Catalogue no. 12-001-XIE

Survey Methodology

June 2004

Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

June 2004

Published by authority of the Minister responsible for Statistics Canada

April 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 30, No. 1, pp. 105-114
Statistics Canada, Catalogue No. 12-001

Longitudinal Analysis of Labour Force Survey Data

Geoff Rowe and Huan Nguyen [1]

Abstract

The Canadian Labour Force Survey (LFS) was not designed to be a longitudinal survey. However, given that respondent households typically remain in the sample for six consecutive months, it is possible to reconstruct six-month fragments of longitudinal data from the monthly records of household members. Such longitudinal micro-data – altogether consisting of millions of person-months of individual and family level data – is useful for analyses of monthly labour market dynamics over relatively long periods of time, 25 years and more.

We make use of these data to estimate hazard functions describing transitions among the labour market states: *self-employed*, *paid employee* and *not employed*. Data on job tenure, for employed respondents, and on the date last worked, for those not employed – together with the date of survey responses – allow the construction of models that include terms reflecting seasonality and macro-economic cycles as well as the duration dependence of each type of transition. In addition, the LFS data permits spouse labour market activity and family composition variables to be included in the hazard models as time-varying covariates. The estimated hazard equations have been incorporated in the LifePaths microsimulation model. In that setting, the equations have been used to simulate lifetime employment activity from past, present and future birth cohorts. Simulation results have been validated by comparison with the age profiles of LFS employment/population ratios for the period 1976 to 2001.

Key Words: Microsimulation; Censoring; Truncation; Employment dynamics.

1. Introduction

In recent years, there has been increased recognition of the importance of studying labour market dynamics using individual level (micro-) data. For this purpose, new panel surveys have been developed, for example, the Survey of Income and Labour Dynamics (SLID) (Statistics Canada 1998). But, existing LFS data (Statistics Canada 2002) provides a virtually untapped historical resource, in the form of many fragmentary event histories. From a conventional standpoint, the data currently comprises a time series of more than 300 cross-sectional surveys that were conducted monthly over more than 25 years. However, from a longitudinal perspective, those same data consist of about 6.5 million fragmentary event histories covering overlapping time intervals within the past quarter century and totalling over 34 million person-months of observation.

The analysis referred to in this paper was specifically directed towards development of hazard models to be incorporated in LifePaths (Statistics Canada 2001) – a micro-simulation model of the Canadian population. Further details on the LifePaths model are available from the Statistics Canada website at www.statcan.ca/english/spsd/index.htm.

The paper is organized in the following way. In section 2, we discuss some features of LFS data when reorganized as longitudinal records and we present three examples comparing estimates derived from the resulting longitudinal file with corresponding estimates from other sources. In section 3, we focus on the use of the data to model employment activity for LifePaths. There, we discuss the use of LFS micro-data in estimating hazard equations that describe employment dynamics. Finally, we present some illustrations of estimation results and a validation of LifePaths simulations that make use of the hazard equations.

2. Longitudinal LFS Data: Distinguishing Features and Proof-of-Concept

A longitudinal version of the LFS data was constructed by concatenating the monthly records of individual respondents into a file containing one record per respondent. Since an LFS respondent normally remains in the LFS sample for six consecutive months, we can obtain six-month histories for most respondents. These histories are not, by themselves, long enough for most longitudinal analyses. However, given the overlapping rotation groups that are part of the LFS design, these six-month fragments may be used in analysis of the experiences of employment cohorts over decades. (In line with the focus of the analysis below, we use the term "cohort" to refer to a relatively homogeneous group for all of whom a specified initial event has occurred. Thus, an "employment cohort" might refer to all persons who started a new job within a specified time period or, more narrowly, to all of those who started their third job

1. Geoff Rowe and Huan Nguyen, Socio-Economic Analysis and Modeling Division, Analysis and Development Branch, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6.

within a specified time period. The data available from the LFS determines how narrowly such a cohort can be defined here).

Figure 1, which illustrates some characteristics of the LFS data after they are formed into longitudinal records, focuses on changes in employment status for the employment cohort who started a job in January 1976. Respondents who were members of this cohort and who entered the sample through rotation 1 contribute data on the first six months, from January 1976, when the job started, to June 1976, when they left the LFS sample. For respondents from rotation 2, the six-month longitudinal data window shifts right one month (starting and ending one month later than those given by rotation 1). The overlapping data windows of respondents from subsequent rotations evolve similarly. Thus, the longitudinal LFS data can be seen as a combination of overlapping sets of panel data, in which respondents from the same rotation constitute a conventional data panel.

Successive six-month fragments of longitudinal LFS data can be combined to provide successive estimates of cumulative attrition from an initial employment cohort and, further, to identify new cohorts defined in terms either of a new job or of a period without employment. Thus, over the long term (currently up to 25 years), many different samples of individuals can contribute information about the same employment cohort observed at different points in time. Even so, month-to-month changes are observed largely from the same sample of individuals. The two shaded areas in Figure 1 illustrate this. The respondents from each of the rotations 2-5 contribute data for both the May-June and the June-July intervals.

This is not the first attempt to use LFS data longitudinally. Stasny (1986) and Lemaître (1988) studied errors in the estimation of "gross flows" between labour force states (*employed*, *unemployed* and *not in the labour force*) over intervals of one month. Lemaître found that problems arose both because of response errors and because "Labour Force Survey concepts, designed for cross-sectional purposes, tend to "create" flows when consecutive months' responses are linked". (Examples include the treatment of o*n-call workers* and of the *self-employed without a business).* Nevertheless, he concluded, "Administrative data have shown that not all sub-groups of status changers are seriously overestimated". Kinack (1991) examined the longitudinal consistency of responses to questions on job search activity that were used to distinguish between the categories *unemployed* and *not in the labour force*. He found substantial inconsistency, particularly when associated with proxy responses from different proxy respondents. These studies have shown that focusing on transitions between the categories *employed* and *not employed* (*i.e.*, without distinguishing between *unemployed* and *not in the labour force*) could help reduce the impact of response error.

