AUSTRALIAN JOURNAL OF LABOUR ECONOMICS
Volume 17 • Number 2 • 2014 • pp 139 - 161

Experimental Estimates of Indigenous Employment from Administrative Data*

Kyle Thomson and Cezary A. Kapuscinski, Australian Government Department of Employment

Abstract

In this paper we test a 'proof of concept' to remedy the lack of timely data on Indigenous labour market outcomes. We utilize Centrelink administrative data and the HILDA survey data to estimate a forecasting equation of the employment to population ratio of Indigenous Australians. In doing so we examine the probabilities of Income Support Recipients and the general population transiting from one labour market state to another. We conclude the paper by discussing the experimental time-series estimates of employment to population ratio of Indigenous Australians derived from our forecasting equation.

Keywords: Indigenous employment, Indigenous disadvantage, Closing the Gap, HILDA data, Transition probabilities between labour states.

JEL Classification: J15, J21, C22, C55

Address for correspondence: Cezary Kapuscinski, Strategy and Economic Branch, Department of Education, GPO Box 9880, Canberra ACT 2601. Email: Cezary.Kapuscinski@education.gov.au

Acknowledgements: This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper reflect the authors' views and do not necessarily represent those of the Department of Employment, Department of Education or the Australian Government and should not be attributed to either the Department of Social Services or the Melbourne Institute. The authors would like to acknowledge the assistance of Michael Thomson and Dr Kabir Chowdhury in working with the Research and Evaluation Dataset. The authors are also indebted to Dr Greg Connolly for initial discussions and suggestions. The authors also express their appreciation for comments and suggestions provided by participants of a Department of Employment seminar. The authors would like to acknowledge comments provided by the workshop participants at the ALMRW in February 2014. We also thank two anonymous referees of this Journal for invaluable comments and suggestions. All errors and omissions, however, remain the responsibility of the authors.

*An earlier version of this paper was presented at the 2013/14 ALMR Workshop at the University of Melbourne in February 2014.

1. Introduction

The defining characteristic of any attempt to measure employment outcomes for Indigenous Australians is the lack of data which is both timely and accurate. This restricts policy responsiveness and makes effective evaluation extremely difficult. The responsiveness and evaluation of even the most ambitious and far reaching Indigenous policy agenda, the Closing the Gap framework, as set out in the National Indigenous Reform Agreement, is no exception.

Agreed to by all Australian State and Territory governments and the Australian Federal government through the Council of Australian Governments in 2008, the agreement outlines six targets with respect to the life expectancy, child mortality, education and employment outcomes of Indigenous Australians. The employment target seeks to halve the gap, as measured in 2008, between the employment to population ratio of Indigenous and non-Indigenous Australians by 2018. At the time the Indigenous employment to population ratio was estimated to be 53.8 per cent (Australian Bureau of Statistics, 2008 A), compared with 75.0 per cent for non-Indigenous Australians (Australian Bureau of Statistics, 2008 B), a gap of 21.2 percentage points. The latest estimate from 2012-13 of the Indigenous employment to population ratio of 47.8 per cent (Australian Bureau of Statistics, 2012-13) compares with 75.6 for non-Indigenous Australians (Australian Bureau of Statistics, 2012), a wider gap of 27.8 percentage points.

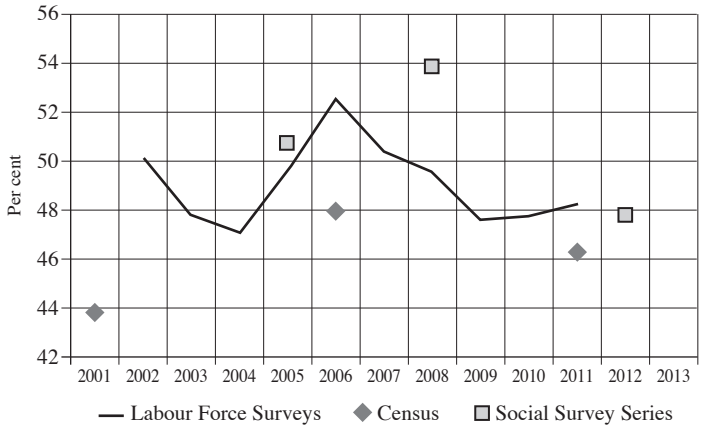
We can assess our progress towards this target only at infrequent intervals and invariably with a long time delay. From 1981, information on Indigenous labour force status has been available from the Census, however, this is only available every five years and usually around 18 months after the Census date. Estimates are also available from the series of National Aboriginal and Torres Strait Islander surveys, beginning in 1994 with a survey of this name and followed alternately by the National Aboriginal and Torres Strait Islander Social Survey (2002, 2008) and the National Aboriginal and Torres Strait Islander Health Survey (2004-05, 2012-2013). These estimates form the basis for the Closing the Gap employment target and provide accurate estimates of Indigenous labour market outcomes every three years, with future surveys planned for 2017 and 2020.

More timely information is available from the ABS publication *Labour Force Characteristics of Aboriginal and Torres Strait Islander Australians, Estimates from the Labour Force Survey* which has been published annually from 2004 (until 2011) by pooling the results of the monthly Labour Force Survey. While results are usually published annually, the estimates contained in this publication are considerably less reliable than either the Census or the Indigenous specific survey series. There is, however, also a long delay between the end of the survey year and publication, as well as difficulty obtaining historical estimates for the 15 to 64 year old age range.

Where these three existing data sources fall short is, firstly, the timeliness of the data release, which usually occurs long after the data collection in the field has taken place, and, secondly, a time series with observations at regular and sufficiently frequent intervals suitable for detailed analysis. The chart below (see figure 1) collects together the time series and cross-sectional data of the employment to population ratio for Indigenous 15 to 64 year olds available from the Australian Bureau of Statistics

since 2001. As the original data collections serve different purposes they have not been matched and analysed together before (apart from an occasional summary table 'Other sources of information' provided by the ABS, as in the ABS 2011, table 8).

Figure 1 - Estimates of the Indigenous employment to population ratio (15 to 64 year olds, per cent)



Sources: ABS (2001, 2006), Census of Population and Housing, Cat. No. 2037.0.55.001; ABS (2011), Census of Population and Housing, Cat. No. 2072.0; ABS, Labour Force Characteristics of Aboriginal and Torres Strait Islander Australians, Estimates from the Labour Force Survey, Cat. No. 6287.0, 2011; ABS (2004-05, 2012-13), National Aboriginal and Torres Strait Islander Health Survey, Cat. No. 4715.0; ABS (2008), National Aboriginal and Torres Strait Islander Social Survey, Cat. No. 4714.0.

In this paper we attempt to construct a model for generating a timely estimate of the employment to population ratio for Indigenous 15 to 64 year olds, using Centrelink data on Income Support Recipients and some widely available macro-economic indicators as explanatory variables. While this exercise does generate estimates for the Indigenous employment to population ratio, at this stage it is intended as a proof of concept, showing that the available administrative data can contribute to a robust estimate of Indigenous labour market outcomes, rather than the creation of definitive estimates. To do this we have used time series information constructed from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The probabilities of Income Support Recipients transiting from one labour force state to another, along with several broader macroeconomic variables, are then used to explain movements in the derived employment to population ratio.

