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 represented 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 around 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 in Australia has been available since early 1980. From time to time researchers (e.g. Foster 1981, Foster and Gregory 1984, Fahrer and Heath 1992, Borland 1996a, Leeves 1997 and Leeves 2000) visit this data with a view to gain-

* Department of Economics, University of Melbourne, Victoria, 3010. I am grateful to Jim Thomson and two referees for very helpful comments. 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.

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

ing 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 direction and source of the bias? The structure of this paper is as follows. Section 2 details 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 unemployment rate for those persons represented in the matched sample as against the rate for the population as a whole.[1] 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 with a discussion of the implications for future research and policy.

2. The LFS and the Matched Sample

The Labour Force Survey (LFS) has been undertaken on a monthly basis since February 1978.[2] 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.[3] In the interviews an attempt is made (inter alia) to

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

establish whether each person is in or out 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 relevant component in the Australian population, 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. The weights are computed in such a way so as to also adjust for any under-enumeration and non-response.

For the LFS, private dwellings (such as houses and flats) and non-private dwellings (non-private dwellings are those that provide a communal or transitory type of accommodation[4] – such as hotels and motels, boarding houses, short-stay caravan parks, hospitals, nursing homes and homes for the aged, educational colleges and boarding schools, and Aboriginal and Torres Strait Islander communities) 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. Since 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), matched sample data can only refer to persons in private dwellings.

As it is not reasonable to retain the same respondents in the survey for a long period of time, one-eighth of the dwellings in the sample are replaced each month.[5] This procedure is known as sample rotation. 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.

We have seen that the rotation procedure used by the ABS is such that seven-eighths of the private dwelling sample from one month is retained for the next month's survey. Persons residing in these dwellings who respond in both months form a 'matched sample' for the later

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

month. The data obtained from the records which can be matched across any two months are expanded up to a 'population figure' which (leaving to one side adjustments for over or under enumeration and for non-response[6]) is equivalent only to the proportion of persons in the sample 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) participated in the Survey in both months (i.e those rotated in or out are not included).

In the next section we look at the size of the 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. In addition, a proportion of the persons in those private dwellings that are included in the sample in successive months cannot be matched. "Normally, those who can be matched represent about 80% of all persons in the survey" (ABS 6203.0, October 2000, p 42).

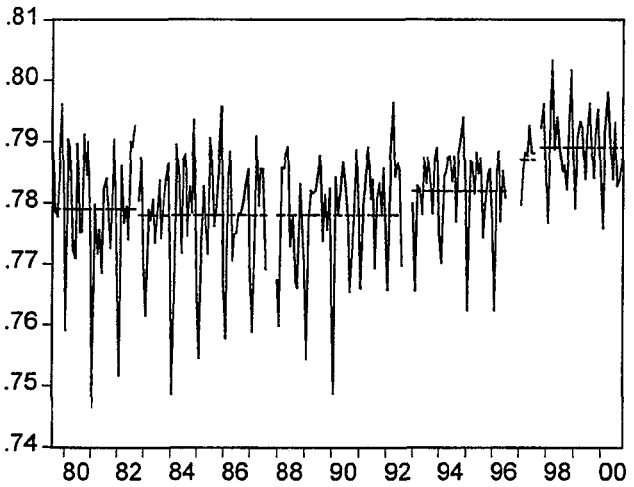
Let *PRMS* denote the size of the "Population Represented by the Matched Sample"[7] and *POP* denote the civilian population aged 15 years and over. We begin by asking: in practice, what proportion of the population is represented by the matched sample? We can establish this proportion (*PRMS/POP*) by 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.[8] Figure 1 shows 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.[9]

The series is quite 'noisy'. A 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 of each of the series 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.[10] The means for the various sub-periods are indicated by the horizontal lines in Figure 1 and are set out in Table 1 below.[11] The information in the Table reinforces the impression obtained by a scan of Figure 1,

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

namely that the proportion of the total population covered by the matched sample rose in the early 90's and rose even further after 1996.

