched sample is consistently below that of the whole population. This is also evident in Table 6 which looks at the mean values of the seasonally adjusted[27] unemployment rate in the whole population (UR_T) and the matched sample (UR_{PM}).

Table 6: Mean values of UR_T and UR_{PM} in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
UR_T	8.3	8.0	9.5	8.7	7.4
UR_{PM}	7.9	7.6	9.3	8.4	7.2
$(UR_T - UR_{PM})$	0.4	0.4	0.2	0.3	0.2

Figure 11 shows (original) series[28] for the unemployment rate for the population enumerated in private dwellings (UR_P) and in non-private dwellings (UR_{NP}) over the period 1984:09 - 2000:10. The two series are poorly correlated, the simple correlation coefficient is (r =) 0.273. The mean value of the unemployment rate for persons in non-private dwellings is 11.1% which is much higher than that for those persons in private dwellings (8.3%). Also, the non-private dwellings series has a standard deviation of 3.0% whereas the private dwellings component has a standard deviation less than one-half of that, 1.4%. In short, the non-private series has a higher average and is more volatile than the private series. Figure 12 shows smoothed values for these two series. Table 7 shows the mean values of the seasonally adjusted unemployment rate for the population not resident in private dwellings (UR_{NP}) and for the population resident in private dwellings (UR_P) in various sub-periods. The two series have markedly different time series profiles with the result that the gap between the two varies over time. Clearly then, one reason the matched sample characteristics

[27] For the reason why means are taken of the seasonally adjusted and not the original series, see n 17.
[28] Source is 6203.0. Published data for these two portions of the survey are only available since September 1984.

will differ from that of the whole population is that the unemployment rate of persons in private as compared with non-private dwellings can differ markedly.[29]

Table 7: Mean values of UR_{NP} and UR_P in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
UR_{NP}	13.6	12.0	10.2	7.5	9.5
UR_P	8.2	7.9	9.5	8.7	7.4
(UR_{NP} - UR_P)	5.4	4.1	0.7	-1.2	2.1

We have noted that the matched sample only refers to (a sub-set of) persons in private dwellings. An obvious question to ask then is - how does the series for the unemployment rate in the matched component of private dwellings compare with that for all persons in private dwellings?

Figure 13 shows (original) series for the unemployment rate for all persons enumerated in private dwellings (UR_P) and for those persons in the matched sample ((UR_{PM})) over the period 1984:09 - 2000:10. The two series are highly correlated, the simple correlation coefficient being (r =) 0.998. The mean value of the unemployment rate for persons in the matched sample is 7.9% which is smaller than that for those in private dwellings (8.3%). Figure 14 shows smoothed values for these two series. Both Figures indicate that the two series have identical time series profiles and that there is a persistent gap between the two rates. This is also brought out in Table 8 which looks at the mean values of the seasonally adjusted unemployment rate for all persons resident in private dwellings (UR_P) and the unemployment rate for those residents of private dwellings who are represented in the matched sample (UR_{PM}) in various sub-periods. The fact that the unemployment rate of persons in private dwellings and who can be matched is consistently below that of all persons who reside in private dwellings implies that the unemployment rate of persons in private dwellings and who are not represented in the matched sample is consistently above that of those residents of private dwellings who are represented in the matched sample.

Table 8: Mean values of UR_P *and* UR_{PM} in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
UR_P	8.2	7.9	9.5	8.7	7.4
UR_{PM}	7.9	7.6	9.3	8.4	7.2
(UR_P - UR_{PM})	0.3	0.3	0.2	0.3	0.2

[29] At the same time we saw in Table 2 that the proportion of the population resident in non-private dwellings is quite small (3.25%, on average) and so in practice the contribution of the this difference to the total will be smaller than would appear from looking at Table 7 and also Figures 11 and 12. (More on this later.)

Now, since we know the labour market characteristics of the persons in the matched sample and we also know the labour market characteristics of all persons in private dwellings, it is possible to form a series for the unemployment rate of those persons enumerated in private dwellings but who are not in the matched sample[30] (UR_{PNM}) and to compare this with the unemployment rate for those represented in the matched sample (UR_{PM}). This is done in Figure 15 which shows (original) series for the unemployment rate for the 'unmatched persons' enumerated in private dwellings (solid line) and for those persons in the matched sample (dotted line) over the period 1984:09 - 2000:10. The two series are highly correlated, the simple correlation coefficient is (r =) 0.943. However, the mean value of the unemployment rate for persons in the matched sample is 7.9% which is well below that for those persons in private dwellings who were not matched (9.9%). Figure 16 shows smoothed values for these two series.

Table 9 reports the mean values of the seasonally adjusted unemployment rate for those persons who are resident in private dwellings and who are represented in the matched sample (UP_M) as against the unemployment rate for those who are not (UR_{PNM}) in various sub-periods. Clearly, the unemployment rate of persons in private dwellings and who are matched is systematically different to those who are in private dwellings but cannot be matched. Now, the unmatched group will be made up of two different groups of persons. One group cannot be matched as they have only just been rotated into the sample. Since the monthly LFS commenced in 1978, one-eighth of the sample has been replaced each month. This replacement sample generally comes from the same geographic area(s) as the outgoing one and for this, and other reasons, "each rotation group is a representative sample of the Australian population in its own right" (Bell, 1998, p 3). Given this, it is unlikely that the mere fact that some members of the private dwelling sample are replaced each month will itself introduce any systematic bias. However, the second group of persons who will be in the unmatched component will be those who have not been rotated out but who have moved or for some other reason did not respond to the survey. It must be the characteristics of this group which differs from the matched group. In other words, it would appear that the non-respondents and/or those who have changed address (or who, for some other reason cannot be contacted by the interviewers) tend to have a higher unemployment rate than the respondents. This is not surprising.[31]

Table 9: Mean values of UR_{PNM} and UR_{PM} (%) in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
UR_{PNM}	9.9	9.4	11.1	10.4	8.7
UR_{PM}	7.9	7.6	9.3	8.4	7.2
(UR_{PNM} - UR_{PM})	2.0	1.8	1.8	2.0	1.5

[30] The unemployment rate for all persons in private dwellings is a weighted average of the unemployment rate for the matched component and the rate for the unmatched component. We have information for the total and also for the matched component and for the relative size of the two components. This allows us to recover the UR_{PNM}.

[31] Nor is it a peculiarly Australian phenomenon. It has also been reported to be the case in the US (Barkume and Horvath, 1995, p 30).

Our findings in this section of the paper may be summarised as follows: The population represented by the matched sample tends to systematically have a lower unemployment rate than does the population as a whole and there seem to be two reasons for this. First, the matched sample only refers to persons resident in private dwellings and it would appear that persons who are not resident (strictly speaking, I should write "not enumerated") in private dwellings tend to have a higher unemployment rate than those who are. Second, it appears to be the case that those persons who are resident in private dwellings and who are in the sample but who are not matched have a higher unemployment rate than those who are matched. In other words, the non-respondents and those who have moved residence have a higher unemployment rate than those who respond and who have not moved and consequently are in the matched sample.