The rest of the paper has been broken up into five sections. Section 2 deals with the method used to extract quarterly time series data from HILDA and the administrative data. The following section compares the transition probabilities of Indigenous and non-Indigenous Income Support Recipients. Section 4 examines the information content of transition probabilities of Indigenous Income Support Recipients in relation to the general population. The fifth section presents an experimental model estimating the Indigenous employment to population ratio. We also compare these

estimates with the Indigenous employment to population ratio from the HILDA survey and the total 15 to 64 year old ratio from the Labour Force survey. Finally, the paper concludes with some remarks on possible directions for future research.

2. Data

HILDA and extraction of time-series data

The HILDA Survey is a household-based panel study which began in 2001 with 7,682 households and 19,914 individuals from non-very remote Australia.[1] The population in wave 11 was topped up with an additional 2,153 households and 5,477 individuals (Melbourne Institute of Applied Economics, 2013). The calendar is collected only from responding persons from the original sample for their own activities and covers all months from 1 July of the preceding year to the date of the current interview (covering between 14 and 18 months) (Watson, 2009). The calendar divides each month into three periods, 'early', 'mid' and 'late' and an activity (such as education, employment, etc.) is recorded for each period.

A consequence of the varying interview dates is a 'ragged edge' to the data with respect to the end date of the calendar in a given wave, and an uneven seam where persons have been interviewed in successive waves. This means that while some information is available for the early months of 2013, the response rate drops off rapidly from late 2012 and so the period of analysis in this paper ends at the third quarter of 2012. As transitions from one labour force state to another can only be derived by knowing the labour force state in two consecutive periods, the data used in this paper is restricted to begin in the first quarter of 2001. For the sake of simplicity and to minimise recall bias, where an overlap exists the response which is closest to an interview date has been used. A full discussion of the effect of overlapping seams in the HILDA calendar can be found in Watson, 2009. The restriction of the calendar data to only responding persons in conjunction with the survey non-response leaves this research with a small in scope Indigenous sample ranging from 221 to 487 persons, which trends upwards over time.

As the respondent numbers are so low and 'very remote Australia' is excluded from the sample, the number of Indigenous participants in Community Development Employment Projects (CDEP) identified in HILDA is near non-existent in the first wave and non-existent thereafter. Thus, including or attempting to exclude participation in these programs as a form of employment in the HILDA survey has no discernable effect on the proportion in employment. Where those who were identified as being CDEP participants in the first wave have recorded employment as a calendar entry they have been by default included in the employed category, as the calendar entry may refer to non-CDEP employment. Due to the negligible effect of CDEP on the employment to population ratio in the first wave and the absence of identified CDEP participants in following waves, the employment to population ratio derived from HILDA and used in this paper is best viewed as not including CDEP participation as employment.[2]

[1] Watson and Wooden, 2002a have described the sample for HILDA as excluding 'people living in remote and sparsely populated areas'. The current terminology refers to these areas as 'very remote Australia'. They suggest that this exclusion resulted in a loss of 80,000 people (both Indigenous and non-Indigenous) from the reference population.

[2] Gray, Hunter and Lohoar, 2012 discuss the reasons why Indigenous employment analysis is more appropriate with data excluding CDEP figures.

The calendar's uneven seam creates a seasonal variation in the number of transitions, which is compounded both by the concentration of interviews around certain dates and the inability to assign a transition at the first period of a wave unless the person responded in the previous wave and the calendars adjoin or overlap. Although recall bias in the calendar is not random with respect to an individual's characteristics (Goode, 2007), no attempt has been made to correct for this in this paper. However, seasonal dummies have been included in the regression analysis which go some way to correcting for the seasonal effect of the calendar's uneven seams.

The Indigenous subsample derived from the HILDA calendar has been re-weighted by sex, age and educational attainment to benchmarks derived from the 2001, 2006 and 2011 Census of Population and Housing unit record files with reference to the entire enumerated population. In between Censuses, interpolation has been used to create benchmarks. The small sample size of the Indigenous population in HILDA restricts the number of characteristics which can reasonably be used in re-weighting without having an unacceptable number of missing cells. The small number of characteristics has been chosen in order to have less than one percent missing cells while still capturing the major factors affecting labour market outcomes. The weights have been applied to the aggregated HILDA calendar observations for each quarter. We have not rescaled the weights to match either the employment to population ratio in the Census or the Labour Force survey as both count some people participating in CDEP in the employed category. This re-weighting, however, makes it possible to treat this employment to population series as referring to the whole of Australia and rather than just the not very remote population.

The cross sectional weights included in HILDA (and derived from the Census) have been used to better approximate the wider Australian non-Indigenous population and to help compensate for biased non-response and attrition (See, Watson, 2012 for a full discussion of the included HILDA weights). Non-response to only the calendar part of the survey is so small as to be insignificant to the analysis (Watson, 2009).

Where only one activity has been recorded in any period the responses have been recoded into the mutually exclusive and collectively exhaustive single labour market states of either employed, unemployed or not in the labour force (NILF). In general more than one state is recorded in a period when there is a transition from one state to another, as transitions usually occur within periods rather than exactly at end of one period and the beginning of another.[3] For this reason, where more than one activity has been reported in a given period, where it exists, the status of the period immediately prior is used.

A consequence of this coding for multiple states is that where multiple states are recorded in more than one consecutive period both the transitions between and within these periods are concealed. The number of transitions is thus underestimated, and the underestimate is greater the more frequently a person moves between states.

[3] For example, a sequences where a person is unemployed in periods 1 through 9 (three months), multiple states are recorded in period 10, and employed in periods 11 to 19 (a further three months) is more common than a sequence where a person is unemployed in periods 1 to 9 and employed in periods 10 to 19.

Administrative data set and time series extraction

The Research and Evaluation Dataset (RED) is a confidentialised (i.e. stripped of any personally identifying information) administrative dataset constructed from the records of Income Support Recipients. It is episodic in nature, in that it contains interactions individuals have had with Centrelink from 1997 onwards. In order to make it resemble the HILDA calendar as closely as possible, the dataset has been sliced into three periods corresponding to the 1st, 10th and 20th days of the month. On the 1st of August 2001 (the first date included in this study) there were 2,609,000 non-Indigenous and 142,000 Indigenous in scope person recorded. In the first quarter of 2013 the end of the HILDA sample there were 2,376,000 non-Indigenous and 197,000 Indigenous in scope persons recorded in this dataset.[4]

Each Income Support Recipient has been assigned a mutually exclusive and collectively exhaustive single labour market state for each period of either employed, unemployed or NILF. The employed status is derived from reported earnings data. Income Support Recipients initially report routine earnings and are required to report any non-routine earnings at the end of each fortnight along with changes to routine earnings. This gives the earnings data an inherent fortnightly periodicity. Where the Income Support Recipient earned more than the national minimum wage in the fortnight from which the slice has been taken they have been assigned a labour market status of employed.[5] The inherent periodicity of the data makes it possible that two slices will fall within the same fortnight and thus reflect the same fortnightly earnings. Deriving employment status by reported earning means that work for a Community Development Employment Project has not been counted as employment.

