

# **Alternative Measures of the Average Duration of Unemployment**

by Miles Corak\* and Andrew Heisz\*\*

**No. 83**

**11F0019MPE No. 83**

**ISSN: 1200-5223**

**ISBN: 0-660-15585-0**

24, R.H. Coats Building, Ottawa, K1A 0T6  
Business and Labour Market Analysis, Statistics Canada

\*(613) 951-9047 Internet: coramil@statcan.ca

\*\* (613) 951-3748

Facsimile Number: (613) 951-5403

The paper is available on Internet: ([www.statcan.ca](http://www.statcan.ca))

December 1995

An earlier version of this paper was presented to the 23rd General Conference of the International Association for Research in Income and Wealth held at St. Andrew's New Brunswick. We thank Denise Dorion, Andrew Harvey, Ian Macredie and Deborah Sunter for helpful comments. The contents of the paper remain the sole responsibility of the authors.

This paper represents the views of the author and does not necessarily reflect the opinions of Statistics Canada.

*Aussi disponible en français*

## **ABSTRACT**

This paper examines the robustness of a measure of the average complete duration of unemployment in Canada to a host of assumptions used in its derivation. In contrast to the average incomplete duration of unemployment, which is a lagging cyclical indicator, this statistic is a coincident indicator of the business cycle. The impact of using a steady state as opposed to a non steady state assumption, as well as the impact of various corrections for response bias are explored. It is concluded that a non steady state estimator would be a valuable compliment to the statistics on unemployment duration that are currently released by many statistical agencies, and particularly Statistics Canada.

Key words: Unemployment, Duration, Dynamics, Canada.

## **Alternative Measures of the Average Duration of Unemployment**

The unemployment rate, while certainly being one of the most closely watched economic indicators, offers on its own a rather incomplete picture of the labour market. An unemployment rate of say 10 per cent may reflect a situation in which 10 per cent of the labour force becomes unemployed each month and spends only a few weeks looking for a job, or a case in which the same 10 per cent is unemployed for the entire year. In the first case the labour market is characterized by a great deal of flux with a spell of unemployment not having serious consequences, while in the latter it is a stagnant market with unemployment implying severe hardship. The welfare implications of these two possibilities may be very different, and to accurately understand the situation requires a reliable indicator of the average duration of a spell of unemployment.

The design of surveys such as the Canadian Labour Force Survey (LFS) or the U.S. Current Population Survey (CPS), like that of similar surveys in other countries, recognizes the dynamic nature of the labour market. Official releases of information, however, are limited to grouped data on the reported duration of unemployment, and the average duration of in-progress (that is incomplete) unemployment spells. Analysts have been able to use this information to develop measures of the average length of a *complete* spell of unemployment. This has been done both non-parametrically and parametrically. Baker and Trivedi (1985) suggest that non-parametric methods, which rely on the results of Kaplan and Meier (1958), are superior. Kaitz (1970) using data from the CPS is an early example. More recent examples from the United States are Sider (1985) and Baker (1992a), while Corak (1993) uses Canadian data.

This well established literature has led to important insights that can be used to reconsider the way duration statistics are officially released. The major objective of this paper is to explore some of these insights and related issues, and to recommend an approach that is sufficiently robust for official use. Canadian data are used throughout, but many of our conclusions and recommendations may extend to data from other countries. We begin by comparing the average incomplete duration of unemployment to a measure of the average complete duration with the intention of illustrating some of the biases in the former. We then discuss the derivation of the latter. In particular, we examine the robustness of the average complete duration of unemployment according to the use of a steady state or a non steady state assumption, as well as to the manner in which the response biases inherent in the data are corrected. We find that the average complete duration of unemployment based upon a steady state assumption, although computationally simpler, may seriously understate spell duration when the unemployment rate is rising (when inflows to unemployment increase), as well as overstate it when unemployment is falling (when inflows decrease). The steady state assumption also distorts the seasonality of the statistic. We conclude that the computational advantages of the steady-state estimator are slight, and do not outweigh these rather serious biases. We also find significant spikes in the frequency distribution of reported unemployment duration associated with a "digit preference" by survey respondents. However, the smoothing of the data to correct this measurement error is only crucial to the level of the average complete duration not its

cyclical variability, and is most important at the densest parts of the distribution. On this basis we recommend an approach that is sufficiently robust for official release and argue that it is a valuable compliment to the sole statistic on the (incomplete) duration of unemployment that is currently released by many statistical agencies.