It is possible to quantify the relative contributions of these two components of bias in the matched sample unemployment rate. This is the task of the next section of the paper.

6. A Simple Model of the Relationship between the Aggregate Unemployment Rate and the Unemployment Rate for Persons in the Matched Sample

Let U denote the number unemployed in any period. It must be true, by definition, that:

$$U_T = U_{NP} + U_{PNM} + U_{PM}$$

As before, the T subscript indicates that the variable refers to the total population, NP refers to persons in non-private dwellings, PNM refers to those persons in private dwellings who are not represented in the matched sample and PM refers to those persons in private dwellings who are represented in the matched sample.

Dividing both sides by the size of the aggregate Labour Force (LF_T) gives an expression for the aggregate unemployment rate:

$$\frac{U_T}{LF_T} = \frac{U_{NP}}{LF_T} + \frac{U_{PNM}}{LF_T} + \frac{U_{PM}}{LF_T}$$

It is possible to convert this into an expression for the aggregate unemployment rate (U_T/LF_T), as a weighted sum of the unemployment rate for persons in each of the three categories we have identified above (persons enumerated in non-private dwellings (NP), persons in private dwellings who are represented in the matched sample (PM) and persons in private dwellings who are not represented in the matched sample (PNM):

$$\frac{U_T}{LF_T} = \frac{U_{NP}}{LF_{NP}}\frac{LF_{NP}}{LF_T} + \frac{U_{PNM}}{LF_{PNM}}\frac{LF_{PNM}}{LF_T} + \frac{U_{PM}}{LF_{PM}}\frac{LF_{PM}}{LF_T} \qquad (1)$$

The weights in the above expression are the proportions of the total labour force which are found in each category. In passing, we might note that their mean values over the period 1984:09-2000:10 are: $LF_{NP}/LF_T = 0.014$, $LF_{PNM}/L_T = 0.188$ and $LF_{PM}/LF_T = 0.798$.

Using the symbol *UR* to denote an unemployment rate, equation (1) may be written as:

$$UR_T = UR_{NP}\frac{LF_{NP}}{LF_T} + UR_{PNM}\frac{LF_{PNM}}{LF_T} + UR_{PM}\frac{LF_{PM}}{LF_T} \tag{2}$$

Our aim is to find an expression for the difference between the aggregate unemployment rate and the unemployment rate for that part of the population which is represented by the matched sample (UR_T - UR_{PM}) in terms of[32] (UR_{NP} - UR_P) and (UR_{PNM} - UR_{PM}).

If we add and subtract UR_P (LF_{NP}/LF_T) to/from the RHS of (2) we find, after some slight rearrangement, that:

$$\begin{aligned} UR_T = (UR_{NP} - UR_P)\frac{LF_{NP}}{LF_T} + UR_P\frac{LF_{NP}}{LF_T} + UR_{PNM}\frac{LF_{PNM}}{LF_T} \\ + UR_{PM}\frac{LF_{PM}}{LF_T} \end{aligned} \tag{3}$$

Now, UR_P is a weighted sum of UR_{PM} and UR_{PNM}, such that:

$$UR_P = UR_{PM}\frac{LF_{PM}}{LF_P} + UR_{PNM}\frac{LF_{PNM}}{LF_P}$$

Substituting this into the second term on the RHS of (3) and rearranging, gives the expression we are looking for. It links the behaviour of (UR_T - UR_{PM}) on the one hand, with the behaviour of (UR_{NP} - UR_P), and (UR_{PNM} - UR_{PM}) on the other:[33]

$$\begin{aligned} UR_T - UR_{PM} = (UR_{NP} - UR_P)\left(\frac{LF_{NP}}{LF_T}\right) \\ + (UR_{PNM} - UR_{PM})\left(\frac{LF_{PNM}}{LF_P}\right) \end{aligned} \tag{4}$$

We now turn to look at the contribution of the various terms on the RHS of equation (4) to the difference between UR_T and UR_{PM}.

To begin with we look at the means for each of the series. Evaluated at the means over the period 1984:09 - 2000:10 we have: (UR_T - UR_{PM}) = 0.41%, (UR_{NP} - UR_P) =

[32] Where UR_P is the unemployment rate for all persons in private dwellings.

[33] A proof of this using the approach mentioned in the here can be obtained from the author. In an Appendix to this paper I provide an alternative derivation of (4).

3.01% and (UR_{PNM} - UR_{PM}) = 1.97%. The means for the weights are: (LF_{NP}/LF_T) = 0.014 and (LF_{PNM}/LF_P) = 0.192. Although the mean of (UR_{NP} - UR_P) is higher than the mean for (UR_{PNM} - UR_{PM}), the weight given to the latter makes it far more important as a determinant of the size of (UR_T - UR_{PM}) than the former. Indeed, the mean value of the two components on the RHS of equation (4) are: 0.04% for [(UR_{NP} - UR_P)*(LF_{NP}/LF_T)] and 0.38% for [(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)].[34] Clearly the latter is the dominant element.

Figures 17 - 21 show the smoothed time paths for each of the elements of equation (4).[35] Figure 17 shows a time series for the difference between the unemployment rate for all persons and that for the matched sample, i.e. (UR_T - UR_{PM}). Figure 18 shows the difference between the unemployment rate for persons resident in non-private dwellings and the unemployment rate for those resident in private dwellings, i.e. (UR_{NP} - UR_P). Figure 19 shows the difference between the unemployment rate for persons resident in private dwellings who are not represented in the matched sample and the unemployment rate for persons resident in private dwellings who are represented in the matched sample, i.e. (UR_{PNM} - UR_{PM}). Figure 20 shows the behaviour over time of the ratio of LF_{NP} to LF_T and Figure 21 shows a time series for the ratio of LF_{PNM} to LF_P. All time series are for the period 1984:09 - 2000:10.

One way to more clearly discern any trends in the series is to compute the means[36] for the seasonally adjusted series for appropriate sub-periods. Table 10 reports mean values of the various unemployment rate difference terms which figure in equation (4). Table 11 gives mean values for the weights on each term. [37] It is noteworthy that the share of the aggregate labour force for the population enumerated in non-private dwellings is much smaller than (indeed, it is only one-half) their share of the total population. We see this by comparing the figures in the first row of Table 11 with those given in Table 3 above.