Figure 1: The Proportion of the Population Represented by the Matched Sample (*PRMS/POP*) over the period 1979:08 – 2000:10[(a)] (the horizontal lines are the mean values for each sub-period).



Note: (a) 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 and 1997:09-1997:10) and when telephone interviewing was being phased in (1996:08-1997:01).

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	1997:11-2000:10
Represented	0.779	0.778	0.778	0.782	0.789
Not Rep.	0.221	0.222	0.222	0.218	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.[12] Another reason is the practice of sample rotation, which has the effect that only 7/8 of the residents of private dwellings can potentially be matched across successive months.[13] In addition, non-response

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

by persons in the potentially matchable private dwellings reduces the size of the population represented by the matched sample below its potential maximum.

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. Over the whole of the period 1984:09 – 2000:10 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 3.3% with a maximum of 3.8% and a minimum of 2.4%. Table 2 shows the means of the series for each of our sub-periods. The proportion of persons in the whole population who were 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 were enumerated in non-private dwellings in various sub-periods

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

Looking at Table 2 we can see that most of the fall in the proportion of the population enumerated in non-private dwellings occurred in the period between 1992 and 1993. This reduction is associated with a redesign of the private dwelling part of the LFS following the 1991 Census (the redesign was implemented in late 1992). The most important change which occurred at that time (a change which was reported in ABS Labour Force publications) was the relocation of predominantly long-stay caravan parks into the private dwelling component of the sample (previously – at least since 1981 – both short and long-stay caravan parks were part of the non-private dwellings sample). The change resulted in "an increase in the matched sample" (ABS, 1992, p 3). Some idea of the extent to which this would push up the size of the matched sample may be obtained from the data published in Social Trends (ABS, 1994, p 163) which shows that in the 1991 Census 0.6 of 1% of the population were permanent residents (i.e. persons who said they "usually reside") in caravan parks and marinas.[14] The reduction in the non-private

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

dwelling component of the LFS between 1996 and 1997 also reflects a reclassification of respondents from the non-private to the private dwelling component of the sample. Specifically, "the proportion of the population enumerated in non-private dwellings declined between 1991 and 1996 due to changes in the enumeration procedures and classification of self-care accommodation for retired/aged and manufactured home estates as private dwellings" (McDonald and Majchrzak-Hamilton, 1999, p 37). Some idea of the numbers involved may be obtained from noting that the number of persons enumerated in the non-private dwellings category 'homes for the aged' dropped by 0.1 of 1% of the population between the 1991 census – when aged self-care accommodation was part of the non-private dwelling component – and the 1996 census – when it was not (ibid, p 37).

There are many other factors, apart from the mere reclassification of dwellings which accounts for the fall in the proportion of the LFS enumerated in non-private dwellings over the long term.[15] In particular there has been a reduction in the proportion of the population resident in boarding houses and hotels.[16] Government policies have also had their effect and especially the policy of deinstitutionalisation in relation to human services delivery. Policy changes in this area have involved the closure or downscaling of institutions and have resulted both in shorter stays in institutions and in fewer persons being in institutions at any moment in time. Associated with this has been an increase in the number of aged persons and people with disabilities and illness (psychiatric illness in particular) living on their own or with relatives in private dwellings.

Earlier we noted that the potentially matchable population will be 7/8 of the population resident in private dwellings.[17] Table 3 below shows how the potentially matchable population (*PMP*) as a proportion of the total population (*POP*) has changed over time.[18] Obviously, the increase in *PMP/POP* reflects a fall in the proportion of the population who are enumerated in non-private dwellings. We discussed the reasons for this in the preceding paragraphs.

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 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*).[19] The result is reported in Table 4 below.