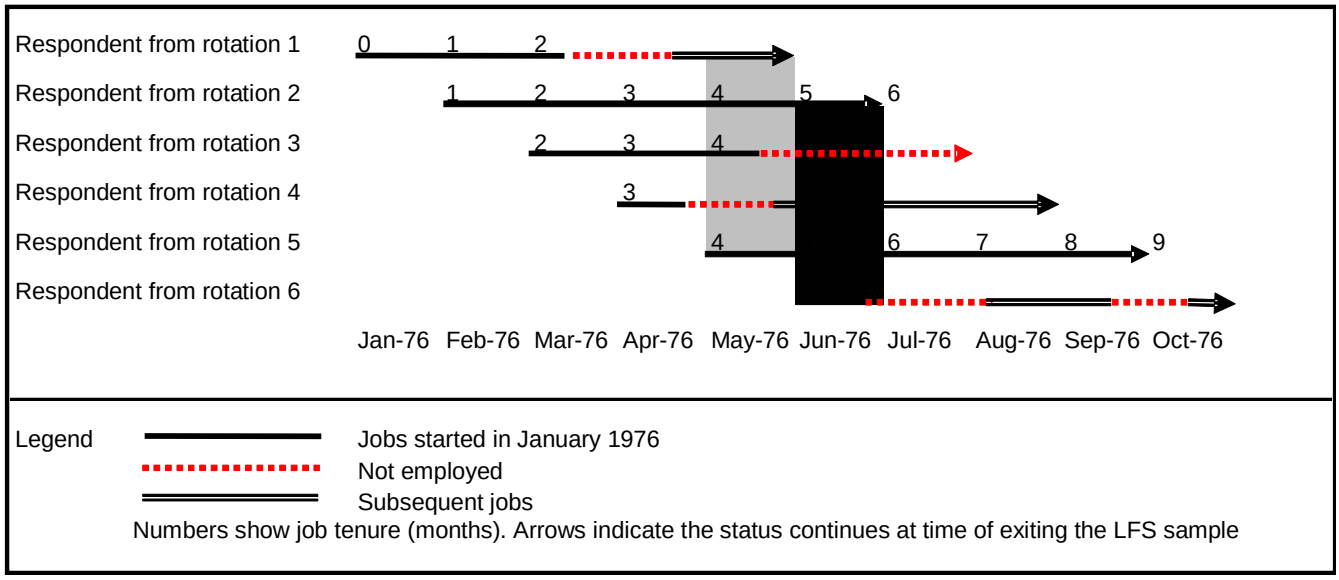

Figure 1. Illustration of LFS fragmentary data on cohort starting jobs in January 1976.

Cross-sectional LFS data have previously been used to estimate frequencies of job hiring and job separation over monthly intervals (Lemaître, Picot and Murray 1992). In that case, hiring was directly observed from the frequency of reported job-tenures of one month or less, while separation was determined residually using aggregate estimates of employment change together with the estimates of hiring. Cross-sectional LFS data have also been used to calculate and compare duration statistics for synthetic-cohorts. For example, Corak and Heisz (1995) use retention rates from a single time interval to represent a hypothetical cohort's experience. Synthetic-cohort retention rates were obtained using the numbers of employed LFS respondents reporting job tenure "t" in month "m" together with those reporting tenure "$t + 1$" the next month. Such uses of cross-sectional data have certain limitations. In particular, because the movement of individuals is not directly observed, destination states are unknown. (Although we may estimate the proportion that separated from a job, we can not estimate the proportion of those that became unemployed rather than dropping out of the labour force or beginning another job immediately). Nevertheless, a time series of synthetic-cohort statistics – for example, the proportions of jobs that might last a certain duration – can serve as an index that is sensitive to changing labour market conditions.

2.1 Proof-of-Concept: Selected Examples of Longitudinal Data Validation

The LFS data were not intended to be used longitudinally and problems can arise with such use (Stasny 1986; Lemaître 1988; Kinack 1991). Consequently, it is important to verify, for each analysis individually, that valid estimates can be obtained by month-to-month comparison of longitudinal responses. We present three examples of the verification of LFS longitudinal estimates below. In Figure 2, we compare estimates of the annual number of job separations in Canada from 1976 to 1995 (separations of all types, permanent and temporary) based on LFS data and on administrative data. The latter are based on Records of Employment (ROE) issued by employers at the time of job separation for Employment Insurance purposes (Statistics Canada 1998).

As may be seen, the number of transitions determined by month-to-month comparison of LFS data corresponds closely to the number from ROE data. Still, there are differences between the two series. Some of these differences could arise because of differences in coverage between the LFS and administrative data, as well as periodic changes in the LFS design or questionnaire. Another source of difference could arise because our counts based on LFS data neglect job separations of multiple job holders who remained employed in at least one job (*i.e.*, we counted only main-job changes). Nevertheless, we regard the degree of agreement between the LFS and administrative data as close enough to justify further analysis of the LFS micro-data. Both data sources imply that the annual rate of job separations was high: based on ROE data between 1978 and 1995, the average annual job separation rate for males was over 38 percent of annual person-jobs. Further analysis of the LFS micro-data can shed light on these dynamics.

Figure 3 goes further in the validation of employment dynamics, comparing "job survival" probabilities for males and females who started a job in 1993, as estimated from the LFS data and from SLID. (Note that 1993 corresponds to the first year of SLID data).