Table 10: Mean values of (UR_T - UR_{PM}), (UR_{NP} - UR_P) and (UR_{PNM} - UR_{PM}) in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
(UR_T - UR_{PM})	0.4	0.4	0.2	0.3	0.2
(UR_{NP} - UR_P)	5.4	4.1	0.7	-1.2	2.1
(UR_{PNM} - UR_{PM})	2.0	1.8	1.8	2.0	1.5

[34] Rounding errors account for the difference between 0.41 and 0.38 + 0.04.
[35] Each series is computed by comparing the smoothed seasonally adjusted values of the individual variables, where the smoothing has been undertaken by using the Henderson method.
[36] Data in Tables 10 -12 are all in percentages.
[37] Note that the values of the differences given in Table 10 are differences in the means given for individual variables in earlier tables (6, 7, and 9).

Table 11: Mean values of (LF_{NP}/L_T) and (LF_{PNM}/L_P) in various sub-periods

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10
LF_{NP}/L_T	0.017	0.016	0.013	0.014	0.010
LF_{PNM}/L_P	0.192	0.194	0.192	0.185	0.186

Both weights seem to be trending downwards. This not only accounts to some (small) extent for the difference between UR_T and UR_{PM} to be falling over time but it also, at the same time, neatly captures the two reasons why *PRMS/POP* has been rising (see section 3 above).

Mean values for the two weighted unemployment rate differences (i.e. [(UR_{NP} - UR_P)*(LF_{NP}/LF_T)] and [(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)]) which appear on the RHS of equation (4) are given in Table 12 whilst Figure 22 shows their behaviour over the period 1984:09 - 2000:10. Clearly it is the second term, the one involving (UR_{PNM} - UR_{PM}), which the most important in determining the extent to which the aggregate unemployment rate differs from that for the matched sample. However, an inspection of the Table and the Figure shows that we need to take into account both terms involving unemployment rate differences (i.e both (UR_{PNM} - UR_{PM}) and (UR_{NP} - UR_P)) if we wish to account for variations in (UR_T - UR_{PM}) over time.

Table 12: Means[38] of [(UR_{NP} - UR_P)*(LF_{NP}/LF_T)], [(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)] and (UR_T - UR_{PM}) in various sub-periods

	1984:09-1987:08	1988:01-1992:08	1993:01-1996:07	1997:02-1997:08	1997:11-2000:10
(UR_{NP} - UR_P)*(LF_{NP}/LF_T)	0.1	0.1	0.0	0.0	0.0
(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)	0.4	0.3	0.3	0.4	0.2
(UR_T - UR_{PM})	0.4	0.4	0.2	0.3	0.2

7. Concluding Remarks

I begin with a summary of the main conclusions arrived at thus far:

(a) The proportion of the population represented by the matched sample has risen over the period since the introduction of the monthly LFS and especially since the introduction of telephone interviewing in late 1996. There seem to be two reasons for this. One reason is that the proportion of the population living in non-private dwellings has fallen. A second reason is that the proportion of the potentially matchable population who provide data which can be matched has risen. This presumably reflects a rise in the response rate of persons in the sample, (again) particularly following the introduction of telephone interviewing at the end of 1996.

[38] Some columns might not add up exactly due to rounding.

(b) The population represented by the matched sample has a lower unemployment rate than does the population as a whole. There are two reasons why this is so. One reason is that the matched sample only refers to persons resident in private dwellings and it would seem that persons who are not resident in private dwellings tend to have a higher unemployment rate than those who are. The second reason is that those persons who are resident in private dwellings and who are in the sample but who are not able to be matched across months, have a higher unemployment rate than those residents in private dwellings who are able to be matched. In other words, the non-respondents and those who have moved have a higher unemployment rate than those who are represented in the matched sample.

(c) A simple model of the relationship between the aggregate unemployment rate and the unemployment rate for persons in the matched sample shows that it is the difference between the unemployment rate for the unmatched persons living in private dwellings and the rate for those persons who can be matched which is, quantitatively, the most important item in determining the extent to which the aggregate unemployment rate differs from that for the matched sample.

The implications of the above for the representativeness of stocks data derived from matched sample estimates for persons not normally resident in private dwellings and for those who reside in private dwellings but who possess characteristics which might lead them to be in the unmatched component are clear.

It is important to consider the implications of this analysis about stocks for the representativeness of the matched sample's flows and thus for the representativeness of transition rates and other calculations based on the flows data. It is tempting to say that we can, without qualification, transfer our findings about the representativeness of stocks to flows, but there is no necessary or simple connection between the two, not least because flows data to the extent that it relates to unemployment rates, is about changes in the rate of unemployment and not the level of the rate, per se. Indeed, without data on flows for each of the unmatched groups, this is not an easy matter to resolve. However, it is reasonable to suppose that if we observe similar volatility in the unemployment rate for any two groups of the population, we might then infer similarity in (relative) inflow and outflow levels and/or transition rates. The corollary of this is that if the volatility of the unemployment rate series differs systematically between (say) the *NP* & *PM* and/or the *PNM* & *PM* groups, then we may conclude that not only the stocks data but also the flows data based on the matched sample is not representative of the other groups. Well, what can we say about the relative volatility of the unemployment rate for these groups? We will use the standard deviation of the (seasonally adjusted) unemployment rate as the measure of volatility. The relevant indices are reported in Table 13.

The figures in the last two rows would suggest that it may be possible to use matched sample flows (i.e. the Gross Flows data as reported by the ABS and figures for transition rates etc derived from them) as a proxy for whole labour force flows (provided of course they are correctly expanded up to allow for non-response, sample rotation, population growth etc). However, comparisons of the standard deviations for the other pairs (*NP* c.f. *PM* and for *PNM* c.f. *PM)* suggest that we should not use

Table 13: Standard deviations in unemployment rate (%) for various sub-groups of the population

	1984:09 - 1987:08	1988:01 - 1992:08	1993:01 - 1996:07	1997:02 - 1997:08	1997:11 - 2000:10		1984:09 - 2000:10
NP	1.56	2.24	2.50	1.29	1.79		2.61
PM	0.27	1.71	1.04	0.13	0.64		1.35
PNM	0.57	1.84	1.22	0.69	0.81		1.53
PM	0.27	1.71	1.04	0.13	0.64		1.35
T	0.28	1.73	1.07	0.10	0.64		1.37
PM	0.27	1.71	1.04	0.13	0.64		1.35

matched sample flows or figures for transition rates etc derived from matched sample flows as a proxy for the rates for persons who tend to be resident in non-private dwellings or for those persons who reside in private dwellings but whose characteristics are such that they cannot be matched (in particular, people who have a propensity to change address). It would therefore appear unwise to use data based on matched sample in the LFS to throw light on the unemployment experience of, or on labour market policy directed towards, groups with a high risk of being unemployed.[39]

[39] One should therefore view with caution the conclusions reached in studies such as those by Foster and Gregory (1982) and Leeves (2000).

References

Australian Bureau Statistics (1987) *A Guide to Smoothing Time Series,* Australian Bureau of Statistics (Cat. No. 1316.0). Canberra.

