SUPPORT DOCUMENT

From education to employment: how long does it take? Support document

DARCY FITZPATRICK, LAURENCE LESTER,

KOSTAS MAVROMARAS, SUE RICHARDSON AND YAN SUN

This document was produced by the authors based on their research for the report *From education to employment: how long does it take?*, and is an added resource for further information. The report is available on NCVER's website: <http://www.ncver.edu.au>

Contents

Tables and figures

Attrition estimations

Attrition in panel data sets is defined as the rate at which people who are interviewed in one wave drop out in the next wave. Attrition is an unavoidable problem of panel data sets (i.e. data sets that interview the same people repeatedly over a longer period of time). People will drop out for many reasons, such as moving without leaving a forwarding address, death, or just because they decide to not respond to any more requests for an interview. Although the study of attrition in the LSAY Y95 cohort data is beyond the remit of the present analysis, we carry out a number of regressions to assess the level and nature of attrition in the LSAY data subsets we use. Attrition can be a severe problem when (i) it is very prevalent (in which case the sample size may be critically reduced) and/or (ii) it has happened in a systematic way (in which case the remaining sample will stop being representative of the surveys target population). Attrition can be more severe for a data set that samples a single cohort and the introduction of replacement/new subjects is not appropriate. Data sets that begin with a targeted sample of young people with the intention to follow them throughout their lives, such as the LSAY Y95, suffer particularly from attrition. By contrast, conventional household surveys, such as the HILDA survey, have various methods to replace their lost subjects and maintain the surveys representativeness of its intended population. The LSAY Y95 data set has suffered from high attrition in terms of its sample size, to the degree that its representativeness may be compromised. From the complete starting sample (i.e. 13 613 Year 9 students), we are only able to analyse the education and first employment experiences of 7641 individuals. However, even a sample size of a few thousand can be sufficiently informative for statistical analysis, so the remaining sample is considered sufficiently large for estimation purposes. The main limitation of the attrition in the LSAY Y95 data is that disaggregation of the data, into sub-categories that are not very prevalent, cause small number problems. There is very little that one can do about this, except exclude the analysis of such small sized sub-categories, or provide a warning about their lack of statistical significance.

Is attrition non-random?

The remaining concern would be that the observed attrition may have happened in a systematic way. If that is the case then the ability of the sample to represent the population will be endangered and any derived estimates may suffer from bias. This bias may occur due to people dropping out of the sample in ways that are observed (e.g. when more men drop out than women and the data reports the gender of the respondents) or unobserved, by the data. The presence of attrition according to observed characteristics can be established and, to a degree, dealt with; whereas attrition according to unobserved characteristics is much harder to detect and also deal with. Having established a high degree of attrition in the LSAY Y95 dataset, this Appendix presents a number of simple estimations as a preliminary attempt to establish the extent to which this attrition may be systematic in accordance with some observed characteristics of the survey respondents. To establish this we present a binary Probit model to estimate the probability of respondents', present at the 1995 interview, completeness of information on their education achievement and their first job search duration. Table 1 presents the Probit estimation and Table 2 presents the descriptive statistics of those respondents. The dependent variable takes the value of 1 for those that stayed in the sample (i.e. interviewed in 1995, have education completion information and post-education employment or last interview information) and 0 for those who left the sample (i.e. interviewed in 1995, but have no education completion information, or no post-education employment or last interview information). Estimation in

Table 1 allows us a first look at the degree of randomness in the LSAY attrition. We include in the estimation a number of core socio-demographic variables, many of which appear to be statistically significant. The implication of this finding is that the way in which attrition occurred was not random. Table 1 shows clearly that males (*male*), indigenous persons (*indig*), individuals who felt unhappy at school in 1995 (*unhappy*), and individuals with low/poor self-concept of overall ability in 1995 (*ability3*), are less likely to stay in the sample.

Table 1 Probit estimation of attrition

```
Probit regression                                 Number of obs   =      10190
                                                  Wald chi2(18)   =     275.69
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -6789.3493                 Pseudo R2       =     0.0202
```

stayers	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
cobf_2	-.0969079	.0422229	-2.30	0.022	-.1796632	-.0141526
cobf_3	-.0984093	.0441989	-2.23	0.026	-.1850375	-.0117811
cobm_2	-.0401238	.0433216	-0.93	0.354	-.1250325	.0447849
cobm_3	.0667551	.046378	1.44	0.150	-.0241441	.1576543
male	-.1244199	.0254262	-4.89	0.000	-.1742542	-.0745855
indig	-.3173326	.0854328	-3.71	0.000	-.4847779	-.1498874
ANU3_F	.0023864	.0005819	4.10	0.000	.0012459	.003527
ability1	.2434406	.0263	9.26	0.000	.1918934	.2949877
ability3	-.2242366	.0728014	-3.08	0.002	-.3669248	-.0815484
private	.0801667	.0274421	2.92	0.003	.0263811	.1339522
unhappy	-.1015951	.0296673	-3.42	0.001	-.159742	-.0434482
VIC	.1658092	.0379114	4.37	0.000	.0915043	.2401141
QLD	.1455459	.0399643	3.64	0.000	.0672172	.2238745
SA	.1623816	.0441603	3.68	0.000	.0758289	.2489342
WA	.1709459	.0434776	3.93	0.000	.0857315	.2561603
TAS	.0991188	.068227	1.45	0.146	-.0346036	.2328412
ACT	.0473841	.0649917	0.73	0.466	-.0799972	.1747655
NT	.1197658	.0845449	1.42	0.157	-.0459392	.2854707
_cons	-.0304031	.0409868	-0.74	0.458	-.1107359	.0499296

However, the estimated results in Table 1 show that the level of explanatory power of the observed characteristics are limited. In precise terms, we find that only 2% of the total variation in the attrition variable (*stayers*) can be explained by all the explanatory variables in the estimation. Although this may appear as a small percentage, it should be accompanied by the caveat that Probit estimation in large samples rarely achieves a high explanatory power, as measured by the Pseudo R^2 estimator.

Table 2 presents the coefficients of the explanatory variables, included in the Probit estimation, in a way that they can be interpreted as probabilities. For example, Table 2 suggests that: (i) males (*male*) are 4.85% more likely to have dropped out of the sample, relative to females; (ii) people that attended a private school in 1995 (*private*) are 3.12% less likely to have dropped out, relative to their publicly educated peers. All other variables can be interpreted in a similar fashion as probabilities. It should be noted that this is an indicative estimation only.

The summary and descriptive statistics of the dependent and explanatory variables included in the attrition probit estimation are shown in Table 3.

To summarise, the probit estimation in Table 1 suggests that non-random attrition is present in the data we analyse, but it also indicates that the resulting bias may not be as damaging as we initially expected. This is further investigated with additional structure to the estimation procedure used, see below.

Table 2 Probit estimation of survey attrition: marginal effects

Marginal effects after probit
y = Pr(stayers) (predict)
= .58237272

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
cobf_2*	-.0380836	.01668	-2.28	0.022	-.070785	-.005382	.120805
cobf_3*	-.0386365	.01744	-2.22	0.027	-.072809	-.004464	.182139
cobm_2*	-.0157125	.01701	-0.92	0.356	-.049058	.017633	.111973
cobm_3*	.0259281	.01791	1.45	0.148	-.009181	.061037	.160844
male*	-.048567	.00992	-4.90	0.000	-.068004	-.02913	.481943
indig*	-.1258684	.03389	-3.71	0.000	-.192296	-.059441	.022866
ANU3_F	.0009317	.00023	4.10	0.000	.000486	.001377	38.6115
ability1*	.0949088	.01021	9.29	0.000	.074891	.114926	.519431
ability3*	-.0887987	.02902	-3.06	0.002	-.14568	-.031918	.032581
private*	.0312139	.01065	2.93	0.003	.010335	.052092	.358292
unhappy*	-.0398384	.01168	-3.41	0.001	-.062724	-.016953	.257802
VIC*	.0639509	.01441	4.44	0.000	.035703	.092199	.21158
QLD*	.0561697	.01522	3.69	0.000	.026347	.085993	.181256
SA*	.0624268	.01667	3.75	0.000	.029758	.095096	.128361
WA*	.0656861	.01639	4.01	0.000	.033563	.097809	.138371
TAS*	.0382776	.02602	1.47	0.141	-.012728	.089283	.04053
ACT*	.0184106	.02512	0.73	0.464	-.030828	.06765	.045535
NT*	.0461075	.03202	1.44	0.150	-.016655	.10887	.025221

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Table 3 Summary statistics of variables used in the Probit estimation of attrition

Variable	Obs	Mean	Std. Dev.	Min	Max
stayers	10190	.5807655	.493458	0	1
cobf_1	10190	.6970559	.4595538	0	1
cobf_2	10190	.1208047	.3259162	0	1
cobf_3	10190	.1821394	.3859783	0	1
cobm_2	10190	.1119725	.3153481	0	1
cobm_2	10190	.1119725	.3153481	0	1
cobm_3	10190	.160844	.367405	0	1
male	10190	.4819431	.4996984	0	1
indig	10190	.0228656	.1494822	0	1
ANU3_F	10190	38.6115	22.7625	0	100
ability1	10190	.5194308	.4996468	0	1
ability2	10190	.4479882	.4973118	0	1
ability3	10190	.032581	.1775459	0	1
private	10190	.3582924	.4795222	0	1
unhappy	10190	.2578018	.4374458	0	1
NSW	10190	.2291462	.4203041	0	1
VIC	10190	.21158	.4084486	0	1
QLD	10190	.1812561	.3852492	0	1
SA	10190	.1283611	.3345079	0	1
WA	10190	.138371	.3453059	0	1
TAS	10190	.0405299	.1972082	0	1
ACT	10190	.0455348	.2084842	0	1
NT	10190	.0252208	.1568028	0	1

Does the non-random attrition influence search duration estimates?

