Australian Bureau of Statistics

1351.0.55.059

Research Paper

Unemployment Duration in Australia: A Longitudinal Analysis with Missing Data

New Issue

Research Paper

Unemployment Duration in Australia: A Longitudinal Analysis with Missing Data

Cristian I. Rotaru

Methodology Division

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) TUE 24 MAY 2016

ABS Catalogue no. 1351.0.55.059

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Ruel Abello, Methodology Transformation Branch, on Canberra (02) 6252 6307.

CONTENTS

ABSTRACT 1

1. INTRODUCTION 2
   - 1.1 Background 2
   - 1.2 Objective and hypotheses 2
   - 1.3 Modelling approach 3
   - 1.4 Contributions 4

2. DATA AND DESCRIPTION OF THE PROBLEM 5
   - 2.1 Data 5
   - 2.2 Description of the missing problem 6

3. THEORETICAL FRAMEWORK 10
   - 3.1 Multiple imputation 10
   - 3.2 The unemployment model 13

4. RESULTS 17
   - 4.1 Explanatory variables 17
   - 4.2 Personal characteristics 18
   - 4.3 Design-related estimates 19
   - 4.4 Baseline hazard function and model comparisons 20

5. CONCLUDING REMARKS 27

REFERENCES 28

ACKNOWLEDGEMENTS 28

UNEMPLOYMENT DURATION IN AUSTRALIA: A LONGITUDINAL ANALYSIS WITH MISSING DATA

Cristian I. Rotaru
Methodology Division

ABSTRACT

This paper analyses the conditional probability of exiting unemployment of Australian individuals, aged 18–65 years, over the 2008–2010 period. The paper makes use of the ABS Longitudinal Labour Force Survey file, which includes more than 1.8 million records of around 150,000 households, which were observed on a month-by-month basis for up to eight months.

Multiple imputation is used to impute the missing values of previous educational attainment – a key covariate included in the analysis – using Bayesian Monte Carlo simulation methods. The subsequent investigation implements discrete-time hazards models to examine the duration of unemployment, with allowance for competing risks. The analysis includes empirical as well as design-related variables.

Keywords: unemployment, hazard rate, survival analysis, multiple imputation, ABS LFS

JEL Classification: J21, J64, C15, C18, C41

1. INTRODUCTION

1.1 Background

This study is an extension of a recently-completed ABS research that made use of the ABS Longitudinal Labour Force Survey (LLFS) file. From this research, two papers emerged, one focused on the methodologies developed to analyse the labour market transitions (Rotaru, 2014b) and the other on an application: understanding the factors that explain these transitions (Rotaru, 2014a).

This paper extends the previous research on a number of fronts. From a methodological perspective, the analysis is based on a longitudinal multinomial framework, instead of the previous longitudinal binary models. The new framework has carefully been built so as to (1) advance the previous methodology and adequately address the specific features of the dataset, which differ substantially from other similar datasets (e.g. HILDA), and (2) to address the empirical question – modelling the duration of unemployment – by accounting for the longitudinal nature of the data and the different types of exits out of unemployment.

Further, the analysis makes use of multiple imputation – a powerful modern tool for handling missing data – adopting a Bayesian approach. This is the first time the method has been implemented in the analysis of data at the ABS. The imputation methodology will play important roles in the Statistical Business Transformation Program, particularly with the use and analysis of administrative and longitudinal datasets.

The analysis includes both empirical and design-related variables and makes a number of adjustments to the previous analyses, such as the inclusion of those aged 18–20 years, the inclusion of an industry variable, and the redefining of the previous labour force status indicator. The hope is that the analysis adds value to the dataset, produces statistically robust results, addresses a topic of key interest to the public, and promotes the potential of the ABS data.

1.2 Objective and hypotheses

Unemployment can be a major life event for an individual, with quite severe consequences. Apart from the immediate financial impact, prolonged unemployment can significantly affect the individual's physical and mental wellbeing, can result in an erosion of skills and consequently decrease the likelihood of securing future employment, increase the risk of lower future incomes, and lead to lower self-esteem. The effects can extend beyond the unemployed, to their families and the wider community, in the form of increased marriage breakdowns, violence, and crime.

High unemployment is also a major concern for the economy in that it can result in a loss of potential growth and in increased government spending. This is because potential productive resources are not being utilised effectively in the economy, there is a corresponding reduction in tax revenues being collected, and there is an increase in the payments of unemployment benefits. (See Sen, 1997 and Rotaru, 2014a for a more comprehensive literature review.)

From an empirical perspective, the objective of this paper is to provide insights on the unemployment duration in Australia and in particular, to improve understanding of a number of important factors that influence this duration. This is done using an important newly available longitudinal dataset: the ABS Longitudinal Labour Force Survey file.

In particular, the analysis examines the conditional probability of exiting unemployment for Australians aged 18–65 years, over the period 2008–2010. The paper makes use of the recently constructed file which is based on a sample of individuals who were observed over a period of up to eight consecutive months. The analysis is focused on those who entered unemployment during the interview period and on their first unemployment spells.

Two important empirical hypotheses are investigated. First, the analysis examines the effects of a number of factors on the conditional probability of exiting unemployment, distinguishing between a number of exits, in particular, employment (full-time and part-time) and out of the labour force. Second, the analysis examines the state dependence of unemployment, i.e., whether the time spent in unemployment affects the probability of remaining in unemployment. The analysis includes both empirical and design-related variables.

1.3 Modelling approach

From a methodological perspective, the paper implements a range of techniques to adequately address the specific features of the dataset and to meet the objectives of the analysis. In particular, the paper adopts the multiple imputation approach to impute the missing values of previous educational attainment – a key covariate included in the analysis. The subsequent investigation then implements discrete-time hazards models to examine the duration of unemployment, taking advantage of the discrete nature of the unemployment data. To distinguish between the different exits from unemployment, the analysis makes allowance for competing risks, via multinomial models.

1.4 Contributions

The paper makes two important contributions to the literature. The first relates to the data used in the analysis – the dataset being a newly constructed longitudinal file built from the ABS Labour Force Survey. Apart from its rich information, the file also has the distinguishing feature of having relatively high-frequency data. This is appealing in that it allows the analyst to include variables that are changing over time and to model actually-reported instead of retrospectively-collected unemployment duration information, which has typically been used in the literature and which can suffer from recollection bias.

The second contribution relates to the actual methodology used in the analysis. Unlike the majority of similar studies, which simply ignore or discard the units with any missing values (missingness being one of the common features of unemployment-related data), this paper instead paid close attention to the issue, closely examined it, and used all the available information in the analysis. A modern statistical tool, namely, multiple imputation, was adopted to impute the missing values. The study also focuses on design-related as well as empirical factors in the analysis; the former factors are often ignored in practice.