### **An Overview**

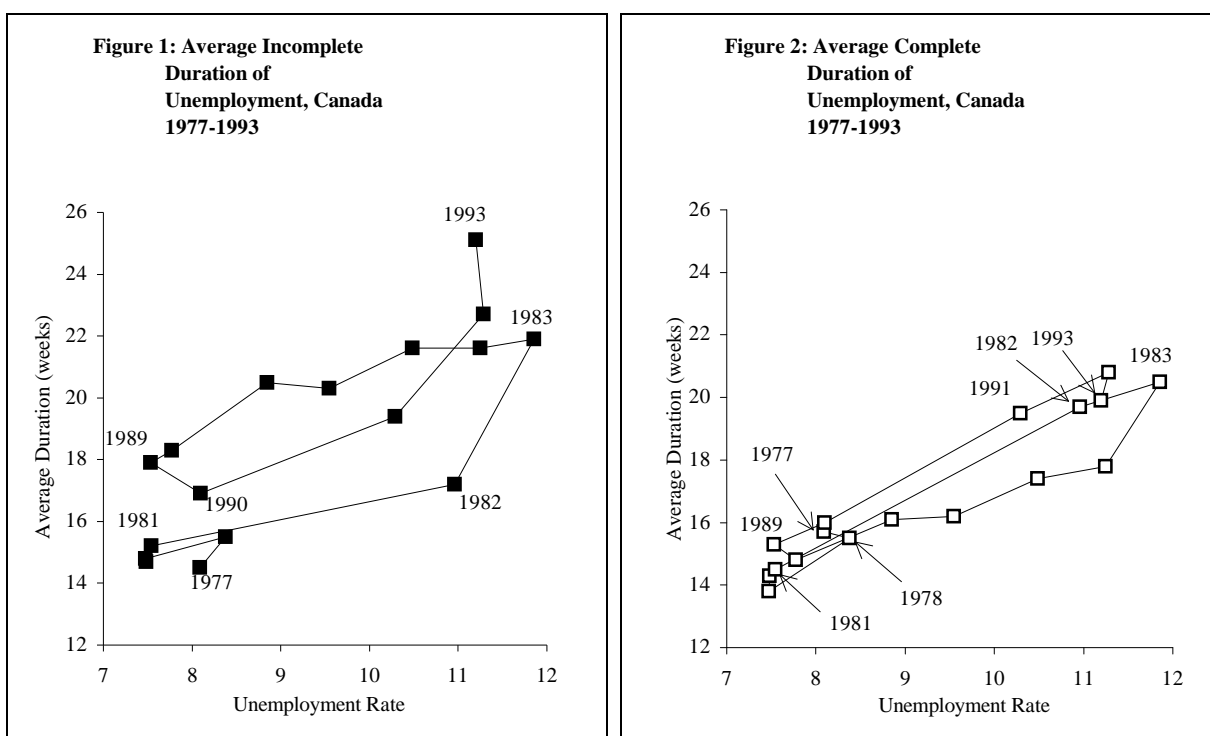
The average duration of unemployment, as it is officially released by Statistics Canada, is derived from a sample of currently unemployed individuals. The LFS does not capture the complete length of an unemployment spell, only the time spent unemployed up to the reference week. The spell may continue for some time afterwards, or it may end the next day. The average duration of unemployment is the sum of all these in-progress spell lengths divided by the number of unemployed.<sup>1</sup> As such the official statistic is the average *incomplete* duration of unemployment for the *currently* unemployed.

The statistic derived and examined in this paper is the average expected complete duration of unemployment for a cohort of individuals that begin their spell of unemployment at the same time. It is a measure of the complete length of an unemployment spell, and is based upon the assumption that the economic conditions prevailing at the time a cohort becomes unemployed will continue throughout the entire spell. In what follows we refer to it as simply "the average complete duration of unemployment."

The average incomplete duration of unemployment is a biased measure of the average complete duration for two reasons: a length bias and a sampling bias. These are clearly presented and discussed in Salant (1977). The length bias arises, obviously enough, from the fact that only in-progress spells are sampled: it implies an underestimate of the complete spell length. The sampling bias refers to the fact that the probability that an unemployed individual will be captured by the survey is proportional to the length of his or her unemployment spell: those experiencing short spells of unemployment will as a result be under-sampled by such point in time surveys. This bias implies that the complete spell length will be overestimated. Salant provides a theorem to show that the average incomplete duration will be greater than the average complete duration if the hazard rates governing the transition out of unemployment decline with time spent unemployed, that is the sampling bias will outweigh the length bias. For the most part this is the case in the data we examine. Over the 1977-93 period the average incomplete duration is 18.7 weeks, while the average complete duration is 16.9 weeks.

---

<sup>1</sup> For the most part the LFS classifies survey respondents as unemployed if they are without work and looking for work. Individuals classified as being on temporary layoff are not required to fulfill the job search requirement in order to be considered unemployed. In this case the duration of unemployment is the number of weeks since the layoff began. Furthermore, individuals may also be deemed unemployed in the LFS if they have found a job and expect to start within four weeks. The duration of unemployment is recorded for these "future starts" only if they also happened to be searching for work in the reference week. See Statistics Canada (1992).



This difference between the two statistics is well known. Another difference that deserves mention concerns their cyclical variation. The relationship between each and the Canada wide unemployment rate is depicted in Figures 1 and 2. The average incomplete duration displays a broad counter-clockwise loop. In large part this is due to the fact the composition of the unemployed changes over the business cycle with the result that the average incomplete duration is a *lagging* cyclical indicator. At the onset of a recession large inflows into unemployment result in the stock of unemployed becoming more heavily weighted with individuals just beginning a spell of unemployment. While these individuals may ultimately experience long spells of unemployment, only the length of unemployment up to the time of the survey is used in calculating the average spell length. For example, between 1981 and 1982 as the economy entered into recession the unemployment rate increased by 3.5 percentage points, but the average LFS duration increased by about only one week. Similarly, as the economy moved from expansion to recession between 1989 and 1990 the unemployment rate increased, but the average incomplete duration actually fell.<sup>2</sup>

