Unemployment is a complex form of human behavior, and there is probably no single analysis which attempts to cover all of its facets. One aspect of this behavior has had little understanding in the press and in public discussions, but has been exposed to an increasing amount of study among economists in connection particularly with job search theory, and that is the phenomenon of turnover among the unemployed and unemployment duration. At the most primitive levels of belief, the fact that unemployment has been at a level of 5 million people for well over a year is translated into a visual image of 5 million people each with over a year of unemployment. It is not known generally--and we share the blame for inadequate communication on this score--that a level of unemployment of 5 million for a year corresponds to at least one week of unemployment experience for perhaps 20 million people in 30 million or more spells, and that the vast majority of these spells--three-quarters or more--are less than five weeks in total duration. Only a small proportion of the people who experience unemployment have extended periods of joblessness. The public policy issues are obviously quite different with respect to the short-term unemployed as compared with the longterm unemployed. However, the nature of turnover among the unemployed is not explicitly given in the data customarily published. Published data present the cross section distribution of the unemployed, i.e., the distribution of the unemployed by the duration of their unemployment through the end of the reference week each month (the week including the 12th). Only a small fraction of these people are actually completing spells of unemployment in the reference week; the remainder will go on to experience additional week(s) of unemployment. Because reference weeks are four or five weeks apart, the intervening movements into and out of unemployment are not evident, and the full nature of unemployment turnover has remained somewhat obscure. A first attempt to deduce the patterns of completed spells of unemployment and of turnover is presented in an article in the Monthly Labor Review for November 1970.

The present paper continues this work by developing a time series of completed spells of unemployment. The details of the estimating procedure are given in the first part below. In the second part certain shortcomings of the under lying data are examined, and a method for eliminating them to a substantial extent is developed. Since the techniques used here are somewhat different from those presented in the earlier paper, some comparisons between the results in the two papers are made, and the validity of certain assumptions in the earlier article is examined. Finally, a time series of completed spells and their average duration from mid-1967 to mid-1972 is presented.

## Derivation of Monthly Series on Completed Spells

Let  $g_i =$  number of people who report they have been unemployed for 1 week through the end of the reference week.

Let N = the number of people completing spells of unemployment in the week prior to the reference week.

Let F = the total number of people unemployed in the week prior to the reference week.

Finally, let G = the total number of people unemployed in the reference week.

It may be readily shown that N = F - G +  $g_1$ .  $\frac{1}{2}$ 

In words rather than symbols, the number of completed spells in the week prior to the reference week is equal to the number of unemployed in that week minus the number of unemployed in the following week (the reference week) plus the number of unemployed who had completed only one week of unemployment in the reference week).

G and g are available from the regular current 1

population survey data. (There is a basic problem in the measurement of g which must be faced;

this is the burden of the next section of this paper.)

F is not available from the current population survey since it refers to a nonsurvey week. However, it can be estimated by interpolation as described a little later. It must be emphasized that the calculation of the number and duration of completed spells requires knowledge of the indicated statistics in adjacent weeks.

In order to estimate the average duration of completed spells in the week prior to the reference week, we need several additional symbols.

Let A = the cumulated number of weeks of unf

employment of those unemployed in the week prior to the reference week.

Let A = the corresponding quantity for those g

unemployed in the reference week.

Let A = the cumulative number of weeks of unemployment of those completing their spells of unemployment in the week prior to the reference

week. If we let  $\bar{x}_f$  and  $\bar{x}_q$  represent the cross section

average durations of unemployment in the week prior to the reference week and in the reference week respectively, and  $\bar{x}_{g}$  represent the average

duration of unemployment of those completing spells in the week prior to the reference week, we have the following relationships by definition:

$$\begin{array}{cccc} A = \bar{x} F; & A = \bar{x} G, & \text{and} & A = \bar{x} N \\ f & f & g & g & s & s \end{array}$$

It can be readily shown that

$$A_{s} = A - A + G. \frac{2}{2}$$

A is obtained directly from current population g

survey data for the reference week. As will be indicated below A is estimated by interpolation f as is F. The average duration of completed spells in the week prior to the reference week can therefore be written as:

 $\bar{x}_{s} = (A_{f} - A_{q} + G) / (F - G + g_{1}).$ 

Under conditions of stability or equilibrium, we will have A = A, and F = G, so that the f g expression for average duration simplifies to

 $x_s = G / g_1$ , a result previously given in the cited article in the Monthly Labor Review.