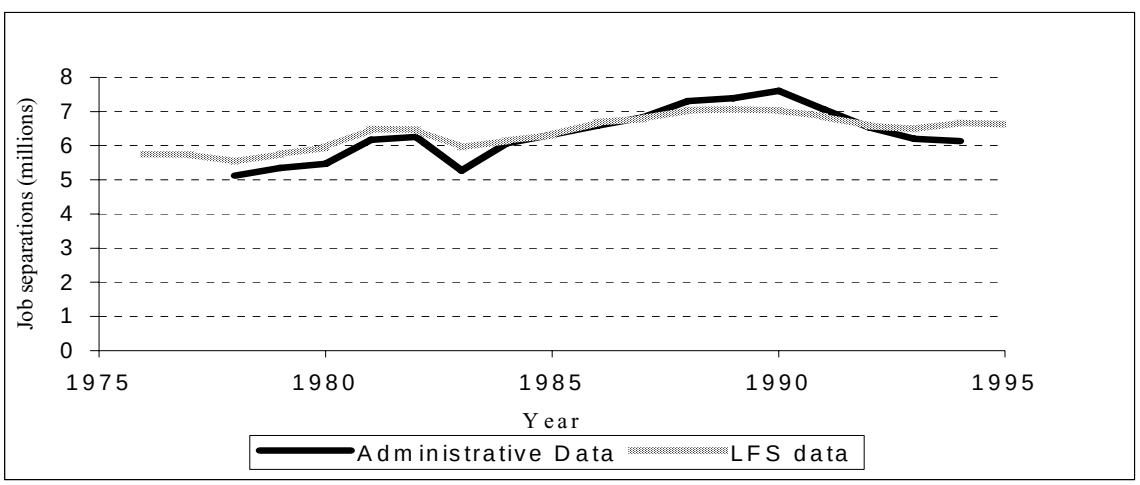

Figure 2. Estimates of Annual Job Separations.

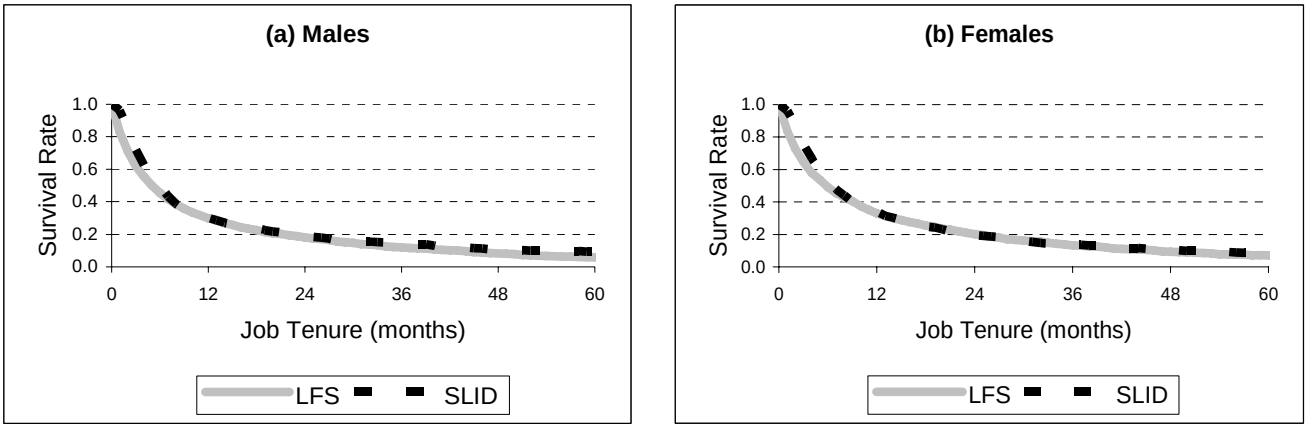

Figure 3. 'Job Survival' Probabilities of the Cohort Starting Jobs in 1993: Comparison of Estimates Based on LFS and SLID.

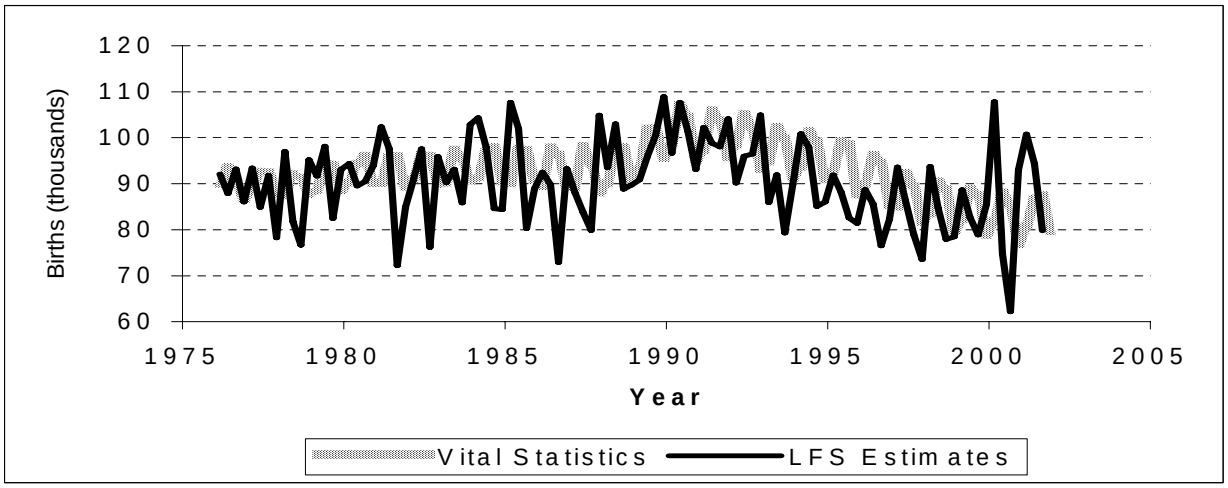

Figure 4. Estimates of Births in Canada by Quarter, 1976 – 2001.

