Paper 216-2012

How Frail Are Great British Immigrants to Find First Job After Arrival?

JuYin Helen Wong, University of Manchester

ABSTRACT

Instead of modeling (repeated) cross sectional or panel data when comparing immigrants' employment patterns, this study employs Cox models to explore the length of time taking immigrants in the Great Britain (GB) to find their first employment to measure whether any transition-duration penalties (i.e., the length of time for a transition to take place) experienced by ethnic minority when compared to the majority groups. Frailty and stratifying terms were tested to take account of unobserved individual and geographic heterogeneities. Besides testing the default option of the frailty term in PHREG, three other frailty models were also tested to validate the significance test results of the frailty term.

Keywords: survival analysis, Cox model, stratifying and frailty models, PROC PHREG, PROC NLMIXED, and PROC GLIMMIX

INTRODUCTION

Survival analysis is a statistical procedure that deals with 'the analysis of data in the form of times from a well-defined "time origin" (or an event)' until the occurrence of another event (Gharibvand and Liu 2009: 1). This procedure has been primarily used in medical and biological research, such as time till death or time till the relapse of a disease, as it has also the ability to deal with incomplete or censored observations. Such technique has also been applied in social science research such as time to (un)employment or time to landing a skill-matching job because modeling cross sectional or panel data only provide partial snapshot pictures of dynamic process, which do no measure the degree of transition duration experienced by our sample of interest (Lockhead 2003; Wooldridge 2008). Since time is an important variable, survival analysis is more appropriate for modeling transition.

However, traditional methods in survival analysis assume populations are homogeneous and that individuals have the same risk of experiencing an event. As a result, they do not take into account the problem of dependence caused by unobserved heterogeneity (Wienke 2009; Allison 2010). Therefore, the standard errors may become too small, and may subsequently lead to incorrect confidence intervals and potentially misleading p-values. Hence, this paper focuses on the analysis of clustered data in survival analysis, and introduces stratifying or frailty terms to taking account unobserved (individual and geographic) heterogeneities.

SURVIVAL MODELS

In this paper, we model the length of time it takes immigrants in GB to find their first employment as an example to model clustered data for survival analysis. The first three models are a Cox's semi-parametric model, using PROC PHREG. This technique has several advantages over other survival models. It requires minimal assumptions about the distribution of event times; allows for modeling time-varying variables; has the ability to handle censored cases or ties for both continuous and discrete data as well as the capacity to fit semi-parametric frailty or stratification models (Allison 1984, 2010; SAS® Institute Inc. 2011). Furthermore, it is a robust model, so that the results generated will "closely approximate correct parametric model" (Kleinbaum and Klein 2005, 96).

The first model is a main effect model (a reference model) that looks at the independent net effect of each predictor on the log cumulative hazard function to measure any transition-duration penalties experienced by visible minority when compared to the white immigrant groups. The hazard rate $h(t)$ is the product of a non-parametric baseline hazard $h_0(t)$ and a parametric function of explanatory covariates X and corresponding parameters β, such that we have (Cantor 2003; Machin et al 2006; Mills 2011; SAS Institute Inc. 2011):

$$h_i(t) = h_0(t)\exp(\beta' X_i)$$

Using the counting process, SAS Code:

```
PROC PHREG DATA=job;
  CLASS ethnicity(ref='4') marital(ref='2');
   MODEL (start, stop)*event(0)= ethnicity age_arrival marital
   /ties=efron;
RUN;
```

The second model is a stratifying model that relaxes the proportionality assumption and looks at whether the underlying pattern of employment rate differs across regions with each stratum having its own baseline log hazard function:

$$h_i(t) = h_z(t)\exp(\beta' X_i)$$

where z represents the arbitrary functions of time between regions. SAS code:

```
PROC PHREG DATA=job;
  CLASS ethnicity(ref='4') marital(ref='2');
  MODEL (start, stop)*event(0)= ethnicity age_arrival marital
  /ties=efron;
  STRATA region;
RUN;
```

The third model is a frailty model (a new feature in SAS® software 9.3) testing whether immigrants residing in some counties are more prone to find first employment than immigrants living in other counties. The hazard rate is:

$$h_{ij}(t) = h_0(t) u_j \exp(\beta' X_{ij})$$

where j represents the j^{th} cluster for individual *i* at time *t*, and u_j is the random effect for cluster *j* (which represents the unobserved heterogeneity or frailty) . SAS code:

```
PROC PHREG DATA=job;
  CLASS ethnicity(ref='4') marital(ref='2') county;
  MODEL (start, stop)*event(0)= ethnicity age_arrival marital
  /ties=efron;
  RANDOM county;
RUN;
```