Even under conditions of instability, if we let G be the average annual total unemployment,  $g_1$  be the average annual number of unemployed

with one week of unemployment (entrants), the average duration of all spells completed in the year is given by

$$\bar{\mathbf{x}}_{s} = \frac{(A_{1} - A_{3}) / 52 + G}{(G_{1} - G_{53}) / 52 + g_{1}}$$

with the subscripts in the parentheses referring to the first and fifty-third weeks. Since A is about 10 times the size of G, and G is about 6 times the size of  $g_1$ , and the differences,  $A_1 - A_{53}$ , and  $G_1 - G_{53}$  are, in most cases, fractions of the sizes of A and G respectively, the nature of the approximation in the formula  $\bar{x}_s = G/g_1$  is pretty good on an average

annual basis, although it is by no means perfect.

The exposition in the preceding paragraph is intended to serve as a rationale for the derivation of completed spell distributions from average annual cross section data in the Monthly Labor Review article. It may also be shown with the same reasoning that the derivation of the distribution of completed spells by intervals of weeks is also generally a reasonable first approximation.

The present paper does not go on however to estimate a time series for the distribution of completed spells by duration intervals for reasons which are briefly discussed in the Monthly Labor Review article, and may profitably be repeated here. The recorded cross section distributions of the unemployed by single weeks of duration are subject to a number of irregularities, most prominently the rounding effect in the process of recall on the respondent's part, which introduces local modes into the data at 4, 8, 13, 26, etc. weeks of duration, because of the tendency to report unemployment in terms of months, quarters, half years and so on. Another irregularity, which may not superficially appear to be one is the local mode at a duration of two weeks. This irregularity is of fundamental importance to this paper and is discussed in more detail in the next section. Until these irregularities can be smoothed in an appropriate way, the derivation of the time series of the distribution of completed spells by duration intervals must be held in abevance.

One other point will conclude this section. As mentioned earlier, it is necessary to estimate  $A_r$  and F, since they do not apply to the

reference week. The simplest thing to do is to use linear interpolation for F between the two G values in the adjacent month reference weeks. In the same way linear interpolation between adjacent month reference week values of A will  $_{\rm G}$ 

yield A approximations.

## A Bias in the Duration Distribution of Unemployment and its Correction

As noted earlier, a major problem in the derivation of the duration distribution of completed spells is in the biased reporting of those unemployed in the reference week who indicate that they have had one, two or three elapsed weeks of unemployment. The bias is evident in the data below for the average annual cross section duration distribution for 1971, but it appears in the data for almost all individual reference weeks, and in all years.

Duration	Number of		
in weeks	Unemployed		
0	10		
1	445		
2	656		
3	532		
4	583		
5	177		

The 12 (000) people who have zero weeks of unemployment need not concern us particularly. Conceptually they can only be people who were out of the labor force for more than half of the reference week before they began looking for a job. For convenience they are included with the number identified with one week of unemployment.

On the other hand, it is conceptually and empirically impossible for those unemployed with two weeks of unemployment to regularly exceed, and usually by a substantial amount, those with one week. This is the phenomenon studied in this section of the paper.

Finally, the local mode at 4 weeks is obviously due to rounding by the respondent or the interviewer (4 weeks equals one month) so that suitable smoothing techniques must be developed to remove this mode. Fortunately this effort is not required for this paper, so it is not considered further here.

A hypothesis for the aberrant pattern among duration groups of 1, 2 and 3 weeks is explored below. In order to develop a reasonable basis for this hypothesis, we must first look at the actual questionnaire used in the Current Population Survey.

The week including the 12th of the month has been called the reference week in the discussion thus far. The following week (which includes the 19th of the month) is the survey week. On each working day of that week the interviewer visits households in his or her part of the sample. On the CPS questionnaire, questions 19, 20 and 21 ask about activities in the preceding week with emphasis on the phrase "LAST WEEK" (the capitals are on the questionnaire form itself). The appropriate responses to questions 20, 21, and 22 classify a person as unemployed during the reference week.

However, the phrasing in question 22C may be subject to some ambiguity in response. It is as follows:

- 22C. 1) How many weeks has . . . been looking for work?
  - 2) How many weeks ago did . . . start looking for work?
  - 3) How many weeks ago was . . . laid off?

These three questions are alternate versions to be used as appropriate, depending on replies to earlier questions.