The pattern is just the opposite during recovery and expansion: flows into unemployment fall, and the stock of unemployed becomes more heavily weighted with individuals who are in the midst of rather long spells that began during the recession and reflect the state of the economy during that period. Thus, as recovery took hold in 1983, the unemployment rate rose by less than one percentage point, but the average duration increased by about

<sup>2</sup> The 1981-82 recession began in July 1981 and ended in November 1982. The turning points for the recession of the 1990s have not been established, but it most probably began in April 1990 and ended in mid 1992.

five weeks. Between 1983 and 1985 expansion was well under way and the unemployment rate fell 1.5 percentage points, but there was virtually no change in the average incomplete duration of unemployment. Similarly between 1992 and 1993 the unemployment rate fell slightly, but the average duration of unemployment increased by almost 2.5 weeks.

In contrast the cyclical variation in the average complete duration of unemployment for those just becoming unemployed is stable throughout the period. There is a loop in the data, but it is a very muted clockwise movement. Furthermore, the turning points in the movement of the statistic correspond to peaks and troughs in labour market conditions. The average complete duration peaks at the same time as the unemployment rate, declines during recovery and expansion, and increases immediately with the onset of recession. The change in this statistic during the recession of the 1990s appears to follow roughly the same path as during the 1981-82 recession. The situation faced by those becoming unemployed during the early 1990s is not much different on average than that faced by those becoming unemployed a decade earlier. During 1983 the average complete duration of unemployment was 20.5 weeks, during 1992 it was 20.8 weeks.<sup>3</sup>

## **Methodology**

Our derivation of the average complete duration follows the work of Sider (1985), Baker (1992a), and Corak (1993) in using a synthetic cohort approach. This approach need not rely upon a steady state assumption, one that characterizes many earlier derivations.

Let  $S(x,t)$  represent the conditional probability that an individual will stay unemployed at least to the  $x$ th month given that he or she has been unemployed for  $x-1$  months.  $S(x,t)$  is one minus the hazard rate, and we refer to it as the continuation rate. It can be estimated from a sample of individuals as  $N(x,t)/N(x-1,t-1)$ , where  $N(x,t)$  represents the number of individuals unemployed at least  $x$  months, but less than  $x+1$  months in period  $t$ . That is, the probability of surviving to the  $x$ th month of unemployment given unemployment of  $x-1$  months is simply the ratio of the number of individuals reporting to be unemployed  $x$  months during period  $t$  to the number of individuals that reported being unemployed  $x-1$  months during the previous month.

We calculate six continuation rates from LFS data on the reported number of weeks of unemployment using progressively wider intervals: one month, two months, three months, four to six months, seven months to 12 months, and greater than 12 months. Wider bands are required at longer durations because of sample size limitations. Specifically the continuation rates are derived as the ratios of the number of individuals in each of the following categories:

---

<sup>3</sup> Corak (1993) points out that in spite of this, unemployment has been more "polarized" during the recent recession: the short term unemployed faring relatively better, but the long-term unemployed faring worse.

5 - 8	weeks in month t	to	< 5	weeks in month t-1
9-12	weeks in month t	to	5-8	weeks in month t-1
13-16	weeks in month t	to	9-12	weeks in month t-1
27-39	weeks in month t	to	13-26	weeks in month t-3
53-78	weeks in month t	to	27-52	weeks in month t-6
99 +	weeks in month t	to	53-98	weeks in month t-12.

The LFS data are top coded at 99 weeks. The fourth, fifth, and sixth continuation rates are converted to monthly equivalents by raising them respectively to the 1/3, 1/6, and 1/12 powers. This assumes that the monthly continuation rates are constant within each interval. These monthly rates are used in the derivation of the average complete duration of unemployment which, for a group of individuals who begin their unemployment spell at time  $t$ , is given as:

$$AvgDur(t) = \sum_{x=1}^n \prod_{i=1}^x S(i, t) \quad (1)$$

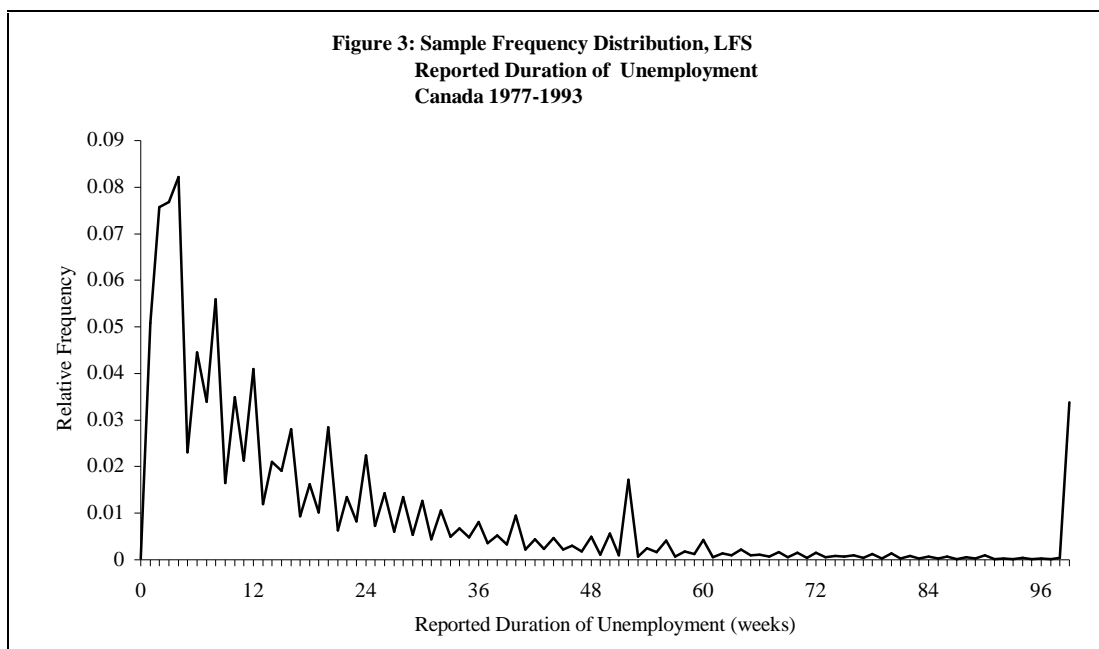
where  $n = 25$  in our data. The first element in this summation is one. Each element is an estimate of a point on the survivor function, and the summation is the discrete time version of the result that in continuous time the average duration is equal to the integral of the survivor function.<sup>4</sup>