The "job survival" probabilities were estimated from LFS data by the chained product of average retention rates derived from monthly main-job separation rates over the period 1993 to 1998. Survival probabilities from the SLID data were estimated in a similar manner using the reported job tenure and dates of job end. Both survival curves display the same characteristic shape; showing relatively high attrition for jobs of duration less than a year, but with much lower attrition rates at job tenures of one to five years. There are discrepancies between the estimates for durations of about six months or less, which may be related to the one-year recall period of SLID interviews and to the restriction of LFS job-tenure data to main-jobs. However, over periods as long as five years, the LFS and SLID provide very similar estimates. And, with the available LFS data, we can track some employment cohorts for as long as 25 years after the employment spell began.

A final illustration of effective longitudinal use of LFS data involves month-to-month comparison of the number of children aged less than one year as reported by female economic family heads or by the spouse of a male head. A infant child that is newly reported by a woman aged between 15 and 50 likely signifies the birth of a child. In order to make direct comparisons between these LFS estimates and vital statistics, we made some straight-forward adjustments to account for the proportion of births occurring to other women living in economic families (*e.g.*, teen lone parents living with their parents) and for births in the Yukon, NWT and Nunavut. A comparison of the resulting LFS monthly estimates of births with the corresponding counts of births registered in vital statistics (Figure 4) demonstrates that the LFS estimates follow secular trends in fertility as well as capturing some of the month-to-month fluctuation in births. Taken together, these three examples indicate that – with careful attention to survey coverage, survey concepts and the possibility of response error – the LFS can provide useful longitudinal micro-data.

3. Using Longitudinal LFS Micro-Data for Modeling Employment Activity in LifePaths

This section focuses on the use of the LFS data to simulate employment activity in LifePaths. Currently, LifePaths uses a 3-category classification of employment status – employee (E), self-employed (SE), and not employed (NE). We have not analyzed transitions involving unemployment. (Unemployment is a complex state requiring additional questions to ascertain and so, as noted above, unemployment transitions are particularly subject to response error).

There are six transitions that can result in a change in employment status (as represented in Figure 5). LifePaths models all of these transitions. In addition, job changes that do not appear to involve an interruption of employment are also modeled by LifePaths (denoted here as $E \Rightarrow E$). The LFS micro-data were used to estimate hazard equations for each of these seven transitions. The estimated coefficients of these equations became parameters in the LifePaths "Career Work" module. Below we discuss some technical issues that arise due to the limitations of the LFS data, followed by an illustration of the estimation results and then of a simulation outcome.

The fragmentary nature of these data poses a challenge for analysis. An important question is whether there are unavoidable biases that result from their fragmentary nature. In general, the answer is that the limitations of these data can be accounted for and potential sources of bias can be avoided with careful analysis.

3.1 Censoring and/or Truncation of Event Histories

One source of concern for an analyst of these data is the absence of retrospective employment information other than the length of the *current* employment spell. We might think of individual employment histories as consisting of a (largely unobserved) succession of contingent employment states (illustrated in Figure 6) with transitions among these states reflecting the process of career development. Thus, given only the transitions observable within the LFS window, the transition rates that can be estimated will inevitably involve pooling data from respondents who have had markedly different prior careers. In contrast, panel surveys like SLID, collect retrospective data at the first interview that, although limited, at least permits some experience rating of respondents in terms of previous extended work interruptions or periods of part-time work.

Another concern, illustrated in Figure 6, is that LFS employment spell durations may be left-truncated and/or right-censored. Right-censoring refers to the circumstance in which a spell ceases to be observed or a respondent ceases to be at risk without a transition occurring of the type being studied. This happens either (1) because the respondent's household "rotated out" of the LFS sample before any transition occurred, or (2) because another transition occurred that was not of the type under active study. Similarly, these data are frequently left-truncated. This refers to the circumstance in which the beginning of a spell is unobserved, because it happened before the respondent's household "rotated in" to the LFS sample. (These data are left-truncated rather than left-censored, because respondents provide the information necessary to determine the elapsed duration of the current spell at the time of the first interview). Since both truncation and censoring are generally independent of employment event processes, neither should lead to bias in the estimation of transition probabilities, if properly accounted for in the likelihood function.

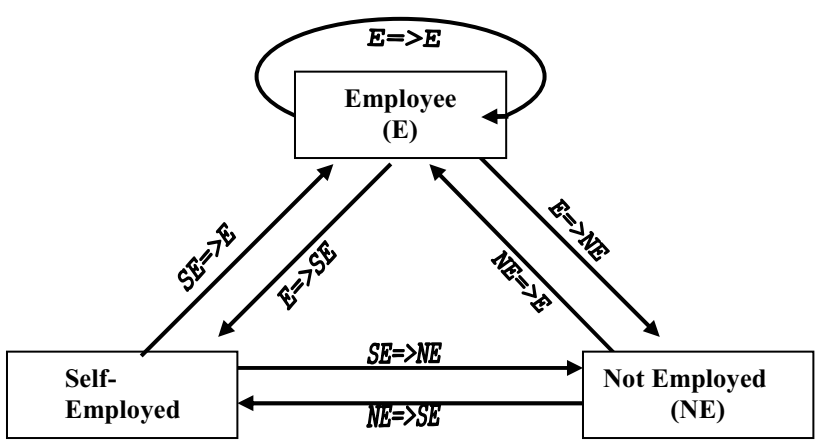

Figure 5. Employment Status and Transitions in LifePaths.