Australian Bureau Statistics (1997) *Information Paper: Labour Force Survey Sample Design,* Australian Bureau of Statistics (Cat. No. 6269.0). Canberra.

Barkume, A. and Horvath, F. (1995) 'Using Gross Flows to Explore Movements in the Labor Force', *Monthly Labor Review*, 118(4), pp. 28-35

Bell, P. (1998), 'Can Labour Force Estimates be Improved Using Matched Sample Estimates?', *Australian Economic Indicators*, May, Australian Bureau of Statistics (Cat. No. 1350.0). Canberra.

Boehm, E. and Summers, P. (1999) 'Analysing and Forecasting Business Cycles with the Aid of Economic Indicators', *International Journal of Management Reviews*, 1(3), pp. 245-277

Borland, J. (1996a) 'Labour Market Flows Data for Australia', *Australian Economic Review*, 29(116), pp. 225-235

Borland, J. (1996b) 'What Can Labour Market Flows Tell Us About Unemployment?' Paper presented to the Macroeconomics Workshop held at the University of Melbourne

Borland, J. and Kennedy, S. (1998) 'Dimensions, Structure and History of Australian Unemployment', in G. Debelle and J. Borland (eds), *Unemployment and the Australian Labour Market*, Reserve Bank of Australia, Sydney, pp. 68-99

Debelle, G. and Swann, T. (1998) 'Stylised Facts of the Australian Labour Market', *Research Discussion Paper 9804*, Reserve Bank of Australia. Sydney

Dixon, R., Lim G. C. and Thomson, J. (2001) 'The Gross Flows Data: The Labour Force Survey and the Size of the Population Represented by the Matched Sample', unpublished paper, Department of Economics at the University of Melbourne

Fahrer, J. & Heath, A. (1992) The Evolution of Employment and Unemployment in Australia, *Research Discussion Paper 9215*, Reserve Bank of Australia. Sydney

Foster, W. (1981) 'Gross Flows in the Australian Labour Market', *Australian Economic Review*, 4th Quarter, pp. 57-64

Foster, W. and Gregory, R. (1982) 'The Contribution of Employment Separation to Teenage Unemployment', *Economic Record*, 58(161), pp. 118-132

Foster, W. and Gregory, R. (1984) 'A Flow Analysis of the Labor Market in Australia', in R Blandy & O Covick, *Understanding Labour Markets in Australia*, Allen & Unwin, Sydney, pp. 111-136

Leeves, G. (1997) 'Labour Market Gross Flows and Transition rates 1980-1992', *Economic and Labour Relations Review*, 8(1), pp. 110-127

Leeves, G. (1999) 'Countercyclical Unemployment Outflows and Labour Reallocation', *Australian Economic Papers*, 38(3), pp. 290-309

Leeves, G. (2000) 'Duration-Specific Unemployment Outflow Rates and Labour Market Programs', *Australian Economic Review*, 33(3), pp. 221-234

Piggott, J. and Chapman, B. (1995) 'Costing the Job Compact', *Economic Record*, 71(215), pp. 313-328

APPENDIX: Alternative derivation of equation (4) in the text.

Begin with the expression for the unemployment rate for all persons:

$$UR_T = UR_{NP}\frac{LF_{NP}}{LF_T} + UR_{PNM}\frac{LF_{PNM}}{LF_T} + UR_{PM}\frac{LF_{PM}}{LF_T}$$

Adding and subtracting both $UR_{PM}\frac{LF_{NP}}{LF_T}$ and $UR_{PM}\frac{LF_{PNM}}{LF_T}$ from the RHS of the above gives, after some rearranging:

$$UR_T = (UR_{NP} - UR_{PM})\frac{LF_{NP}}{LF_T} + (UR_{PNM} - UR_{PM})\frac{LF_{PNM}}{LF_T} + UR_{PM}\left(\frac{LF_{PM}}{LF_T} + \frac{LF_{NP}}{LF_T} + \frac{LF_{PNM}}{LF_T}\right)$$

Since $LF_{NP} + LF_{PNM} + LF_{PM} = LF_T$, we may manipulate this to give an expression for the difference between the unemployment rate for all persons and the unemployment rate for persons in the matched sample:

$$UR_T - UR_{PM} = (UR_{NP} - UR_{PM})\frac{LF_{NP}}{LF_T} + (UR_{PNM} - UR_{PM})\frac{LF_{PNM}}{LF_T}$$

Note that this may be written as

$$UR_T - UR_{PM} = ((UR_{NP} - UR_P) + (UR_P - UR_{PM}))\frac{LF_{NP}}{LF_T} + (UR_{PNM} - UR_{PM})\frac{LF_{PNM}}{LF_T}$$

Now, UR_P is a weighted sum of UR_{PM} and UR_{PNM}, such that:

$$UR_P = UR_{PM}\frac{LF_{PM}}{LF_P} + UR_{PNM}\frac{LF_{PNM}}{LF_P}$$

Substituting this into the above and rearranging gives:

$$UR_T - UR_{PM} = (UR_{NP} - UR_P)\left(\frac{LF_{NP}}{LF_T}\right) + (UR_{PNM} - UR_{PM})\left(\frac{LF_{PNM}}{LF_P}\frac{LF_{NP}}{LF_T} + \frac{LF_{PNM}}{LF_T}\right)$$

which may be written as

$$UR_T - UR_{PM} = (UR_{NP} - UR_P)\left(\frac{LF_{NP}}{LF_T}\right) + (UR_{PNM} - UR_{PM})\left(\frac{LF_{PNM}}{LF_P}\right)$$

This is equation (4) in the text.

Figure 1: The Proportion of the Population Represented by the Matched Sample (*PRMS/POP*) over the period 1979:08 - 2000:10[40] (Solid line is the seasonally adjusted series; the dotted line is the original series.)