This derivation does not rely on a steady state assumption, but if such an assumption were made the derivation of the statistic would, from a data manipulation perspective, be simplified. In a steady state both the rate at which individuals become unemployed and the continuation rates are constant so that inflows into unemployment equal outflows. In this case  $N(x-1, t)$  would equal  $N(x-1, t-1)$  and  $S(x, t)$  would simplify to  $N(x, t)/N(x-1, t)$ , the probability of surviving to the  $x$ th month of unemployment is the ratio of those reporting to be unemployed  $x$  months during period  $t$  to those reporting  $x-1$  months in the same period. It is in this way possible to derive the average complete duration of unemployment using only one month of data. Once again, it should be noted that the derivation of the continuation used in equation (1) assumes that conditions prevailing in the recent past will continue to hold into the future.

The LFS requires unemployed survey respondents to report the duration of their unemployment spells in weeks. We use the monthly survey results from 1977 through 1993. The frequency distribution for the entire sample reveals significant spikes in the data at two, and especially four week intervals. There are also notable spikes at 52 weeks and 99 weeks (see Figure 3). In reporting their unemployment spells survey respondents seem

---

<sup>4</sup> Let  $\tau$  index the complete duration of an unemployment spell, and let  $f(\tau)$  represent the associated density function. Then  $F(\tau) = \int_0^{\tau} f(u) du$  is the cumulative distribution function, and  $\delta(\tau) = 1 - F(\tau)$  is the survivor function. The average duration of unemployment is  $\int_0^{\infty} \tau f(\tau) d\tau$ . Integrating this expression by parts yields  $\int_0^{\infty} \delta(\tau) d\tau$ . See Baker and Trivedi (1985) for more details.



to prefer even numbers to odd, and months to part months. This digit preference has also been observed in CPS data. Sider (1985) suggests that the data be smoothed before the average duration is calculated and Baker (1992b), also using the CPS, explores the implications of various smoothing assumptions. Since broad intervals are being used in the derivation, smoothing need only occur for those weeks on the interval boundaries: some fraction of individuals reporting a spell length that coincides with these boundaries need to be reallocated to the next interval. Sider reallocates 50 per cent, and Baker (1992a) reallocates 30, 40, and 50 per cent at progressively longer intervals. Corak (1993) follows Baker's algorithm with LFS data. Baker and Trivedi (1985) note that while it is preferable to use the narrowest possible intervals in the derivation of the average duration, there may actually be a trade off: the narrower the intervals the more apparent the digit preference, and hence the more sensitive the results to the (arbitrary) smoothing assumptions adopted. In other words, wider intervals may reduce the efficiency of the statistic but they may also reduce the distortion caused by measurement error.

The spike at 99 weeks represents both the effect of digit preference and the truncation of the distribution due to top coding by survey administrators. An assumption must be made regarding how this spike is allocated among adjacent intervals. Sider (1985), Baker (1992a, 1992b) and Corak (1993) base their smoothing on the assumption that half of the respondents are at 99 weeks because of a response bias.

The choice of smoothing weights in the existing literature is made arbitrarily. However, the design of the LFS permits a closer analysis of the nature of this response bias. The LFS has a rotational design with respondents being surveyed for six consecutive months before being dropped. Paul (1986) uses linked records of individual responses between adjacent survey months to examine the response bias inherent in the reported weeks of unemployment. She finds that the unemployed who are classified as job seekers (about 89



per cent of unemployed respondents) show consistent month to month responses 67.8 per cent of the time (over the 1979-1982 period). Paul codes the duration responses into four week intervals, and defines a consistent response to be a linked record which shows an increase by one interval between adjacent survey months. The change in duration from one month to the next was on average 2.9 weeks for inconsistent reporters, less than the expected four weeks. Thus, there is on average (for about 30 per cent of the sample) a tendency to under report unemployment duration. Similar results were found for those unemployed because of temporary layoff. As a rough rule of thumb these findings suggest that the appropriate weight for smoothing LFS data might be in the range of 30 per cent and that the redistribution should be towards longer durations.

