Economics Letters 111 (2011) 84–87

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Estimating labour market transitions and continuations using repeated cross sectional data ☆

Pierre Brochu *

Department of Economics, University of Ottawa, 55 Laurier Avenue East, Desmarais Building, Room 10105, Ottawa (Ontario) K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 1 June 2009
Received in revised form 23 December 2010
Accepted 5 January 2011
Available online 13 January 2011

JEL classification:
C41
J64

Keywords:
Repeated cross-section data
Duration analysis
Employment
Job stability

ABSTRACT

This paper proposes a population cohort approach for estimating labour market continuations (or transitions) using repeated cross sectional data. This approach allows for the construction of consistent standard errors that account for the full variability of cross sectional data.

1. Introduction

There is a long tradition of exploring labour market transitions in economics. Although the unemployment–employment transition has been the most frequently explored, other transitions or continuations have also been examined, such as the transition out of the labour force (e.g. Jones and Riddell (1999)) and the continuation of a job (job stability, e.g. Brochu (2010), Heisz (2005), Neumark et al. (1999)).

While using panel data to estimate these labour market, transitions is generally the preferred approach, there are circumstances where that approach is problematic. For example, the limited historical coverage (Canadian panels) and data limitations (U.S. panels) make it difficult to differentiate between cyclical and secular changes in job stability.[1] With the absence of this differentiation, one cannot address the real question of interest in the job stability literature: how and why has job stability changed? In such instances, repeated cross sectional data sets offer a valid alternative.

In this paper, I propose a population cohort approach for estimating the continuation (or transition) probability when using repeated cross-section data. The proposed non-parametric approach is empirically tractable, and its identifying assumptions are relatively mild and easy to interpret. Using the proposed population cohort framework, I also re-examine the non-parametric estimator used in the job stability literature. I propose a consistent estimator for its standard errors—one that accounts for the full variability of cross sectional data.

Finally, I use Current Population Survey (CPS) data to show that the existing approaches tend to underestimate the true standard errors. This can lead the researcher to (incorrectly) conclude that job stability has changed.

2. Existing approach

Following the existing cross sectional literature (e.g. Neumark et al. (1999), Heisz (2005)), one can present the retention rate simply as the fraction of at-risk individuals in the *population* that remains with the same employer in the next period

$$R_t^{s,c} = \frac{N_{t+1}^{s+1,c}}{N_t^{s,c}} \tag{1}$$

where $N_t^{s,c}$ is the number of people in the population with time-invariant characteristics c who have been employed for s periods at time t.[2] Researchers (e.g. Baker (1992), Neumark et al. (1999)) take advantage of the fact that base weights of representative cross-

☆ I would like to thank David Green, Thomas Lemieux, Louis-Philippe Morin and Joris Pinkse for their helpful comments.
* Tel.: +1 613 562 5800x1424; fax: +1 613 562 5999.
E-mail address: pbrochu@uottawa.ca.
[1] See Brochu (2006) for details.

[2] For the remainder of this paper, the at-risk group consists of individuals with characteristics c who have been with the same employer for s periods as of time t.

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.01.007

sections, like the CPS (U.S.), sum up to their respective populations. The existing estimator is

$$\hat{Q}_t^{s,c} = \frac{\tilde{n}_{t+1}^{s+1,c}}{\tilde{n}_t^{s,c}} \tag{2}$$

where $\tilde{n}_t^{s,c}$ is the sum of the base weights of all individuals with characteristics c who have been employed for s periods with the same employer as of period t. By using weights as counts, the denominator (numerator) of Eq. (2) directly estimates the denominator (numerator) of Eq. (1).

Estimating population counts is very intuitive, and it is reasonable to think that the accuracy of the estimator will improve with larger samples; yet, this cannot be proven in any statistical sense. In addition, one cannot lay bare all underlying identifying assumptions without such a proof. Most importantly, the lack of precision carries over to the inference stage. Given the functional form of the estimator, there is no standard way to construct standard errors. In Section 4, I show that these approaches tend to systematically underestimate the true standard errors.

3. Proposed approach

I start with a population cohort.[3] Having a population cohort simply means that there is more than one period of information for each individual in the *population*. I assume that the repeated cross-sections are drawn from this population cohort, i.e. that each sample (cross-section) be drawn from the same population, but at different moments in time.

Let X_{it} be a vector of characteristics of individual i in period t; the characteristics are, for now, assumed to be time-invariant. Further, let TEN_{it} represent the length of tenure, i.e. the number of periods the worker has been employed with the same employer as of period t. The retention rate for the at-risk group, $R_t^{s,c}$, can be written as

$$R_t^{s,c} = \frac{E(1[TEN_{it+1} = s+1, X_{it+1} = c])}{E(1[TEN_{it} = s, X_{it} = c])} \tag{3}$$

where $1[\cdot]$ is an indicator function that equals 1 if the conditions inside the bracket hold, and zero otherwise.[4] One can estimate this retention rate using two repeated cross-sections, i.e.

$$\hat{R}_t^{s,c} = \frac{\sum_{i=1}^{n_{t+1}} 1[TEN_{it+1} = s+1, X_{it+1} = c] / n_{t+1}}{\sum_{i=1}^{n_t} 1[TEN_{it} = s, X_{it} = c] / n_t} \tag{4}$$

where n_t is the sample size in year t.