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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

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

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	1997:11-2000:10
0.922	0.921	0.922	0.929

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). We now see that there are essentially two reasons for this rise. 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, coincides with the introduction of telephone interviewing at the end of 1996 and may reflect a rise in the LFS response rate associated with this.[20]

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 2 shows (smoothed[21] seasonally adjusted) series for the unemployment rate in the whole population (UR_T) and for the matched sample (UR_{PM}) over the period 1978:09 – 2000:10.[22] The two series are highly correlated, the simple correlation coefficient being (r =) 0.968. However, the mean and median values of

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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).

As previously mentioned, one 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 unemployment rate for the population as a whole are above those for the population represented by the matched sample in every sub-period.

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

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

Figure 2: Smoothed[(b)] seasonally adjusted series for the unemployment rate for all persons (UR_T) – dashed line – c.f. the unemployment rate for the matched sample (UR_{PM}) – solid line – 1979:08 – 2000:10.

Note: (b) In all figures, smoothed values have been obtained by the application to seasonally adjusted data of a 13-term Henderson moving average as described in ABS (1987).

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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 one reason why the labour force characteristics of the matched and total population would differ is if the characteristics of persons in private and non-private dwellings were to differ. A second reason they would differ is 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 (they may not be represented because of sample rotation or because of mobility and/or 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 were enumerated in non-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 who were enumerated in private dwellings (denoted by the subscript *P*). This second group may be further divided into those persons who were in private dwellings and who are represented in the matched sample (these persons will be

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

denoted by the subscript *PM*); and, those who were in 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*).

Figures 3 through 6 show the behaviour over the period 1984:09 – 2000:10 of the unemployment rates for the various groups we have identified.

Figure 3 shows smoothed seasonally adjusted series for the unemployment rate in the whole population (UR_T) and for the matched sample (UR_{PM}). As we noted in our discussion of Table 5 above, the unemployment rate for the matched sample has the same time series profile as the unemployment rate for the whole population but its level is consistently below that of the whole population.

Figure 4 shows smoothed, seasonally adjusted series for the unemployment rate for the population enumerated in private dwellings (UR_P) and in non-private dwellings (UR_{NP}).[23] 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 dwellings series has a higher average and is more volatile than the private dwellings series.

Table 6 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 will differ from that of the whole population is that the unemployment rate of persons in private as compared with non-private dwellings differs markedly.[24]

Table 6. Mean values of UR_{NP} and UR_P in various sub-periods

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

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

Figure 3: Smoothed seasonally adjusted series for the unemployment rate for all persons (UR_T) – dashed line – c.f. the unemployment rate for the matched sample (UR_{PM}) – solid line – 1984:09 – 2000:10.

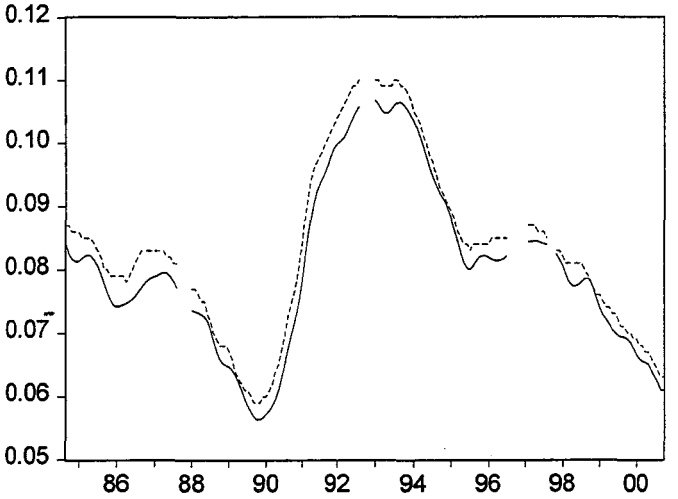
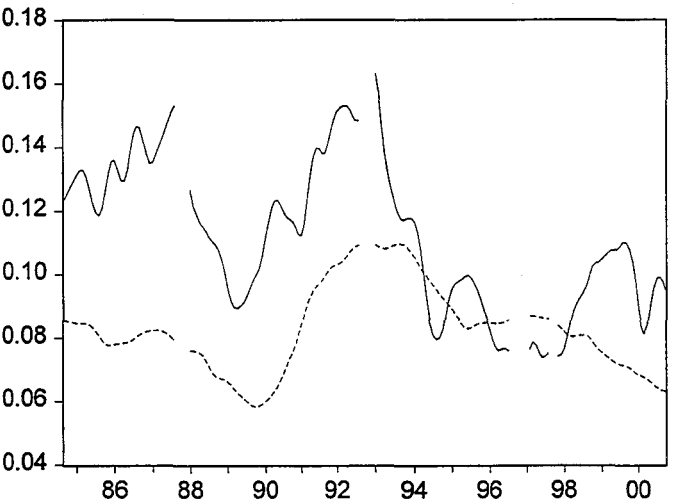