The remaining sections of this paper are organised as follows. Section 2 describes the data. The analytical framework adopted is presented in Section 3. Section 4 presents the results and Section 5 concludes.

2. DATA AND DESCRIPTION OF THE MISSING DATA PROBLEM

2.1 Data

The data set consists of a random sample of close to 14,000 individuals aged 18–65 years, included in the ABS Longitudinal Labour Force Survey (LLFS) file. The survey covers the period from the beginning of 2008 to the end of 2010. As the survey has an eight-month rotation panel, each individual is being captured for a period of up to eight months. The spells are measured in months, corresponding to the monthly collection of the data, and the demographic variables included in the analysis refer to the beginning of the unemployment spells. Each completed spell ends in one of three competing destinations: employed full-time, employed part-time, or out of the labour force. If the spell is not completed, then it is right-censored.[1]

The analysis is based on the individuals who have entered unemployment during the eight-month period in which they were interviewed and is restricted to private dwellings. The analysis is also only focused on the first unemployment spells, which last from the time the individuals become unemployed until the end of their first unemployment spell, or until censoring.

Tables 2.1–2.4 provide some summary statistics of the sample included in the analysis.

2.1 Distribution of Initial LFS status by Age-groups

	Age-groups				
Initial LFS status	*18–24 years*	*25–34 years*	*35–44 years*	*45–54 years*	*55–65 years*
			%		
Full-time employment	22.5	29.3	28.7	30.9	24.6
Part-time employment	24.3	17.6	18.6	20.1	21.5
Out of the Labour Force	32.7	37.0	35.3	33.5	35.1
First-time job seeker	16.7	6.0	3.5	3.3	1.6
Last worked 2 years ago	3.8	10.1	13.9	12.2	17.2
Total	100.0	100.0	100.0	100.0	100.0

2.2 Distributions by Exit state

Exit state	*Frequency*	*Percentage*
Full-time employment	2,837	20.7
Part-time employment	3,170	23.1
Unemployment	3,612	26.3
Out of the Labour Force	4,109	29.9
Total	13,728	100.0

1 Note that the right-censored spells belong to those individuals who did not exit unemployed at the end of the eight-month interview period as well as those who withdrew or were dropped from the sample prior to the completion of the interview period. This latter group includes those individuals who had moved during the LFS waves and those who did not respond in subsequent waves after becoming unemployed.

2.3 Distributions of Exit state by initial LFS status

	Initial LFS status				
Exit state	*Full-time employment*	*Part-time employment*	*Out of the Labour Force*	*First-time job seeker*	*Last worked 2 years ago*
	%	%	%	%	%
Full-time employment	44.4	14.3	15.4	2.5	1.9
Part-time employment	16.2	40.0	26.4	9.4	6.5
Unemployment	24.0	25.2	24.4	33.8	35.6
Out of the Labour Force	15.4	20.5	33.8	54.3	56.0
Total	100.0	100.0	100.0	100.0	100.0

2.2 Description of the missing data problem

The monthly ABS Labour Force Survey (LFS), conducted over the period January 2008 to December 2010, forms the backbone of the LLFS file. Using information provided by the households interviewed during this period, the file includes the LFS person-linked responses to form a detailed and rich longitudinal dataset. This is possible since each household included in the LFS is interviewed for a number of (up to eight) consecutive months.

The file has some special characteristics, which distinguishes it from a typical longitudinal dataset. The file's specific aspects also require specifically-tailored models for analysis (which will be explained in more detail in the next section). One such characteristic is the design of the file, the file being constructed from the consecutive panels of a non-longitudinal survey that has been designed to produce cross-sectional estimates. It is this panel structure of the LFS that permits the construction of a longitudinal dataset.

Another specific characteristic is its relatively short individual-specific time length. The analyst is thus limited to observing what happens within an eight-month window, with no knowledge of what happens outside this time interval. This, however, is compensated by the large amount of information available in the file and by the high-frequency of the data.

Another aspect relates to the source of the data, which is not a sole survey. In addition to the LFS, the file is supplemented with information collected in other supplementary surveys; information which is usually not available in the LFS. An example is the information on education, which mainly comes from four supplementary surveys:

- the Survey of Education and Work (SEW),
- the Survey of Underemployed Workers (UEW),
- the Labour Mobility Survey (LabMob), and
- the Multipurpose Household Survey (MPHS).

One complexity that arises then is in dealing with the fact that these supplementary surveys do not have the same frequency as the LFS – they are not being run on a monthly basis. In the case of education, for example, the supplemented data only cover the months of May (SEW), September (UEW), and February (LabMob and MPHS), for each year. As this additional information is often essential in the analysis, the analyst is then faced with the task of incorporating the available, yet incomplete, information in the analysis – which is a missing data problem. This is not a straight-forward task, however, since most classical and modern statistical techniques and most statistical packages assume completeness of data. In the case of the common statistical packages, most of them automatically delete the records with any missing values, despite the potential valuable information in the non-missing reported values.

In this paper, previous educational attainment is one of the key covariates used in the analysis. As it is widely acknowledged, education generally plays an important role in explaining the duration of unemployment. However, due to the way data were collected, and as explained before, the education information was available only for February, May, and September. And even for these three months the education information was missing for some observations. However, no such data were available at all for the individuals who were not observed during these three months. A moderate proportion (24%) of responses fall into this latter category.

This latter category presents some challenges. First, discarding all these observations will certainly result in the loss of a lot of valuable information. More serious consequences however will be due to the impact on the model results, which could be biased. It is clear, for example, that the education information is more likely to be completely missing (i.e. not available in any of the three months) for those individuals that experience an earlier exit from the survey. This implies that there is likely to be an association between the attrition in the survey, and therefore the characteristics associated with this attrition, and the missingness for education.

Table 2.4 indicates that the missing values are not missing completely at random. In particular, the results suggest that those who have not worked in the last two years, and therefore do not have a previous occupation or industry, and those who exited the labour force are more likely to have missing values.

The problem was addressed in two parts. First, where previous educational attainment data were available for some, but not all time points, the study used the available information and assumed that the individuals had not changed their status in the subsequent waves of the survey included in the analysis. As those included are aged 18 years and above, and as they are being kept in the surveys for a period of up to eight months, this seems to be a reasonable assumption. Second, where no education information was available, the study has opted to use multiple imputation. It is one of the aims of this paper to showcase the application of multiple imputation in analysing ABS survey data.