Where a person is not employed but is required to look for work as a condition of receiving social security benefits, they have been assigned the labour market status of unemployed. All those who are neither employed nor required to look for work are assigned a labour market status of NILF. These definitions rest on the assumption that those who are compelled to look for employment will actually actively look for a job and that all those who are not compelled to look for work will not do so.[6]

[4] The social security payments cover transfers such as the Newstart allowance for the unemployed, family supplement, or the aged pension. The numbers of social security recipients can therefore vary not just with the business cycle but also with other social or biological forces, e.g. the ageing of the population or the changing fertility. For the Indigenous population, there may be additional variability as a result of increasing propensity to self-identify as an Indigenous person especially after the Sorry Speech by the Prime Minister in 2008 (Rudd, 2008). Note, however, that both HILDA and RED have the procedures for correcting personal information backwards to minimize the issues of self-identification. In addition, our focus on ratios does, to a certain degree, mitigate the problem of this non-biological growth rate of the Indigenous population, but it still remains a conjecture that the newly self-identified Indigenous people display similar labour force characteristics as their peers. Future work with the Longitudinal Census may throw some light on this issue.

[5] There is a long-standing discourse in the literature about the differences between the measures of labour force states as implemented by the ABS and as provided by various government agencies administering social security payments (see, for example, two studies 20 years apart: Junankar and Kapuscinski, 1990, and ABS, 2009). Our approach does not resolve these differences but it allows us to derive series that are conceptually closely related, although some gaps remain (e.g. people working without pay in a family business).

[6] This assumption as supported by the evidence presented in section 3 regarding the differences in employment probabilities of these categories.

Definitions and notation

Probabilities provide a number of advantages over raw number measures. The normalized ratios of the flows between states reduce the effect of the difference in the absolute size of each stock in the two data sets. Probabilities are also expected to have less of a lag, especially for flows from employment to other states. There is also evidence from our initial investigation of the relationship between the RED and the HILDA sample that the characteristics of those in income support who transit between states are more similar to the general population than the average stock of persons in income support.[7]

A transition is reported where the state of the person in the previous period is different to the state of the person in the current period. If the state in either period is unknown then no transition is recorded. The data has been aggregated from thirds of a month to quarterly data as a form of smoothing. To arrive at the quarterly figure the number of transitions is summed from all periods in a quarter to a total for that quarter.

The probabilities analysed in this paper are the average conditional probabilities of a person transiting from state A in time t-1 to state B in time t calculated as:

$$\frac{\text{Transitions from state A to state B in quarter t'}}{\text{All transitions from state A in quarter t'}}$$

This can be interpreted as average probability of the (hypothetical) average person transiting from state A to state B in any third of a month period within the quarter.

In subsequent discussions, the suffix Y is used to denote HILDA data, the suffix X is used to denote RED and the suffix Z refers to other data sources. The suffix A is used where the data refers to Indigenous persons. Notation for the transition probabilities is outlined in table 1.

Table 1 - Key to probability notation

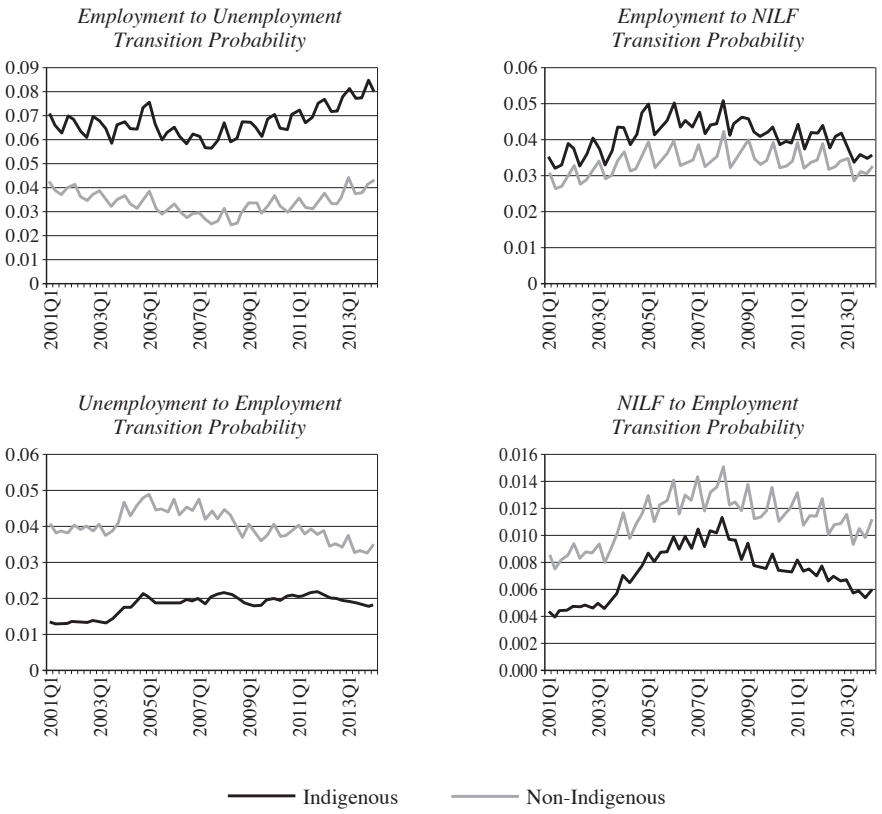
State at period t-1	*State at period t*	*Transition at period t*	*Notation*
Employed	Employed	Employed to employed	EE
Employed	Unemployed	Employed to unemployed	EU
Employed	NILF	Employed to NILF	EN
Unemployed	Employed	Unemployed to employed	UE
Unemployed	Unemployed	Unemployed to unemployed	UU
Unemployed	NILF	Unemployed to NILF	UN
NILF	Employed	NILF to employed	NE
NILF	Unemployed	NILF to unemployed	NU
NILF	NILF	NILF to NILF	NN

[7] Also see Budd, Levine and Smith (1987), Chapman, Junankar, Kapuscinski (1992).

3. Transition probabilities of Income Support Recipients by Indigenous status

Having derived the transition probabilities for Income Support Recipients it is of interest to investigate their behaviour over time. The primary issue is the comparison of these transition probabilities for the Indigenous and non-Indigenous populations. Overall there appears to be a broadly similar pattern in the transition probabilities involving the employment state for Indigenous and non-Indigenous Income Support Recipients (see figure 2).

Figure 2 - Selected empirical transition rates for the Indigenous and non-Indigenous Income Support Recipients



Source: Authors' calculations based on the RED.

Among both Indigenous and non-Indigenous Income Support Recipients, the employment to unemployment probability steadily decreased from the early 2000s up until the onset of the Global Financial Crisis (GFC), after which it has trended upwards. The employment to unemployment probability for the Indigenous Income

Support Recipients has remained consistently above the level for non-Indigenous Income Support Recipients and this gap appears to be widening slightly. However, this does not provide substantial evidence that in the most recent downturn Indigenous Income Support Recipients were laid off at a disproportionally greater rate than non-Indigenous Income Support Recipients, given their initial higher probability of moving from employment to unemployment.

The probability of moving from employment to NILF follows a similar pattern for both Indigenous and non-Indigenous Income Support Recipients; however, in this case both appear to rise in the period leading up to 2005 before becoming steadier and then gradually declining. Due to the definition of NILF being those not required to look for work as a condition of receiving their payment, rather than those who choose not to look for work, this probability more closely reflects income support policy than labour market conditions, as those who stop working tend to move to unemployment rather than to NILF as activity requirements have become more stringent.