## Results

Table 1 offers an overview of our results. The average value of the duration statistic between 1977 and 1993 is presented according to whether or not a steady state assumption is used, and according to the degree to which the underlying data are smoothed. The smoothing weight is applied to all data on the boundaries of the duration intervals used, except at 99 weeks which is not smoothed at all. On the basis of Paul's results we consider a weight of 30 per cent to be preferred. Since the responses at 99 weeks represent both a response bias and top coding the smoothing of this datum is addressed separately. For the sample period used the steady state estimate is about 0.4 to 0.6 weeks less than the non steady state estimate, while the degree of smoothing can make a difference of more than three weeks to the estimate obtained. In what follows we examine these two dimensions in more detail.

**Table 1: Average Complete Duration of Unemployment: Steady State and Non Steady State (weeks)**

Weight Used in Correcting Response Bias	Non Steady State	Steady State
0.0	15.2	14.8
0.2	16.3	15.8
0.3	16.9	16.4
0.4	17.6	17.0
0.5	18.3	17.7

Data are averages over 1977 to 1993.

### A) *Steady State Versus Non Steady State Estimators*

The annual differences between the steady state and non steady state estimators (based upon a smoothing weight of 0.3) are depicted in Table 2. The average duration of unemployment over the entire period using the non steady state estimator is 16.9 weeks with a standard error of 2.28. The steady state estimator leads to an overall average of 16.4 weeks with a standard error of 1.97, the smaller standard error illustrating a dampened cyclical affect. Changes in the incidence of unemployment associated with recession and expansion cause the cross section of in progress spells used in the steady state calculation to become too heavily or too lightly weighted by shorter duration spells. As a result the steady state estimate is lower than the non steady state estimate at the onset of recessions (as incidence rises), and above it during recovery and expansion (as incidence falls). The largest difference between the two measures occurs in 1982, at the height of the 1981-82 recession, when the steady state estimate is 3.1 weeks below the non steady state. There are also large differences between the two measures during the recession of the 1990s, with the steady state estimate as many as 1.9 weeks below the non steady state estimate. During the recovery and expansion of the mid to late 1980s the steady state estimate is longer than the non-steady state estimate, but only at most by 0.4 weeks.

A comparison of the elasticity of each measure with respect to the unemployment rate offers a ready illustration of the differences in their cyclical variation. These elasticities are derived by a least squares regression of  $\ln(\text{AvgDur})$  against  $\ln(\text{Unemployment Rate})$  using 204 monthly observations from January 1977 to December 1993, where Unemployment Rate is the national seasonally unadjusted unemployment rate. (The regressions also include a constant, seasonal indicators, a time trend). The elasticity of the non steady state estimate with respect to the unemployment rate is 0.71, but only 0.57 for the steady state statistic.<sup>5</sup> These estimates clearly support the observation that the latter follows a dampened cyclical pattern

The steady state assumption has a bearing not only on the cyclical nature of the estimate but also on its seasonality. We explore this issue through a series of least squares regressions using monthly data on each of the non steady state estimator, the steady state estimator, and difference between the two estimates (non steady state less steady state) as the regressand. The regressors are the unemployment rate (seasonally unadjusted), a time trend, and twelve monthly indicator variables. The predicted values for each month of the year are presented in Table 3, holding the unemployment rate at its average of 9.3 per cent over the period and setting the time trend to zero. In many months the difference is quite substantial ranging from 6.1 weeks in January to -2.6 weeks in April. The differences are negligible only in the minority of months: for example, in February, June, and December.

---

<sup>5</sup> The standard errors associated with these estimates are respectively 0.033 and 0.026.

**Table 2: Average Complete Duration of Unemployment, Steady State and Non Steady State: 1977-93**

Year	Non Steady State (weeks)	Steady State (weeks)	Non Steady State less Steady State (weeks)
1977	15.7	14.8	0.9
1978	15.5	15.2	0.3
1979	13.8	14.2	-0.4
1980	14.3	14.0	0.3
1981	14.5	13.8	0.7
1982	19.7	16.6	3.1
1983	20.5	19.9	0.6
1984	17.8	18.0	-0.2
1985	17.4	17.8	-0.4
1986	16.2	16.5	-0.3
1987	16.1	16.4	-0.3
1988	14.8	15.2	-0.4
1989	15.3	15.0	0.3
1990	16.0	14.9	1.1
1991	19.5	17.6	1.9
1992	20.8	18.9	1.9
1993	19.9	20.2	-0.3
1977-93 mean	16.9 (2.28)	16.4 (1.97)	0.5

Table entries are annual averages of monthly data and are based upon a smoothing weight of 0.3.  
( ) indicates the sample standard error.

**Table 3: Predicted Seasonal Variations in Average Complete Duration of Unemployment, Non Steady State and Steady State**

Month	Non Steady State (weeks)	Steady State (weeks)	Non Steady State less Steady State (weeks)
January	20.6	14.5	6.1
February	18.4	18.7	-0.3
March	19.5	17.1	2.4
April	14.9	17.5	-2.6
May	15.2	13.7	1.5
June	15.0	14.4	0.6
July	15.6	14.2	1.4
August	15.5	17.5	-2.0
September	12.4	13.6	-1.2
October	15.6	14.8	0.8
November	16.3	13.7	2.6
December	17.0	17.5	-0.5

Table entries are derived from least squares estimates of each regressand against monthly indicators, the unemployment rate, and a time trend using monthly data from 1977 to 1993.