2.4 Distributions of variables (Reported vs Missing)

	Educational attainment	
	Reported	Missing
	%	%
Exit state		
Full-time employment	21.6	17.5
Part-time employment	24.5	18.8
Out of the Labour Force	24.6	31.7
Censored	29.3	31.8
Previous Labour Force status		
Full-time employment	28.4	22.6
Part-time employment	23.3	12.1
Out of the Labour Force	33.5	38.0
First time looking for work	5.9	13.0
Worked 2 years ago	8.9	14.3
Duration of uemployment		
1 month	39.1	41.5
2 months	22.8	25.0
3 months	15.3	15.2
4 months	10.4	8.8
5 months	7.0	5.5
6+ months	5.4	4.0
Age-group		
18–24 years	27.9	30.9
25–34 years	22.0	22.7
35–44 years	21.8	20.1
45–54 years	17.6	16.1
55–65 years	10.7	10.2
Sex		
Female	47.8	50.1
Male	52.2	49.9
Marital status		
Not married	54.6	55.4
Married	45.4	44.6
Born overseas		
No	70.6	64.9
Yes	29.4	35.1
Children		
No children under 4 years	87.5	88.4
Children under 4 years	12.5	11.6
State or Territory		
New South Wales	24.5	26.2
Victoria	22.2	22.6
Queensland	17.8	17.5
South Australia	11.7	12.1
Western Australia	11.3	10.6
Tasmania	6.2	5.0
Aust. Capital Territory	3.4	3.2
Northern Territory	2.9	2.8

2.4 Distributions of variables (Reported vs Missing) – continued

	Educational attainment	
	Reported	*Missing*
	%	%
Previous occupation		
No occupation	14.8	27.3
Lower-skilled occupation	34.5	31.5
Medium-skilled occupation	34.0	28.0
High-skilled occupation	16.7	13.2
Industry		
No Industry	14.8	27.3
Agriculture, Forestry and Fishing	2.6	2.2
Mining	1.6	1.7
Manufacturing	9.5	8.9
Electricity, Gas and Water Services	0.8	0.8
Construction	9.2	7.6
Wholesale Trade	3.5	2.4
Retail Trade	10.8	9.6
Accommodation and Food Services	9.1	8.1
Transport, Postal and Warehousing	3.9	3.3
Information Media and Telecommunications	1.7	1.2
Financial and Insurance Services	2.3	1.8
Rental, Hiring and Real Estate Services	1.6	1.4
Professional, Scientific and Technical Services	4.9	3.9
Administrative and Support Services	4.4	4.8
Public Administration and Safety	3.1	2.6
Education and Training	4.6	3.5
Health Care and Social Assistance	6.7	5.2
Arts and Recreation Services	1.6	0.9
Other Services	3.3	2.8
Overall	75.8	24.2

3. THEORETICAL FRAMEWORK

This section is divided into two parts. The first describes the methodology and the algorithm adopted to impute the missing values of previous educational attainment, one of the variables included in the modelling. The second presents the analytical framework and describes the unemployment model.

In summary, the analysis follows a four-step approach:

1. Start with an incomplete data set and specify the posterior density from which the imputations will be drawn;
2. Create m complete data sets from m sets of imputations;
3. Conduct the analysis on each of the m complete data sets; and
4. Combine the results from the m analyses and make inferences.

It should be noted that in effect, the analysis accounts for two types of missingness: the missingness in one of the covariates used in the modelling (i.e., educational attainment) and the missingness due to right-censoring. The former type of missingness is addressed via multiple imputation, whereas the second type via hazard modelling techniques.

3.1 Multiple imputation

Let Y be an $n \times p$ matrix that represents the partially observed data on p variables of interest collected from a sample of n units. Partition Y such that the completely observed values are collectively denoted by Y_{obs} and the missing values by Y_{mis}. In the case of this paper, Y_{mis} is one-dimensional and it includes the missing educational attainment responses.

Denote by y_{ij} the entries of $Y = (Y_{\text{obs}}, Y_{\text{mis}})$ where $i = 1,\ldots,n$ and $j = 1,\ldots,p$. Let R be a corresponding $n \times p$ binary matrix of 0–1 response indicators with $r_{ij} = 1$ if y_{ij} is observed and $r_{ij} = 0$ if y_{ij} is missing.

Denote by Q the quantity of interest to the analyst, which is to be estimated. The estimand Q generally refers to a vector of population quantities, such as population means, correlations, or as is the case in this study, regression coefficients. If the complete data (Y) were available and the sampling design was known, the analyst could estimate Q with the sample counterpart $\hat{Q}(Y)$, which is assumed to be a function of Y. There are two issues however. First, the sample information is not complete as the information in Y is missing for the cases where $r_{ij} = 0$. Second, the results from the previous section indicate that the values in Y_{mis} are not missing completely at random, i.e.

$$P\left(R = 0 \middle| Y_{\text{obs}}, Y_{\text{mis}}, \varphi\right) \neq P\left(R = 0 \middle| \varphi\right)$$

where φ denotes the parameters of the missing data model $P\left(R \mid Y_{\text{obs}}, Y_{\text{mis}}, \varphi\right)$.

It is this second issue that usually makes the analysis more complex. As the missingness depends on the data values, the analyst cannot just ignore the missing data mechanism if valid inferences are to be made about Q (which is a function of Y) , although this is often done in practice.

Multiple imputation provides a neat solution to this problem. Its aim is to replace Y_{mis} with $m > 1$ independently simulated draws from the posterior distribution of the missing data given the observed values and the missing data mechanism, i.e. $\left[Y_{\text{mis}} \mid Y_{\text{obs}}, R\right]$, and to use these draws to make inferences about Q.

Denote the draws by $Y_{\text{mis}}^{(l)}$, where $l = 1, \ldots, m$ and m stands for the number of imputations. Note that only the missing values (i.e., where $r_{ij} = 0$) of Y are replaced. For each of these imputed datasets, one then estimates Q with $\hat{Q}\,(Y^{(l)})$, where $Y^{(l)} = (Y_{\text{obs}}, Y_{\text{mis}}^{(l)})$. For brevity, let $\hat{Q}^{(l)}$ denote $\hat{Q}\,(Y^{(l)})$; let also its variance be denoted by $\hat{U}^{(l)}$. As described in the seminal work of Rubin (1987), the analyst can then make valid inferences about the estimand Q by combining $\hat{Q}^{(l)}$ and $\hat{U}^{(l)}$.
In particular, the overall estimate of a scalar Q is given by:

$$\bar{Q} = m^{-1} \sum_{l=1}^{m} \hat{Q}^{(l)}$$

and the variance of $\bar{Q}$ by

$$T = (1 + m^{-1})\, B + \bar{U} \quad ,$$

where

$$\bar{U} = m^{-1} \sum_{l} \hat{U}^{(l)}$$

is the within-imputation variance and

$$B = (m-1)^{-1} \sum_{l} \left(\hat{Q}^{(l)} - \bar{Q}\right)^2$$

is the between-imputation variance.