Both the Indigenous and non-Indigenous Income Support Recipients experienced an increase in the unemployment to employment probability around 2004 and a decline after 2008. The non-Indigenous unemployment to employment probability has fallen faster and this has caused some level of convergence, although the unemployment to employment probability for Indigenous Income Support Recipients remains significantly lower than for the non-Indigenous. Unsurprisingly, the NILF to employment probabilities show a similar pattern to the unemployment to employment probabilities, all be it on a much smaller scale.

These probabilities show that both the employment to unemployment probability and the unemployment to employment probability react to changing labour market conditions and influence the volume of stocks in each state. The counter cyclical pattern in the employment to unemployment transition probability for Income Support Recipients is in contrast to much recent research on transition probabilities, which generally find that the unemployment to employment transition probability is fairly stable over the business cycle.

On the one hand this may be due to aggregation bias. The employment to unemployment transition probability may appear counter cyclical due to the increased likelihood of observing a person who is unemployed as labour market conditions become more difficult (Shimmer, 2012). Certainly no attempt has been made to correct for aggregation bias in the transition probabilities presented in this paper. However, if we compare the employment to unemployment transition probability of Income Support Recipients presented here to those derived from gross flows data, it seems unlikely that this strong counter cyclical trend is purely related to aggregation bias.

Despite the longer observation period in the monthly gross flows data derived from the ABS Labour Force survey, the similarly uncorrected unemployment to employment transition probabilities reported in Chindamo and Uren (2010), Chindamo (2010) do not show a counter cyclical pattern. The observation period for RED is considerably shorter (roughly a third of a month), and so should exhibit less aggregation bias. Despite this, the employment to unemployment transition probability clearly appears to increase as labour market conditions deteriorate. This leads to the conclusion that while some aggregation bias may exist, there is a counter

cyclical pattern in the employment to unemployment transition probability among Income Support Recipients which is not observed in the wider Australian population as represented by gross flows data. Rather, this and other differences in transition probabilities are more likely due to differences in the average characteristics of RED population compared with the population captured by the gross flows data.

Heterogeneity and state dependence

Within a given labour force state, heterogeneity with respect to characteristics that impact on the probability of transiting from one labour force state to another cause the average transition probabilities to vary depending of the composition of the group at a given time. The heterogeneity hypothesis is based on two observations. First that peoples' characteristics differ and, second, that some people have characteristics which make them more likely to be in one labour force state rather than another. Certainly there are characteristics which have been found to affect transitions from one state to anther (see, for example, Chapman and Smith (1993); Stromback, Dockery and Ying (1998) and Cai, Ghantous and Wilkins (2006)).

The second observation is that those with certain characteristics, for example, employability skills such as good communication or reliability, have a higher probability of transiting to employment, and that after such a transfer the average level of these characteristics remaining in the pool of those not employed is then less than before. In downturns we then expect the average likelihood of becoming employed of those not in the labour force or unemployed to increase as people who have characteristics which make them relatively more likely to be employed are either separated from employment or are unable to transition to employment. Conversely we expect that during upturns those who have characteristics which make them more likely to be employed will more quickly become employed, causing the average likelihood of becoming employed of those left in the pool of not employed to fall.

Varying levels of labour force attachment is a particularly important form of heterogeneity, especially with respect to the transition probabilities of Indigenous Australians (Taylor and Hunter, 1998; Hunter and Gray, 2001). For this reason it is worth briefly exploring attachment and the limitations of the RED as presented in this paper. More than usual caution is needed when comparing transitions derived from responses to the ABS Labour Force survey or National Aboriginal and Torres Strait Islander Survey, and those presented here for Income Support Recipients as there are differences in the way in which unemployment and not in the labour force status are defined.

In the RED some people who may not want a job are defined as unemployed because they are compelled to search for one. In other words, their search activity just satisfies the activity test but probably does not constitute a purposeful search for employment. In such cases there is no guarantee that they are in fact actively searching, and so would be defined, under the ABS classification, as not in the labour force. In contrast, people who want a job might not be required to search for one, and are so classified as not in the labour force, even though they may in fact be searching for one (as shown by the positive not in the labour force to employed transition probability), but would be classified as unemployed in the ABS Labour Force survey. This makes comparison with traditional marginal attachment literature along the lines of Jones

and Riddle (1999), where the marginally attached are part of the measured not in the labour force category, problematic.

This 'blurring' will likely lower the average attachment of the pool of people described as unemployed in RED compared with the average of those described as unemployed by the ABS Labour Force survey. The consistently lower probability of unemployed to employed transitions in the RED relative to the ABS Labour Force survey's gross flows data is consistent with this hypothesis, although again caution must be used due to the incompleteness of the transitions captured in the gross flows data.

The effect on those described as not in the labour force is more ambiguous, as, on the one hand, there are likely to be fewer people available to work since they would likely be compelled to look for work and thus be defined as unemployed, making this group on average less attached. On the other hand, there may be people who are not compelled to work but who both want a job and are actively looking for a job, making this group on average more attached. The fact that the NILF to employed transition probability is lower by an order of magnitude lends weight to the dominance of the effect causing lower attachment, although again the incompleteness of the transitions captured in the gross flows data suggests a need for caution in such a comparison.

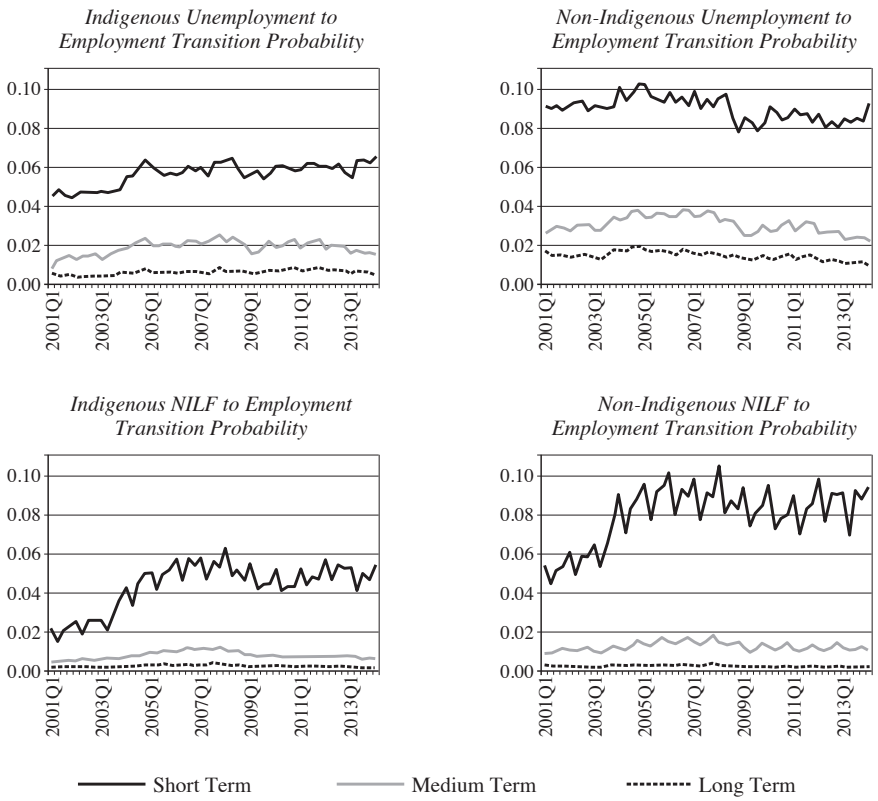
While differences in attachment may explain the lower probability of entering employment by Income Support Recipients, regardless of their initial state, it appears less credible as an explanation of why Income Support Recipients have a much higher probability of leaving employment which is also sensitive to the state of the labour market. This would be the case if those who choose to leave their jobs because they no longer want a job make up a sufficiently large proportion of those leaving employment in the RED. This is due to their being classified as unemployed, rather than not in the labour force as they would be in, for example, the ABS Labour Force survey. However, the effect of this difference in classification does not seem likely to be the sole cause of the difference, as those who transit from employed to unemployed were attached enough to the labour force to be employed in the first place.