Having established the non-random nature of the attrition in the LSAY Y95 data, the pertinent question is the degree to which the attrition may bias our subsequent analysis of search duration. Selection in duration estimation can be extremely complex and is best handled with double hurdle models. However, such an econometric investigation is beyond the scope of this analysis. Instead, we first estimate a simple selection-correction model (often referred to as the Heckman correction model) to provide a simple indication about the likelihood that the non-random selection revealed in the attrition estimation may bias the results of the subsequent estimation of duration of the first job search. The estimated results of the two-step procedure are presented in Table 4. The first step of the estimation is the same as the aforementioned single step probit estimation, in Table 1 (numbers will not agree completely as this estimation is solved numerically and not analytically). The Heckman procedure uses the results from the first step to calculate a correction term, commonly referred to as the Inverse Mills Ratio (IMR), which is then included in the second stage as an explanatory variable. The specification in the second step of the Heckman procedure is an OLS estimation of first job search duration (i.e. the length of time from the completion of highest education to the first period of employment).

While the econometrics behind this result may be too complex for the non-technical reader, the interpretation is very simple: one only has to look at the statistical significance of the IMR variable in the second step. A significant IMR suggests that there is sufficient selection bias and that the inclusion of the IMR has corrected it. Where we see a significant IMR it is always advisable to check if the remaining estimated coefficients in the second step change as a result of its inclusion/exclusion. Table 4 very clearly suggests that the IMR variable (under the name of *lambda* at the bottom of the table) is clearly not respectively significant (with a t-ratio of 0.63 which translates into a p-value of 0.53).

Table 4 Heckman two-step selection model of attrition and subsequent job search duration to first period of employment

Heckman selection model -- two-step estimates	Number of obs	=	10190
(regression model with sample selection)	Censored obs	=	4272
	Uncensored obs	=	5918
	Wald chi2(29)	=	1737.32
	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
duration1						
postgrad	-25.97976	1.371197	-18.95	0.000	-28.66726	-23.29227
grad_dip	-25.20488	1.182977	-21.31	0.000	-27.52347	-22.88629
bachelor	-25.22728	.6843934	-36.86	0.000	-26.56866	-23.88589
adv_dip	-25.12225	.8859897	-28.36	0.000	-26.85876	-23.38574
cert_IV	-24.54597	1.115456	-22.01	0.000	-26.73222	-22.35971
cert_III	-24.79276	.9440943	-26.26	0.000	-26.64315	-22.94237
cert_II	-25.31707	1.133399	-22.34	0.000	-27.53849	-23.09564
cert_I	-24.40302	1.476252	-16.53	0.000	-27.29642	-21.50962
yr_12	-22.93544	.6254582	-36.67	0.000	-24.16131	-21.70956
yr_11	-23.05058	.7749219	-29.75	0.000	-24.5694	-21.53176
cobf_2	-.9765072	.5948116	-1.64	0.101	-2.142317	.1893022
cobf_3	.270485	.6004803	0.45	0.652	-.9064348	1.447405
cobm_2	-.2387972	.564913	-0.42	0.673	-1.346006	.8684119
cobm_3	1.88792	.5941033	3.18	0.001	.7234992	3.052341
male	-.1474016	.4785637	-0.31	0.758	-1.085369	.790566
indig	3.140461	1.5732	2.00	0.046	.0570443	6.223877
ANU3_F	.0005165	.0102746	0.05	0.960	-.0196212	.0206543
ability1	.4985223	.7672108	0.65	0.516	-1.005183	2.002228
ability3	-.2835531	1.296241	-0.22	0.827	-2.824139	2.257033
DISAB	.3054125	1.203402	0.25	0.800	-2.053211	2.664036
VIC	-.6880096	.6560626	-1.05	0.294	-1.973869	.5978494
QLD	.08414	.6590547	0.13	0.898	-1.207583	1.375863
SA	-.1771601	.7195486	-0.25	0.806	-1.587449	1.233129
WA	-.2875259	.7226738	-0.40	0.691	-1.703941	1.128889
TAS	-.4590264	.9287292	-0.49	0.621	-2.279302	1.361249
ACT	-.6056951	.8469323	-0.72	0.475	-2.265652	1.054262
NT	-2.779581	1.126437	-2.47	0.014	-4.987357	-.5718056
regional	.7378336	.4032834	1.83	0.067	-.0525874	1.528255
rural	.8280449	.4384366	1.89	0.059	-.0312751	1.687365
_cons	24.08369	3.968181	6.07	0.000	16.3062	31.86118
stayers						
cobf_2	-.0969079	.0421348	-2.30	0.021	-.1794906	-.0143252
cobf_3	-.0984093	.0443901	-2.22	0.027	-.1854124	-.0114062
cobm_2	-.0401238	.0434318	-0.92	0.356	-.1252486	.0450009
cobm_3	.0667551	.0465892	1.43	0.152	-.0245582	.1580683
male	-.1244199	.025433	-4.89	0.000	-.1742676	-.0745722
indig	-.3173326	.0851437	-3.73	0.000	-.4842112	-.1504541
ANU3_F	.0023864	.0005805	4.11	0.000	.0012487	.0035242
ability1	.2434406	.0263034	9.26	0.000	.1918868	.2949943
ability3	-.2242366	.0724857	-3.09	0.002	-.3663061	-.0821672
private	.0801667	.0274178	2.92	0.003	.0264289	.1339045
unhappy	-.1015951	.0296451	-3.43	0.001	-.1596984	-.0434918
VIC	.1658092	.0380228	4.36	0.000	.0912859	.2403325
QLD	.1455459	.0399995	3.64	0.000	.0671482	.2239435
SA	.1623816	.0442257	3.67	0.000	.0757009	.2490623
WA	.1709459	.0432403	3.95	0.000	.0861964	.2556954
TAS	.0991188	.0684666	1.45	0.148	-.0350732	.2333108
ACT	.0473841	.0644603	0.74	0.462	-.0789557	.173724
NT	.1197658	.0834853	1.43	0.151	-.0438624	.2833939
_cons	-.0304031	.0409304	-0.74	0.458	-.1106252	.0498189
mills						
lambda	2.936336	4.677263	0.63	0.530	-6.23093	12.1036
rho	0.24106					
sigma	12.180947					
lambda	2.9363364	4.677263				

The Heckman procedure shows that (i) where there is selection bias (which in this case could be resulting from attrition) that is due to observable characteristics and (ii) where these characteristics have been correctly included in both steps of the estimation[1], the inclusion of the IMR, in the second step, corrects for the selection bias.

The implication of the estimation results in Table 4 is that the non-randomness of the attrition in the LSAY Y95 data is of no consequence on the estimated coefficients in the first period of job search duration specification. One caveat to this is that there could be a large number of people transiting directly from their highest level of education attainment to their first period of employment. Hence, there are many durations that take the value of 0 in the data (i.e. no job search took place). Therefore, we also estimated the Heckman two-step procedure using a Tobit estimation technique in the second step and found that the significance of the IMR variable was equally low (t-ratio of 0.73, which translates into a p-value of 0.47). The advantage of using a Tobit estimation is that it considers the bunching up of many zeros in the dependent variable (*duration1*). This estimation was repeated using different combinations of explanatory variables[2], only to find that the results were largely consistent with the main result of this Appendix. Finally, as stated in the introduction of this appendix, these estimations should be treated as very preliminary results. A more comprehensive analysis of attrition in the context of first job search duration would be recommended, although it is not clear at this stage how far the information contained in the data would be able to support it. The problem of attrition, however, is best prevented through maintaining sample sizes during the survey period, rather than corrected in retrospect.

[1] Some observable characteristics may only be suitable for one-step, but not the other: typically, it is helpful to have such characteristics—called exclusion restrictions—for correct estimation.

[2] The Tobit estimation technique can on occasion be sensitive to the model specification.

Table 5 Education attainment by indigenous status, % (2006)

	Male		Female		Total	
	Non-indigenous	Indigenous	Non-indigenous	Indigenous	Non-indigenous	Indigenous
Postgrad	3	1	4	2	4	2
Bachelor	20	6	27	4	24	5
Adv dip, dip	5	3	6	2	5	3
Cert IV	2	1	3	3	3	3
Cert III	3	1	6	2	4	2
Cert I & II	4	10	4	6	4	8
Year 12	40	36	38	30	39	33
Year 11	12	14	6	16	8	15
Year 10	11	26	7	34	9	30
Total (%)	100	100	100	100	100	100
Total (number)	3258	69	3770	89	7028	158

Table 6 Education attainment by disability status, % (2006)

	Male		Female		Total	
	No disability	Disabled	No disability	Disabled	No disability	Disabled
Postgrad	3	0	4	8	4	3
Bachelor	19	7	27	13	23	10
Adv dip, dip	5	2	6	6	5	4
Cert IV	2	2	3	4	3	3
Cert III	3	9	5	8	4	8
Cert I & II	4	6	4	6	4	6
Year 12	40	32	38	32	39	32
Year 11	12	21	6	8	9	16
Year 10	11	20	7	17	9	19
Total (%)	100	100	100	100	100	100
Total (number)	3296	81	3801	53	7097	134

Table 7a Education attainment by socioeconomic status of father, % (2006)