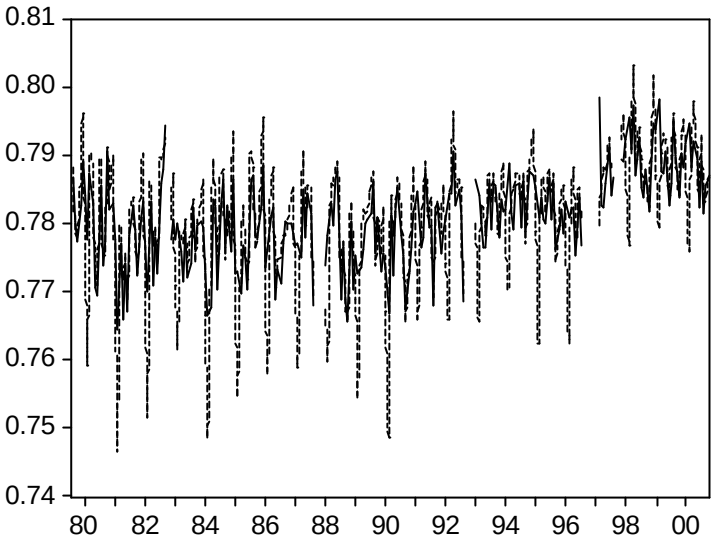
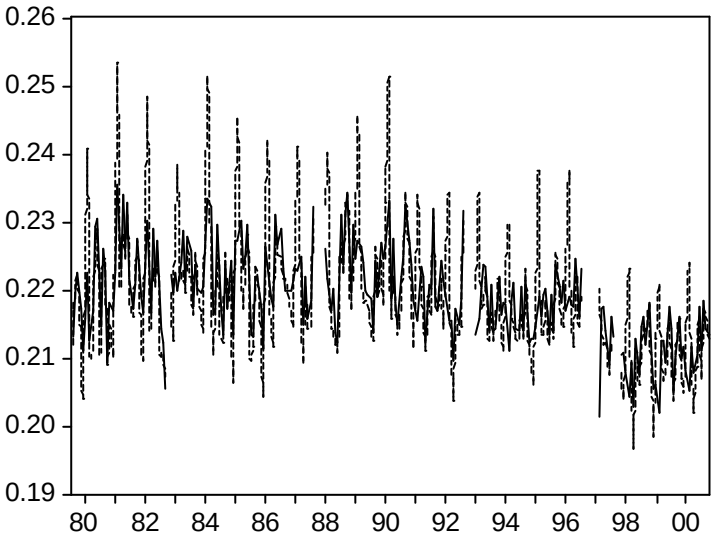


Figure 2: The proportion of persons in the population who are not 'in' the Matched Sample (*1-(MSS/POP)*) over the period 1979:08 - 2000:10. (Solid line is the seasonally adjusted series; the dotted line is the original series.)



[40] Data is not reported in any figures for those periods where the sample was being redesigned (1982:10, 1987:09–1987:12, 1992:09–1992:12 & 1997:09–1997:10) and when telephone interviewing was being phased in (1996:08 - 1997:01).

Figure 3: The proportion of persons in the population who are 'in' the Matched Sample (*1-(MSS/POP)*) over the period 1979:08 - 2000:10. Smoothed values (using a Hodrick - Prescott filter) solid line, seasonally adjusted values dotted line.

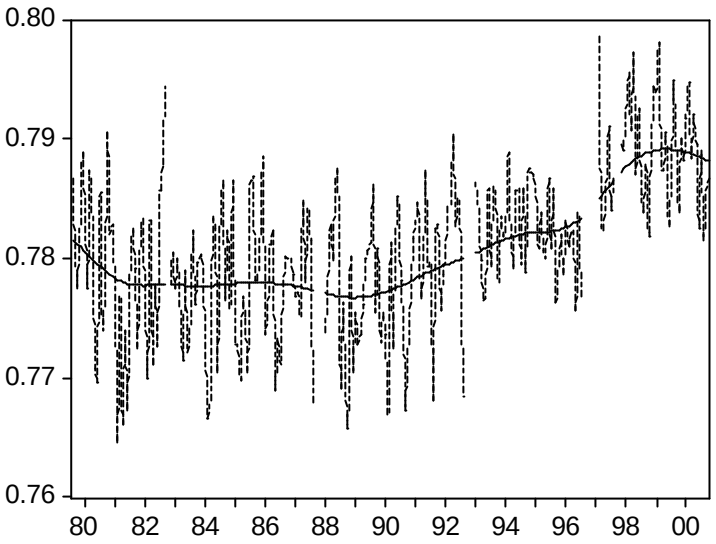


Figure 4: Seasonally adjusted (solid line) and original values (dotted line) of the ratio of the size of the population enumerated in non-private dwellings to the size of the total civilian population aged 15 years and over for the period 1984:09 - 2000:10

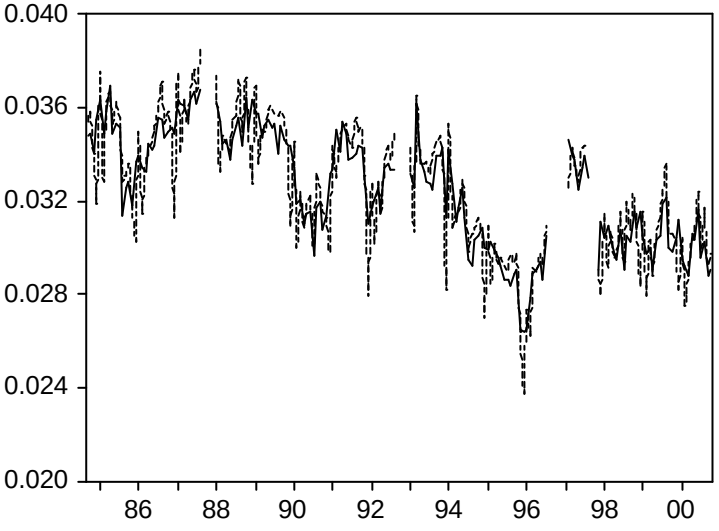


Figure 5: Time series for the proportion of the total civilian population aged 15 years which was enumerated in non-private dwellings, smoothed Solid line) and seasonally adjusted series (dotted line), 1984:09 - 2000:10. (Smoothed series with HP filter on the right, 13-term Henderson on the left)

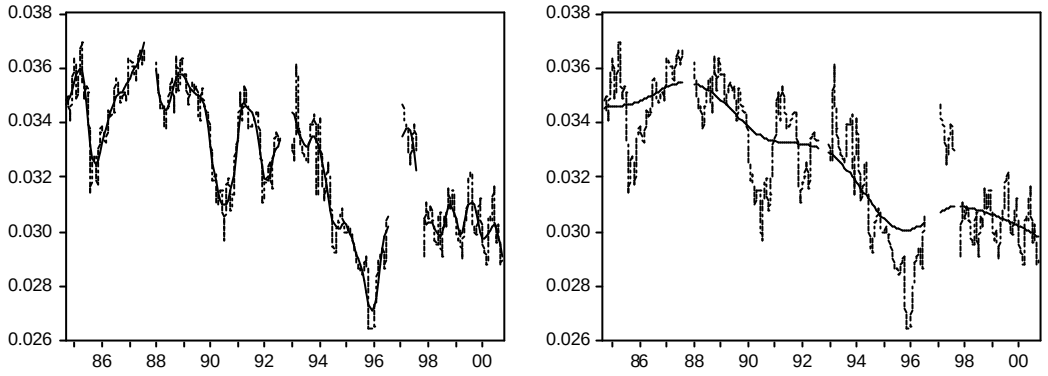


Figure 6: All persons unemployment rate (dotted line) c.f. the unemployment rate for the matched sample (solid line) 1979:08 - 2000:10 (original series)