The difference is more likely due to the fact that there is no employed but away from work category in the RED. If a person does not earn more than the hourly minimum wage in any Centrelink reporting period corresponding the one third of a month observation period, they transit from employed to unemployed. This trend might then be a reflection of an increasing number of spells where Income Support Recipients do not work any hours in a two week period as the labour market deteriorates.

Various studies have also observed state dependence, that is, that the time spent in a given state reduces the probability of changing state, especially with respect to unemployment, on the basis of skill atrophy or stigma (see, for example, Brooks and Volker (1986); Stromback, Dockery and Ying (1998) and Black, Tseng and Wilkins (2005)).

In practice the relative importance of heterogeneity and state dependence on transition probabilities are not readily separated, as those who have characteristics which make them less likely to become employed also spend longer in states other than employment. For the purposes of modelling the employment to population ratio, if variables are included to simulate the overall labour market along with transition probabilities, it is necessary to take these two effects into account. The existence and magnitude of these two effects is shown below by the differences in transition probabilities depending on duration in a given state (see figure 3).

Figure 3 - Transition probabilities to employment for the Indigenous and non-Indigenous Income Support Recipients by duration in a non-employment state



Source: Authors' calculations based on the RED.

4. Income Support Recipients and the general population

The next step involves evaluating the similarity (or, technically, the degree of correlation) between the transition probabilities derived for the Income Support Recipients and the employment to population ratio derived from HILDA for both the Indigenous and non-Indigenous populations. These regressions (see tables 2 and 3) test the ability of administrative data to explain (or correlate with) the employment outcomes of the general population. In other words, they provide an important empirical test of the usefulness of administrative data in explaining the variations over time of employment outcomes of the general population (i.e., people both inside and outside the income support system). This type of test has a long tradition in economics going back at least half a century to the work done at the National Bureau of Economic Research (Mincer and Zarnowitz, 1969).

Tables 2 and 3 report the results of regressing the employment to population ratios for the general non-Indigenous and Indigenous populations, both derived from HILDA, on a subset of transition probabilities derived from RED. Both of these regressions have been estimated on the sample from the first quarter of 2001 to the third quarter of 2012.

Table 2 - Regression of the non-Indigenous employment to population ratio on RED transition probabilities

Dependent variable: EPRY			
	coefficient	*t-ratio*	*p-value*
PNEX	9.83E+00	10.732	0.000
PNUX	7.54E+00	1.319	0.195
PUEX	-2.22E+00	-6.137	0.000
PUNX	-8.75E-01	-0.676	0.503
NE3MX	-2.38E-01	-1.773	0.084
Q1	9.76E-04	0.151	0.881
Q2	1.49E-02	2.449	0.019
Q3	7.39E-03	1.592	0.120
CONSTANT	9.02E-01	7.195	0.000
Diagnostic	*value*		
Adj. R-square	0.926		
SEE	4.97E-03		
Std. dev. (regressand)	1.82E-02		
LM(1)	2.296		
DW	1.315		
DW - p-value	0.003		
ACF: lag 1	2.236		
ACF: lag 2	1.121		
ACF: lag 3	-0.960		
ACF: lag 4	0.675		
Harvey test	1.398		
RESET (2)	1.654		

Source: Authors' calculations from the HILDA survey and RED.

Notes:
1. Adj. R-square is the adjusted regression coefficient of determination. SEE is the standard error of the regression, while Std.dev. (regressand) is the standard deviation of the dependent variable (EPRY).
2. LM(1) is the LM test for serial correlation of order one. The critical value at 1 per cent level is 2.57.
3. DW is the Durbin-Watson test for serial correlation. The DW-p-value is the calculated marginal significance level of the Durbin-Watson statistic.
4. ACF is the residual autocorrelation function. The entries in the table are the t-ratios of the first four coefficients of the estimated ACF. The approximate critical value at 1 per cent level is 2.71.
5. Harvey test is a test for heteroscedasticity in the residuals. The approximate critical value at 1 per cent level and 8 degrees of freedom is 20.09.
6. Reset(2) is a test for regression misspecification. The approximate critical value at 1 per cent level is 7.37.

From table 2 we can see that the combination of just a few variables derived from RED (two probabilities of transition from unemployment, two probabilities of transition from NILF and the proportion of all unemployed and NILF who have not been employed in the past 3 months – NE3MX explains a majority (over 90 per cent) of variation in the non-Indigenous employment to population ratio.[8] From a statistical point of view, the estimated model provides an adequate specification with no evidence of major problems, although there is some evidence of first order serial correlation at the five per cent significance level. The estimated specification yields a standard error of the regression which is an order of magnitude smaller than the standard deviation of the employment-population ratio over the sample period. In other words, this simple specification significantly explains a substantial proportion of the behaviour over time of the non-Indigenous employment to population ratio.

Table 3 - Regression of the Indigenous employment to population ratio on RED transition probabilities

Dependent variable: EPRWYA			
	coefficient	*t-ratio*	*p-value*
PNEXA	2.96E+01	3.939	0.000
PNUXA	-3.18E+01	-2.379	0.023
PUEXA	-6.13E+00	-2.830	0.007
PUNXA	6.25E-01	0.144	0.886
NE3MXA	4.51E-01	0.399	0.692
Q1	-3.91E-02	-1.502	0.141
Q2	-2.28E-02	-1.511	0.139
Q3	-3.14E-02	-1.996	0.053
CONSTANT	1.30E-01	0.121	0.904
Diagnostic	*Value*		
Adj. R-square	0.837		
SEE	1.66E-02		
Std. dev. (regressand)	4.10E-02		
LM(1)	2.251		
DW	1.375		
DW - p-value	0.006		
ACF: lag 1	2.131		
ACF: lag 2	0.202		
ACF: lag 3	-2.614		
ACF: lag 4	-2.184		
Harvey test	7.906		
RESET (2)	0.714		

Source: Authors' calculations from the HILDA survey and RED.
Notes: 1. The dependent variable (EPRWYA) has been derived from the reweighted sample, as described in section 3.
2. Additional explanatory notes are given below table 2.

[8] The ratios of NE3MX and NE3MXA are included as proxies for the changing average characteristics of the stock of Income Support Recipients.

