Techniques for Estimating the Number of Workers With Wage Credits Toward Unemployment Compensation

CHARLES M. LARSON *

ALL State unemployment compensation agencies except those permitting employers to report on a wage and separation basis,[1] annually submit to the Bureau of Employment Security reports indicating the number of workers who, during the preceding calendar year, earned any wages in employment covered by that State's unemployment compensation law. The number of workers with wage credits is generally regarded as an administrative figure, valuable when used in connection with other data for estimating the work load in the State agencies.

The concept of workers with wage credits is also useful in economic analyses. For example, chart 1 shows the ratio of the number of covered workers starting a benefit series during 1940 to the number of covered workers earning wage credits during 1940. The relative sizes of these groups are of interest to economists and students in the field of unemployment insurance. In making comparisons between the number of workers starting a benefit series and the number of workers with wage credits, allowance should be made for the fact that certain State unemployment compensation laws have more severe benefit qualifying provisions than others, so that the percentage of workers with wage credits who are eligible for benefits varies from State to State. In addition, in using the data shown in chart 1 it should be understood that all workers who initiated a benefit series in 1940 did not necessarily earn wage credits in 1940. The concepts of lag quarter, partial quarter, base period, benefit year, and others introduced by the provisions of the various State unemployment compensation laws make it impractical to attempt to determine a relationship between wages earned

in one period and benefits paid at a later period on the basis of those same wages. As an alternative, it is customary as in chart 1 to compare wage factors and benefit factors that have been measured during the same time period.

Workers with wage credits may also be compared with the number of workers filing initial claims or with the amount of wages earned in covered employment or with the number of workers

Chart 1.—*Ratio of workers receiving a first benefit payment in 1940 to workers with wage credits in 1940, 49 States*[1]
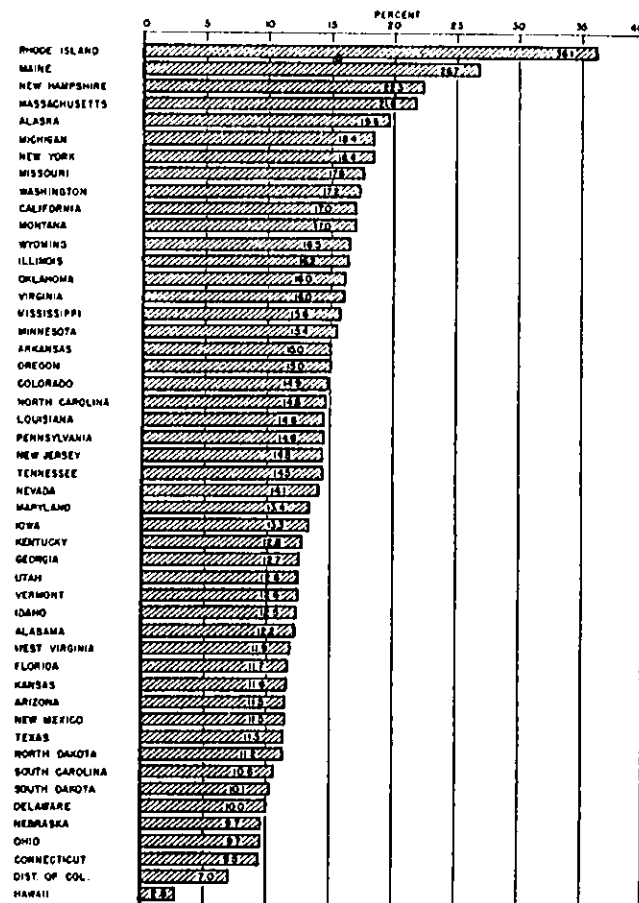
[1] Data for Indiana and Wisconsin not comparable.

Table 1.—*Workers with wage credits under State unemployment compensation systems, and type of 1940 sample, by State, 1939 and 1940*

State	Size-of-firm Inclusion (number of workers) [1]	Workers with wage credits		Type of 1940 sample
		1939	1940	
Total		[2]30,086,000	[2]31,947,000	
Alabama	8 or more	377,300	448,600	End-digit.
Alaska	do	23,700	23,700	Ledger card.
Arizona	3 or more	109,700	112,700	Random wage-record.
Arkansas	1 or more	237,600	271,800	Do.
California	4 or more	2,080,000	2,191,000	End-digit.
Colorado	8 or more	210,800	[3]224,000	Regression.
Connecticut	5 or more	[3]667,100	[3]755,400	Do.
Delaware	1 or more	[3]114,800	[3]127,100	Do.
District of Columbia	do	[3][4]255,800	[3]292,100	Do.
Florida	8 or more	412,000	503,300	First wage-record.
Georgia	do	523,500	570,500	End-digit.
Hawaii	1 or more	[3]145,800	[3]135,000	Regression.
Idaho	1 or more [4]	[3]119,200	[3]121,800	Do.
Illinois	6 or more [7]	2,398,000	2,470,500	End-digit.
Indiana	8 or more	787,000	901,100	Block.
Iowa	do	356,400	394,800	Random wage-record.
Kansas	do	238,000	254,000	Block.
Kentucky	4 or more [6]	399,300	427,700	End-digit.
Louisiana	4 or more [6]	434,600	547,200	Do.
Maine	8 or more	214,200	221,200	Actual count.
Maryland	4 or more	533,900	570,000	Random wage-record.
Massachusetts	4 or more [10]	1,410,600	1,446,600	Block.
Michigan	8 or more	1,472,000	1,550,000	End-digit.
Minnesota	1 or more [11]	541,000	528,100	Do.
Mississippi	8 or more	225,700	245,300	Do.
Missouri	do	745,000	804,100	Block.
Montana	1 or more [12]	116,000	131,000	End-digit.
Nebraska	8 or more	172,000	177,000	Random wage-record.
Nevada	1 or more [11]	46,700	48,400	End-digit.
New Hampshire	4 or more	150,000	159,000	Actual count.
New Jersey	8 or more	[3]1,273,000	[3]1,402,100	Regression.
New Mexico	2 or more [11]	85,900	95,800	Random wage-record.
New York [11]	4 or more	4,450,000	4,700,000	Actual count.
North Carolina	8 or more	644,700	714,100	End-digit.
North Dakota	do	50,500	50,700	Do.
Ohio	3 or more	[3]1,964,100	[3]2,118,100	Regression.
Oklahoma	8 or more	281,000	308,000	End-digit.
Oregon	4 or more [10]	278,500	313,200	Do.
Pennsylvania	1 or more	[3]3,053,000	[3]3,185,600	Regression.
Rhode Island	4 or more	204,000	288,000	Block.
South Carolina	8 or more	322,000	346,500	End-digit.
South Dakota	do	56,900	58,300	Actual count.
Tennessee	do	452,100	489,400	End-digit.
Texas	do	1,000,600	1,185,300	Worker.
Utah	1 or more [17]	124,800	130,300	Actual count.
Vermont	8 or more	78,100	79,500	Do.
Virginia	do	540,200	507,000	Ledger card.
Washington	do	407,900	452,500	Actual count.
West Virginia	do	411,500	418,800	End-digit.
Wisconsin [14]	6 or more [16]	[3]644,100	[3]682,800	Regression.
Wyoming	1 or more [16]	58,000	61,300	End-digit.