Before proceeding, there is one more issue that needs to be settled. As mentioned, the aim is to draw synthetic observations from the posterior distribution of $(Y_{\text{mis}} \mid Y_{\text{obs}}, R)$. However, in the general case when $P(Y_{\text{mis}} \mid Y_{\text{obs}}, R = 0) \neq P(Y_{\text{mis}} \mid Y_{\text{obs}}, R = 1)$, modelling $P(Y_{\text{mis}} \mid Y_{\text{obs}}, R = 0)$ solely from the observed data presents a challenge, since by assumption the values in Y are all missing in the cases when $r_{ij} = 0$. The traditional way to address this (which has also been adopted by this study) is to assume an ignorable nonresponse process, which means that $P(Y_{\text{mis}} \mid Y_{\text{obs}}, R) = P(Y_{\text{mis}} \mid Y_{\text{obs}})$. This is often a reasonable assumption and is frequently assumed in practice (see van Buuren, 2012, Chapter 2).

In the case of this study, this means that after conditioning on the observed data, the distribution of the educational attainment variable is the same across the two groups (response and nonresponse). This assumption implies that the imputations can be drawn from the posterior distribution of $(Y_{\text{mis}} \mid Y_{\text{obs}})$, which can be modelled directly from the observed data.

The imputation algorithm

In this paper Y_{mis} is a categorical variable and the draws from $(Y_{\text{mis}} \mid Y_{\text{obs}})$ were obtained using a multinomial logit model.

In the general case, let $Y_{\text{mis}} = (y_{\text{mis},1}, \ldots, y_{\text{mis},n_{\text{mis}}})^{\text{T}}$, where $y_{\text{mis},i} \in \{1, \ldots, K\}$, such that each $y_{\text{mis},i}$ belongs to one of the K distinct and mutually-exclusive categories and n_{mis} stands for the number of sample units whose education information was not available. Denote by X the variables in Y_{obs} that are included in the modelling of $P(Y_{\text{mis}} \mid Y_{\text{obs}})$ and by $\beta = (\beta_1^{\text{T}}, \ldots, \beta_K^{\text{T}})^{\text{T}}$ the unknown parameters to be estimated.

The posterior density of interest then becomes

$$P\left(y_{\text{mis},i} = k \mid X_i, \beta\right) = \exp(X_i \beta_k) \Big/ \sum\nolimits_k \exp(X_i \beta_k) \tag{1}$$

where $k = 1, \ldots, K$ and $i = 1, \ldots, n_{\text{mis}}$. To identify the model, the vector β_1 is set to zero.

Note that the posterior distribution of $(Y_{\text{mis}} \mid Y_{\text{obs}})$ can be written as:

$$P\left(Y_{\text{mis}} \mid Y_{\text{obs}}\right) \equiv P\left(Y_{\text{mis}} \mid X\right) = \int P\left(Y_{\text{mis}} \mid X, \beta\right) P\left(\beta \mid X\right) d\beta .$$

Therefore to obtain the desired draws, one can proceed as follows:

- Obtain a draw $\dot{\beta}$ from the posterior distribution $P(\beta \mid X)$. In order to do this, assume first a uniform prior for β (i.e. $P(\beta) \propto$ constant). Note that, in this case, the mode of the posterior distribution $P(\beta \mid X)$ corresponds to the maximum likelihood estimate. Denote this by $\hat{\beta}$. Appeal next to the large sample properties of the maximum likelihood estimate (assuming that the model is correctly specified) and use the fact that $(\beta - \hat{\beta}) \sim N(0, \hat{V})$, where $\hat{V}$ is the corresponding asymptotic covariance matrix for $(\beta - \hat{\beta})$.

  To obtain the maximum likelihood estimate $(\hat{\beta})$ run model (1) on the observed data, i.e., on the observations with $r_{ij} = 1$. $\hat{V}$ is obtained by computing the inverse of the observed Fisher information matrix and evaluating it at $\hat{\beta}$. Finally, $\dot{\beta}$ is drawn from $N(\hat{\beta}, \hat{V})$.

- Next, draw Y_{mis} from its posterior distribution conditional on the aforementioned draw, i.e. $P(Y_{\text{mis}} \mid X, \beta = \dot{\beta})$.

- Repeat the process m times.

The algorithm implemented to obtain the simulations follows the above strategy and is based on the suggestions of Raghunathan *et al.* (2001) and van Buuren (2012). It is as follows:

1. Apply the maximum likelihood and estimate model (1) using the observed data and obtain the regression coefficients $\hat{\beta}$.
2. Compute $\hat{V}$, the asymptotic covariance matrix of $\hat{\beta}$, and calculate its Cholesky decomposition – denote this by $\hat{V}^{1/2}$.
3. Assume a non-informative prior for β , $p(\beta) \propto$ constant, and draw $\dot{\beta}$ from the asymptotic distribution $N(\hat{\beta}, \hat{V})$. This is achieved in two steps.
   First, obtain a vector $\dot{z}$ of independent draws from the multivariate normal distribution $N(0, I)$. Second, calculate $\dot{\beta} = \hat{\beta} + \dot{z}\,\hat{V}^{1/2}$. This will give a draw from $N(\hat{\beta}, \hat{V})$.
4. Plug in $\dot{\beta}$ in (1) and compute K predicted probabilities for each of the values in Y_{mis} . Denote these probabilities by $\dot{p}_1, \ldots, \dot{p}_K$, where $\dot{p}_k$ denotes the probability of the missing value being in category k, where $k = 1, \ldots, K$.
5. Compute the cumulative sums S_k such that $S_0 = 0$ and $S_k = \sum_{i=1}^{k} \dot{p}_i$, where $k = 1, \ldots, K$. Note that $S_K = 1$.
6. Draw μ from the uniform distribution $U(0,1)$ and assign the missing value to category k if $S_{k-1} \le \mu < S_k$, for $k = 1, \ldots, K-1$, and to category K otherwise.
7. Finally, repeat steps 3–6 and create m versions of Y_{mis} . Denote these by $Y_{\text{mis}}^{(1)}, \ldots, Y_{\text{mis}}^{(m)}$.

3.2 The unemployment model

The paper estimates a duration model using the available information about close to 14,000 individuals' unemployment spells included in the ABS LLFS file. The overall goal of the analysis is to model the conditional probability of exiting unemployment, distinguishing between the different exits, and to determine the effects of the observed characteristics.

In building the model consideration is given to three important aspects: (1) the two types of missingness aforementioned; (2) the different modes of exiting unemployment, namely, whether the individual exits into full-time or part-time employment, or whether the individual exits the labour force; and (3) the highly discrete unemployment duration information, as the data were collected on a monthly basis.