However, because the baseline hazard for Cox model is an unspecified, non-negative function of time and that the true underlying distribution of frailty is unknown, this paper proposes to cross-validate the frailty term and examine whether it is significant or not in other frailty models. Thus, three other frailty models were tested. Model four uses PROC NLMIXED to estimate a Weibull random-effects model:

$$h_{ij}(t) = p u_j \lambda t^{p-1}$$

where p is the shape parameter (when p=1, it becomes an exponential model and the hazard becomes constant hazard) and λ is a positive scale parameter, such that $\lambda=\exp(p^*\log(t) + \beta' X_{ij})$. SAS code (Allison 2011, 271):

```
PROC NLMIXED DATA=job;
  lambda=exp(b0+bW*eW+bE*eE+bBl*eBl+bP*eP+bBa*eBa+bO*eO+bage_arrival*age_arrival+
    bmm*mm+u);
  ll=-lambda*stop**(alpha+1)+event*(LOG(alpha+1)+alpha*LOG(stop)+log(lambda));
  MODEL stop~GENERAL(ll);
  RANDOM u~NORMAL(0,s2u) SUBJECT=county;
  PARMS b0=1 bW=0 bE=0 bBl=0 bP=0 bBa=0 bO=0 bage_arrival=0 bmm=0;
RUN;
```

The fifth and sixth models use PROC GLIMMIX to estimate discrete-time logit and complementary log log (cloglog) survival models with random effects, respectively. The logistic regression (proportional odds) model is written as:

$$\mathrm{logit}[h_j(t)] = \log[p_j(t)/(1-p_j(t))] = \alpha(t)+\beta' X_j(t)$$

where t=1, 2, 3, ….. and α(t) is the logit of the baseline hazard function. SAS code (Allison 2011, 279):

```
PROC GLIMMIX DATA=job METHOD=QUAD;
  MODEL event=eW eE eBl eP eBa eO age_arrival mm
  /DIST=BIN SOLUTION LINK=LOGIT;
  RANDOM INTERCEPT/ SUBJECT=county;
RUN;
```

The cloglog (discrete time proportional hazard) regression model is written as:

$$\mathrm{cloglog}[h_j(t)] = \log[-\log(1-p_j)] = \alpha(t)+\beta' X_j(t)$$

SAS code:

```
PROC GLIMMIX DATA=job METHOD=QUAD;
  MODEL event=eW eE eBl eP eBa eO age_arrival mm
  /DIST=BIN SOLUTION LINK=CLOGLOG;
  RANDOM INTERCEPT/ SUBJECT=county;
RUN;
```

DATA

To be consistent, the six models in this paper use the same dataset: the Family and Working Lives Survey (FWLS), which can be downloaded from Economic and Social Data Service at http://www.esds.ac.uk/findingData/snDescription.asp?sn=3704. The FWLS sample consists information of 11, 237 respondents from people living in the Great Britain (GB) between 1994 and 1995, and born between 1924 and 1978 (Rohwer 1996; McKay 1997). It is a cross-sectional survey with a retrospective longitudinal design that provides information about where the respondents were born and lived since birth for non-immigrants and entry for immigrants; when they lived in GB if born outside; the ethnic group they belonged to as well as other individual characteristics. Furthermore, it contains information about respondents' life and work history data such as changing family structures; levels of training and education; spells of (un)employment and occupational statuses. For this paper, we only look at immigrant sample (n=2094) and those who have not experienced any employment in the GB, but who are people looking for work the first time and having the possibility of experiencing first employment (i.e. economically active the first time in the GB and never get employed in the GB before). There are a total of 910 of these individuals. Because this paper focuses on techniques modeling unobserved heterogeneity within survival analysis, only three explanatory variables are selected for illustrating purposes and to model immigrants' time to landing first employment: 1) ethnicity (1=White, 2=Eire, 3=Black, 4=Indian, 5=Pakistani, 6=Bangladeshi and 7=Other); 2) age at arrival; and 3) marital status (1=married/cohabiting and 2=independent/single or divorced/separated/widowed; a time-varying variable). Other variables input into SAS are: 1) start (the time that immigrants start looking for a job); 2) stop (the time that immigrants find a job or censoring time, which is measured in months); 3) event (1=finds a job, 0=otherwise); 4) region (the stratifying term); and 5) county (the random or frailty term).

RESULTS

Table 1 Comparison Outputs from Six Models

Variables	Model 1 Cox Main Effect	Model 2 Cox Stratifying	Model 3 Cox Frailty
	B(SE)	B (SE)	B(SE)
Ethnicity (Ref: Indian)			
White	0.0656(0.1056)	0.0700(0.1114)	0.0626(0.1064)
Eire	-0.0922(0.1654)	-0.0410(0.1679)	-0.0789(0.1661)
Black	-0.1287(0.1059)	-0.0945(0.1075)	-0.1158(0.1067)
Pakistani	-02634(0.1105)	-0.2273(0.1155)	-0.2687(0.1116)
Bangladeshi	-0.3630(0.1081)	-0.2975(0.1125)	-0.3372(0.1103)
Other	-0.2129(0.2006)	-0.2077(0.2069)	-0.1912(0.2015)
Age at Arrival	0.0098(0.0049)	0.0101(0.0050)	0.0102(0.0049)
Marital Status (Ref: Independent/Single or Divorved/Widowed)			
Married/cohabiting	-0.5622(0.1170)	-0.5384(0.1196)	-0.5621(0.1172)
Random Effects			0.0048(0.0101)
AIC	10239.084	6756.794	
BIC	10277.223	6794/933	