	Males				Females			
	High: Upper Prof and Managers	Lower Prof and Managers, Para Profs Technical	Trades, Clerks, Sales Reps and Farmers	Low: Sales Assistants, Plant Ops, Labs	High: Upper Prof and Managers	Lower Prof and Managers, Para Profs Technical	Trades, Clerks, Sales Reps and Farmers	Low: Sales Assistants, Plant Ops, Labs
Postgrad	4	5	2	3	8	5	4	3
Bachelor	34	24	16	13	40	32	25	20
Adv dip, dip	4	5	5	5	4	6	6	7
Cert IV	3	2	3	2	3	2	3	3
Cert III	3	3	3	3	3	5	6	7
Cert I & II	2	2	5	5	2	3	4	4
Year 12	42	45	40	36	35	37	39	39
Year 11	5	7	13	17	2	5	7	8
Year 10	3	7	13	16	3	4	7	9
Total (%)	100	100	100	100	100	100	100	100
Total (number)	473	705	1,130	744	502	794	1281	912

Table 7b Education attainment by socioeconomic status of mother, % (2006)

	Males				Females			
	High: Upper Prof and Managers	Lower Prof and Managers, Para Profs Technical	Trades, Clerks, Sales Reps and Farmers	Low: Sales Assistants, Plant Ops, Labs	High: Upper Prof and Managers	Lower Prof and Managers, Para Profs Technical	Trades, Clerks, Sales Reps and Farmers	Low: Sales Assistants, Plant Ops, Labs
Postgrad	6	5	2	2	4	6	5	4
Bachelor	29	26	19	15	39	37	27	19
Adv dip, dip	2	5	6	6	5	5	6	7
Cert IV	3	2	3	3	2	1	3	4
Cert III	4	3	3	4	3	4	6	6
Cert I & II	4	4	4	5	1	3	3	3
Year 12	41	41	40	38	38	36	38	38
Year 11	8	8	13	14	4	5	6	8
Year 10	4	7	11	14	5	3	6	10
Total (%)	100	100	100	100	100	100	100	100
Total (number)	140	796	937	596	137	887	1184	728

Table 8a Education attainment by father's level of education, % (2006)

	Male					Female				
	No secondary school	Some secondary school	All years of secondary school	Trade or technical qualification	Degree or dip	No secondary school	Some secondary school	All years of secondary school	Trade or technical qualification	Degree or dip
Postgrad	3	1	3	2	5	1	4	5	5	7
Bachelor	14	13	18	18	33	19	20	26	26	42
Adv dip, dip	7	5	5	6	3	9	6	7	5	4
Cert IV	1	3	1	3	3	6	3	3	4	1
Cert III	2	4	3	4	2	3	7	5	4	3
Cert I & II	4	4	4	5	3	4	5	4	3	1
Year 12	36	41	43	38	44	36	38	39	38	38
Year 11	18	15	11	12	4	12	7	4	8	3
Year 10	14	14	12	12	4	9	10	7	6	1
Total (%)	100	100	100	100	100	100	100	100	100	100
Total (number)	91	636	426	596	767	95	860	501	608	789

Table 8b Education attainment by mother's level of education, % (2006)

	Male					Female				
	No secondary school	Some secondary school	All years of secondary school	Trade or technical qualification	Degree or dip	No secondary school	Some secondary school	All years of secondary school	Trade or technical qualification	Degree or dip
Postgrad	1	2	3	1	5	1	4	6	5	6
Bachelor	10	16	20	24	29	17	21	25	32	41
Adv dip, dip	10	5	5	4	4	6	7	5	7	5
Cert IV	1	2	3	2	3	5	4	2	2	1
Cert III	4	3	3	2	4	4	6	7	3	3
Cert I & II	6	4	4	4	3	2	4	5	3	2
Year 12	36	42	40	40	43	43	40	38	37	35
Year 11	10	14	10	14	6	10	6	6	6	4
Year 10	21	11	13	8	4	11	8	7	5	2
Total (%)	100	100	100	100	100	100	100	100	100	100
Total (number)	80	710	799	204	643	93	1112	774	241	796

Table 9 Education attainment by gross hourly pay

	Male					Female				
	5-9	10	11-15	16-30	31-50	5-9	10	11-15	16-30	31-50
Postgrad	0	0	3	9	27	0	0	3	14	50
Bachelor	1	1	22	55	35	6	4	32	62	42
Adv dip, dip	1	2	9	10	12	2	2	13	6	4
Cert IV	0	1	5	3	0	1	2	4	4	4
Cert III	1	1	7	5	0	5	4	8	5	0
Cert I & II	4	2	6	5	8	3	2	6	2	0
Year 12	49	62	43	12	19	56	66	30	7	0
Year 11	22	15	4	1	0	13	11	1	0	0
Year 10	21	14	1	1	0	14	10	1	0	0
Total (%)	100	100	100	100	100	100	100	100	100	100
Total (number)	830	842	627	513	26	774	924	856	655	26

Detailed TPS tabulations

As discussed in the survival and multivariate regression analyses (see main report), there is a clear distinction between the outcomes for the various levels of education completion and across gender. This discussion examines further the influence of these levels of disaggregation on the Total Proportion of Search (TPS)[3].

Building on the discussion of the influence of school and post-school (i.e. university and VET) graduates on the TPS measures (for all persons), disaggregation by gender is investigated. Figure 1 and Figure 2 illustrate the TPS measures by broad education category (university, VET and school graduates), for males and females respectively. Table 10 and Table 11 provide the exact percentages that correspond to Figure 1 and Figure 2, respectively. In general, at the post-school levels of education completion, the proportion of time spent not in employment is greater for males than females. Furthermore, the gender disparity is more persistent at the university level of education completion compared with the VET level. Interestingly, for both males and females, the effectiveness of the university and VET education pathways on employment outcomes also vary.

At the school levels of education completion, the measures of TPS are approximately equal for both the male and female sub-samples, proportionally. Furthermore, as illustrated in Figure 1 and Figure 2, the TPS measures for both male and female school graduates are well below their post-school graduate counterparts. This indicates that a school level education only has a lower effectiveness on sustaining employment, within a three-year period, compared with post-school levels of education completion.

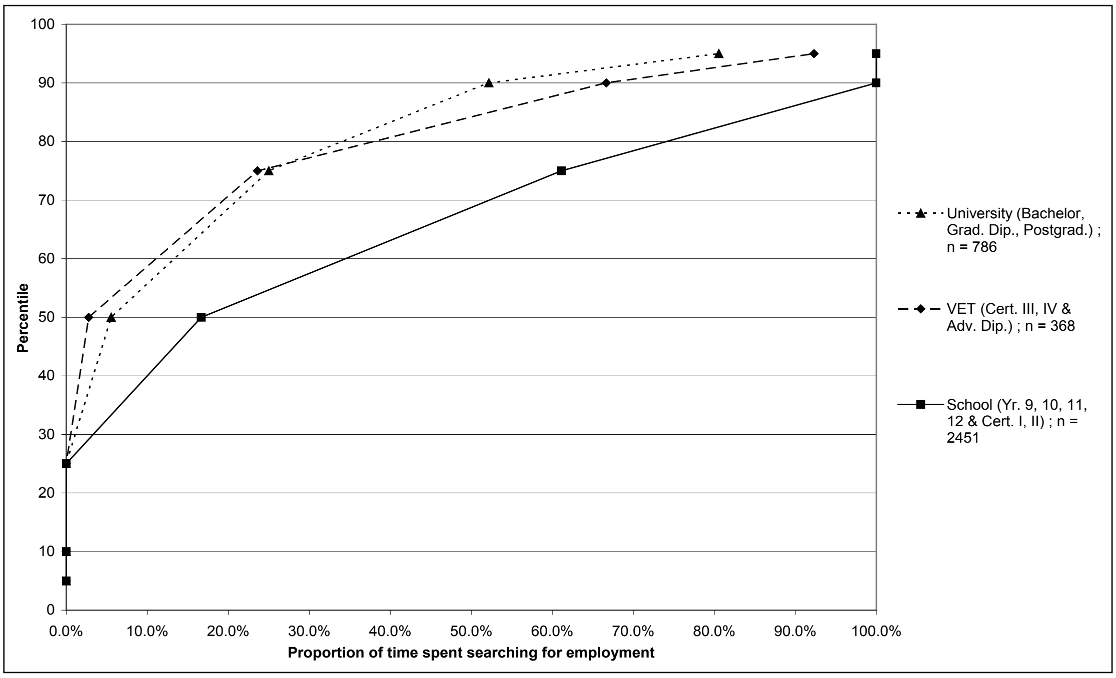
At the university and VET levels of education, there are a number of differences in the TPS measures between the two levels of education and across gender. For female VET graduates, 50% moved immediately (i.e. approximately zero months searching) from education to employment, shown in Table 11; whereas, males spent approximately one month or less (i.e. 2.8% of 36 months) searching for employment, shown in Table 10. Similarly, for female university graduates, 50% spent approximately one month or less searching for employment (see Table 11), while the equivalent proportion of males spent approximately two months or less searching (see Table 10). Therefore, at the 50% level of each sub-sample, firstly, the TPS measures for female university and VET graduates were consistently lower than their male counterparts, and, secondly, for both males and females, the effectiveness of a VET qualification in sustaining employment was higher in comparison to the university level of education.

However, for the university and VET levels of education, there is a switch in the effectiveness of university and VET qualifications on employment outcomes. While the gender gap in the TPS measures persists, the proportion of total job search time becomes lower for both female and male university graduates compared with VET graduates, at the 75% and 90% sub-sample proportions, respectively. For example, out of all the female (male) VET graduates, 50 (75)% spent a total of approximately zero (eight) months (or less) searching for employment, while 50 (75) percent of female (male) university graduates spent approximately one (six) months or less.

[3] The proportion of (total) time spent by respondents not employed (and assumed to be time spent searching for employment) during the first 36 months (three years) since completing their highest level of education within their survey timeframe (i.e. the timeframe that respondents are observed for).