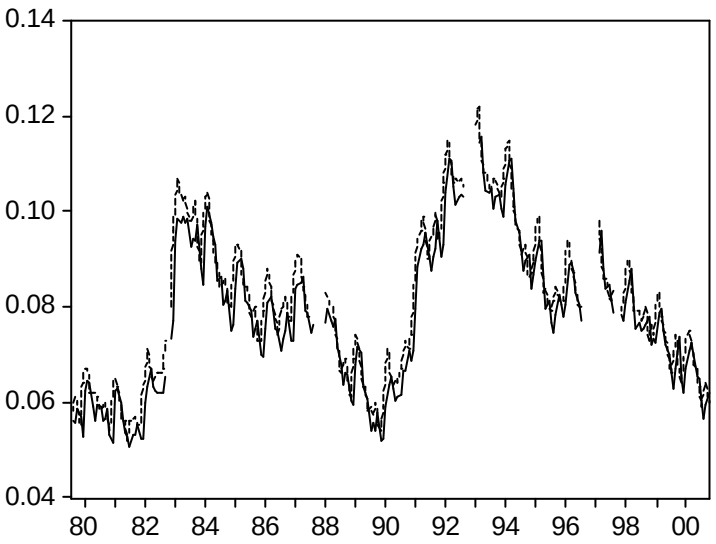


Figure 7: All persons unemployment rate (dotted line) c.f. the unemployment rate for the matched sample (solid line) 1979:08 - 2000:10 Seasonally adjusted series

Figure 8: All persons unemployment rate (dotted line) c.f. the the unemployment rate for matched sample (solid line) 1979:08 - 2000:10 (smoothed - using 13-term Henderson left and HP filter - right)

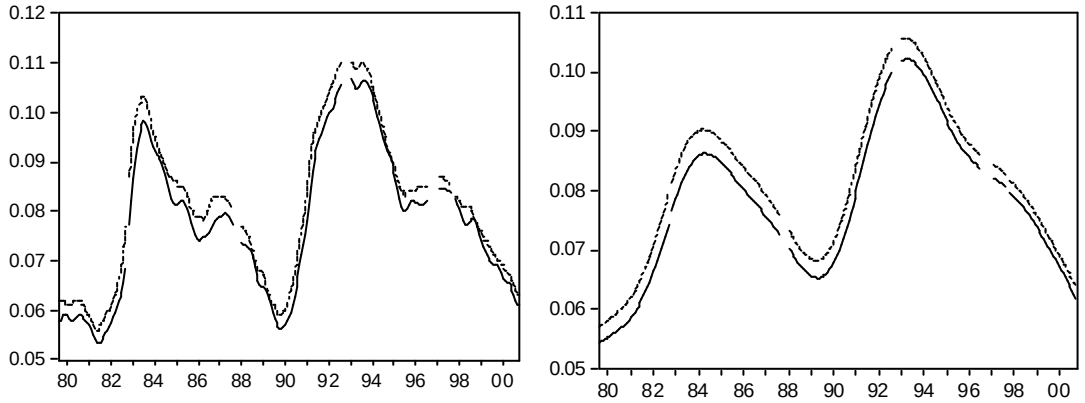


Figure 9: All persons unemployment rate (UP_T) (dotted line) c.f. the unemployment rate for the matched sample (UP_{PM}) (solid line) 1984:09 - 2000:10.

Figure 10: Smoothed seasonally adjusted series for the unemployment rate for all persons (UP_T) - dotted line - c.f. the unemployment rate for the matched sample (UP_{PM}) - solid line - (smoothed - using 13-term Henderson left and HP filter - right) 1984:09 - 2000:10.

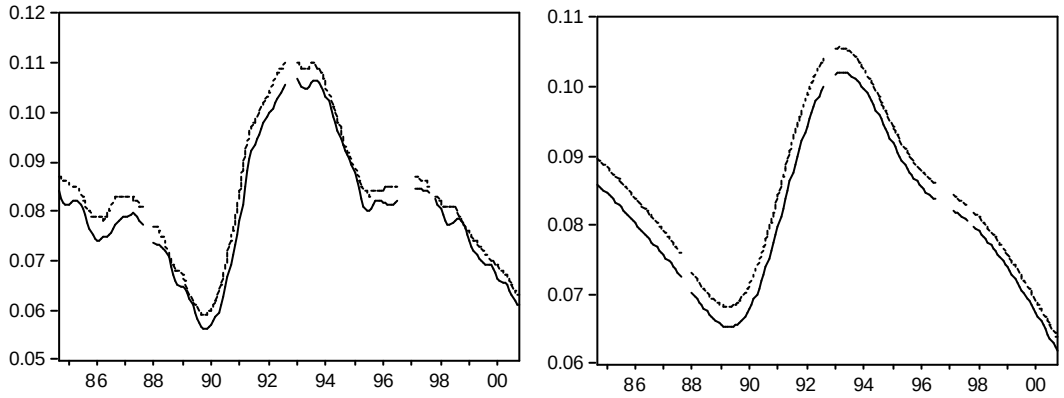


Figure 11: Actual unemployment rate for persons enumerated in private dwellings (UP_P) - dotted line - as against the unemployment rate for persons enumerated in non-private dwellings (UP_{NP}) - solid line - 1984:09 - 2000:10. (Original series)

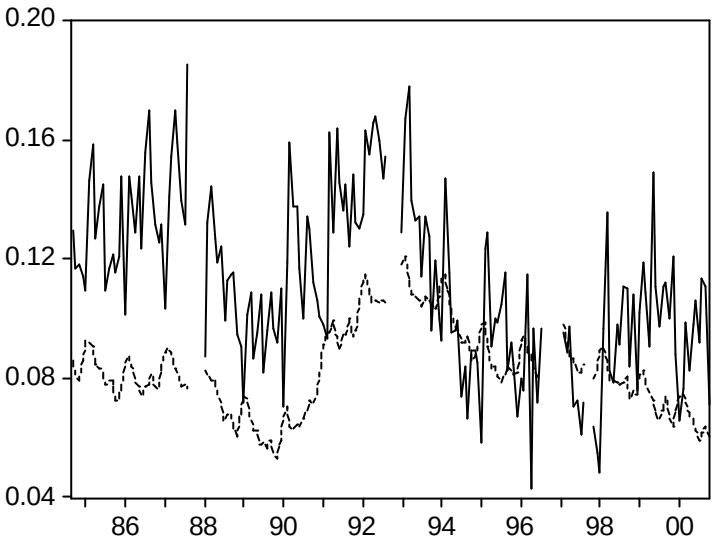


Figure 12: Smoothed seasonally adjusted series for the unemployment rate of persons enumerated in private dwellings (UP_P) - dotted line - as against the unemployment rate for persons enumerated in non-private dwellings (UP_{NP}) - solid line - 1984:09 - 2000:10. (13-term Henderson on left and HP filter on right)