Figure 4: Smoothed seasonally adjusted series for the unemployment rate of persons enumerated in private dwellings (UR_P) – dashed line – as against the unemployment rate for persons enumerated in non-private dwellings (UR_{NP}) – solid line – 1984:09 – 2000:10.



https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

Figure 5: Smoothed seasonally adjusted series for the unemployment rate for persons enumerated in all private dwellings (UR_P) – dashed line – as against the unemployment rate for persons in the matched sample (UR_{PM}) – solid line – 1984:09 – 2000:10.

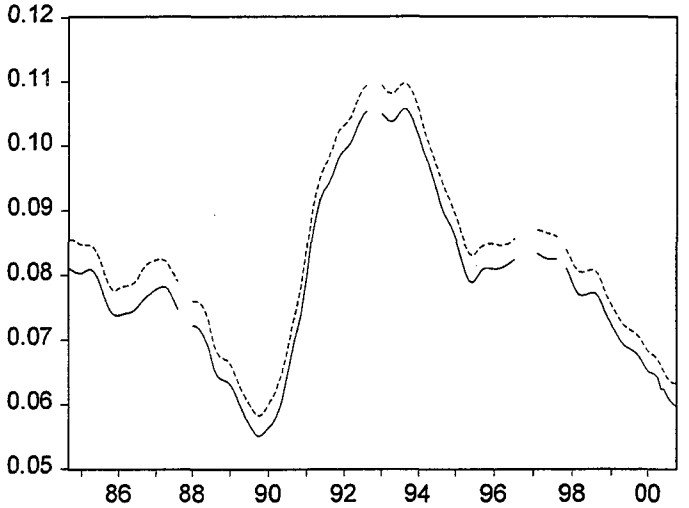
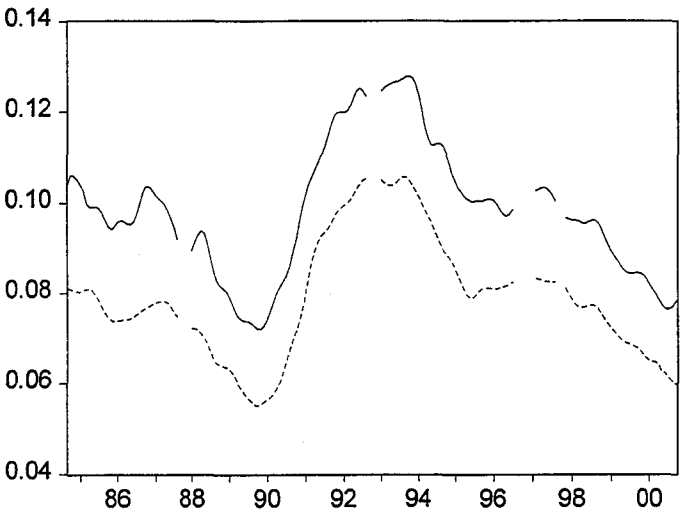