Table 3 reports the results of a similar regression estimated for the Indigenous employment to population ratio. While preliminary data explorations showed that the non-Indigenous general population and its subset of Income Support Recipients are more closely correlated than the equivalent Indigenous populations, the regression in table 3 suggests that the Indigenous Income Support Recipient population and the general Indigenous population have enough common characteristics to allow good statistical link between the behaviour of variables derived for the Income Support Recipient population and the employment-population ratio derived from HILDA. The regression reported in table 3 explains over 80 per cent of the variation in the Indigenous employment to population ratio. Again, the statistical properties of this regression are adequate (with the first order serial correlation being present at the five per cent level but not at the one per cent significance level).

The strong conclusion from these results is that there is a significant correlation, for both Indigenous and non-Indigenous populations, between the employment-population ratio for the general population and selected probabilities of transitions between non-employment labour market states for the Income Support Recipient population.

In the previous section we have also presented strong results regarding the similarity of the patterns of behaviour of the transition probabilities involving the employment state for Indigenous and non-Indigenous Income Support Recipients. Overall, therefore, these results provide an opportunity to construct a model explaining variations in the Indigenous employment to population ratio with variables derived from a readily available administrative data set but which only relate to a subset of the population. We will now turn to constructing a prototype of such a model which can then be used for forecasting the Indigenous employment to population ratio on a quarterly basis.

5. Experimental model explaining Indigenous employment to population ratio

The last step in our journey to construct a model suitable for predicting the Indigenous employment to population ratio requires a statistical structure linking this ratio to a set of variables which can be derived from the RED and other relevant macro-economic variables. Such a structure can be technical, based on methods developed for deriving the Department of Employment's Leading Indicator of Employment (see, Connolly and Lee, 2005), which essentially means finding the best-fitting relationship between the employment to population ratio and a range of variables. While such a procedure can yield a robust relationship for a given sample, there is no guarantee given our short time series that it will remain strong in the post-sample period, which is essential for our purposes. Another procedure involves developing a theoretical model of the Indigenous labour market, which can then be used to generate Indigenous employment figures. While this second procedure is attractive where sufficient information is available, estimation of such a model is unlikely to yield useful results in the current context due to the lack of relevant data.

An alternative to these two procedures entails utilizing an economic relationship, which can then be made operational by quantitative analysis (i.e.

estimation of a multiple regression equation). One such long-standing relationship which is simple, involves fundamental macroeconomic variables and appears to be robust in over fifty years of testing is 'Okun's law' (Okun, 1962). This relationship reflects the observation that in order to produce more output an increased amount of labour is required (or a reduced level of idle capacity – unemployment). Amongst the many variants of Okun's law, one of the simplest is a relationship linking the change in the unemployment rate to the growth rate of national output. A little algebraic manipulation yields a relationship between employment-population ratio and the lagged unemployment rate, labour force participation rate and output growth.

Our experimental model of the Indigenous employment to population ratio uses an ordinary least squares estimate of a transformed Okun's law with the additional variables being the variables derived from RED. During the process to select the preferred specification we have dropped the labour force participation rate, which was insignificant. The preferred specification includes the annual change in the unemployment rate, the level of GDP and its annual growth rate and three transition probabilities derived from RED (NILF to employment, unemployment to employment and unemployment to NILF). The probabilities of an Indigenous income support recipient transiting from either NILF or unemployment to employment (PNEXA, PUEXA) reflect the contemporaneous labour market. Similarly, the probability of an Indigenous income support recipient transiting from unemployment to NILF (PUNXA) is a contemporaneous reflection of the tightness of the activity requirements of the income support system.

We also present a reduced version of this specification (the minimal model) which only includes the annual change in the unemployment rate as a relevant macroeconomic variable but also includes another variable derived from RED - the proportion of all unemployed and NILF who have not been employed in the past 3 months (NE3MXA). We include this variable to control for heterogeneity by simulating the average characteristics of the pool of Income Support Recipients. Both of these specifications also include a dummy for the May quarter.[9]

The estimation period of these regressions is the first quarter of 2001 to the third quarter of 2012. We should note that since information in RED is fully updated on a monthly basis and the HILDA data is available annually, it is unlikely that a production version of this model would need to predict the employment to population ratio more for than four quarters ahead. Nevertheless, we have utilized our prototype model to predict six quarters ahead – the maximum span which could be derived at the time of writing. Table 4 presents these estimated models as well as some basic diagnostic tests carried out on these models.

[9] See appendix for sources and definitions of variables.

Table 4 - Estimated forecasting models

	Preferred model			*Minimal model*		
	coefficient	*t-ratio*	*p-value*	*coefficient*	*t-ratio*	*p-value*
PNEXA	2.30E+01	9.247	0.000	1.34E+01	4.204	0.000
PUEXA	-7.13E+00	-2.719	0.010	-6.09E+00	-3.282	0.002
PUNXA	-3.64E+00	-1.719	0.094	-1.51E+00	-0.726	0.472
NE3MXA				-1.90E+00	-2.975	0.005
D4PUR	-1.12E-02	-2.108	0.042	-7.95E-03	-1.801	0.079
GDP	3.37E-01	1.876	0.068			
GR4GDP	-3.58E-03	-1.406	0.168			
Q2	1.35E-02	2.213	0.033	8.65E-03	1.444	0.157
CONSTANT	3.73E-01	7.810	0.000	2.27E+00	3.648	0.001

Diagnostic	*Preferred model*	*Minimal model*
Adj. R-square	0.820	0.834
SEE	1.74E-02	1.67E-02
Std. dev. (regressand)	4.10E-02	4.10E-02
LM(1)	1.022	2.514
DW	1.710	1.284
DW - p-value	0.059	0.002
ACF: lag 1	0.933	2.405
ACF: lag 2	0.331	0.180
ACF: lag 3	-2.933	-2.274
ACF: lag 4	-1.413	-2.243
Harvey test	10.009	15.483
RESET (2)	1.128	0.023

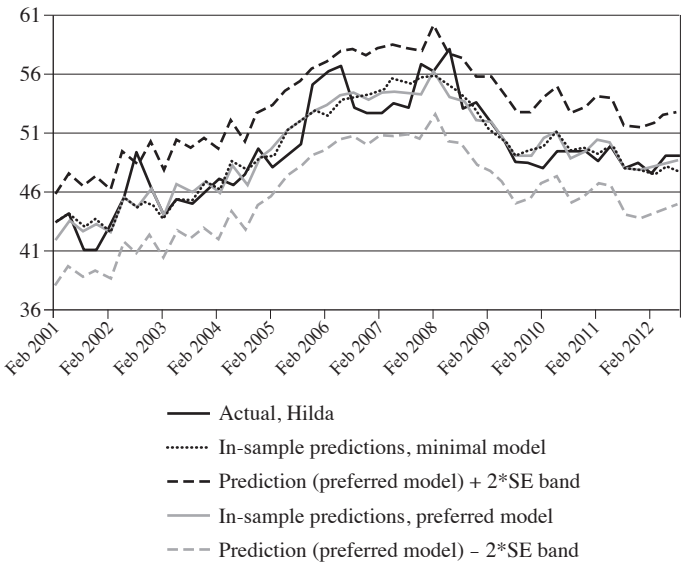
Source: Authors' estimates based on HILDA survey, RED and ABS (see Data Appendix).
Notes: 1. The dependent variable (EPRWYA) has been derived from the reweighted sample, as described in Section 3.
2. Additional explanatory notes are given below table 2.
3. The approximate critical values for the Harvey test at 1 per cent level and 8 degrees of freedom are 18.48 (preferred model) and 16.81 (minimal model).