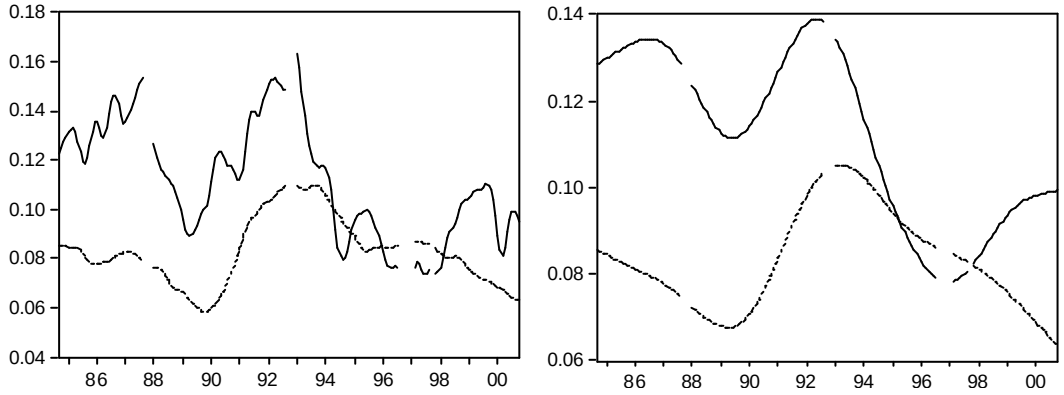


Figure 13: Actual unemployment rate persons enumerated in all private dwellings (UP_P) - dotted line - as against the unemployment rate for persons in the matched sample (UP_{PM}) - solid line - 1984:09 - 2000:10.

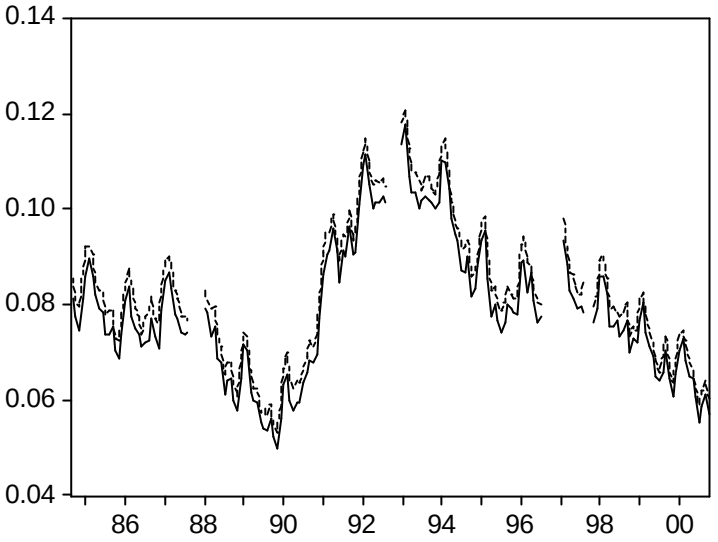


Figure 14: Smoothed seasonally adjusted series for the unemployment rate for persons enumerated in all private dwellings (UP_P) - dotted line - as against the unemployment rate for persons in the matched sample (UP_{PM}) - solid line - (13-term Henderson moving average left, HP filter, right) 1984:09 - 2000:10.

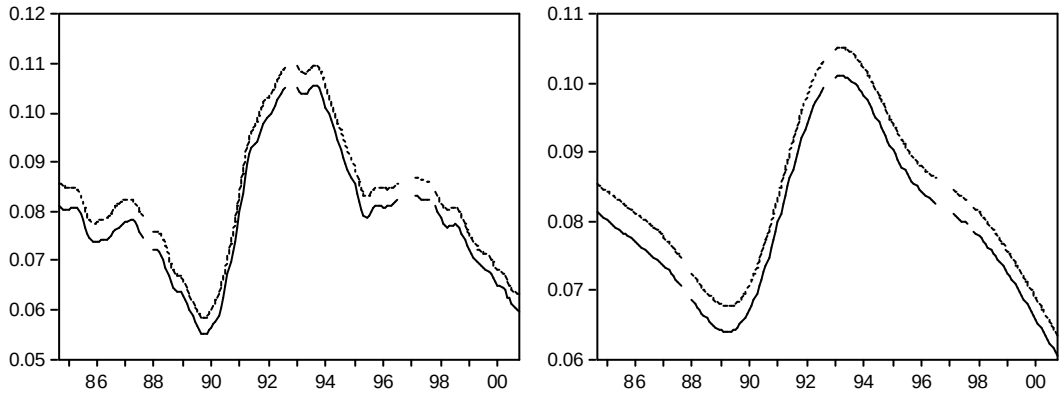


Figure 15: Unemployment rate for unmatched persons in private dwellings (UP_{PNM}) - solid line - as against that for matched persons in private dwellings (UP_{PM}) - dotted line - 1984:09 - 2000:10.

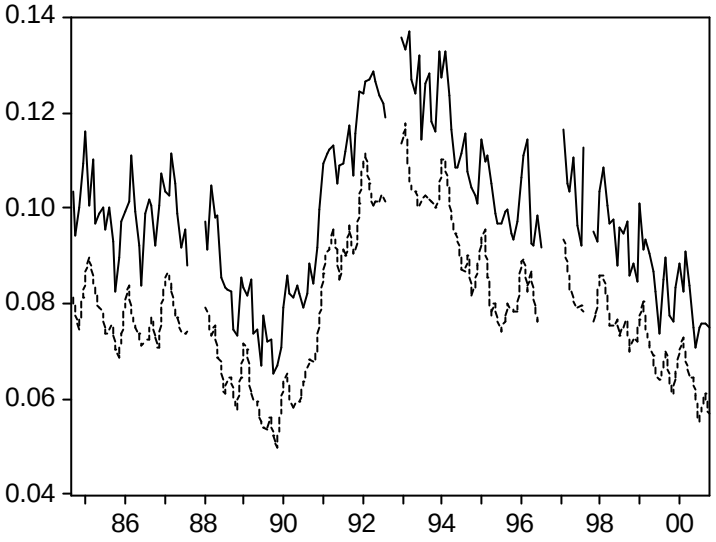


Figure 16: Unemployment rate for unmatched persons in private dwellings (UP_{PNM}) - solid line - as against that for matched persons in private dwellings (UP_{PM}) - dotted line - smoothed seasonally adjusted series (13-term Henderson at left and HP filter right) 1984:09 - 2000:10.