The switch in the effectiveness of the university and VET levels of education are clearly illustrated in Figure 1 for males, between the 70 and 80 percentile indicators, and in Figure 2 for females, between the 50 and 70 percentile indicators. Therefore, a university qualification had a higher effectiveness in reducing the proportion of time spent out of employment compared with a VET qualification, for most people (90%), and both were superior to a school level education.

Figure 1 Total job search duration as a proportion of total survey time (capped at 36 months) for males, by broad education category



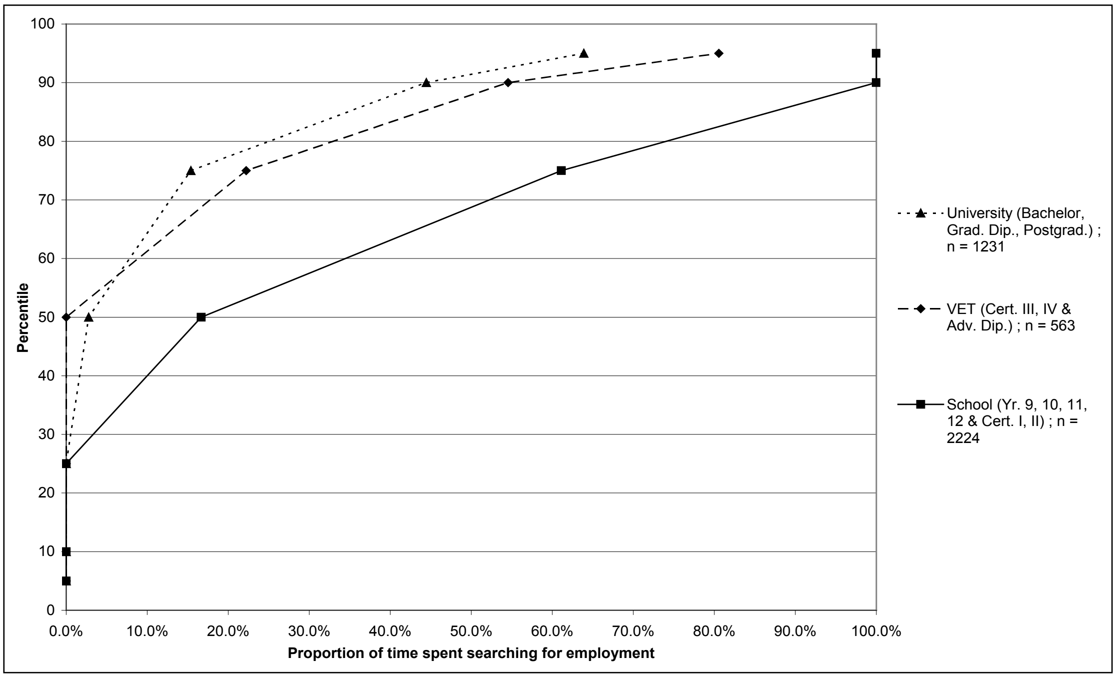
Note: Includes respondents who exited the survey before obtaining employment, i.e. observations censored at the date of the last interview. These comprise approximately 4% of the total.

Table 10 Total job search duration as a proportion of total survey time (capped at 36 months) for males, by broad education category

Highest level of education attainment	Sample Percentile					Total
	10	25	50	75	90	
School (Yr. 9, 10, 11, 12 & Cert. I, II)	0.0%	0.0%	16.7%	61.1%	100.0%	2451
VET (Cert. III, IV & Adv. Dip.)	0.0%	0.0%	2.8%	23.6%	66.7%	368
University (Bachelor, Grad. Dip., Postgrad.)	0.0%	0.0%	5.6%	25.0%	52.2%	786
Total	0.0%	0.0%	11.1%	45.8%	100.0%	3605

Note: Total excludes nine observations where 'total survey time' (i.e. denominator of proportion) equalled zero.

Figure 2 Total job search duration as a proportion of total survey time (capped at 36 months) for females, by broad education category



Note: Includes respondents who exited the survey before obtaining employment, i.e. observations censored at the date of the last interview. These comprise approximately 4% of the total.

Table 11 Total job search duration as a proportion of total survey time (capped at 36 months) for females, by broad education category

Highest level of education attainment	Sample Percentile					Total
	10	25	50	75	90	
School (Yr. 9, 10, 11, 12 & Cert. I, II)	0.0%	0.0%	16.7%	61.1%	100.0%	2224
VET (Cert. III, IV & Adv. Dip.)	0.0%	0.0%	0.0%	22.2%	54.5%	563
University (Bachelor, Grad. Dip., Postgrad.)	0.0%	0.0%	2.8%	15.4%	44.4%	1231
Total	0.0%	0.0%	8.3%	38.5%	90.0%	4018

Note: Total excludes nine observations where 'total survey time' (i.e. denominator of proportion) equalled zero.

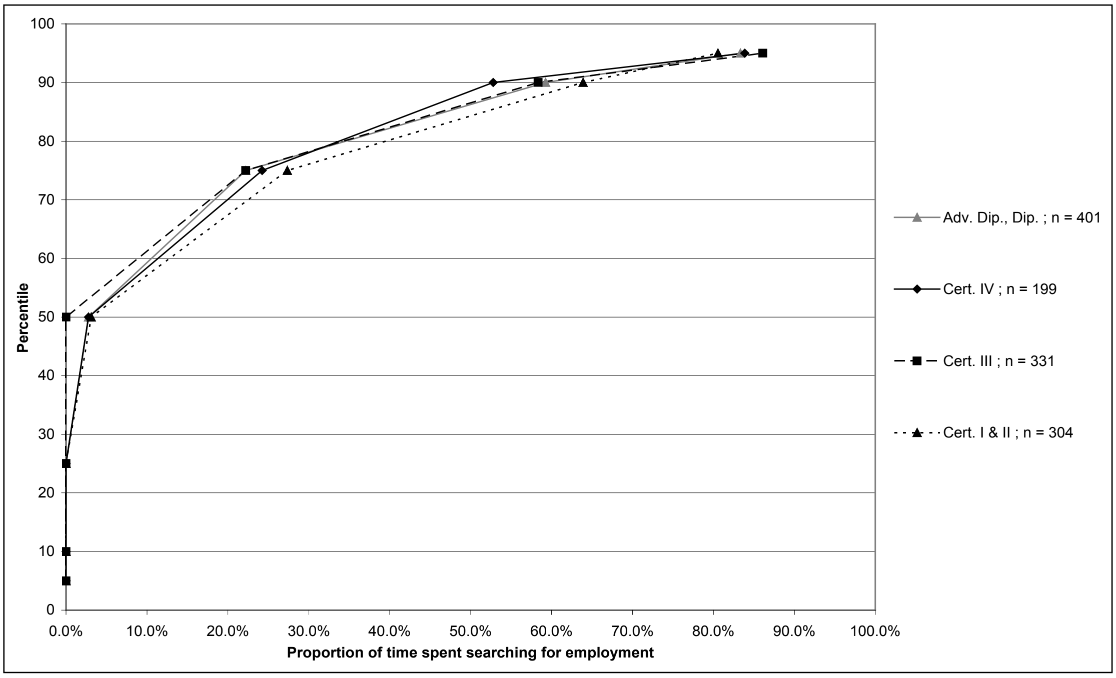
To complete the analysis of the influence of education on the proportion of total job search time, the qualifications within the VET level of education are disaggregated and examined. Unfortunately, further disaggregation of the data by gender caused results to suffer a small number problems and are not discussed or presented. Figure 3 illustrates the TPS measures by the VET qualifications: certificates I and II, certificate III, certificate IV, and advanced diploma and diploma. Table 12 provides the exact percentages that correspond to Figure 3.

As illustrated in Figure 3, the VET qualifications are closely aligned, with only the certificates I and II qualifications consistently providing the least effective pathway to a successful employment outcome. The influence of the higher three qualifications on job search is ambiguous. For example, from the TPS measures shown in Table 12, out of all certificate III graduates, 50% transited immediately from education to employment; whereas, for the same proportion of advanced diploma and certificate IV graduates, they spent approximately one month or less searching. When considering the sub-samples at their 75% levels, it is the advanced diploma and certificate III graduates that have the lower total search durations (i.e. approximately

8 months or less), closely followed by the certificate IV graduates (i.e. approximately 9 months or less).

Overall, the differences in search time between the higher VET qualifications and, hence, their effectiveness on employment outcomes, are negligible. However, the similarities in the TPS measures between the certificate III, certificate IV and advanced diploma and diploma qualifications may disguise differences across gender, as seen in the results of the survival and multivariate regression analyses, but are unable to be considered due to data restrictions.

Figure 3 Total job search duration as a proportion of total survey time (capped at 36 months) by VET qualification



Note: Includes respondents who exited the survey before obtaining employment, i.e. observations censored at the date of the last interview. These comprise approximately 4% of the total.

Table 12 Total job search duration as a proportion of total survey time (capped at 36 months) by VET qualification

Highest level of education attainment	Sample Percentile					Total
	10	25	50	75	90	
Cert. I & II	0.0%	0.0%	3.1%	27.4%	63.9%	304
Cert. III	0.0%	0.0%	0.0%	22.2%	58.3%	331
Cert. IV	0.0%	0.0%	2.8%	24.2%	52.8%	199
Adv. Dip., Dip.	0.0%	0.0%	2.8%	22.2%	59.3%	401
Total	0.0%	0.0%	2.8%	25.0%	59.3%	1235

Note: Total excludes seven observations where ‘total survey time’ (i.e. denominator of proportion) equalled zero.