In summary, both of these models present adequate fit with over 80 per cent of variation in the employment to population ratio derived from HILDA being explained by the included variables. This is also confirmed by the standard error of the estimate in both regressions being substantially lower than the standard deviation of the dependent variable. The RESET test indicates the adequacy of the functional form of the model while the Harvey test also finds no evidence of heteroscedasticity in the residuals. There is some indication of the presence of first-order autocorrelation in the minimal model and the tests for structural stability utilizing recursive residuals (not reported in the table) also signal (at the five per cent level) a possible break around 2006. The Durbin-Watson statistic used as a test for co-integration (Engle and Granger, 1987) indicates that there is evidence of a long run relationship between the employment to population ratio and the set of regressors in both specifications.

While the theoretical derivation and the overall fit favour the preferred model we have used both models to generate in-sample predictions of the Indigenous employment to population ratio (see figure 4). Comparison of these two sets of predictions with the

actual observations derived from HILDA suggest that both models track the actual data accurately (all actual employment to population ratio observations, except one, lie within two standard errors bounds of the in-sample predictions). The preferred model has then been used to generate out-of-sample forecasts up to the first quarter of 2014. These forecasts are presented in figure 5.

Figure 4 - Employment to Population ratio – actual values, in-sample predictions and their standard error band (15 to 64 year olds, per cent)

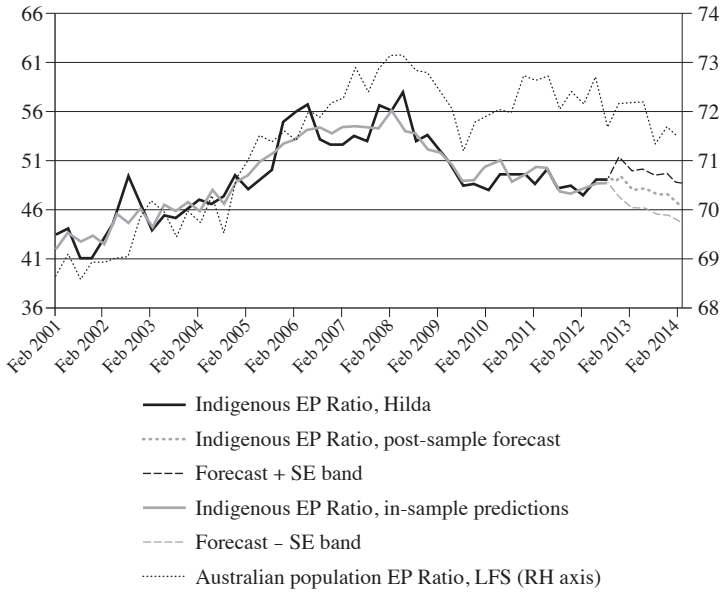


Source: Authors' calculations based on HILDA survey; Authors' estimates.

As discussed earlier, there is very little data on Indigenous employment which could be used to validate these predictions. There is also very little data on non-Indigenous employment which could be used for quantifying the employment gap. The employment to population ratio for the total Australian working age population is slightly lower than the employment to population ratio for non-Indigenous people, but it does reflect the trends of the non-Indigenous population as Indigenous people make up a relatively small proportion of overall employment. We have, therefore, included this series in figure 5 as an indicator of the non-Indigenous trend in employment.

As figure 5 illustrates, the predicted and the HILDA survey employment to population ratios move together very closely over the estimation period. The out of sample forecast also reflects the trend of the employment to population ratio for the total population and appears to mirror its seasonal movements. There is, however, a large difference in the magnitude of change over the sample period: the total employment to population ratio has dropped only a quarter of a percent between August 2012 and February 2014 while the predicted employment to population ratio has dropped by 4.2 per cent.

Figure 5 - Indigenous Employment to Population ratio – actual values, in-sample predictions, post-sample predictions and their standard error band and total employment to population ratio for the Australian population (15 to 64 year olds, per cent)



Source: Authors' calculations based on HILDA survey; Authors' estimates; ABS (2014), Labour Force, Australia (Table 18. Labour force status by Sex - Persons aged 15 to 64 years), ABS Cat. No. 6202.0 (Australian population employment to population ratio).
Note: EP in the legend stands for employment to population.

It is also worth mentioning that coinciding with this downturn in the labour market, a series of government policies came into effect which increased the number of Income Support Recipients who were required to search for employment. The most important and widely reported of these changes was the addition of job search requirements to single parents with the youngest child five years of age or older. This has substantially increased the number of unemployed persons as defined here and reduced the number classified as not in the labour force, flowing through into the probabilities of changing between these states. Our framework has been able to account for such shifts by inclusion of variables linked to administrative data which reflect such policy changes. While a model utilizing macro-economic variables would not be able to accommodate such changes, our framework incorporating administrative data has been able to add additional dimension to a forecasting ability of an econometric model.

6. Concluding comments

The comparison of the transition probabilities of Indigenous and non-Indigenous Income Support Recipients shows that they have a very similar pattern, with the difference in magnitude remaining fairly constant. The comparison of the transition probabilities between the Indigenous and non-Indigenous Income Support Recipients and the HILDA sample of each group provides some evidence that Indigenous Income Support Recipients and the general Indigenous populations are less similar to each other than are non-Indigenous Income Support Recipients and the general non-Indigenous community.

Despite these differences the experimental model we have presented in this paper demonstrates the proof of concept – that it is possible to use administrative data to construct a robust estimate of Indigenous employment outcomes. While the estimates presented in this paper should be considered very cautiously and, as they stand, are not suggested as more than an indicative and experimental measure of Indigenous employment, a downturn in Indigenous employment, causing a widening of the employment gap, is supported by the most recent ABS estimate (Australian Bureau of Statistics 2012-13).

Several areas present themselves as avenues for future research. Having constructed a time series for the Indigenous employment to population ratio from the HILDA survey a further exploration of the macroeconomic influences on Indigenous employment would be valuable both for policy and for revisions of the experimental model presented above.

At present the RED gives us the capability to evaluate outcomes for Indigenous Income Support Recipients; however it is unable to provide insights into the circumstances of those outside the welfare system. A time series such as the one presented above provides a tool for policy evaluation in terms of overall employment outcomes, and this presents a potentially very valuable area of further research.

Unlike the non-Indigenous population, a very large proportion of Indigenous people are in some way in contact with Centrelink at a given time. The characteristics of this group can be assessed though administrative data, however, there is still a large proportion of the population which is not in contact with Centrelink and there is some evidence that these two groups are more different than the corresponding non-Indigenous populations. An investigation into the difference in the characteristics of Indigenous Income Support Recipients and the wider Indigenous population, especially in terms of demographics, geography and human capital could prove useful for the design and implementation of Indigenous employment policy and would certainly benefit future revisions of the above model.