Figure 6: Unemployment rate for unmatched persons in private dwellings (UR_{PNM}) – solid line – as against that for matched persons in private dwellings (UR_{PM}) – dashed line – smoothed seasonally adjusted series: 1984:09 – 2000:10.



https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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 5 shows smoothed seasonally adjusted series for the unemployment rate for all persons enumerated in private dwellings (UR_P) and for those persons in the matched sample (UR_{PM}). 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 all persons in private dwellings (8.3%). Figure 5 indicates that the two unemployment rates have identical time series profiles and that there is a persistent gap between the two rates. This is also brought out in Table 7 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 7. Mean values of UR_P *and* UR_{PM} in various sub-periods

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

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[25] (UR_{PNM}) and to compare this with the unemployment rate for those represented in the matched sample (UR_{PM}). This is done in Figure 6 which shows (smoothed seasonally ad-

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

justed) series for the unemployment rate for the 'unmatched persons' enumerated in private dwellings (solid line) and for those persons in the matched sample (dashed line). 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 (7.9%) is well below that for those persons in private dwellings who were not matched (9.9%).

Table 8. Mean values of UR_{PNM} and UR_{PM} (%) in various sub-periods

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

Table 8 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. First, there is a group who 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 due to sample rotation will itself introduce any systematic bias into labour force estimates.[26] However, a 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 could not be contacted and/or did not cooperate with the interviewers. 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

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

a (much) higher unemployment rate than the respondents. This is not surprising.[27]

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.

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 – i.e.

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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:

$$UR_T = UR_{NP}\frac{LF_{NP}}{LF_T} + UR_{PNM}\frac{LF_{PNM}}{LF_T} + UR_{PM}\frac{LF_{PM}}{LF_T} \qquad (2)$$

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 ($UR_{NP} - UR_P$) and ($UR_{PNM} - UR_{PM}$).[28]

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

$$UR_T = (UR_{NP} - UR_P)\frac{LF_{NP}}{LF_T} + UR_P\frac{LF_{NP}}{LF_T} + UR_{PNM}\frac{LF_{PNM}}{LF_T} + UR_{PM}\frac{LF_{PM}}{LF_T} \qquad (3)$$

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:[29]

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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

We turn now to examine 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 over the whole period. Evaluated at the means over the period 1984:09 – 2000:10 we have: $(UR_T - UR_{PM}) = 0.4\%$, $(UR_{NP} - UR_P) = 3.0\%$ and $(UR_{PNM} - UR_{PM}) = 2.0\%$. 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 $[(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)]$.[30] Clearly the latter is the dominant element.[31]

Table 9 reports mean values of the various unemployment rate difference terms which figure in equation (4).[32] Table 10 gives mean values for the weights on each term.[33] 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 10 with those given in Table 2 above.

Table 9. 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	1997:11-2000:10
$(UR_T - UR_{PM})$	0.4	0.4	0.2	0.2
$(UR_{NP} - UR_P)$	5.4	4.1	0.7	2.1
$(UR_{PNM} - UR_{PM})$	2.0	1.8	1.8	1.5

Both weights seem to be trending downwards.[34] 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).

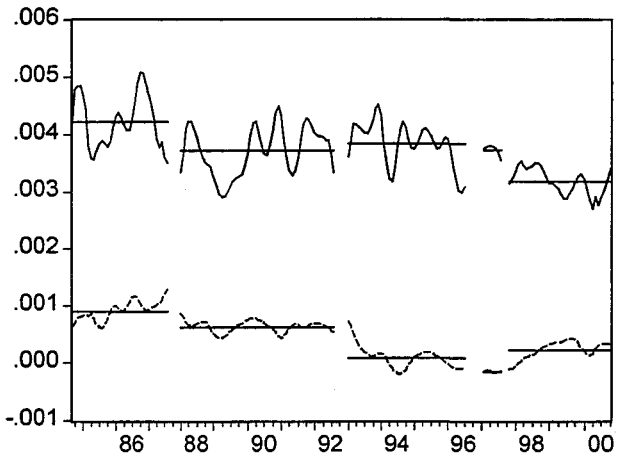
https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

Table 10. 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	1997:11-2000:10
LF_{NP}/L_T	0.017	0.016	0.013	0.010
LF_{PNM}/L_P	0.192	0.194	0.192	0.186

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 11 whilst Figure 7 shows their behaviour over the period 1984:09 – 2000:10. Clearly it is the second term, the one involving ($UR_{PNM} - UR_{PM}$), which is 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.