Econometric models of duration

Introduction

The empirical analysis examines the duration of the (first) spell of non-employment after completion of the highest level of education the individual has achieved using the hazard function. The hazard rate represents the probability of leaving unemployment (a particular state) within a specific interval (which depends on the frequency with which data are collected) conditional on not having left unemployment (the initial state) up to the starting time of the interval.

The duration variable is referred to as grouped duration data as duration is only known to fall into a certain time interval, the exact duration is unknown (i.e. data are reported at the month level not daily).

Concepts in Hazard Model Analysis

The survival function, S(t), is the cumulative frequency of the proportion of the sample who do not experience the event by time t (i.e. individuals in the LSAY do not exit non-employment/unemployment and enter employment—they survive in non-employment). S(t) can therefore be interpreted as the probability employment will not occur until time t. Plots of S(t) can compare the survival rate for various groups and statistical differences can be examined (see e.g. Kaplan-Meier below).

The cumulative probability function, F(t), is simply 1 - S(t) and S(t) = 1 – F(t).

Probability density function, f(t), is the area under the curve representing the (unconditional—not dependent on explanatory variables) instantaneous probability (at time t) of the event (exit to employment).

Hazard rate, h(t), is the probability that the event (exit to employment) will occur in the next time period (t+1) given that the individual is not employed at time t (where h(t) is also commonly referred to as the failure rate or transition rate). The hazard rate can also be expressed as h(t) = f(t)/S(t).

The Hazard function is the specification of the shape across time of h(t).

The cumulative hazard function is the integration of H(t) from time zero to time t.

The baseline hazard is the form the hazard ratio takes before explanatory variables are accounted for. The baseline therefore applies to all individuals in the sample, they differ from each other due to the influence of explanatory variables.

Non-parametric methods

Common non-parametric methods for examining the survivor and cumulative hazard function are the Kaplan-Meier and Nelson-Aalen methods respectively. For these models all that is required is an ordering of the duration data—models do not make assumptions about the shape of the hazard function, or the influence of explanatory variables. Influence of other variables can be shown by stratification of the data into groups (e.g. gender); but the models are further restricted as they cannot handle continuous data. They are referred to as event history analysis where time is the only salient variable.

The Kaplan-Meier model is specified:

$$\hat{S}(t) = \prod \left(\frac{n_i - d_i}{n_i} \right) \quad (1)$$

where d_i where *i* are individuals who exit to employment (duration ends), n_i is the number of individuals at risk (who are currently not employed but have exited education), and S(t) is the Kaplan-Meier 'survivor' function (i.e. the probability an individual will not exit to employment).

Semi-parametric methods

A number of hazard model specifications are available: for example, a discrete-time Cox proportional hazard semi-parametric model does not require detailed knowledge of the distribution of the hazard function model to examine duration or time-to-first job. The discrete time proportional hazard model approach is practical because it allows investigation of the duration dependence, i.e. how the probability of getting a job changes with search duration. Moreover, proportional hazard model allows the hazard function to be conditional on explanatory variables such as gender, ethnicity, education level, and family background.

Complementary log-log model (cloglog)

The discrete-time proportional hazard model can be specified in the complementary log-log form:

$$h_j(x_{ij}) = 1 - \exp\left[-\exp\left(x_{ij}^{'}\beta + \gamma_j + \log(v_i) \right) \right] \quad (2)$$

where the logarithm of the integral of the baseline hazard, γ_j, over the interval $[t_{j-1}, t_j]$ is specified as:

$$\gamma_j = \log \int_{t_{j-1}}^{t_j} h_o(\tau) d\tau \quad (3)$$

The complementary log-log hazard function can be interpreted as the discrete-time model corresponding to an underlying continuous time proportional hazard model. It is similar to a logistic regression, but tends to have a fatter tail as the function approaches zero and by being 'steeper' in the vicinity of one (which means a 'one' response is a rare event).

Where:

- *i* denotes the individual.
- *j* denotes the time interval.
- **X** is a vector of explanatory variables.
- β is the vector of unknown parameters to be estimated, they show the effects of the explanatory variables X on the hazard rate.
- v_i is the model error term (a Gamma distributed random variable with unit mean and variance) which estimates the part of unobserved heterogeneity between individuals that is constant over time.

- γ_j is the baseline hazard or probability of exiting unemployment given 'survival' at the start of the interval denoted as *j*. It is the 'value' of hazard from which individuals in the data deviate at each time period, i.e. it is common to all individuals. Intuitively explained, γ_j is like a dummy variable which estimates the way unobserved heterogeneity in the data varies by the duration of the spell.

We interpret the dependent variable, duration dependence $h_j(x_{ij})$, as the probability that an individual *i* will exit unemployment for employment between the beginning and the end of the j^{th} period, given his/her attribute set , X at that time.

CllogLog discrete-time flexible hazard model extended results

Table 13 Full-Time Permanent ('Good') Job MALES: Complementary log-log regression (Education in six categories)

Variable	Coefficient	Std. Err	z-statistic	p-value
Education < YR12 (Base)				
Education YR12	0.4349	0.0569	7.6500	0.0000
Education Cert I or II	0.9373	0.1239	7.5600	0.0000
Education Cert III or IV	0.9764	0.1052	9.2800	0.0000
Education Diploma or Adv Dip	0.7796	0.1192	6.5400	0.0000
Education University	1.0030	0.0716	14.0100	0.0000
Age	-0.0291	0.0567	-0.5100	0.6080
Indigenous (Not Indigenous)	-0.2049	0.1660	-1.2300	0.2170
Disability (No Disability)	0.0301	0.1227	0.2500	0.8060
ACT (Base)				
NSW	-0.0217	0.1270	-0.1700	0.8640
VIC	-0.0185	0.1258	-0.1500	0.8830
QLD	0.0736	0.1296	0.5700	0.5700
SA	-0.0448	0.1353	-0.3300	0.7410
WA	0.1058	0.1359	0.7800	0.4370
TAS	-0.0801	0.1590	-0.5000	0.6150
NT	0.2403	0.1884	1.2800	0.2020
Metropolitan (Base)				
Regional	0.1301	0.0572	2.2700	0.0230
Rural/Remote	0.1097	0.0633	1.7300	0.0830
School Government (Base)e				
School Catholic	-0.1057	0.0598	-1.7700	0.0770
School Independent	-0.2593	0.0629	-4.1200	0.0000
Country of Birth Australia (Base)				
Country of Birth _ESB	-0.2659	0.1480	-1.8000	0.0720
Country of Birth _NESB	-0.2704	0.1106	-2.4500	0.0140
Country of Birth Mother Aust (Base)				
Country of Birth Mother ESB	-0.0733	0.0786	-0.9300	0.3510
Country of Birth Mother NESB	-0.1577	0.0878	-1.8000	0.0720
Country of Birth Father Aust (Base)				
Country of Birth Father ESB	0.0490	0.0747	0.6600	0.5120
Country of Birth Father NESB	-0.0752	0.0802	-0.9400	0.3480
Baseline Coefficients				
Sep-95	-2.2028	0.8607	-2.5600	0.0100
Oct-95	-2.9824	0.8623	-3.4600	0.0010
Nov-95	-1.5679	0.8584	-1.8300	0.0680
Dec-95	-3.0876	0.8625	-3.5800	0.0000
Jan-96	-3.5911	0.8660	-4.1500	0.0000
Feb-96	-4.2765	0.8792	-4.8600	0.0000

Variable	Coefficient	Std. Err	z-statistic	p-value
Mar-96	-4.1087	0.8740	-4.7000	0.0000
Apr-96	-3.8180	0.8700	-4.3900	0.0000
May-96	-4.3154	0.8793	-4.9100	0.0000
Jun-96	-4.8032	0.8940	-5.3700	0.0000
Jul-96	-4.7370	0.8921	-5.3100	0.0000
Aug-96	-5.2221	0.9140	-5.7100	0.0000
Sep-96	-5.9857	0.9681	-6.1800	0.0000
Oct-96	-3.4482	0.8696	-3.9700	0.0000
Nov-96	-2.0232	0.8618	-2.3500	0.0190
Dec-96	-3.9010	0.8791	-4.4400	0.0000
Jan-97	-4.1025	0.8863	-4.6300	0.0000
Feb-97	-4.7016	0.9060	-5.1900	0.0000
Mar-97	-5.0607	0.9219	-5.4900	0.0000
Apr-97	-4.1532	0.8873	-4.6800	0.0000
May-97	-4.7453	0.9101	-5.2100	0.0000
Jun-97	-5.2779	0.9432	-5.6000	0.0000
Jul-97	-5.0219	0.9159	-5.4800	0.0000
Aug-98	-4.9103	0.9143	-5.3700	0.0000
Aug-99	-4.4290	0.8635	-5.1300	0.0000
Aug-00	-4.6071	0.8679	-5.3100	0.0000
Aug-01	-5.1795	0.8742	-5.9200	0.0000
Aug-02	-5.2745	0.8824	-5.9800	0.0000
Aug-03	-5.5049	0.8939	-6.1600	0.0000
Aug-04 Plus	-6.8435	0.9310	-7.3500	0.0000

Note: Log pseudo likelihood = -8679.1594; Prob > chi2 = 0.0000; Wald chi2(55) = 21950.87; Std. Err. adjusted for 3922 clusters in id.