The hypothesis advanced here is as follows: In answering question 22C, the respondent calculates the length of time he (or she) has been looking for work up to the time of the interview, and not up to the end of the preceding week (the reference week). Since answers are recorded in whole weeks (rounding is called for), it is likely that persons in households which are interviewed in the second half of the survey week tend to round their answers up one week. For example, suppose the respondent has only been unemployed in the reference week (one week duration), but is still unemployed in the survey week. The interviewer visits his household on Thursday of that week and asks: "How many weeks have you been looking for work?", and he replies: "About two", which is an appropriate reply from his perspective. Now the interviewer's instruction manual says: "In computing the weeks a person has been looking for work, count the number of weeks from the time he started looking for work through the end of the reference\* week for the current month." (Emphasis supplied in original) However, this is the instruction to the interviewer alone. Question 22C does not in its own phrasing pin down the actual duration through the end of the reference week. The hypothesis offered here is that respondents in fact are likely to use the date of the interview as the end of the elapsed duration period.

The balance of this section examines the effect on the data if this hypothesis is true and estimates how this bias may be removed.

Let the "true" number of those unemployed people with an elapsed duration of 1 week through the end of the reference period be denoted by  $g_1$ , those with 2 weeks by  $g_2$  and

those with 3 weeks by  $g_3$  . Let the correspond-

ing measured quantities be denoted by the same symbols with primes.

Let the symbol f with appropriate subscripts refer to the corresponding values for the week prior to the reference week and the symbol h apply to the week after the reference week.

Now let r be the fraction of those with one elapsed week of unemployment to the end of the reference period, who are still unemployed

\* The manual uses the word "survey" here to designate what I have called the "reference" week. in the survey week and who are interviewed in the latter part of that week and consequently report themselves with two elapsed weeks of unemployment. Let the same fraction hold for those with two elapsed weeks who report themselves with three weeks (and for those with three elapsed weeks who report four). We will therefore find the following relationships between the true and measured quantities:

$$f_{1} = f_{1} - rg_{2}$$

$$f_{2}' = f_{2} + rg_{2} - rg_{3}$$

$$f_{3}' = f_{3} + rg_{3} - rg_{4}$$

1

Of the f, people who experience one week of unemployment by the end of the reference week,  $g_2$  go on to experience a second week of unemployment. When these  $g_2$  people are visited in their households,  $r g_2$  of them report themselves with two weeks of unemployment up to the end of the

reference week (the preceding week). The reasoning is the same for the other groups to whom the second and third equations above refer.

In the same way relationships for the following week may be specified:

$$g'_1 = g_1 - rh_2$$
  
 $g'_2 = g_2 + rh_2 - rh_3$   
 $g'_3 = g_3 + rh_3 - rh_4$ 

We have assumed that the fraction of people (r) who round their length of unemployment up to the next week is the same in the two adjacent weeks as it is for the first few duration categories.

The relationships which follow from this assumption are still indeterminate, because there are more unknowns than relationships. We need therefore to introduce an additional assumption in order to permit a solution. This assumption is that the continuation (survival) rate is also constant over the first several weeks of duration, and in the two adjacent weeks.

The continuation rate is defines as  $p_{i+1} =$ 

g / f, i.e., the proportion of people unemployed i weeks who go on to have at least one additional week of unemployment. p has a subscript to identify its position in the duration scale, and is not, as a rule, constant over the entire duration (see MLR article, <u>op. cit.</u>) but it does not usually change very much within a duration range of two or three weeks. The assumption used here is that, within our area of consideration, the continuation rate is constant.

In other words,

$$p = g_2 f_1 = g_3 f_2 = g_4 f_3 = h_2 g_1 = h_3 g_2 = h_4 g_3$$

An earlier equation may be rewritten as

 $g'_{3} = pf_{2} + rp g_{2} - rp g_{3}$ We also have the earlier relationship

 $f_2 = f_2 + rg_2 - rg_3$ .

If we divide the first of these by the second we get  $g'_3 / f_2 = p$ , a direct estimate of the continuation rate.

Another of the earlier relationships may be rewritten as

 $g'_{2} = p (f_{1} - rp f_{1} + r g_{1})$ 

Two other equations may be expressed as follows:

$$f_{1} = f_{1} (1 - rp) ; g_{1} = g_{1} (1 - rp) or$$
  
 $f_{1} / f_{1} = 1 - rp = g_{1} / g_{1}, whence f_{1} / g_{1} =$   
 $f_{1} / g_{1}$ 