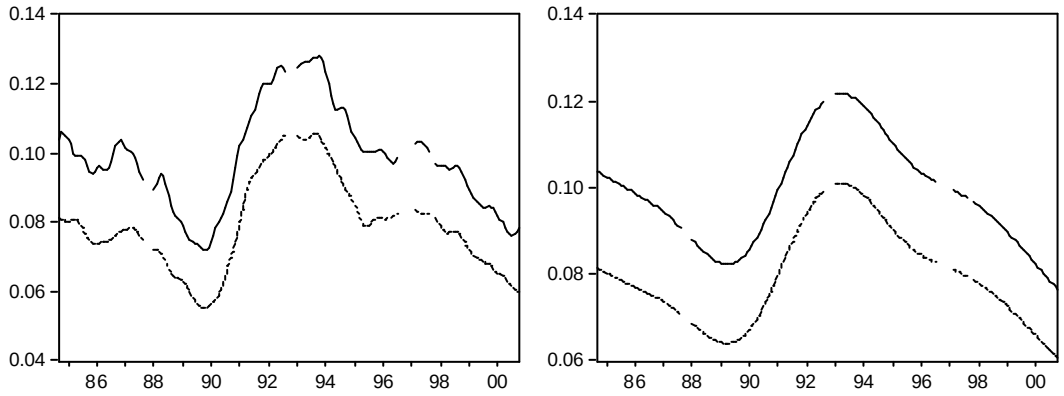


Figure 17: Smoothed seasonally adjusted series for UR_T - UR_{PM}

Figure 18: Smoothed seasonally adjusted series for UR_{NP} - UR_P

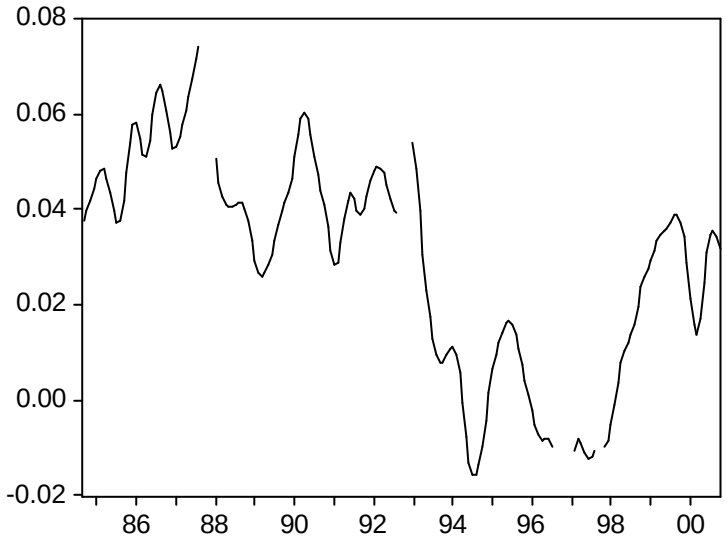


Figure 19: Smoothed seasonally adjusted series for UR_{PNM} - UR_{PM}

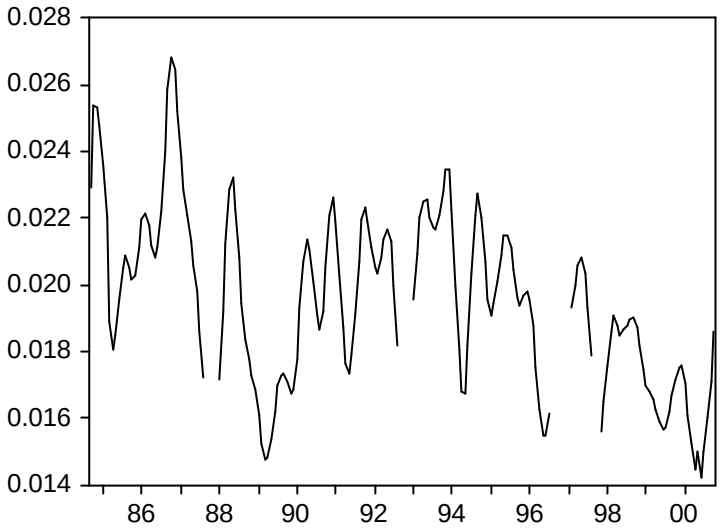


Figure 20: Smoothed seasonally adjusted series for LF_{NP}/LF_T

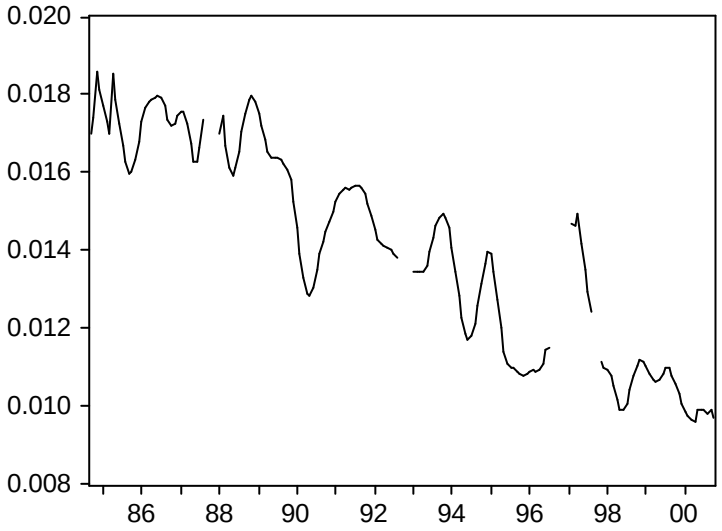


Figure 21: Smoothed seasonally adjusted series for LF_{PNM}/LF_P

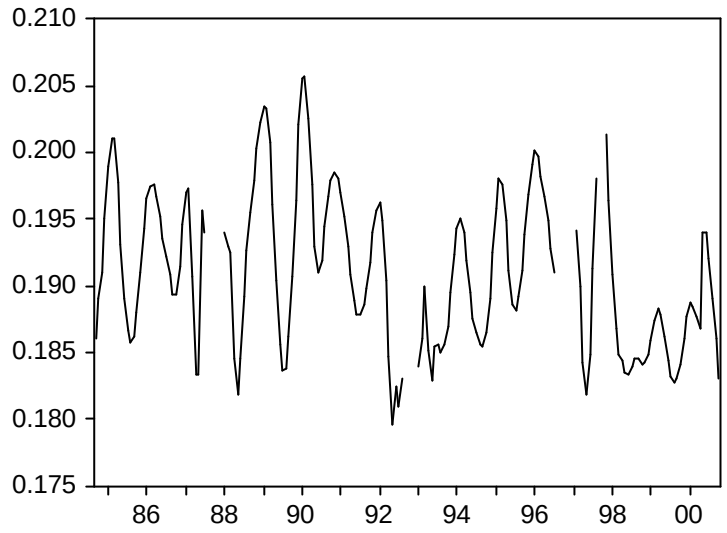


Figure 22: Smoothed seasonally adjusted values of $[(UR_{NP} - UR_P)*(LF_{NP}/LF_T)]$ - dotted line - and $[(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)]$ - solid line.