B) *Corrections for Digit Preference*

The information presented in Table 4 illustrates the effect of various smoothing assumptions on the non steady state estimates of average complete duration. In all cases we assume that no smoothing is required at 99 weeks. As described in Table 1 varying the weight used in the smoothing algorithm from 0 to 0.5 increases the average spell duration by about 3 weeks, from 15.2 to 18.3 weeks. This effect is pretty well constant through time. The elasticity of each measure with respect to the unemployment rate confirms that the cyclical variation of the statistic is not significantly affected. These elasticities range from 0.69 for a weight of 0.0 to 0.72 when the weight is 0.5. The effect of changing the smoothing assumption upon the steady state estimate is similar, increasing the overall average by about 3 weeks from 14.8 weeks with no smoothing to 17.7 weeks with a weight of 0.5. Likewise, the elasticity of the steady state estimate with respect to the unemployment rate is affected only slightly and ranges from 0.56 when the weight is 0.0 to 0.58 when it is 0.5. On the other hand the seasonal variation of the statistic is affected slightly by changes in the smoothing weight used. Changing the smoothing weight from 0.0 to 0.5 increases the estimated duration from 1.9 weeks in August to as much as 4.3 weeks in January.<sup>6</sup> The seasonal variation caused by increasing the smoothing weight from 0.0 to 0.5 is about  $\pm 1.2$  weeks around the annual average.

The choice of the smoothing weight is most critical for those weeks representing the densest part of the distribution. In particular, the magnitude of the estimate obtained is influenced in the first instance by the weight chosen to reallocate the number of respondents reporting four weeks of unemployment. This is illustrated in the second column of Table 5, which presents the results of allocating 30 per cent of those reporting four weeks to the next interval but making no other reallocations. These data should be contrasted with the results associated with a weight of 0.3 for all of the transition weeks from Table 4, and repeated as the first column of Table 5. The overall average duration obtained changes by only 0.3 weeks. The maximum difference between the two estimates is 0.5 weeks and occurs in 1982 and 1993. The smoothing of this one point changes the overall average duration by 1.4 weeks, from 15.2 to 16.6 weeks (see first column of Table 4). Thus, the duration estimate is highly dependent upon the weight chosen for the fourth week, but not very sensitive for other transition weeks associated with longer spell durations. The cyclical variation of the estimate, however, is not affected at all.

<sup>6</sup> The actual predicted seasonal variations for smoothing weights of 0.0 and 0.5 applied to the non steady state estimator are as follows.

	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
0.0	18.2	17.1	17.2	13.4	13.4	13.4	13.6	14.4	10.6	14.3	14.7	15.6
0.5	22.5	19.3	21.2	16.1	16.6	16.2	17.1	16.3	13.3	16.5	17.4	17.9

**Table 4: Average Complete Duration of Unemployment by Year and Alternative Smoothing Weights**

Year	Alternative Weights in Correcting Response Bias				
	0.0	0.2	0.3	0.4	0.5
1977	14.1	15.1	15.7	16.4	17.0
1978	13.9	14.9	15.5	16.1	16.7
1979	12.5	13.3	13.8	14.2	14.8
1980	12.9	13.9	14.3	14.9	15.4
1981	13.2	14.1	14.5	15.0	15.6
1982	17.7	19.0	19.7	20.5	21.3
1983	18.5	19.8	20.5	21.3	22.2
1984	15.9	17.1	17.8	18.5	19.2
1985	15.6	16.7	17.4	18.1	18.8
1986	14.6	15.6	16.2	16.8	17.5
1987	14.5	15.5	16.1	16.7	17.4
1988	13.3	14.3	14.8	15.3	15.9
1989	13.7	14.8	15.3	16.0	16.6
1990	14.4	15.4	16.0	16.6	17.2
1991	17.6	18.8	19.5	20.3	21.1
1992	18.5	20.0	20.8	21.7	22.6
1993	17.7	19.1	19.9	20.8	21.7
1977-93 mean	15.2	16.3	16.9	17.6	18.3

The effect of smoothing the data spike at the top code of 99+ weeks is also presented in Table 5. The data in column 3 show the effect of smoothing 50 per cent of respondents in this category into the preceding interval. This is the procedure followed in Corak (1993). When compared to the data in the first column the average duration is 0.7 weeks shorter on average, and can differ by as much as 1.4 weeks or as little as 0.2 weeks. The cyclicity of the estimate is mildly affected, reducing the elasticity with respect to the unemployment rate from 0.7 to 0.65. As illustrated in Figure 3 this data point represents fully 3 per cent of the sample, and thus any adjustments made to it have a significant influence on the overall estimate. The decision of how much to smooth at this interval centers upon how much of the response is due to the truncation of the distribution, and how much is due to response bias. If response bias causes an insignificant fraction then no smoothing is required. Since smoothing of responses beyond the fourth week has little effect upon the duration statistic we expect that smoothing for response bias at 99 weeks may also have little effect. Even if the response bias at 99+ weeks is as large as that at 52 weeks, the smoothing of it still would not significantly affect the final estimate.