In other words, the ratio of the observed number of people with one week of unemployment in two adjacent weeks is equal to the ratio of the "true" values of these numbers. Using this

result in the equation for  $g'_2$ , we arrive at a solution for r:

$$r = \frac{(g'_{2}/f'_{1})/p - 1}{g'_{2}/f'_{1} + g'_{1}/f'_{1} - p}$$

The observed number of people entering the unemployed in the reference week can be corrected to the "true" number by use of

$$g_1 = g'_1 / (1 - rp)$$

For 1969 annual average data, letting

$$f'_1 = g'_1$$
, and estimating p by use of  $g'_3 / g'_2$ ,

we have p = 0.733, and r = 0.459. The estimated continuation rate of 0.73 for low duration values in 1969 compares with estimates in the .70 - .80 range for 1969 estimated by a different approach (MLR article, p. 13, Table 4). The r value of 0.46 is consistent with the rather simple and plausible notion that about half the  $g_2$  people round their responses on

duration to one week, and the other half round to two weeks.

As indicated in the last formula, we now have a way of estimating  $g_1$  values month by month,

which are essential to the calculation of completed spell statistics.

Because the monthly series on completed spells constitutes only between 15 and 20 percent of the total level of unemployment, it is subject to much greater irregularity. Consequently the derived time series shown in the table below is given in terms of quarterly averages only.

Time	Series	of	Completed	Spel1	ls c	Σf	Unemployment
	(Wee	ekly	Averages	Per Q	)uar	:te	er)

Year-Quarter	Spells Completed (000)	Spells Begun (000)	Average Duration (in weeks) of Completed Spells
1		<i>coc</i>	
1967 - 111	746	686	4.4
IV	511	492	5.3
1968 - I	480	488	6.0
II	559	616	6.0
III	703	627	4.6
IV	504	487	4.8
1969 - I	493	513	5.2
II	665	724	4.5
III	709	678	4.2
IV	515	489	5.3
1970 - I	500	580	5.3
II	685	766	5.8
TTT	871	843	5.0
IV	705	734	5.0
1971 - T	711	747	6.0
TT	761	769	7.3
	796	745	6.6
 TV	733	713	6.2
1072 - T	, 55 660	694	6.2
	725	725	Q 1
11	/35	100	0.1

At this point it may be useful to compare the average annual data on completed spells estimated directly from the average annual cross section duration distributions of the unemployed (A) with the average annual data derived from the twelve monthly observations in each year (M). The table below presents these comparisons.

		Average Weekly	
		No. of Com-	Average Dura-
	Type of	pleted Spells	tion of Com-
Year	Estimate	(000)	pleted Spells
1968	М	562	5.3
	A	554	5.1
1969	м	596	4.7
	A	601	4.7
1970	м	690	5.3
	А	734	5.6
1971	м	750	6.6
	А	742	6.7

The difference between the results obtained from time series versus those obtained from cross section data are the greatest for the year 1970, when the seasonally adjusted rate of unemployment rose from 3.9 percent in January to 6.1 percent in December. Even here, the estimates based on cross section data may be considered to be reasonable approximations of the estimates based on time series. In the other three years the two estimates for both number of completed spells and average duration are very close to each other and the approximation is excellent.

Two additional comments about these estimates may be made in closing:

1. The average duration of completed spell is substantially and uniformily lower than the average duration of cross section data regularly published. The table below compares these two averages in annual average form.

Average Duration of Unemployment

Year	Cross Section	Completed Spell
1968	8.5	5.3
1969	8.0	4.7
1970	8.8	5.3
1971	11.4	6.6

A discussion of the reasons for these differences is found in the MLR article, <u>op</u>. <u>cit</u>., and will not be repeated here.

2. Average weekly additions to unemployment for new entrants as a percent of the average level of unemployment for these four years is as follows:

Year	New Entrants/Total	Unemployment
1968	.20	
1969	.21	
1970	.18	
1971	.15	

No extensive analysis of this phenomenon will be undertaken here, but it is interesting to note that the rise in unemployment in a recession partakes of two elements:

- a rise in the number of entrants to unemployment and
- a decline in the rate of mobility into and out of the ranks of the unemployed.

Again, there is some discussion of this phenomenon in the reference cited, but a full study of it remains to be made.

## Footnotes

 Let C = number of unemployed with unemployment in both the reference week and the prior week.

Then F = C + NG = C + g

By subtraction we get  $N = F - G + g_1$ 

Then  $A_f = A_C + A_S$ ;  $A_g = A_C + C + g_1$ , but  $G = C + g_1$  so  $A_g = A_C + G$ . By subtracting this last equation from the first, we get the desired result:

$$A_{s} = A_{f} - A_{q} + G.$$

## Reference