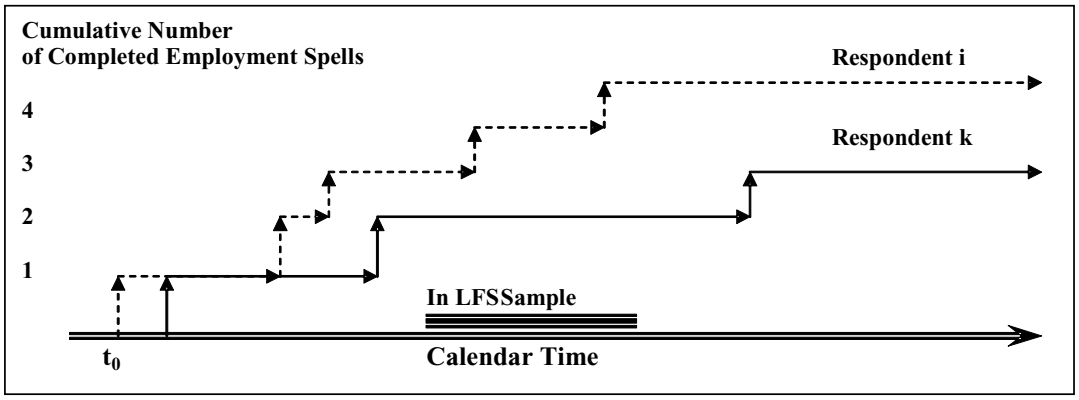

Figure 6. Recurrent Events and Employment Spell Durations Observable within the LFS Sample Window.

The combination of full and partial information provided by left-truncated and right-censored data can be represented in a conditional likelihood (Wang 1991). In a competing risks framework, the likelihood of an employment transition type *j* involving respondent *i* may be expressed in terms of the spell duration observed *k* months after *i* was first observed to be at risk of transition *j*. Let t_i denote the year and month of the LFS interview in which *i*'s current employment state was first observed (*i.e.*, often the first interview). Based on information collected at each interview, we can determine the length of the current spell of employment or spell not employed (m_{t_i}). Then $m_{t_i+k} = m_{t_i} + k$ would denote the elapsed spell duration in the state as assessed *k* months after the first observation – assuming no intervening events – and the likelihood of a transition of type *j* (*i.e.*, L_{j,t_i+k}) can be expressed in terms of m_{t_i+k}. Terms in the likelihood function comprise: the probability density of durations leading up to transitions of type *j* ($f_j(m_{t_i+k})$), the corresponding cumulative probability ($F_j(m_{t_i+k})$), a binary variable indicating whether or not censoring has occurred (C_{j,t_i+k}), and a further binary variable indicating whether or not the current spell was left-truncated ($LT_{i,j}$). Note that, in the competing risks framework, the density $f_j(m_{t_i+k})$ relates to a latent variable – the waiting time leading specifically to transition *j* – and that we must assume there is one such density for each competing event. In principle, the completed spell duration (observed when a transition occurs) will correspond to the minimum of competing, latent waiting times.

To account for left truncation, the likelihood is expressed in terms of conditional probabilities given the spell duration first observed (m_{t_i}): these probabilities take the form either of conditional probabilities evaluated at the time of an observed transition ($f_j(m_{t_i+k}/m_{t_i})$) or of conditional probabilities of surviving – without the occurrence specifically of transition *j* – to the observed duration ($1 - F_j(m_{t_i+k}/m_{t_i})$), depending on whether or not censoring has occurred.

$$L_{i,j,t+k} = f_j\left(m_{t_i+k} \middle| m_{t_i}\right)^{1-C_{j,t_i+k}} \left(1 - F_j\left(m_{t_i+k} \middle| m_{t_i}\right)\right)^{C_{j,t_i+k}}$$

$$= \frac{f_j\left(m_{t_i+k}\right)^{1-C_{j,t_i+k}} \left(1 - F_j\left(m_{t_i+k}\right)\right)^{C_{j,t_i+k}}}{\left(1 - F_j\left(m_{t_i}\right)\right)^{LT_{i,j}}}. \qquad (1)$$

This likelihood accounts for all of the information we have regarding the specific risk of transition *j* and can incorporate the effect of other competing risks by treating them as censoring events that are in addition to censorship by "rotating out" of the sample. Competing risks problems are commonly formulated in terms of such latent waiting times, especially in epidemiology and biostatistics, but also in economics (*e.g.*, Heckman and Honoré 1989). However, while providing a mathematically convenient motivation for the likelihood, the approach has been criticized "on the basis of unwarranted assumptions, lack of physical interpretation and identifiability problems" (Prentice, Kalbfleisch, Peterson, Flournoy, Farewell and Breslow 1978).

The conditional likelihood (1) can be approximated by a Poisson likelihood (Holford 1980; Laird and Olivier 1981), thereby also acknowledging the discreteness of the data (*i.e.*, transitions are generally "observed" in the one month interval between successive interviews). Equation (1) can be re-expressed in terms of a binary variable (Y_{j,t_i+k}) that represents occurrence or non-occurrence of a transition in a particular time interval (note that $Y_{j,t_i+k} = 1 - C_{j,t_i+k}$). Then, Y_{j,t_i+k} is treated as a Poisson random variable having an expected value equal to the hazard "h_{j,t_i+k}" which is assumed piecewise constant. Under this model, the contribution from *i* to the log-likelihood over *n* periods (using $h_{j,t_i+k} = f_j(m_{t_i+k}) / (1 - F_j(m_{t_i+k})) = -\partial \ln(1 - F_j(m_{t_i+k})) / \partial m_{t_i+k}$ together with (1)) is approximately:

$$\ln(L_{i,j}) \approx \sum_{k=1}^{n} \left[Y_{j,t_i+k} \ \ln(\hat{h}_{j,t_i+k}) - \hat{h}_{j,t_i+k} \right]. \qquad (2)$$