**Table 5: Average Complete Duration of Unemployment by Year and Alternate Smoothing Methods**

Year	Alternative Weights in Correcting for Response Bias			Wide Interval
	0.3	0.3, 0.0*	0.3, 0.5+	
1977	15.7	15.4	15.4	17.7
1978	15.5	15.2	15.2	17.8
1979	13.8	13.6	13.6	15.8
1980	14.3	14.1	14.0	16.4
1981	14.5	14.3	14.2	17.0
1982	19.7	19.2	18.7	23.1
1983	20.5	20.2	19.4	24.1
1984	17.8	17.5	17.1	21.7
1985	17.4	17.1	16.6	20.9
1986	16.2	15.9	15.5	19.7
1987	16.1	15.8	15.3	19.8
1988	14.8	14.6	14.3	18.1
1989	15.3	15.0	14.7	18.5
1990	16.0	15.7	15.3	19.2
1991	19.5	19.3	18.3	23.4
1992	20.8	20.4	19.4	25.4
1993	19.9	19.4	18.7	24.1
1977-93 mean	16.9	16.6	16.2	20.2

\* indicates a weight of 0.3 for the first interval, and 0 for all others.

+ indicates a weight of 0.3 for all intervals and 0.5 for the 99+ category.

Wide interval refers to a grouping of the first three months into a single interval and using a weight of 0.3

Finally, Table 5 also illustrates the results of using wider intervals in the calculation of the continuation rates. The first three months are grouped into a single interval. In this way the sensitivity of the results to the choice of the smoothing weight will be attenuated. The standard smoothing assumptions are used for the remaining intervals (0.3 for all interval boundaries, and 0.0 for 99+ weeks). While the resulting estimate does not differ cyclically from the standard estimate, its level is significantly increased to 20.2 weeks (from 16.9 weeks). In fact, this approach increases the level to the extent that for many years it is greater than the average incomplete duration, and hence leads to a contradiction of the theorem put forward by Salant that was referred to earlier. The extent of this increase underscores the importance of using as narrow an interval specification as possible, at least for the shorter tail of the distribution. Since the continuation rates tend to rise with length of time spent unemployed grouping the first three months together leads to an overestimate of the rate for the first month and this causes a longer average duration estimate. In our data the loss of information in widening the intervals overrides any possible improvements through an attenuation of the response bias. It should be noted, however, that Baker and Trivedi (1985) reach the opposite conclusion with Australian data.

## A Caveat

The non steady state estimator is based upon the assumption that current economic conditions will continue into the future. In particular, it is assumed that the continuation rates, calculated on the basis of the labour market experience of the unemployed in the recent past (that is up to one year ago), will prevail for the duration of the unemployment spell. This is not as restrictive as a steady state assumption, but it may nonetheless imply that the statistic will not be perfectly accurate. If the labour market is deteriorating continuation rates should be expected to rise over time with the result that the average complete duration will understate the true average, while if they are improving continuation rates should be expected to decrease with the result that the estimated value will overstate the true figure.

To evaluate the extent of the error introduced by this assumption we require the average duration of unemployment based upon the actual experience of a cohort of unemployed. We calculate such a statistic by incrementing the reference continuation rates in the following manner:

$$TrueAvgDur(t) = \sum_{x=1}^n \prod_{i=1}^x S(i, t+i) \quad (2)$$

In contrast equation (1), which may be referred to as a "backward-looking" estimator, this is a "forward-looking" estimator. It can be computed from the same set of continuation rates used earlier, but by following the experience of a cohort forward through time. Since  $n-1$  future intervals are required it is impossible to carry the computation through for the most recent time periods. This makes the resulting statistic of historical interest only.

Annual averages are presented in Table 6 (based upon a smoothing assumption of 0.3 for all intervals and 0.0 for the last). The comparable non steady state (that is, backward looking) estimate from Table 2 is also provided for reference. The "true" average complete duration tracks the business cycle closely, rising sharply in 1981 and peaking at 19.1 weeks during 1982 as well as reaching a low of 14.8 weeks during 1987. In contrast the turning points of the non steady state estimator are a year later, the peak at 20.5 weeks in 1983, and the low of 14.8 weeks in 1988. As expected the backward looking estimator underestimates average duration during the recession and overestimates it during recovery. The differences between the two statistics can be substantial at the business cycle turning points. During 1981 the backward looking estimate is 3.0 weeks shorter than the forward, but 2.5 weeks longer during 1983. Similarly at the onset of the recent recession in 1990 the backward looking estimate was 3.4 weeks below the forward looking estimate.

Table 7 presents results similar to Table 3 on the seasonal variation of the statistics. At certain points during the year, particularly between January and April but also during the autumn, the difference between the two estimates can be quite significant. The backward looking estimate overstates the forward looking by 3 weeks in January, and by 5.5 weeks during March, while understating it by 3 weeks in September. These differences are similar

in nature to those associated with the business cycle, namely both the bleak winter conditions and robust summer conditions are assumed to last longer than they actually do.