Table 14 Full-Time & Permanent (‘Good’) Job FEMALES: Complementary log-log regression (Education in six categories)

Variable	Coefficient	Std Err	z-statistic	p-value
Education < YR12 (Base)				
Education YR12	0.7552	0.0774	9.76	0.0000
Education Cert I or II	1.6411	0.1407	11.66	0.0000
Education Cert III or IV	1.4733	0.1079	13.65	0.0000
Education Diploma or Adv Dip	1.4369	0.1128	12.74	0.0000
Education University	1.5970	0.0800	19.95	0.0000
Age	-0.0369	0.0580	-0.64	0.5240
Indigenous (Not Indigenous)	-0.0520	0.1658	-0.31	0.7540
Disability (No Disability)	-0.5234	0.2341	-2.24	0.0250
ACT (Base)	-	-	-	-
NSW	0.1086	0.1172	0.93	0.3540
VIC	0.0205	0.1146	0.18	0.8580
QLD	-0.0637	0.1238	-0.51	0.6070
SA	-0.0652	0.1214	-0.54	0.5910
WA	0.0718	0.1254	0.57	0.5670
TAS	0.1177	0.1474	0.80	0.4250
NT	0.3553	0.1847	1.92	0.0540

Variable	Coefficient	Std Err	z-statistic	p-value
Metropolitan (Base)	-	-	-	-
Regional	-0.0305	0.0561	-0.54	0.5860
Rural/Remote	-0.0275	0.0617	-0.45	0.6550
School Government (Base)e	-	-	-	-
School Catholic	-0.0597	0.0549	-1.09	0.2770
School Independent	-0.1833	0.0630	-2.91	0.0040
Country of Birth Australia (Base)	-	-	-	-
Country of Birth _ESB	-0.1210	0.1381	-0.88	0.3810
Country of Birth _NESB	-0.2674	0.1069	-2.50	0.0120
Country of Birth Mother Aust (Base)				
Country of Birth Mother ESB	-0.0457	0.0737	-0.62	0.5360
Country of Birth Mother NESB	-0.0532	0.0815	-0.65	0.5140
Country of Birth Father Aust (Base)	-	-	-	-
Country of Birth Father ESB	-0.0888	0.0785	-1.13	0.2580
Country of Birth Father NESB	-0.1591	0.0746	-2.13	0.0330
Baseline Coefficients	-2.4923	0.8722	-2.86	0.0040
Sep-95	-3.4366	0.8728	-3.94	0.0000
Oct-95	-2.0000	0.8707	-2.30	0.0220
Nov-95	-3.4538	0.8749	-3.95	0.0000
Dec-95	-4.2028	0.8811	-4.77	0.0000
Jan-96	-4.8555	0.8892	-5.46	0.0000
Feb-96	-4.1956	0.8824	-4.75	0.0000
Mar-96	-4.7091	0.8877	-5.30	0.0000
Apr-96	-6.0526	0.9412	-6.43	0.0000
May-96	-5.7299	0.9204	-6.23	0.0000
Jun-96	-5.4140	0.9033	-5.99	0.0000
Jul-96	-6.1711	0.9503	-6.49	0.0000
Aug-96	-5.9038	0.9343	-6.32	0.0000
Sep-96	-3.9906	0.8799	-4.54	0.0000
Oct-96	-2.5492	0.8731	-2.92	0.0040
Nov-96	-4.4452	0.8906	-4.99	0.0000
Dec-96	-5.2988	0.9130	-5.80	0.0000
Jan-97	-5.5525	0.9274	-5.99	0.0000
Feb-97	-5.2071	0.9107	-5.72	0.0000
Mar-97	-4.7898	0.8969	-5.34	0.0000
Apr-97	-5.5209	0.9269	-5.96	0.0000
May-97	-5.8728	0.9463	-6.21	0.0000
Jun-97	-6.0248	0.9639	-6.25	0.0000
Jul-97	-6.2049	0.9779	-6.34	0.0000
Aug-98	-4.8403	0.8745	-5.53	0.0000
Aug-99	-5.3360	0.8780	-6.08	0.0000
Aug-00	-5.7342	0.8833	-6.49	0.0000
Aug-01	-5.6395	0.8864	-6.36	0.0000
Aug-02	-6.7395	0.9244	-7.29	0.0000
Aug-03 plus	-8.0781	0.9853	-8.20	0.0000

Log pseudo likelihood = -9173.3045; Prob > chi2 = 0.0000; Wald chi2(55) = 22377.11; Std. Err. adjusted for 4378 clusters in id.

Table 15 Any Job MALES: Complementary log-log regression (Education in six categories)

Variable	Coefficient	Std Err	z-statistic	p-value
Education < YR12 (Base)				
Education YR12	1.2597	0.0464	27.15	0.0000
Education Cert I or II	1.4343	0.1428	10.04	0.0000
Education Cert III or IV	1.6153	0.0895	18.06	0.0000
Education Diploma or Adv Dip	1.4738	0.1115	13.21	0.0000
Education University	1.6780	0.0619	27.12	0.0000
Age	-0.0257	0.0444	-0.58	0.5620
Indigenous (Not Indigenous)	-0.3620	0.1299	-2.79	0.0050
Disability (No Disability)	-0.1127	0.1240	-0.91	0.3640
ACT (Base)				
NSW	-0.0671	0.0975	-0.69	0.4910
VIC	-0.0141	0.0961	-0.15	0.8830
QLD	-0.0016	0.1017	-0.02	0.9870
SA	-0.0100	0.1080	-0.09	0.9260
WA	-0.0082	0.1070	-0.08	0.9390
TAS	-0.0589	0.1259	-0.47	0.6400
NT	0.2035	0.1540	1.32	0.1860
Metropolitan (Base)				
Regional	-0.0382	0.0489	-0.78	0.4340
Rural/Remote	-0.0621	0.0530	-1.17	0.2410
School Government (Base)e				
School Catholic	0.0909	0.0482	1.89	0.0590
School Independent	-0.0133	0.0515	-0.26	0.7960
Country of Birth Australia (Base)				
Country of Birth _ESB	-0.0910	0.1069	-0.85	0.3940
Country of Birth _NESB	-0.2759	0.0776	-3.55	0.0000
Country of Birth Mother Aust (Base)				
Country of Birth Mother ESB	0.0577	0.0632	0.91	0.3610
Country of Birth Mother NESB	-0.1993	0.0675	-2.95	0.0030
Country of Birth Father Aust (Base)				
Country of Birth Father ESB	0.0060	0.0650	0.09	0.9270
Country of Birth Father NESB	-0.1133	0.0647	-1.75	0.0800
Baseline Coefficients				
Sep-95	-1.0909	0.6747	-1.62	0.1060
Oct-95	-2.1520	0.6780	-3.17	0.0020
Nov-95	-2.1389	0.6770	-3.16	0.0020
Dec-95	-2.5368	0.6808	-3.73	0.0000
Jan-96	-2.7025	0.6841	-3.95	0.0000
Feb-96	-3.0665	0.6887	-4.45	0.0000
Mar-96	-3.3212	0.6963	-4.77	0.0000
Apr-96	-2.9711	0.6908	-4.30	0.0000
May-96	-2.9475	0.6920	-4.26	0.0000
Jun-96	-3.3885	0.7022	-4.83	0.0000
Jul-96	-2.3019	0.6836	-3.37	0.0010
Aug-96	-2.2977	0.6844	-3.36	0.0010
Sep-96	-2.1523	0.6904	-3.12	0.0020
Oct-96	-2.7793	0.6977	-3.98	0.0000

Variable	Coefficient	Std Err	z-statistic	p-value
Nov-96	-3.0488	0.7198	-4.24	0.0000
Dec-96	-3.3745	0.7239	-4.66	0.0000
Jan-97	-3.4000	0.7348	-4.63	0.0000
Feb-97	-3.8046	0.7740	-4.92	0.0000
Mar-97	-3.7606	0.7662	-4.91	0.0000
Apr-97	-3.8640	0.7937	-4.87	0.0000
May-97	-3.2088	0.7300	-4.40	0.0000
Jun-97	-2.2417	0.7031	-3.19	0.0010
Jul-97	-2.2159	0.7024	-3.15	0.0020
Aug-97	-2.8539	0.7389	-3.86	0.0000
Aug-98	-3.4634	0.6740	-5.14	0.0000
Aug-99	-3.0281	0.6827	-4.44	0.0000
Aug-00	-3.9087	0.7456	-5.24	0.0000
Aug-01	-4.5662	0.8044	-5.68	0.0000
Aug-02	-3.3190	0.7293	-4.55	0.0000
Aug-03	-4.0572	0.7908	-5.13	0.0000
Aug-04	-4.4671	0.8657	-5.16	0.0000
Aug-05	-2.0022	0.6685	-2.99	0.0030

Log pseudo likelihood = -7164.7229; Prob > chi2 = 0.0000; Wald chi2(57) = 9981.63; Std. Err. adjusted for 4906 clusters in id.