It is common practice to account for a complex survey design by means of a "pseudo" likelihood that incorporates the survey weight. Maximizing the "pseudo" likelihood corresponds to minimization of a weighted sum of deviance terms (*i.e.*, terms representing the difference between estimated likelihood contributions and their maximum possible values). Thus, the full-sample, conditional log-likelihood for transition j may be transformed into a weighted deviance D_j (note that W is derived from the survey weights and, since transitions are typically identified by comparing employment states between interviews, we use averages of consecutive cross-sectional survey weights to obtain W):

$$D_j \approx -2 \left(\begin{array}{l} \sum_i \left[\sum_{k=1}^{n_i} W_{t_i+k} Y_{j,t_i+k} \ \ln\left(\hat{h}_{j,t_i+k}\right) \right] \\ + \sum_i \left[\sum_{k=1}^{n_i} W_{t_i+k} \left[Y_{j,t_i+k} - \hat{h}_{j,t_i+k} \right] \right] \end{array} \right). \qquad (3)$$

In the analysis of each transition type j, we treat other events (*i.e.*, non-j events occurring to the same population-at-risk) as censoring, and so the deviance for a set of such events will be the sum of component deviances (*i.e.*, if the overall hazard is the sum of competing hazards, then the competing risks may be treated as independent (Prentice *et al.* 1978)).

A more direct motivation of the same deviance takes Poisson processes as its starting point (Borgan 1984; Andersen 1985; Andersen and Borgan 1985; Lawless 1987), rather than starting with postulated event-specific, latent, duration densities like $f_j(m_{t_i+k})$. In this case, we can model sampled multivariate counting processes that represent the number of occurrences of each specific transition in time intervals $[t_0, t)$. Sample counting processes, represented by the step functions in Figure 6, are observable counterparts of cumulative hazard functions. The assumption that the underlying hazard functions are approximately piecewise constant leads directly to the Poisson deviance as an approximation (Lindsey 1995). To limit bias, the principal concerns are that the population-at-risk can be identified, that censoring or truncation mechanisms are conditionally independent of the underlying employment processes and that the intervals over which hazards are assumed constant are not too large.

It is possible to obtain simple averaged estimates of employment hazard functions (such as those displayed in Figure 3) by implicitly splicing together all available information on members of a defined cohort from the longitudinal LFS samples. (That is, maximizing likelihood (1), but without considering any covariates). Making allowance for censoring and truncation in this way is a relatively simple example of such problems compared with the more complex observation schemes considered by Alioum and Commenges (1996). This implicit splicing of information is apparent in the deviance (3) which has two components: the first component is non-zero only at observed transitions, while the second component reflects the weighted differences between cumulative events and cumulative hazards (accumulated over all durations prior to the events or to censoring times). To the extent that the LFS cross-sections are representative samples for each reference week, then – taken together – they will provide an accurate estimate of the numbers of events occurring over the "life" of an employment cohort. Similarly, within samples from employment cohorts, we can expect to find left-truncated and right-censored respondent spells that might fill-in the missing prior histories of those left-truncated spells that terminate with a transition. As such, the first component of the deviance will accurately reflect whether hazard estimates tend to be large over periods where observed events are frequent. And the second component, summed over all respondent-months, may have a value similar to that which we might have obtained had there been no left-truncation. So, for data as extensive as these, the conditional likelihood may be almost equivalent to an unconditional likelihood.

3.2 Estimating Employment Transition Hazard Equations

Patterns of employment transition differ significantly among different demographic groups. For example, full-time students are most active in the labour market during their summer break, whereas the maternity leave that an employed pregnant woman takes may be largely determined by Employment Insurance regulations. Accordingly, LifePaths distinguishes among the following groups and models their employment activities separately:

- Those who are full-time students;
- Those who have just graduated or left school and are in a transition to an after-school job;
- Pregnant women for whom a maternity-leave may apply;
- Those who are in prime ages of employment; and
- Older workers in transition to retirement.

We discuss here only the estimation for the fourth group, comprising individuals who are in what is referred to in LifePaths as their "career employment" phase (the most important phase in terms of impact on the economy). Particulars for the other groups are available from the Statistics Canada website noted above.

For implementation in LifePaths, our hazard model uses a log-linear form of regression equation – one equation for each of the 7 transitions and for each sex separately, giving a total of 14 equations:

$$E\left(Y_{j,t_i+k}\right) \approx \hat{h}_{j,t_i+k} = \exp\begin{pmatrix} \hat{g}\left(m_{j,t_i+k}\right) \\ + X_{j,t_i+k}\,\hat{\beta} \end{pmatrix} \qquad (4)$$

where $E(\,)$ is the expectation operator, $g(m)$ is a log-linear spell duration spline, X is a vector of time-varying covariates and β is a vector of regression coefficients. The term $g(m)$ corresponds to a piecewise Weibull baseline hazard, which, in our specification, distinguishes employment transition risks at durations of less than a year from risks at durations of more than a year. The covariates, X, include variables representing individual age, education, province of residence, presence of children by age group, spouse's employment status, calendar month and calendar year, as well as interactions among some of these factors. Final estimates of β and $g(m)$ minimize the deviance (3).

The only example of detailed results that we present here involves the mutual influence of husband's and wife's employment status on each other's respective transition hazards. Figure 7 compares coefficient estimates from the seven equations that correspond to the seven transitions we specified. The two panels correspond to the separate sets of equations for males and females. The category "no spouse present" was treated as the reference category and the spouse's employment status was classified into "with paid employment", "self-employed", and "not employed". The estimated coefficients are presented here in terms of risk relative to the reference group. Thus, with other covariates controlled, the hazard of becoming self-employed for female employees whose husbands are self-employed is about 2.5 times higher than the hazard of their counterparts who do not have a spouse (see tallest bar in the top panel).