Eq. (3) is a key insight of this paper. It is conducive to cross sectional analysis because the numerator does not condition on period t events. This holds true because an individual who has been with the same employer for $s+1$ periods as of time $t+1$, had to have been with the same employer in the previous period (and has one less period of tenure).

Conditioning on only time-invariant characteristics is a sufficient but not a necessary condition for this result to hold. One only needs to be able to infer—from a period $t+1$ cross-section—whether an individual who remained with the same employer would have been part of the at-risk group in period t. Said differently, one needs to identify whose indicator function is a "1" in period $t+1$. One can, therefore, estimate a broad range of retention rates. One can not only condition on gender, race, education, but also on age, industry and occupation. I elaborate on the latter three categories below.

The above method can easily deal with ageing when $[t, t+1]$ spans one or more years. For the 1-year rate, identifying a "1" in $t+1$ is straightforward: the worker is simply one year older than he was in period t. Industry affiliation is job related, and as such, will change over time. Yet, for the proposed approach to work, one must only be able to identify the industry affiliation if he stayed with the same employer. This is possible if we assume that job tenure (i.e. the employer–employer relationship) ends when the individual switches industry. It is a similar story for occupation. This is a relatively mild assumption as long as the occupations/industries are not too narrowly defined.

Given the simple functional form of the proposed retention rate estimator, $\hat{R}_t^{s,c}$, one can easily generate consistent standard errors. In Proposition 2, I show how to do so by first deriving the asymptotic properties of $\hat{R}_t^{s,c}$. Finally, applying Eq. (4) to survey data where the probability of being selected is not the same across observations is straightforward. One replaces the sample means with weighted ones.[5]

I now relax the assumption of drawing from a population cohort because there are situations where this assumption is too restrictive. For example, the American job stability literature (e.g. Swinnerton and Wial (1995); Neumark et al. (1999)) estimated 4-year retention rates because the job tenure question was not part of the regular CPS question, but only included in select supplements. It would be untenable to assume that the two cross-sections—drawn 4-years apart—come from the same underlying population; the working-age population will have changed due to deaths, emigration and immigration. The population-cohort framework can be extended to deal with such compositional changes. Assuming that compositional changes break the tenure spell, one can write the retention rate as a function of two population means

$$R_t^{s,c} = \frac{adj_t E(1[TEN_{it+1} = s+1, X_{it+1} = c])}{E(1[TEN_{it} = s, X_{it} = c])} \tag{5}$$

where adj_t is the population growth (or adjustment) factor. The population growth factor would be 1.2 if, for example, the population size increased by 20%. An intuitive proof is left to Proposition 3.

A death easily meets the identifying assumption. For changes due to immigration and emigration, one requires that the migrant changes employer upon arrival in his new country. This empirical strategy would be appropriate if job transfers (where workers stay with the same employer) are not the driving force behind migration patterns.

The existing approach, i.e. Eq. (2), is in fact an estimator of $R_t^{s,c}$ as presented in Eq. (5). This becomes apparent if one rewrites Eq. (2) as[6]

$$\hat{Q}_t^{s,c} = \hat{adj}_t \left(\frac{\sum_{i=1}^{n_{t+1}} nw_{it+1} 1[TEN_{it+1} = s+1, X_{it+1} = c] / n_{t+1}}{\sum_{i=1}^{n_t} nw_{it} 1[TEN_{it} = s, X_{it} = c] / n_t} \right) \tag{6}$$

where nw_{it} is the normalized base weight of individual i in year t,[7] and $\hat{adj}_t = \frac{\sum_{i=1}^{n_{t+1}} bw_{it+1}}{\sum_{i=1}^{n_t} bw_{it}}$ (with bw_{it} representing the base weight).

Given that the sum of the base weights add up to the target population in the CPS, $\hat{adj}_t$ is an estimate of the population growth. The second term of Eq. (6) is simply the weighted version of $\hat{R}_t^{s,c}$.

By rewriting the existing estimator as a function of the proposed one, I can identify its underlying assumptions—namely that changes in population must break the tenure spell. Second, I can also easily construct consistent standard errors. They will be similar to those of $\hat{R}_t^{s,c}$, but with an adjustment made for the population change.[8]

---

[3] Other researchers (e.g. Deaton (1985) and Moffitt (1993)) who have estimated dynamic models using repeated cross-sections have also relied on this assumption.

[4] The proof can found in Proposition 1 in the Appendix A.

[5] Where the weights are normalized to sum up to 1 in each sample period.

[6] In Proposition 4, I show that Eqs. (2) and (6) are numerically equivalent.

[7] Where the base weights are normalized to sum up to 1 in each sample period.

[8] If one treats $\hat{adj}_t$ as a constant, then $se(\hat{Q}_t^{s,c}) = \hat{adj}_t \cdot se(\hat{R}_t^{s,c})$.