Table 1 Comparison Outputs from Six Models continued…

Variables	Model 4 Weibull Frailty	Model 5 LOGIT Frailty	Model 6 CLOGLOG Frailty
	B(SE)	B (SE)	B(SE)
b0	-1.2029(0.1697)	3.8130(0.5640)	1.9638(0.3490)
Ethnicity (Ref: Indian)			
White	0.2725(0.1827)	1.3844(0.4651)	1.5237(0.4170)
Eire	-0.1710(0.2711)	-0.4228(0.5397)	-0.2139(0.6443)
Black	0.2855(0.1664)	0.6447(0.3339)	0.5016(0.3044)
Pakistani	-0.3562(0.2022)	-0.2116(0.3442)	-0.1208(0.2838)
Bangladeshi	-1.1271(0.2580)	-1.2362(0.4234)	-1.0447(0.3685)

Other	0.323(0.2815)	0.6237(0.5966)	0.2315(0.5256)
Age at Arrival	0.0292(0.0076)	0.0420(0.0128)	0.0404(0.0113)
Marital Status (Ref: Independent/Single or Divorved/Widowed)			
Married/cohabiting	-1.9621(0.1778)	-2.3843(0.2414)	-2.0683(02095)
Alpha	0.2021(0.0388)		
Random Effects	2.5078(0.2330)	12.2968(2.1739)	7.1081(1.3907)
AIC	4852.9	1956.45	1942.19
BIC	4896.2	1955.80	1981.54

Note: There is a convergence issue modeling time and time squared in the discrete-time models. This is likely because a majority of the immigrant sample finds jobs within month 1; therefore the time terms are not included in the discrete-time models.

Table 1 shows the estimates of covariate effects of the six survival models. The estimates from models one to three show similar results. The random term of the Cox frailty model is not significant at the 0.05 significance level. The estimates produced by models four to six show similar patterns, but display slight discrepancies from the Cox models. Such discrepancies are likely because the baseline hazards of Cox models are unspecified, and that they are approximation of the (correct) parametric survival models. Random terms of the Weibull, logit and cloglog models are, however, all significant at the 0.05 level, hence proving that it is essential to cross-validate the frailty term in other frailty models before concluding that unobserved heterogeneity is not present in a survival model (Keele 2007).

REFERENCES

Allison, P. D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. (Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-046). Thousand Oaks, CA: Sage.

_________. 2010. Survival *Analysis Using SAS: A Practical Guide, Second Edition*. Cary, NC: SAS Institute Inc.

Cantor, A.B. 2003 *SAS Survival Analysis Techniques for Medical Research, Second Edition*. Cary, NC: SAS Institute Inc.

Gharibvand, L. and L.Liu. 2009. Analysis of survival data with clustered events. *SAS Global Forum 2009 Paper* (237-2009): 1-11.

Keele, L. 2007. Cross validation tests for frailty models. Columbus, OH: Ohio State University.

Kleinbaum, D.G. and M. Klein. *Survival Analysis: A Self-Learning Text, Second Edition*. Springer.

Lochhead, C. 2003. The transition penalty: Unemployment among recent immigrants to Canada. Canadian Labor and Business Center.

Machin, D., Y.B. Cheung, and M.K.B. Parmar. 2006. *Survival Analysis: A Practical Approach, Second Edition*. Hokoken, New Jersey: Wiley.

Mckay, S. 1997. Final steps in constructing the family and working lives survey data archive: A research note. Social Security Unit. Loughborough University: the Center for Research in Social Policy.

Mills, M. 2011. Introducing Survival and Event History Analysis. Thousand Oaks, CA: Sage.

Rohwer, G. 1996. A practical introduction to the family and working lives survey: Part I introduction and data handling. Transition Data Analysis User's Manual. London: the Department for Education and Employment.

SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide, Cary, NC: SAS Institute Inc.

Wienke, A. 2009. *Frailty Models in Survival Analysis*. CRC Press.

Wooldridge, J. M. 2008. *Econometric Analysis of Cross Section and Panel Data, Second Edition*. MIT Press Books.

ACKNOWLEDGMENTS

I would like to thank Nate Derby at Stakana Analytics and Julie Petlick at SAS Student Programs for encouraging me to submit a paper to the 2012 Student Ambassador Program. I would also like to thank Murphy Choy and Erik Larsen for their support during the submission process.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

JuYin Helen Wong
Work Phone: 425-533-1455
E-mail: wong_ju@hotmail.com