To set the stage, consider a target population of size N, from which samples are drawn, at regular intervals, over a specified period of time. The survey has an eight-month rotation panel, which makes it feasible to link the person-specific responses over time to create a quasi-longitudinal dataset. To put things into context, in this analysis, the period starts at the beginning of 2008 and ends with the last month of 2010; it is also measured on a monthly basis, thus there are 36 such samples. The target population is the Australians aged 18–65 years, who become unemployed during the observed period.

From this population, during each month, a representative sample s is selected. Let $I_{is} = 1$ if individual i is selected in sample s and $I_{is} = 0$ otherwise, where $i = 1, 2, \ldots, N$ and $1 \le s \le 36$. Let $I_i = (I_{i,1}, \ldots, I_{i,36})$ and $I = (I_1, \ldots, I_N)$. For simplicity, assume that N is fixed during the period under consideration. Note that $\sum_{s=1}^{36} I_{is} \le 8$ and that, in the context of this analysis, the overall sample (which is a combination of all 36 samples) is composed of individuals that are initially either employed or out of the labour force.

At some point during the interview period, these individuals become unemployed. The person-specific unemployment clock is started at this point and ideally one would like to follow the individuals from the first tick, when the individuals become unemployed, until the final tick, the time when they first exit unemployment. (Here, the unit of measurement is months.) However this is not always possible, as some individuals do not exit the state of unemployment during the time they are observed.

There are a number of ways of dealing with these censored cases. A naïve approach would be to drop all these observations, however this could dramatically reduce the sample size and more severely, it could bias the results. Instead, this paper includes these censored cases in analysis and adopts hazard modelling techniques.

Before proceeding, it's worth making some notes regarding the LFS sample design, which could have implications in the modelling. First, the LFS, which forms the core of the LLFS file, does not have a simple random sample design. Instead, the design is based on a stratified-multistage selection approach on geographic areas. Second, the design is different for non-private dwellings. Third, there is a secondary design for the allocation of the sample to the states and territories. Finally, during the July 2008 – November 2009 period there was a "24% reduction in the size of the LFS sample as part of a savings initiative" (ABS, 2013).

The analysis accounts for these aspects in a number of ways: (1) by restricting the analysis to private dwellings; (2) by including design-specific variables in the model, such as the geographical region of the household; and (3) by including an indicator for the period associated with the reduction in the sample size.

Modelling

For each individual i, let T_i^* be a random variable capturing the time of exit from unemployment and let the interval of interest be given by

$$\left(t_0, t_{n_i}\right] = \left(t_0, t_1\right] \cup \bigcup_{k=1}^{n_i - 1} \left(t_k, t_{k+1}\right].$$

Here $(t_0, t_1]$ is the interval during which individual i becomes unemployed, $(t_{n_i - 1}, t_{n_i}]$ is the interval in which they either exit unemployment or are censored, and $(t_0, t_{n_i}]$ is the interval over which the individual is observed. t_{n_i} indicates the last time their responses are recorded (for the censored cases) or the first time the individuals have indicated that they have exited unemployment. Note that the intervals correspond to the period between two consecutive interviews and that the first interval, i.e. $(t_0, t_1]$, is common to all individuals – each person-specific clock measuring the unemployment duration is set to zero during this period.

As the information is collected monthly, T_i^* cannot be observed exactly. In particular, the analyst only knows that $T_i^* \in (t_{n_i - 1}, t_{n_i}]$, for the cases when the individual exits unemployment during $(t_0, t_{n_i}]$, and that $T_i^* > t_{n_i}$,for the censored cases. To model this duration, one could then use either a continuous- or discrete-time hazard model. The discrete recording of the data and the large number of ties (i.e. as a large number of individuals exit unemployment during the same month) favours the latter approach.

To this end, let $T_i \in \{1, 2, \ldots, n_i\}$ be a discrete variable such that $T_i = j(i)$ if $T_i^* \in (t_{j(i)-1}, \ldots, t_{j(i)}]$, for $j(i) = 1, \ldots, n_i$. Let $Z^{(l)} = (Z_1, Z_2, Y_{\text{mis}}^{(l)}, D)$ be the relevant information to be included in the model. Its first component, Z_1, is a subset of Y_{obs} and includes those variables that are associated with either the duration of unemployment or the type of exit from unemployment. Z_2 is also a subset of Y_{obs} and it includes those variable associated with the design of the sample (i.e., they are associated with the selection mechanism I). $Y_{\text{mis}}^{(l)}$ refers to the l-th imputed set of values for the educational attainment variable, where $1 \le l \le m$. Finally, D controls for the duration of unemployment in the model; it includes the period indicators (duration dummies) for each individual. Note that the specification is flexible and it allows for the inclusion of variables that vary across individuals as well as time.

(Note that this strategy of transforming the analysis from a continuous T_i^* to a discrete T_i is based on that of Rotaru, 2014a and Rotaru, 2014b. The readers are referred to these articles for a more detailed exposition.)

Assuming that each individual can exit unemployment into one of the four exits, the analyst can then consider a multinomial model. In particular the "still unemployed" or censored category can be taken as the baseline and the other categories of interest be referenced to this. As each individual is observed for a period of up to eight months, this paper opted for a flexible semi-parametric specification of the multinomial model, one where the duration effect is given by the sum of period indicators.

Denoting by E_i individual i 's exit state, the hazard rate of exiting into state r , at time j , for individual i , given a set of covariates Z_i , is given by:

$$\pi_{irj} = P\left(T_i = j, E_i = r \,\middle|\, T_i \geq j, Z_i\right),$$

where, to simplify the notation, superscript (l) was dropped.

The multinomial model is then given by:

$$\log\left(\frac{\pi_{irj}}{\pi_{ir_0j}}\right) = \phi_{r,0} + \phi_{r,1}^{\mathrm{T}} Z_i^* + \sum_{l=1}^{L} \alpha_{r,l} D_{i,l} \quad ,$$

where π_{irj} is the hazard rate for exiting into state r , π_{ir_0j} is the hazard rate of remaining unemployed, which is treated as the base category, $(\phi_{r,0}, \phi_{r,1}^{\mathrm{T}}, \alpha_{r,1}, \ldots, \alpha_{r,L})^{\mathrm{T}}$ is a vector of unknown parameters to be estimated, $D_{i,l}$ are period indicators, and L stands for the number of risk periods.

Note also that for a given l , $Z^* = (Z_1, Z_2, Y_{\mathrm{mis}}^{(l)})$ and that $Z = (Z^*, D)$.