Figure 7: Smoothed seasonally adjusted values of $[(UR_{NP} - UR_P)*(LF_{NP}/LF_T)]$ – dashed line – and $[(UR_{PNM} - UR_{PM})*(LF_{PNM}/LF_P)]$ – solid line – 1984:09 – 2000:10 (the horizontal lines are the mean values for each sub-period).



https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

Table 11: Means 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[35]

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

7. Concluding Remarks

I begin with a summary of the main conclusions arrived at thus far in relation to the representedness of matched sample data. I then move to highlight the implications of this study for both providers and users of LFS data.

The main findings may be summarised as follows: First, the proportion of the population represented by the matched sample has risen over the period since the introduction of the monthly LFS and especially in recent years. There seem to be two reasons for this. One is that the proportion of the population living in non-private dwellings has fallen. In part this has been due to actions of the ABS involving the reclassification of certain types of dwellings from the non-private to the private component of the sample. A second reason is that the proportion of the potentially matchable population who provide data that can be matched has risen, especially since 1996. This may indicate a rise in the response rate of persons in the sample, particularly following the introduction of telephone interviewing. Second, indices for the labour market characteristics of the matched sample are biased (or 'unrepresentative') in the sense that the population represented by the matched sample is less likely to be unemployed than is the population as a whole. There are two reasons why this is so. One is that the matched sample refers only to persons resident in private dwellings and it would seem that persons who are not enumerated in private dwellings have a higher unemployment rate than those who are. The other reason is that those persons who are resident in private dwellings and who are in the scope of the sample but who are not able to be matched across two 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 (or for some other reason cannot be contacted in the interview

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

period in successive months) have a higher unemployment rate than those who are represented in the matched sample. A third finding is that it is the difference between the unemployment rate for the unmatched persons living in private dwellings and the rate for those persons living in private dwellings 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.

These findings have implications not only for users of LFS data but also for the providers of that data. I begin with the latter.

There are three points which must be made in connection with data collection and dissemination. First, it is important that users of LFS data are informed of changes to the classification dwellings between the non-private and private components of the sample. Information on these matters is important not simply for users of gross flows data but for anyone interested in data pertaining to characteristics of the population residing in private dwellings and/or non-private dwellings. In late 1992 there was a relocation of predominantly long-stay caravan parks into the private dwelling component of the sample from the non-private dwelling component of the sample. In various documents the ABS reported that this had occurred and noted that the change resulted in an increase in the matched sample. In 1997 there was another change involving the movement of self-care accommodation for the retired/aged and dwellings in manufactured home estates form the non-private to the private dwellings component of the LFS. This change was not reported in any LFS documents at the time (or since), despite the self-evident difference in labour market characteristics of at least one of the groups concerned (the retired/aged). Second, we noted that the ABS has a policy of sample rotation and that, in forming matched sample estimates for the population, there is no attempt to expand up the matched records to 'cover' the population equivalent of the size of the group newly rotated into the sample. This is puzzling. Since the 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, it is most unlikely that the mere fact that some members of the private dwelling sample belong to the newly introduced rotation group will itself introduce any systematic bias. Given this, I suggest that the ABS commence the practice of expanding the matched records up to a population number which (putting to one side issues of under or over enumeration or non-response by those in the