Table 16 Any Job FEMALES: Complementary log-log regression (Education in six categories)

Variable	Coefficient	Std Err	z-statistic	p-value
Education < YR12 (Base)				
Education YR12	1.4970	0.0527	28.39	0.000
Education Cert I or II	2.0105	0.1040	19.34	0.000
Education Cert III or IV	1.9776	0.0884	22.38	0.000
Education Diploma or Adv Dip	2.1587	0.0955	22.61	0.000
Education University	2.0840	0.0572	36.46	0.000
Age	0.0138	0.0444	0.31	0.756
Indigenous (Not Indigenous)	-0.3662	0.1274	-2.88	0.004
Disability (No Disability)	0.0275	0.1390	0.20	0.843
ACT (Base)				
NSW	-0.0827	0.1110	-0.75	0.456
VIC	-0.0902	0.1108	-0.81	0.416
QLD	-0.1338	0.1155	-1.16	0.246
SA	-0.0840	0.1146	-0.73	0.463
WA	0.0509	0.1164	0.44	0.662
TAS	-0.2620	0.1352	-1.94	0.053
NT	-0.0126	0.1544	-0.08	0.935
Metropolitan (Base)				
Regional	-0.1217	0.0467	-2.61	0.009
Rural/Remote	-0.1205	0.0497	-2.42	0.015
School Government (Base)e				
School Catholic	0.0780	0.0455	1.71	0.087
School Independent	-0.0623	0.0482	-1.29	0.196
Country of Birth Australia (Base)				

Variable	Coefficient	Std Err	z-statistic	p-value
Country of Birth _ESB	-0.2156	0.0928	-2.32	0.020
Country of Birth _NESB	-0.2612	0.0799	-3.27	0.001
Country of Birth Mother Aust (Base)				
Country of Birth Mother ESB	-0.0305	0.0744	-0.41	0.682
Country of Birth Mother NESB	-0.1075	0.0636	-1.69	0.091
Country of Birth Father Aust (Base)				
Country of Birth Father ESB	0.1927	0.0632	3.05	0.002
Country of Birth Father NESB	-0.2479	0.0599	-4.14	0.000
Baseline Coefficients				
Sep-95	-1.7282	0.6712	-2.57	0.010
Oct-95	-3.2004	0.6766	-4.73	0.000
Nov-95	-3.1271	0.6760	-4.63	0.000
Dec-95	-3.2186	0.6793	-4.74	0.000
Jan-96	-3.5544	0.6847	-5.19	0.000
Feb-96	-3.9094	0.6901	-5.66	0.000
Mar-96	-3.7040	0.6950	-5.33	0.000
Apr-96	-4.0899	0.6990	-5.85	0.000
May-96	-4.1081	0.7028	-5.85	0.000
Jun-96	-4.1366	0.7048	-5.87	0.000
Jul-96	-3.3848	0.6920	-4.89	0.000
Aug-96	-3.2918	0.6889	-4.78	0.000
Sep-96	-2.8192	0.6844	-4.12	0.000
Oct-96	-3.5444	0.6940	-5.11	0.000
Nov-96	-3.7285	0.7113	-5.24	0.000
Dec-96	-4.5768	0.7582	-6.04	0.000
Jan-97	-4.2146	0.7312	-5.76	0.000
Feb-97	-4.1570	0.7455	-5.58	0.000
Mar-97	-4.1885	0.7628	-5.49	0.000
Apr-97	-3.8472	0.7171	-5.36	0.000
May-97	-3.6811	0.7280	-5.06	0.000
Jun-97	-3.3530	0.7168	-4.68	0.000
Jul-97	-3.1624	0.7129	-4.44	0.000
Aug-98	-2.9152	0.7078	-4.12	0.000
Aug-99	-4.0968	0.7001	-5.85	0.000
Aug-00	-3.9832	0.7017	-5.68	0.000
Aug-01	-4.1125	0.7113	-5.78	0.000
Aug-02	-4.2570	0.7527	-5.66	0.000
Aug-03	-4.3514	0.7906	-5.50	0.000
Aug-04	-3.8485	0.7226	-5.33	0.000
Aug-05	-4.1668	0.8392	-4.97	0.000
Aug-06	-2.4113	0.6749	-3.57	0.000

Log pseudo likelihood = -6901.0974; Prob > chi2 = 0.0000; Wald chi2(57) = 9782.53; Std. Err. adjusted for 5199 clusters in id.

Sensitivity analysis of grouping education levels

As noted in the text, there are 12 (usable) education levels reported in the LSAY. First, it is not possible (see text in body of the main report) or necessary (e.g. differentiating between labour market outcome for year 10 or below will not be particularly instructive) to maintain the complete set.

In addition, as noted previously, there are a number of ways the educational categories can be aggregated and it is the purpose of this section to assess whether aggregation of education levels alters results to an extent where interpretation changes.

Complementary log-log (cloglog) models are estimated for those with 'any job' and those with a 'good job', separately for males and females and the conclusion reached is that although there are differences in the coefficients for education, they are not material to interpretation.

Twelve education groups: Data as presented in the LSAY

- Postgraduate (masters or doctorate)
- Graduate diploma or graduate certificate
- Bachelor degree
- Advanced diploma or diploma
- Certificate IV
- Certificate III
- Certificate II
- Certificate I
- Year 12 high school
- Year 11 high school
- Year 10 high school
- The initial Year 9 at the start of the survey

Six education groups

- University (bachelors to post-graduate qualification)
- Diploma or advanced diploma
- Certificate III or certificate IV
- Certificate I or certificate II
- Year 12
- Less than Year12 high school (the base case in the regression models)

Five education groups

- University (bachelors to post-graduate qualification)
- Diploma or advanced diploma
- Certificate III or certificate IV

- Certificate I or certificate II
- Year 12 (the base case in the regression models)

Three education groups (A):

- University (bachelors to post-graduate qualification), diploma or advanced diploma
- VET (certificate III or certificate IV)
- School, certificate I, or certificate II (the base case in the regression models)

Three education groups (B):

- University (bachelors to post-graduate qualification), diploma or advanced diploma
- VET (certificate III or certificate IV)
- Year 12, certificate I, or certificate II (the base case in the regression models).

Estimated exponential coefficient for education level in the various model are provided in Table 17 below; *p*-values are not reported as they are always reported as $p = 0.000$.

Table 17 Duration models, various groupings of highest education level

	AJ Males Exp(β)	**AJ Fem** Exp(β)	**FTP Males** Exp(β)	**FTP Males** Exp(β)
Model: 6-Categories				
Less than YR12 (Base)	-		-	
Year 12	3.5244	4.4681	1.5449	2.1281
Certificate I or Cert II	4.1965	7.4673	2.5530	5.1609
Certificate III or Cert IV	5.0293	7.2251	2.6549	4.3638
Diploma or Adv Dip	4.3659	8.6600	2.1806	4.2076
University (Bachelors to Post-Graduate)	5.3549	8.0364	2.7264	4.9383
Model: 5-Categories				
Year 12 (Base)	-	-	-	-
Certificate I or Certificate II	1.2169	1.6621	1.6535	2.4678
Certificate III or Cert IV	1.4543	1.0028	1.7394	2.0558
Diploma or Adv Diploma	1.2198	1.8506	1.4578	1.9806
University (Bachelors to Post-Graduate)	1.4965	1.7216	1.7625	2.3059
Model: 3A-Categories				
School, Certificate I, or Certificate II (Base case)			-	
VET (Cert III or IV)	1.3460	1.6194	1.7037	2.1743
University or diploma/advanced dip	1.4353	1.7403	1.7015	2.2507
Model: 3B-Categories				
Year 12, Certificate I, or Certificate II (Base case)-			-	
VET (Cert III or Cert IV)	2.5611	3.0298	2.000	2.6359
University or diploma/advanced dip	2.5787	3.2448	1.9609	2.7363

Notes: (1) p-values are not reported as they are always reported as p = 0.000. (2) AJ represents “any job”; FTP represents Full-time permanent employment (a “good” job).

As expected, in models in which those with less than Year 12 education are excluded (Model 3B and Model 5), the impact of education is smaller in size. But, relative impacts between education levels are consistent and the relationship between ‘any job’ and a ‘good’ job is also maintained. We conclude the results are both robust and meaningful.

Complete estimation

Table 18 below is the full version of Table 4 in the main report. It is also used to calculate the profiles in Box 1 in the main report.

Table 18 Duration estimation: From education to employment

Variable	**Model 1:** ANY job (Males)	**Model 2:** ANY job (Females)	**Model 3:** FTP Job (Males)	**Model 4:** FTP job (Females)
Education < YR12 (Base)				
Education YR12	3.52***	4.47***	1.54***	2.13***
Education Cert I or II	4.20***	7.47***	2.55***	5.16***
Education Cert III or IV	5.03***	7.23***	2.65***	4.36***
Education Diploma or Adv Dip	4.37***	8.66***	2.18***	4.21***
Education University	5.35***	8.04***	2.73***	4.94***
Age	0.97	1.01	0.97	0.96
Indigenous (Base, Not Indigenous)	0.70**	0.69**	0.81	0.95
Disability (Base, No Disability)	0.89	1.03	1.03	0.59*
ACT (Base)				
NSW	0.94	0.92	0.98	1.11
VIC	0.99	0.91	0.98	1.02
QLD	1.00	0.87	1.08	0.94
SA	0.99	0.92	0.96	0.94
WA	0.99	1.05	1.11	1.07
TAS	0.94	0.77	0.92	1.12
NT	1.23	0.99	1.27	1.43
Metropolitan (Base)				
Regional	0.96	0.89**	1.14*	0.97
Rural/Remote	0.94	0.89*	1.12	0.97
School Government (Base)e				
School Catholic	1.10	1.08	0.90	0.94
School Independent	0.99	0.94	0.77***	0.83**
Country of Birth Australia (Base)				
Country of Birth _ESB	0.91	0.81*	0.77	0.89
Country of Birth _NESB	0.76***	0.77**	0.76*	0.77*
Country of Birth Mother Aust (Base)				
Country of Birth Mother ESB	1.06	0.97	0.93	0.96
Country of Birth Mother NESB	0.82**	0.90	0.85	0.95
Country of Birth Father Aust (Base)				
Country of Birth Father ESB	1.01	1.21**	1.05	0.91
Country of Birth Father NESB	0.89	0.78***	0.93	0.85*
Base average duration	435	550	481	584
Wald Chi2 test (p-value)	9 982 (0.00)	9 783 (0.00)	21 951 (0.00)	22 377 (0.00)

Notes: (1) stars denote statistical significance * $p<0.05$; ** $p<0.01$; *** $p<0.001$. (3) The model is a complementary log-log (cloglog) model. Reported estimates are the “odds ratio” calculated as e^{β}. (4) Sample sizes do not match across models due to differing rates of missing values (or item non-response).