4. RESULTS

This section discusses the model results. The estimates are included in tables 4.1, 4.2 and 4.3. Table 4.1 presents the main results which were calculated by combining the estimates of 20 imputed data sets.[2] Unless otherwise stated, the discussion below refers to these results. Table 4.2 reports the listwise deletion estimates, where the records with missing values were excluded from the analysis. Finally, the results in table 4.3 distinguish between two types of exits into employment, full-time and part-time employment. Apart from the different categories of the dependent variable, this latter model, which produced the results in table 4.3, is similar to that used to produce the estimates in table 4.1 – the results are thus closely comparable.

The section has three main objectives: to assess the effects of the different factors on the conditional probability of exiting unemployment into the different exits, to assess the state dependence of unemployment, and to assess the effects of imputation on the results. The estimates are displayed as hazard ratios with 'remaining unemployed' as the baseline category. As such, the estimates will be interpreted in terms of relative odds – i.e., relative to the odds of remaining unemployed. For example, statements such as 'variable x is associated with an increase in the relative odds of exiting into employment' would mean that, controlling for the effects of the other variables, variable x is associated with an increase in the odds of exiting into employment *vis-a-vis* the odds of remaining unemployed. The base for each categorical variable is included in brackets. Before detailing the estimates, the next subsection provides a short description of the variables included in the modelling.

4.1 Explanatory variables

The variables were selected based on the information available in the file, the objectives of the analysis, relevant literature, and data quality. For example, the analysis does not include the limited (yet, potentially useful) information on previous weekly income and on the reasons for ceasing the last job as this information was available for only one month.

The variables can be divided into four, not necessarily mutually exclusive, groups: (1) personal characteristics (age, gender, marital status, presence of children under four, education, last occupation, last industry, ethnicity, and geographic location); (2) design-specific variables (geographic location, year of interview, and an indicator for the period corresponding to the change in the sample size); (3) a missing variable (education); and (4) period indicators for the estimation of state dependence, which was specified non-parametrically.

2 50 and 100 imputations were also investigated and the results were similar.

4.2 Personal characteristics

Age-groups: The estimates are all significant at the 0.05 significance level. Compared to the oldest group (those aged 55–65 years), all age-groups are associated with higher relative odds of exiting into employment and lower relative odds of exiting the labour force. Overall, the relative odds of exiting into employment tend to decrease with age. The results in table 4.3 indicate that those aged 18–24 years are associated with the highest relative odds of exiting into part-time employment, whereas those aged 25–34 years with the highest relative odds of exiting into full-time employment. The estimates for exiting the labour force do not vary much across the first four age-groups; they are on average 26 per cent lower than the estimates of the oldest group.

Gender: Being male does not have a significant effect on the relative odds of exiting into employment at the 0.1 significance level, however, when employment is disaggregated into full- and part-time employment (see table 4.3), both estimates are significant. Keeping the other variables fixed, being male increases the relative odds of exiting into full-time employment by 47 per cent and decreases the relative odds of exiting the labour force by 22 per cent. Females have relative odds of exiting into part-time employment that are 43 per cent higher than those of males.

Marital Status and Presence of Children: On average, being married is associated with a 36 per cent increase in the relative odds of securing employment, whereas having one or more children under the age of four, increases the relative odds of exiting the labour force by 37 per cent. The results in table 4.3 indicate that having children under the age of four increases the relative odds of exiting into part-time employment by 22 per cent; the estimates are not significant for exiting into full-time employment.

Educational Attainment: The results indicate that having higher education (in particular, Bachelor and above) is associated with higher relative odds of exiting into employment, in particular full-time employment. Compared to those with Secondary education or lower, those with at least a Bachelor degree have relative odds of exiting into employment that are on average 20 per cent higher. Also, compared to those with Secondary education or lower, those with a TAFE education or higher have lower relative odds of exiting the labour force, although only the TAFE coefficient is significant.

Last Occupation:[3] The results indicate that having worked in a high-skilled or middle-skilled previous occupation increases the relative odds of securing employment by 15 and 17 per cent, respectively (i.e., relative to having worked in a lower-skilled previous

---

3 This study adopts the categories defined in the Productivity Commission (2014) report. They are as follows: high-skilled occupations: managers and professionals; middle-skilled occupations: technicians and trade workers, community and personal services workers, and clerical and administrative workers; and lower-skilled occupations: sales workers, machinery operators and drivers, and labourers.

occupation). The results from table 4.3 indicate that these increases are mainly for the relative odds of exiting into full-time employment. Compared to those with a lower-skilled previous occupation, those with middle- or high-skilled previous occupations have relative odds of exiting into full-time employment that are more than 40 per cent higher.

Industry of Last Job: Based solely on the magnitude of the coefficients, having worked in Education and Training increases the relative odds of exiting into employment by the highest margin. This is followed by the Health Care and Social Assistance and Construction industries.

Born Overseas: The coefficient is highly significant and it indicates that being born overseas decreases the relative odds of exiting into employment by around 14 per cent. In the light of the estimates included in table 4.3, the decrease is for both, the relative odds of exiting into full-time as well as part-time employment. The estimates for exiting out of the labour force, on the other hand, are not statistically significant.

Previous Labour Force Status: Compared to those who were previously employed, those who were previously out of the labour force, have significantly much lower odds of securing employment and much higher odds of re-entering the 'out of the labour force' state. Based on the estimates, and keeping the other variables constant, those who last worked more than two years ago and those who are looking for work for the first time have relative odds of exiting into employment that are around 90 per cent lower than the odds of those who were previously employed. Their relative odds of exiting the labour force are also substantially different, with the relative odds being more than 80 per cent higher than the odds of those previously employed.

4.3 Design-related estimates

State of Residence: Compared to the rest of the states and territories, and referring to the period under consideration, residing in Western Australia, Northern Territory, or the Australian Capital Territory increases the relative odds of exiting into employment by at least 20 per cent. These states, together with Queensland, are also associated with the highest relative odds of exiting into full-time employment (see table 4.3). The estimates for exiting out of the labour force indicate that, relative to the other states and territories, residing in South Australia or Queensland decreases the relative odds by at least 11 per cent. Western Australia and the Northern Territory, on the other hand, are associated with the highest increases.

Year of the interview: Compared to 2008, the next two years are associated with lower relative odds of exiting into employment. Compared to 2008, 2010 is also associated with significantly lower relative odds of exiting the labour force.

Change in sample size: Based on the main estimates, the sample size reduction during 2008–2009 does not significantly (i.e., at the 0.05 significance level) affect the relative odds of exiting into employment or out of the labour force.

4.4 Baseline hazard function and model comparisons

The results indicate that the conditional probability of exiting unemployment depends on the current spell of unemployment. The estimates are almost all highly significant at the 0.01 significance level. The relative odds of exiting into employment decrease with the time spent in unemployment and the rate of change is quite steep at first and then tends to smooth out. Generally, the relative odds of exiting the labour force also decrease over time.