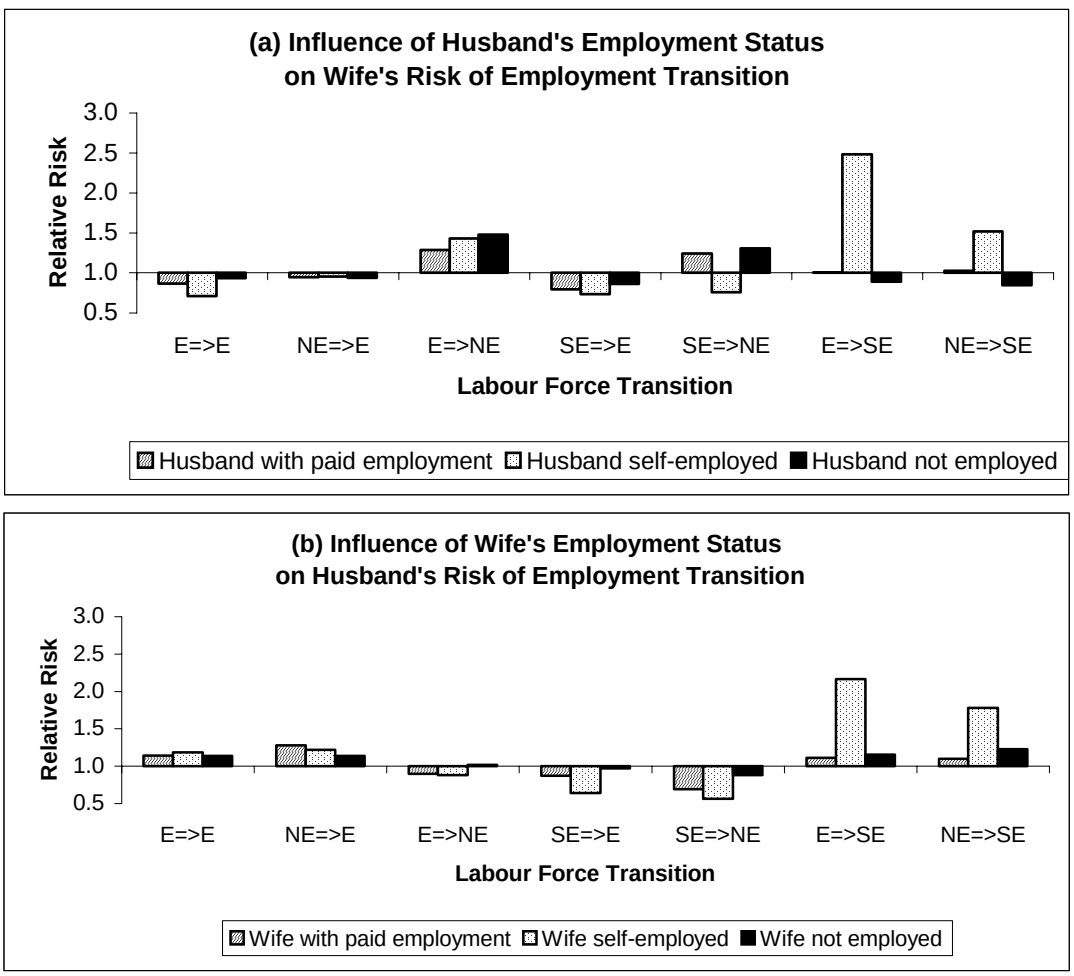

Figure 7. Impact of Spouse's Employment Status on Employment Transition Risks.

Figure 7 shows that the very presence of a spouse can work in opposite directions for males and females. The most frequent transitions for both sexes are *E* => *E*, *NE* => *E* and *E* => *NE*. For females, the first two of those transitions are less likely to occur to married women than to single women, while the transition to "not employed" is more likely. (The presence of children is not the reason for this, as their presence is accounted for by other terms in the equation). For males, the pattern is reversed. Thus, these results appear consistent with conventional gender roles. However, taking account of the magnitudes of these relative risks, we are not given the impression that gender roles have a particularly strong influence after the influence of other variables is credited.

Figure 7 reveals another conspicuous pattern. First, the relative risks of a transition into self-employment, for spouses with husbands/wives in self-employment, stand out as the highest among all other transitions. In addition, spouses with husbands/wives in self-employment have the lowest relative risks of a transition out of self-employment. Thus, self-employment status seems to be mutually reinforcing within families. These observations are consistent with forms of joint self-employment involving a family business (*e.g.*, a corner store) or involving endogamy among professionals (*e.g.*, lawyers marrying other lawyers).

4. From Estimated Parameters to the Simulation Results: An Illustration

Our example of the role of spouse's employment status points to the need for family context in the simulation of employment activities. It is a challenge for LifePaths to integrate these relationships into the simulation process. For example, if individual education progression or the effects of education on employment transitions are not modeled appropriately and accurately, then the consequences will cascade from direct education-employment relationships to a chain of indirect impacts, involving relationships between education and marriage, fertility, interprovincial migration, *etc*. These impacts would then spill over to the simulated spouse, as indicated above. It is not difficult to see that, unless these relationships are specified appropriately and the parameters are estimated with reasonable accuracy, bias would be spread over a wide range of simulated outcomes.

An overall validation of the LifePaths employment hazard equations was obtained by comparing simulated annual average employment/population ratios with direct cross-sectional estimates from the LFS. The simulated employment/population ratios were obtained from a synthetic population whose members were exposed appropriately to one or other of the seven types of employment hazards over the course of each simulated year. The simulated employment/population ratios were calculated from the resulting annual person-years of employment in the synthetic population: that is, these ratios are an outcome of simulated flows into and out of employment. The simulations necessarily involved generating appropriate distributions of covariates that in turn determine the distributions of employment transition hazards. As may be seen in Figure 8, LifePaths accurately reflects the age patterns of female employment in both 1976 and 2001 and correspondingly accounts for the dramatic change observed in those age patterns over the past quarter century.