Flexible baseline hazard estimates

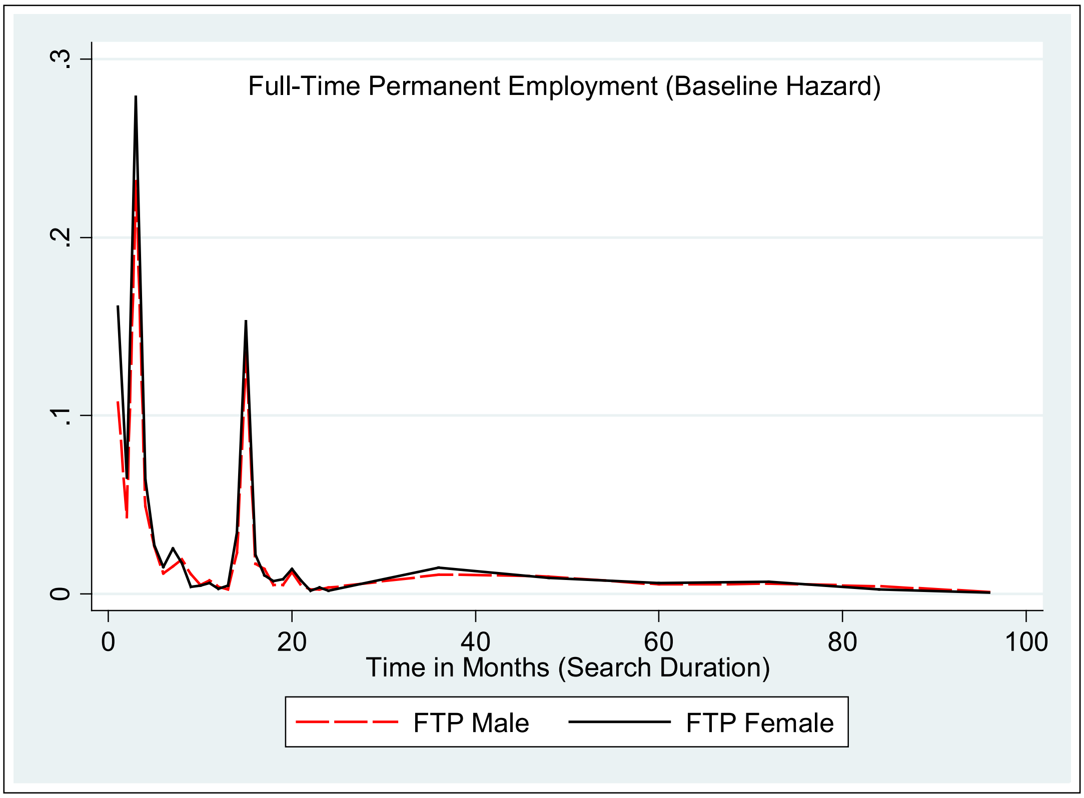
The baseline hazard provides a view of the impact of the passage of time on the probability of obtaining a job when the influence of educational level, and the significant control (or explanatory) variables have been take into account.

There are a number of alternative to specifying the functional form of the baseline hazard function (e.g. time, time squared, time cubed) but preliminary models indicated no need to go beyond the easily interpreted time in observed units.

Given the distribution of time to employment, the most satisfactory representation of time is to include variables for the first 24 months (monthly), followed by an annual variable for the next 5 years, and to complete the specification with a variable covering the remained of the period.

Figure 4 below provides a representation of the baseline hazard for those whose first job is a 'good job' (full-time permanent), and Figure 5 which follows the baseline hazard for those whose first job is designated as 'any job'. Both figures are the time-variables from the models based on educational attainment in six categories.

Figure 4 Baseline hazard: full-time and permanent employment (education in six categories)



A comparison of the models demonstrates that there is some differences in timing of the peaks that indicated an increase in the probability of remaining out of employment—given we have controlled for educational attainment and other control variables. An interesting factor that emerges from the figures is that for employment in a 'good job' there are few differences in the baseline hazard between males and females, and differences tend to favour females. There is an increased likelihood that males will remain out of employment—with a definite advantage to females at those peaks.

The peaks appear to coincide with the end of the calendar year and may possibly be attributed to both labour market supply and demand factors. For example:

- ✧ Many students leave school at the end of the year in the early years of education and hence there are an increased number seeking employment compared to those who gain employment.
- ✧ The end of the year coincides with the recruitment programme in the retail sector which increases the demand for labour and probably increases the supply.
- ✧ In addition, generally, interviews in the LSAY take place at the end of the year and this may influence the distribution of employed and not employed.

Figure 5 Baseline hazard: any form of employment (education in six categories)

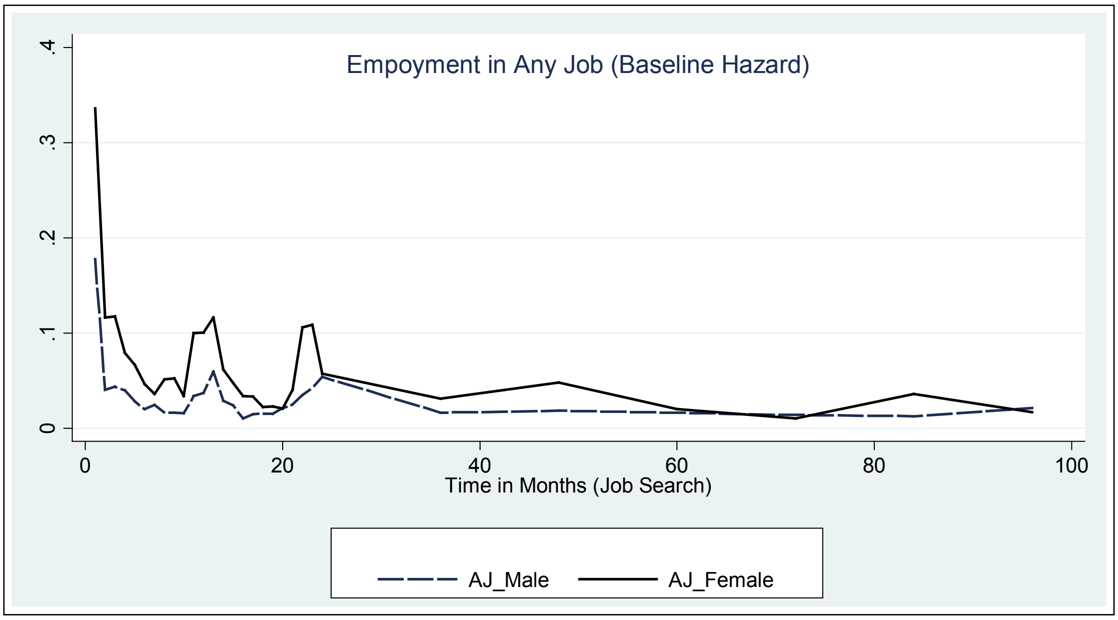


Table 19 below provides the estimated coefficient for the time-variable in the cloglog models for 'any job' and for the first 'good job' (by gender). The coefficients are converted from the cloglog 'log odds ratio' to the more accessible 'odds ratio' or 'hazard rates' (interpreted in this case as the 'hazard' that individuals will remain not employed. Clearly, by the time an individual (males or female) reaches about two years from the time they finishing their highest level of education, the probability of remaining not employed is very small (conversely, the probability of finding a job is high).

Table 19 Baseline Hazard (exp(β))

Date	AJ Male	AJ Female	FTP Males	FTP Females
Sep-95	0.3359	0.1776	0.1105	0.0827
Oct-95	0.1162	0.0407	0.0507	0.0322
Nov-95	0.1178	0.0438	0.2085	0.1353
Dec-95	0.0791	0.0400	0.0456	0.0316
Jan-96	0.0670	0.0286	0.0276	0.0150
Feb-96	0.0466	0.0201	0.0139	0.0078
Mar-96	0.0361	0.0246	0.0164	0.0151

Date	AJ Male	AJ Female	FTP Males	FTP Females
Apr-96	0.0512	0.0167	0.0220	0.0090
May-96	0.0525	0.0164	0.0134	0.0024
Jun-96	0.0338	0.0160	0.0082	0.0032
Jul-96	0.1001	0.0339	0.0088	0.0045
Aug-96	0.1005	0.0372	0.0054	0.0021
Sep-96	0.1162	0.0597	0.0025	0.0027
Oct-96	0.0621	0.0289	0.0318	0.0185
Nov-96	0.0474	0.0240	0.1322	0.0781
Dec-96	0.0342	0.0103	0.0202	0.0117
Jan-97	0.0334	0.0148	0.0165	0.0050
Feb-97	0.0223	0.0157	0.0091	0.0039
Mar-97	0.0233	0.0152	0.0063	0.0055
Apr-97	0.0210	0.0213	0.0157	0.0083
May-97	0.0404	0.0252	0.0087	0.0040
Jun-97	0.1063	0.0350	0.0051	0.0028
Jul-97	0.1091	0.0423	0.0066	0.0024
Aug-97	0.0576	0.0542	0.0074	0.0020
1998	0.0313	0.0166	0.0119	0.0079
1999	0.0484	0.0186	0.0099	0.0048
2000	0.0201	0.0164	0.0056	0.0032
2001	0.0104	0.0142	0.0051	0.0036
2002	0.0362	0.0129	0.0041	0.0012
2003	0.0173	0.0213	0.0011	0.0003

Notes: (1) The baseline hazard coefficient is the exponential of the estimated coefficient (exp(β)). (2) All time specific dummy variables are significant at the 0.000% level. (3) Baseline hazard is give and including 2003 (e.g. about 100 months).