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

sample in successive months) is equivalent to the estimated size of the population resident in private dwellings. The third point related to data collection concerns our finding that if there is a 'single' group for whom it would be desirable to have more information it is those included in the scope of the private dwelling component of the LFS but who are non-respondents or who have moved (or for some other reason cannot be contacted in the interview period in successive months). This could be achieved a number of ways but I can think of only two which may be cost-efficient. One way would be to ask retrospective questions of people who have moved into a dwelling included in the private component of the sample between interviews. Provided the responses were reliable, those people could be regarded as providing 'matched records'. Another way (but less cost-efficient one would expect) would be to take steps to follow up and obtain information from those who have moved out of a dwelling included in the private component of the sample between interviews.[36]

Turning now to the implications of our research for labour market researchers (beyond those implicit in the previous paragraph), the most important thing to note is that, taken together, our findings caution against the uncritical use of the gross flows data to analyse the dynamic behaviour of the subset of individuals at highest risk of becoming unemployed. In particular, we should be careful when using transition rates etc derived from matched sample flows as a proxy for transition rates for persons who tend to be resident in non-private dwellings and/or for those persons who reside in private dwellings but for whom a change in their labour market status – and especially a spell of unemployment – is associated with a change of geographic location.

Notes

1 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.

2 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. Since 1992 it has covered approximately one-half of one percent of the population.

3 The ABS reports that "[d]uring the period of implementation [of telephone

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

interviewing], the new method produced different estimates than would have been obtained under the old methodology. However 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.

4 See ABS (2001, 17.11).

5 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.

6 The 'expansions factors' used by the ABS are such that the matched records are automatically expanded up to compensate for non-response in the second month of any pair, but only the second month. See Dixon, Lim and Thomson (2001) for details of this and for comment on the wisdom of this procedure.

7 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. Estimates are provided in each issue of 6203.0.

8 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).

9 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).

10 Averages reported here and in the other Tables refer to seasonally adjusted data because all of the series have clear seasonal components and the sub-periods are not identical in the months they contain.

11 While the Figures display data for the (relatively short) period 1997:02-08, I have not recorded the means for this period in any of the Tables for the simple reason that all of the other sub-periods span 3 – 5 years and it would be misleading to include an extra column where the means referred to a period as short as 6 months.

12 Strictly speaking I should write "the population enumerated in non-private dwellings".

13 Of course there is no reason to expect the characteristics of the new rotation group to differ from those for other rotation groups in private dwellings. We will return to this point later in the paper.

14 These dwellings contain a relatively high number of persons living alone and a relatively high ratio of males to females. Labour force participation is lower and the unemployment rate higher (almost 3 times higher) than in the total population (ABS, 1994, pp 163-6 and ABS, 2000, pp 179-83).

15 Surprisingly, neither of the reclassifications were noted in Borland's survey of labour market flows data (Borland, 1996b).

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

16 Nowadays the most prevalent forms of living in non-private dwellings are homes for the aged and educational institutions whereas for some time a high proportion lived as 'usual residents' of hotels and boarding houses. (King, 2001, p 728).

17 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.

18 The figures in Table 3 are arrived at by subtracting the figures in Table 2 from unity and taking 7/8 of the number that results.

19 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.

20 It is tempting to assert that it "is" or that it "must be" due to "a rise in the LFS response rate" but one cannot conclude that as the numbers in Table 4 reflect the matched responses of households in two successive surveys and, so insofar as survey response is relevant at all, 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 these relationships.

21 All smoothed series (trends) computed in this paper are 13-term Henderson-weighted moving averages calculated using the procedures described in ABS (1987).

22 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 n3 above.

23 Source is 6203.0. Published data for these two portions of the survey are only available since September 1984.

24 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 6 and Figure 4 (more on this later).

25 The unemployment rate for all persons in private dwellings can be thought of as a weighted average of the unemployment rate for the matched component and the unemployment rate for the unmatched component. We have information for the total and also for the matched component and we have information on the relative size of the two components. It is a simple matter then to use that information to recover the unemployment rate for unmatched persons in private dwellings.