When compared, the multiple imputation estimates (table 4.1) and the listwise deletion estimates (table 4.2) are similar, with respect to both their significance and magnitude. The estimates for education, the variable imputed, also almost mirror each other. When they differ, the multiple imputation results, in general, tend to be more significant. The additional sample size used in the multiple imputation model may explain this. In the context of this paper, the multiple imputation approach is appealing in that it incorporates the missingness of education in the analysis and therefore addresses one of the limitations of the listwise deletion model.

4.1 Multiple imputation modelling results

Variables	*Employment*	*Out of the LF*
	Coefficient	Coefficient
Age-group (*55–65 years*)		
18–24 years	1.56 ***	0.74 ***
25–34 years	1.36 ***	0.72 ***
35–44 years	1.18 **	0.77 ***
45–54 years	1.18 **	0.73 ***
Sex (*Female*)		
Male	0.97	0.79 ***
Marital Status (*Not married*)		
Married	1.36 ***	1.07
Children (*No children*)		
Children	1.10	1.37 ***
Education (*Secondary)*		
TAFE	1.07	0.79 ***
Bachelor+	1.20 **	0.88
Last Occupation (*Lower-skilled Occupation*)		
Middle-skilled Occupation	1.15 **	1.04
High-skilled Occupation	1.17 **	0.96
State (*Victoria*)		
New South Wales	0.98	0.96
Queensland	1.02	0.82 **
South Australia	0.88 *	0.85 *
Western Australia	1.23 ***	1.19 **
Tasmania	1.00	1.04
Aust. Capital Territory	1.25 *	1.05
Northern Territory	1.46 ***	1.34 **
Year (*2008*)		
2009	0.79 ***	0.93
2010	0.85 **	0.80 ***
Previous Labour Force Status (*Employed*)		
Out of the Labour Force (Other)	0.62 ***	1.65 ***
Last worked more than 2 years ago	0.09 ***	1.90 ***
First time looking for work	0.11 ***	1.80 ***
Born Overseas (*No*)		
Yes	0.86 ***	0.96

4.1 Multiple imputation modelling results – continued

Variables	*Employment*	*Out of the LF*
	Coefficient	Coefficient
Industry (*Construction*)		
Agriculture	0.90	0.91
Mining	0.94	0.72
Manufacturing	0.67 ***	0.68 ***
Professional, Scientific & Technical Services	0.79 **	0.83
Electricity	0.51 ***	0.78
Wholesale	0.68 ***	0.67 **
Retail	0.72 ***	0.89
Accommodation	0.75 ***	1.06
Transport	0.79 **	0.84
Telecommunications	0.74 *	0.66 *
Financial	0.84	0.71
Real Estate	0.73 *	0.68
Administrative and Support Services	0.92	1.03
Public Administration	0.66 ***	0.99
Education and Training	1.20	1.03
Health Care	1.00	0.86
Arts and Recreation Services	0.95	1.11
Other Services	0.93	0.96
Sample size reduction (*No*)		
Yes	0.89 *	0.88 *
Time interval		
1	1.05	0.67 **
2	0.57 ***	0.41 ***
3	0.43 ***	0.40 ***
4	0.38 ***	0.34 ***
5	0.32 ***	0.27 ***
6	0.22 ***	0.29 ***

Significance levels: *=0.1, **=0.05, ***=0.01

4.2 Listwise deletion modelling results

Variables	*Employment*		*Out of the LF*	
	Coefficient		Coefficient	
Age-group (*55–65 years*)				
18–24 years	1.47	***	0.74	***
25–34 years	1.32	***	0.69	***
35–44 years	1.17		0.72	***
45–54 years	1.21	**	0.66	***
Sex (*Female*)				
Male	0.92		0.74	***
Marital Status (*Not married*)				
Married	1.33	***	1.09	
Children (*No children*)				
Children	1.09		1.29	***
Education (*Secondary)*				
TAFE	1.05		0.78	***
Bachelor+	1.21	**	0.89	
Last Occupation (*Lower-skilled Occupation*)				
Middle-skilled Occupation	1.14	**	0.96	
High-skilled Occupation	1.12		0.91	
State (*Victoria*)				
New South Wales	1.01		1.03	
Queensland	1.02		0.86	*
South Australia	0.90		0.98	
Western Australia	1.27	***	1.36	***
Tasmania	1.02		1.06	
Aust. Capital Territory	1.35	**	1.11	
Northern Territory	1.61	***	1.61	***
Year (*2008*)				
2009	0.81	***	0.92	
2010	0.88	*	0.78	***
Previous Labour Force Status (*Employed*)				
Out of the Labour Force (Other)	0.67	***	1.70	***
Last worked more than 2 years ago	0.08	***	1.92	***
First time looking for work	0.10	***	1.65	***
Born Overseas (*No*)				
Yes	0.88	**	0.98	

4.2 Listwise deletion modelling results – continued

Variables	*Employment*		*Out of the LF*	
	Coefficient		Coefficient	
Industry (*Construction*)				
Agriculture	0.79		0.78	
Mining	0.85		0.58	*
Manufacturing	0.63	***	0.68	***
Professional, Scientific & Technical Services	0.73	**	0.86	
Electricity	0.57	**	0.95	
Wholesale	0.65	***	0.61	**
Retail	0.67	***	0.87	
Accommodation	0.69	***	1.04	
Transport	0.78	*	0.99	
Telecommunications	0.65	**	0.66	
Financial	0.76	*	0.65	*
Real Estate	0.79		0.75	
Administrative and Support Services	0.75	**	0.92	
Public Administration	0.65	***	0.93	
Education and Training	1.17		0.98	
Health Care	0.93		0.85	
Arts and Recreation Services	0.82		0.95	
Other Services	0.86		1.05	
Sample size reduction (*No*)				
Yes	0.89		0.84	**
Time Interval				
1	1.13		0.74	*
2	0.59	***	0.44	***
3	0.48	***	0.45	***
4	0.42	***	0.40	***
5	0.32	***	0.28	***
6	0.19	***	0.33	***