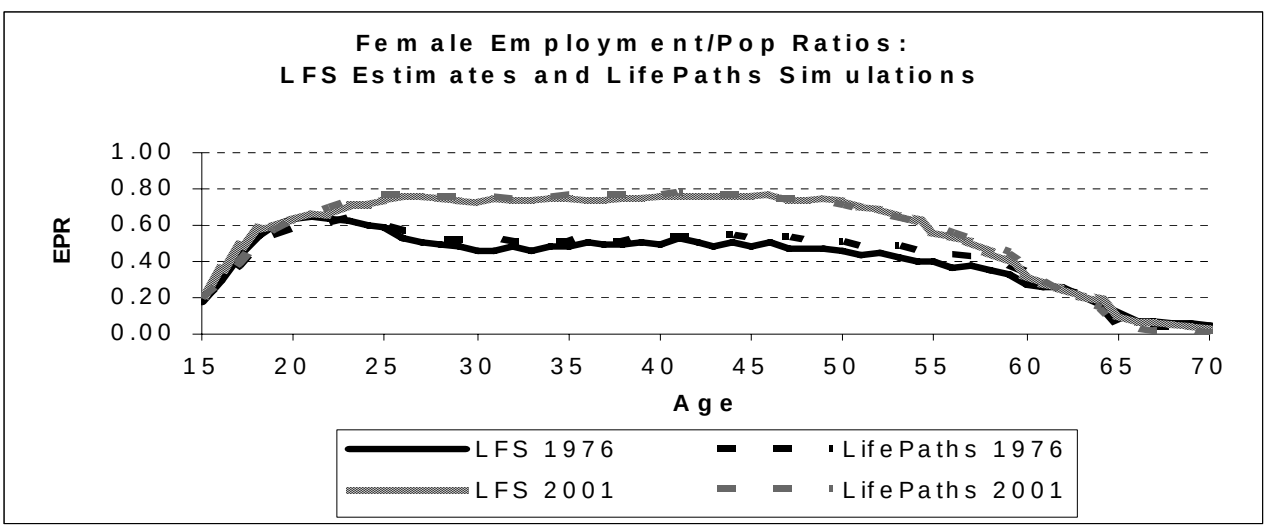

Figure 8. Validating hazard equations using LifePaths.

5. Conclusions

We have demonstrated that the LFS data – when organized into the fragmentary event histories collected over the six-month periods that most respondents spend in the sample – represents a significant longitudinal micro-data asset. There is sufficient sample and breadth of content to provide for important analysis of labour market dynamics and, conceivably, of demographic processes such as fertility. Moreover, the data is monthly and spans more than a quarter century, so that analysis based on it has uninterrupted time depth that is unique in Canada.

In our main application (employment transitions), other results (not reported here) appear to confirm the influence of a range of explanatory variables on an individual's chances of an employment transition. These covariates include age, job tenure (or duration not employed), educational attainment, presence of young children (especially for women), province of residence, seasonality, and business cycles. However, this work is still in its initial stage and, to date, our approach to inference has been informal. Future work will involve extending and refining our models and establishing a more rigorous basis for evaluation of the models.

References

Alioum, A., and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52, 512-524.

Andersen, P. (1985). Statistical models for longitudinal labor market data based on counting processes. In *Longitudinal Analysis of Labor Market Data*, (Eds. James J. Heckman and Burton Singer). Cambridge University Press, Cambridge.

Andersen, P., and Borgan, Ø. (1985). Counting process models for life history data: A review. *Scandinavian Journal of Statistics*, 12, 97-158.

Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, 11, 1-16.

Corak, M., and Heisz, A. (1995). The duration of unemployment: A user guide. Research Paper Series No. 84, Analytical Studies Branch, Statistics Canada.

Heckman, J.J., and Honoré, Bo.E. (1989). The identi- fiability of the competing risks model. *Biometrika*, 76(2), 325-330.

Holford, T.R. (1980). The analysis of rates and survivorship using log-linear models. *Biometrics*, 36, 299-305.

Kinack, M. (1991). Measuring data quality with longitudinal data. *1991 Proceedings of the Section on Survey Research Methods*, American Statistical Association, 514-519.

Laird, N., and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76, 231-240.

Lawless, J.F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 82, 808-815.

Lemaître, G. (1988). The measurement and analysis of gross flows. Working Paper, Labour and Household Surveys Analysis Division, Statistics Canada.

Lemaître, G., Picot, G. and Murray, S. (1992). Workers on the move: An overview of labour turnover. *Perspectives on Labour and Income*, 4(2), Statistics Canada.

Lindsey, J.K. (1995). Fitting parametric counting processes by using log-linear models. *Applied Statistics*, 44, 201-212.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, Jr. N., Farewell, V.T. and Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34, 541-554.

Stasny, E.A. (1986). Estimating gross flows using panel data with nonresponse: An example from the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 81, 42-47.

Statistics Canada (1998). Permanent layoffs, quits and hirings in the Canadian Economy: 1978-1995. Catalogue # 71-539-XIB.

Statistics Canada (1998). Survey of Income and Labour Dynamics: A Survey Overview. Catalogue # 75F0011XPB, http://www.statcan.ca/ english/ freepub/ 75F0011XIE/free.htm.

Statistics Canada (2001). The LifePaths Microsimulation Model: An Overview. http://statcan.ca/english/spsd/ LifePaths.htm.

Statistics Canada (2002). Guide to the Labour Force Survey. Catalogue # 71-543-GIE, http://www.statcan.ca/english/IPS/ Data/71-543-GIE.htm.

Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86, 130-143.