26 It is this group that the ABS has in mind when they write in the 'Explanatory Notes' to 6203.0 that "about two-thirds of the unmatched 20% of persons in the survey are likely to have characteristics similar to those in the matched group" (ABS 6203.0, October 2000, p 43). They go on to say that "the characteristics of the other third are likely to be somewhat different."

27 Nor is it a peculiarly Australian phenomenon. It has also been reported to be the case in the US (Flaim and Hogue, 1985, p 11; Barkume and Horvath, 1995, p 30).

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

28 Where UR_P is the unemployment rate for all persons in private dwellings.

29 In an Appendix to this paper I provide a step-by-step derivation of (4).

30 Rounding errors account for the difference between 0.41 and 0.38 + 0.04.

31 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 (see n 21 above).

32 Data in Tables 10 -12 are all in percentages.

33 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).

34 We have discussed possible reasons for this in earlier sections of the paper.

35 Some columns might not add up exactly due to rounding.

36 The use of telephone interviewing might facilitate this.

References

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

Australian Bureau Statistics (1992) *Information Paper: Labour Force Survey Sample Design*, Australian Bureau of Statistics (Cat. No. 6269.0). Canberra.

Australian Bureau Statistics (1994) *Australian Social Trends 1994*, Australian Bureau of Statistics (Cat. No. 4102.0). Canberra.

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

Australian Bureau Statistics (2000) *Australian Social Trends 2000*, Australian Bureau of Statistics (Cat. No. 4102.0). Canberra.

Australian Bureau Statistics (2001) *Labour Statistics: Concepts, Sources and Methods*, Australian Bureau of Statistics (Cat. No. 6102.0). Canberra.

Australian Bureau Statistics, various dates *Labour Force: Australia*, Australian Bureau of Statistics (Cat. No. 6203.0). 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.

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

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

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

Flaim, P. and Hogue, C. (1985) 'Measuring Labor Force Flows', *Monthly Labor Review*, 108(7), pp. 7-17.

https://doi.org/10.1111/1035304601812002216 Published online by Cambridge University Press

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

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.

King, A. (2001) 'The Australian Housing Stock: 1911 and 1996', in *2001 Year Book Australia*, ABS, Canberra.

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

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

McDonald, S. and Majchrzak-Hamilton, G. (1999) '1996 Census Data Quality: Housing', Census Working Paper 99/3, ABS, Canberra.

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 $UR_{PM}\left(LF_{NP}/LF_T\right)$ and $UR_{PM}\left(LF_{PNM}/LF_T\right)$ from the RHS of the above gives, after some re-arranging:

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

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

Since $LF_{NP} + LF_{PNM} + LF_{PM} = LF_T$, the expression in the brackets on the second line of the above will equal unity. We can then take UR_{PM} over to the LHS to give an expression for the difference between the unemployment rate for all persons and the unemployment rate for persons in the matched sample:

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

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

Note that this may be written as

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

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 middle term on the RHS of the above the above and rearranging gives:

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

Since $LF_{PNM}/LF_T = (LF_{PNM}/LF_P)(LF_P/LF_T)$, the second and third terms in the above may be rewritten so that the equation becomes

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

This may simplified if we recall that $LF_{NP} + LF_P = LF_T$ (and so the expression in brackets at the end of the first line of the above will simply

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press

equal unity) and we note that all of the elements in brackets in the second line have a common denominator, so we can rewrite the above as:

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

Since $LF_P - LF_{PM} = LF_{PNM}$, we may write

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

and given that $LF_{PNM} LF_{NP} + LF_{PNM} LF_P = LF_{PNM} LF_T$, the term in brackets on the second line of the above will equal unity, in which event the above becomes

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

This is equation (4) in the text.

https://doi.org/10.1177/103530460101200210 Published online by Cambridge University Press