**Table 6: Average Complete Duration of Unemployment: Forward and Backward Looking Estimates, 1977-91**

Year	Forward Looking (weeks)	Backward Looking (weeks)	Backward less Forward
1977	15.2	15.7	0.5
1978	14.8	15.5	0.7
1979	14.0	13.8	-0.2
1980	14.3	14.3	0.0
1981	17.5	14.5	-3.0
1982	19.1	19.7	0.6
1983	18.0	20.5	2.5
1984	16.8	17.8	1.0
1985	15.8	17.4	1.6
1986	15.2	16.2	1.0
1987	14.8	16.1	1.3
1988	14.9	14.8	-0.1
1989	16.6	15.3	-1.3
1990	19.4	16.0	-3.4
1991	21.1	19.5	-1.6
1977-91	16.5	16.9	0.4

Table entries are annual averages of monthly data and are based upon a smoothing weight of 0.3.

**Table 7: Predicted Seasonal Variations in Average Complete Duration of Unemployment, Forward and Backward Looking Estimates**

Month	Forward Looking (weeks)	Backward Looking (weeks)	Backward less Forward
January	17.6	20.6	3.0
February	17.3	18.4	1.1
March	14.0	19.5	5.5
April	13.7	14.9	1.2
May	15.1	15.2	0.1
June	15.4	15.0	-0.4
July	15.9	15.6	-0.3
August	14.8	15.5	0.7
September	15.4	12.4	-3.0
October	17.3	15.6	-1.7
November	16.9	16.3	-0.6
December	16.7	17.0	0.3

Table entries are derived from least squares estimates of each regressand against monthly indicators, the unemployment rate, and a time trend using monthly data from 1977 to 1993 for the backward looking estimate and from 1977 to 1991 for the forward looking estimate.



## Conclusions

This paper examines several issues associated with the derivation of the average complete duration of unemployment. There are three major results. First, the average complete duration of unemployment is a superior indicator of prevailing labour market conditions than the average incomplete duration. The former offers an indication of the duration of unemployment that those becoming unemployed can be expected to experience, and its fluctuations correspond with turning points in the business cycle. The average incomplete duration, on the other hand, is a lagging cyclical indicator. Second, the non steady state estimator of the average complete duration is preferred to the steady state estimator. The latter is significantly shorter than the former during the onset of recessions, and also displays a different seasonal pattern. While there are computational advantages associated with the use of a steady state assumption they are not that great. Third, and finally, the assumptions used in the smoothing of the underlying data in order to correct for response errors in reported duration have a significant influence on the level of the duration statistic and (to a lesser extent) its seasonal variability, but not on its cyclical variability. Smoothing is most important for the densest part of the distribution, generally the fourth week but also those points that correspond to top coding of the data by survey administrators.

On the basis of these results we conclude that the preferred statistic for official release, at least by Statistics Canada, is the non steady state estimate using a smoothing weight of 0.3. It is crucial that this weight be applied to the fourth week, but of little consequence at other points. As a result we also suggest that no smoothing for measurement error be done at the top code of 99+ weeks. Even though the non steady state estimator is superior to the steady state estimator its construction entails an assumption that current economic conditions (proxied by those in the recent past) will prevail through the duration of the unemployment spell. An examination of this assumption suggests that the non steady state estimator will overstate the duration of unemployment at the onset of a recovery, and understate it during the onset of a recession. There are similar seasonal bias as well. A certain caution is therefore required in interpreting the statistic.

Caution is also required in extending these recommendations to data from other countries. The choice between the steady state and non steady state is clearly in favour of the latter, but the choice of a smoothing weight is less clear cut. The existing literature appears to make an arbitrary decision. In our data a weight of 0.3 is defensible on the basis of an analysis of the reporting patterns of survey respondents, something that may vary from country to country. This issue, and the treatment of top coding, would require further exploration before any universal recommendations could be made.

## **Bibliography**

- BAKER, G.M. and P.K. Trivedi (1985). "Estimation of Unemployment Duration from Grouped Data: A Comparative Study." *Journal of Labor Economics*. Vol. 3, pp. 153-174.
- BAKER, Michael (1992a). "Unemployment Duration: Compositional Effects and Cyclical Variability." *American Economic Review*. Vol. 82, pp. 313-21.
- \_\_\_\_\_ (1992b). "Digit Preference in CPS unemployment Data." *Economics Letters*. Vol. 39, pp. 117-21.
- CORAK, Miles (1993). "The Duration of Unemployment during Boom and Bust." *Canadian Economic Observer*. Statistics Canada Catalogue No. 11-010. (September), pp. 4.1-4.20.
- KAITZ, H.B. (1970). "Analysing the Lengths of Spells of Unemployment." *Monthly Labor Review*. Vol. 93, pp. 11-20.
- KAPLAN, E.L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*. Vol. 53, pp. 457-81.
- PAUL, Betty Clayton (1986). "A Compendium of Studies on Data Quality of LFS Data on the Duration of Unemployment." Statistics Canada, Labour and Household Surveys Analysis Division Staff Reports, No. 86-2.
- SALANT, Stephen (1977). "Search Theory and Duration Data: A Theory of Sorts." *Quarterly Journal of Economics*. Vol. 91, pp. 39-57.
- SIDER, Hal (1985). "Unemployment Duration and Incidence: 1968-82." *American Economic Review*. Vol. 75, pp. 461-72.
- STATISTICS CANADA (1992). *Guide to Labour Force Survey Data*. Catalogue Number 71-528.