Significance levels: *=0.1, **=0.05, ***=0.01

4.3 Multiple imputation modelling results – Exits (Full-time / Part-time / Out of the LF)

Variables	*Full-time*	*Part-time*	*Out of the LF*
	Coefficient	*Coefficient*	*Coefficient*
Age-group (*55–65 years*)			
18–24 years	1.88 ***	1.33 ***	0.74 ***
25–34 years	2.03 ***	0.96	0.72 ***
35–44 years	1.57 ***	0.93	0.77 ***
45–54 years	1.52 ***	0.96	0.73 ***
Sex (*Female*)			
Male	1.47 ***	0.70 ***	0.78 ***
Marital status (*Not married*)			
Married	1.49 ***	1.26 ***	1.07
Children (*No children*)			
Children	0.94	1.22 **	1.37 ***
Education (*Secondary)*			
TAFE	1.08	1.05	0.79 ***
Bachelor+	1.26 **	1.16	0.88
Last occupation (*Lower-skilled occupation*)			
Middle-skilled occupation	1.42 ***	0.94	1.03
High-skilled occupation	1.50 ***	0.91	0.95
State (*Victoria*)			
New South Wales	1.11	0.89	0.96
Queensland	1.23 **	0.87 *	0.82 ***
South Australia	0.90	0.87	0.85 *
Western Australia	1.62 ***	0.95	1.18 *
Tasmania	0.95	1.00	1.04
Aust. Capital Territory	1.56 ***	1.06	1.04
Northern Territory	2.39 ***	0.89	1.33 *
Year (*2008*)			
2009	0.76 ***	0.82 ***	0.93
2010	0.78 ***	0.91	0.81 ***
Previous Labour Force status (*Employed*)			
Out of the Labour Force (Other)	0.46 ***	0.78 ***	1.65 ***
Last worked more than 2 years ago	0.04 ***	0.16 ***	1.90 ***
First time looking for work	0.05 ***	0.20 ***	1.80 ***
Born overseas (*No*)			
Yes	0.83 ***	0.89 *	0.96

4.3 Multiple imputation modelling results – Exits (Full-time / Part-time / Out of the LF) – continued

Variables	*Full-time*	*Part-time*	*Out of the LF*
	Coefficient	*Coefficient*	*Coefficient*
Industry (*Construction*)			
Agriculture	0.76	1.19	0.91
Mining	1.16	0.50 **	0.72
Manufacturing	0.71 ***	0.68 ***	0.68 ***
Professional, Scientific & Technical Services	0.72 **	0.94	0.83
Electricity	0.71	0.26 ***	0.77
Wholesale	0.70 **	0.72 *	0.67 **
Retail	0.59 ***	0.97	0.89
Accommodation	0.39 ***	1.32 **	1.07
Transport	0.68 ***	1.02	0.84
Telecommunications	0.68 *	0.90	0.66 *
Financial	0.90	0.82	0.71 *
Real Estate	0.77	0.80	0.68
Administrative and Support Services	0.72 **	1.31 *	1.03
Public Administration	0.50 ***	0.98	1.00
Education and Training	0.50 ***	2.38 ***	1.04
Health Care	0.59 ***	1.70 ***	0.86
Arts and Recreation Services	0.47 ***	1.77 ***	1.11
Other Services	0.81	1.20	0.96
Sample size reduction (*No*)			
Yes	0.87	0.91	0.88 *
Time interval			
1	0.31 ***	0.66 **	0.68 **
2	0.17 ***	0.35 ***	0.41 ***
3	0.12 ***	0.28 ***	0.41 ***
4	0.12 ***	0.23 ***	0.34 ***
5	0.10 ***	0.18 ***	0.27 ***
6	0.02 ***	0.22 ***	0.29 ***

Significance levels: *=0.1, **=0.05, ***=0.01

5. CONCLUDING REMARKS

This paper examined the duration of unemployment for Australians aged 18–65 years, over a three-year period, from the beginning of 2008 to the end of 2010. The analysis was conducted on a sample of more than 14,000 individuals, obtained from the ABS LFS Longitudinal file. These individuals were followed for up to eight consecutive months. The focus was on the first unemployment spells.

Empirically, the paper had two main objectives: (1) to examine the effects of a number of factors on the conditional probability of exiting unemployment, distinguishing between the different exits (full-time/part-time employment and out of the labour force), and (2) to examine the state dependence of unemployment. One contribution of this paper is its inclusion of design-related variables in the analysis, alongside other empirical factors. Although the design-related effects are often ignored in practice, they could play important roles in the analysis of unemployment duration.

Methodologically, the study adopted a multinomial framework and opted for a non-parametric specification of the baseline hazard function. As the responses for one of the key variables were missing for around 24 per cent of cases, the paper implemented a multiple imputation approach to impute the missing values. The imputations were obtained using a Bayesian simulation algorithm. This constitutes another contribution to the literature in that, unlike the majority of studies which discard all records with missing information, this study paid considerable attention to this issue, closely examined it, and showcased a plausible way of utilising all the available information in the analysis.

The multiple imputation provided two important benefits to the analysis. The first relates to the preservation of the sample size, which led to more precise estimates. As a second benefit, multiple imputation addressed the fact that the missing values were not missing completely at random.

The results indicate that the demographic factors as well as the information regarding the period prior to unemployment play important roles in explaining the exits out of unemployment. Of particular interest are the estimates of previous labour force status. These indicate that, compared to those who were previously employed and controlling for the effects of the other variables, those who last worked more than two years ago and those who are looking for work for the first time have considerably lower relative odds of exiting into employment and considerably higher relative odds of exiting the labour force. These results merit some further investigation.

REFERENCES

Australian Bureau of Statistics (2013) *Information Paper: Labour Force Survey Sample Design, May 2013*, cat. no. 6269.0, ABS, Canberra. < http://www.abs.gov.au/ausstats/abs@.nsf/mf/6269.0 >

Productivity Commission (2014) *Geographic Labour Mobility*, Research Report, Canberra.

Rotaru, C.I. (2014a) "Transitioning out of Unemployment: Analysis using the ABS Longitudinal Labour Force Survey File', *Australian Journal of Labour Economics*, 17(2), pp. 111–137.

Rotaru, C.I. (2014b) "Modelling the Short-Term Dynamics of Unemployment Using the ABS Longitudinal Labour Force Survey File", *Methodology Advisory Committee Papers*, cat. no. 1352.0.55.142, Australian Bureau of Statistics, Canberra. < http://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.142 >

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York.

van Buuren, S. (2012) *Flexible Imputation of Missing Data*, Chapman & Hall/CRC, Boca Raton, FL.

Raghunathan, T.E.; Lepkowski, J.M.; van Hoewyk, J. and Solenberger, P. (2001) "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models", *Survey Methodology*, 27(1), pp. 85–95.

Sen, A. (1997) "Inequality, Unemployment and Contemporary Europe", *International Labour Review*, 136(2), pp. 155–171.

ACKNOWLEDGEMENTS

The author is grateful to Ruel Abello, Siu-Ming Tam, Sybille McKeown, Bernard Williams, Franklin Soriano, Professor James Brown, and Professor David Steel for their valuable suggestions and comments. Thanks also go to Peter Rossiter for his help in editing and formatting the paper. Remaining errors are the author's.

FOR MORE INFORMATION . . .

INTERNET **www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free
of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS www.abs.gov.au