Data Appendix

Table A1 - Definitions and sources of variables used in the regressions

Variable	*Definition*	*Source*
EPRWYA, EPRY	Average employment to population ratio of HILDA participants aged15-64 years old (Indigenous and non-Indigenous sub-populations)	HILDA survey
PNEXA, PNEX	Probability of transition from NILF to employment (Indigenous and non-Indigenous sub-populations)	RED
PNUXA, PNUX	Probability of transition from NILF to unemployment (Indigenous and non-Indigenous sub-populations)	RED
PUEXA, PUEX	Probability of transition from unemployment to employment (Indigenous and non-Indigenous sub-populations)	RED
PUNXA, PUNX	Probability of transition from unemployment to NILF (Indigenous and non-Indigenous sub-populations)	RED
NE3MXA, NE3MX	The proportion of all unemployed and NILF who have not been employed in the past 3 months (Indigenous and non-Indigenous sub-populations)	RED
D4PUR	Four-quarter difference in total unemployment rate	ABS, Labour Force, Australia, ABS Cat. No. 6202.0
GDP	Gross Domestic Product	ABS, Australian National Accounts: National Income, Expenditure and Product, ABS Cat. No. 5206.0.
GR4GDP	Annual growth of Gross Domestic Product	As above

References

Australian Bureau of Statistics (2001), *Census of Population and Housing*, ABS Cat. No. 2037.0.55.001.

Australian Bureau of Statistics (2004-05), *National Aboriginal and Torres Strait Islander Health Survey*, ABS Cat. No 4715.0.

Australian Bureau of Statistics (2006), *Census of Population and Housing*, ABS Cat. No. 2037.0.55.001.

Australian Bureau of Statistics (2008a), *National Aboriginal and Torres Strait Islander Social Survey*, ABS Cat. No. 4714.0.

Australian Bureau of Statistics (2008b), *Survey of Education and Work*, ABS Cat. No. 6227.0.

Australian Bureau of Statistics (2009), *Australian Labour Market Statistics*, January 2009, ABS Cat. No. 6105.0.

Australian Bureau of Statistics (2011), *Census of Population and Housing,* Accessed through 2011 TableBuilder Basic, ABS Cat. No. 2072.0.

Australian Bureau of Statistics (2011), *Labour Force Characteristics of Aboriginal and Torres Strait Islander Australians*, Estimates from the Labour Force Survey, ABS Cat. No. 6287.0.

Australian Bureau of Statistics (2012), *Survey of Education and Work*, Cat. No. 6227.0.

Australian Bureau of Statistics (2012-13), *National Aboriginal and Torres Strait Islander Health Survey*, ABS Cat. No 4715.0.

Black, D., Tseng, Y. and Wilkins, R. (2005), *'The Causes of Long-Term Income Support Receipt Associated with Unemployment'*. Report prepared for the Australian Government Departments of Family and Community Services and Employment and Workplace Relations, Melbourne Institute of Applied Economic and Social Research.

Brooks, C. and Volker, P. (1986), 'The Probability of Leaving Unemployment: The Influence of Duration, Destination and Demographics', *Economic Record*, 62(3), 296-309.

Budd, A., Levine, P. and Smith, P. (1987), 'Long-term Unemployment and the Shifting UV Curve: A Multi-country Study', *European Economic Review*, 31(2), 296-305.

Cai, L., Ghantous, S. and Wilkins, R. (2006), *'Duration Analysis of Income Support Spells Initiated by Unemployment'*, Report prepared for the Australian Government Department of Employment and Workplace Relations, Melbourne Institute of Applied Economic and Social Research.

Chapman, B.J. and Smith, P.N. (1993), *Predicting the Long-term Unemployed: a Primer for the Commonwealth Employment Service*. No. 285. Centre for Economic Policy Research, Research School of Economics, Australian National University.

Chapman, B.J., Junankar, P.N. and Kapuscinski, C.A. (1992), Projections of Long-term Unemployment, *Australian Bulletin of Labour,* 18(3), 195-207.

Chindamo, P. and Uren, L. (2010), 'Vacancies and Unemployment in Australia', *Australian Economic Review,* 43(2), 136-152.

Chindamo, P. (2010), 'Australian Labour Market Flows Over the Business Cycle', *Australian Bulletin of Labour*, 127-137.

Connolly, G. and Lee, J. (2005), Review of the DEWR Leading Indicator of Employment, Paper presented at the Australian Labour Market Research Workshop, Canberra, December.

Council of Australian Governments (2008), 'National Indigenous Reform Agreement (Closing the Gap)', Intergovernmental Agreement on Federal Financial Relations.

Engle, R.F. and Granger, C.W.J. (1987), Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55(2), 251-276.

Goode, A. (2007), *'Recall Bias in Reported Events: HILDA, Waves 1-5',* HILDA Project Discussion Paper Series, No 2/07, June.

Gray, M.C., Hunter, B.H. and Lohoar, S. (2012), *Increasing Indigenous Employment Rates*, Issues Paper No. 3, Closing the Gap Clearinghouse, Australian Institute of Health and Welfare, Canberra.

Hunter, B.H. and Gray, M.C. (2001), 'Indigenous Labour Force Status Re-Visited: Factors Associated With the Discouraged Worker Phenomenon', Australian *Journal of Labour Economics*, 4(2), 115-37.

Jones, S.R.G. and Riddell, W.C. (1999), 'The Measurement of Unemployment: An Empirical Approach', *Econometrica*, 67(1), 147-162.

Junankar, P.N. and Kapuscinski, C.A. (1990), *The Duration of Unemployment in Australia: Memory and Measurement*, ANU, CEPR Discussion Paper No. 238.

Melbourne Institute of Applied Economic and Social Research (2013), *HILDA Survey*. 2013. http://www.melbourneinstitute.com/hilda/ (accessed November 11, 2013).

Mincer, J. and Zarnowitz, V. (1969), The Evaluation of Economic Forecasts in J. Mincer (ed.), *Economic Forecasts and Expectations*, New York: NBER.

Okun, A.M. (1962), 'Potential GNP: Its Measurement and Significance', *American Statistical Association*, Proceedings of the Business and Economics Statistics Section, 98-104.

Rudd, K. (2008), 'Apology to Australia's Indigenous Peoples' (2008), Speech by the Prime Minister to Parliament, 13 February.

Shimer, R. (2012), 'Reassessing the Ins and Outs of Unemployment', *Review of Economic Dynamics*, 15(2), 127-148.

Stromback, T., Dockery, M. and Ying, W. (1998), *Transitions in the Labour Market: Evidence from the Survey of Employment and Unemployment Patterns*. Discussion Paper No. 1-98. Centre for Labour Market Research.

Taylor, J. and Hunter, B.H. (1998), *The Job Still Ahead: Economic Costs of Continuing Indigenous Employment Disparity,* Canberra, Office of Public Affairs, Aboriginal and Torres Strait Islander Commission.

Watson, N. (2009), 'Disentangling Overlapping Seams', *HILDA Conference Paper.*

Watson, N. (2012), *'Longitudinal and Cross-sectional Weighting Methodology for the HILDA Survey'*, HILDA Project Technical Paper Series No. 2/12 December.

Watson, N. and Wooden, M. (2002a), *The Household, Income and Labour Dynamics in Australia* (HILDA) Survey: Wave 1 Survey Methodology, HILDA Project Technical Paper Series No. 1/02 May 2002.