[1] Represents number of workers an employer must have for a specified period to be subject to State law.

[2] Adjusted by a reduction of 6 percent in 1939 and 7 percent in 1940 to allow for duplication caused by employment of individual workers in more than 1 State during the same year. Without these reductions, the total is 32,006,200 for 1939 and 34,351,800 for 1940. Basis for estimates of duplication are contained in tables 1 and 2 of the September 1941 Bulletin, pp. 4–5.

[3] Estimate based on State's high month of employment during year.

[4] Estimate reduced 5 percent to adjust for reporting procedures which showed number of workers employed during month instead of during last pay-roll periods in month.

[5] Effective Jan. 1, 1940, many food-processing workers removed from coverage by revision of definition of agricultural employment.

[6] And total wages of $78 or more in a quarter, effective Jan. 1, 1939.

[7] Coverage changed from 8 or more to 6 or more, effective Jan. 1, 1940.

[8] Wages of at least $50 to each of at least 4 workers, during each of 3 quarters; or 8 or more workers within 20 weeks of calendar year.

[9] January–September 1938, 8 or more in 20 weeks; effective Oct. 1, 1938, 4 or more in 20 weeks or 12 or more in 10 weeks.

[10] Coverage changed from 8 or more to 4 or more, effective Jan. 1, 1939.

[11] Effective Jan. 1, 1940, employers of less than 8 located outside the corporate limits of a city, village, or borough of less than 10,000 population are excluded from coverage.

[12] And total annual wages of $500 or more, effective Jan. 1, 1939.

[13] And total wages of $225 or more in a quarter, effective Jan. 1, 1939.

(See next column for rest of footnotes.)

declared eligible for benefits, or with other similar data for a given year. Employment figures showing the number of workers in covered employment during one pay period in each month cannot take the place of data on workers with wage credits for the purpose of making aggregate comparisons, since they are especially designed to measure accurately only the level of employment at various periods. They represent the number of workers engaged during such a short period of time that they may not be compared directly with aggregate economic data, such as wages, benefits, and man-hours of employment except when they can be validly treated as an average value for the period in which they are centered. In such a case the comparison has a slightly different meaning from a comparable one which uses workers with wage credits. For example, the annual total of covered wages divided by the number of workers having wage credits in the same year gives the annual average covered earnings per covered worker; on the other hand, the annual total of covered wages divided by the average monthly employment for the same year will approximate the annual average covered wage per covered worker that would have existed if all covered workers had worked full time throughout the year.

Since State figures on the number of workers with wage credits are measured over a 1-year period they cannot, because of duplication, be totaled to give the number of workers with wage credits in the United States. The sum of the 51 State figures on workers with wage credits as shown in table 1 must be reduced by an amount estimated [2] at 6 percent for 1939 and 7 percent for 1940 in order to approximate the actual number

[1] Estimates based on old-age and survivors insurance data in tables 1 and 2, Merriam, Ida C., and Bliss, Elizabeth T., "Effects of Migration on Unemployment Benefit Rights," *Social Security Bulletin*, September 1941, pp. 3–11. These data which apply to 1938 were assumed to increase slightly in 1939 and 1940 because of increased worker migration in connection with the war effort.

[14] January 1938–June 1939, 4 or more in each of 20 different weeks; effective July 1, 1939, 2 or more in 13 weeks of a calendar year, or wages of $450 or more in a quarter.

[15] Includes domestic service in private homes, which service is excluded under laws of other States.

[16] And total wages of $500 or more in a quarter.

[17] January 1938–June 1939, 4 or more; effective July 1, 1939, total wages of $140 or more in a quarter.

[18] Includes non-Federal governmental service covered by State law, which service is excluded from coverage in other States. State amendment effective July 1, 1939, made compliance optional for certain government units, excepting only State departments and cities of the first class.

[19] 1938: 8 or more in 1938 or 7 or more in 1937, or, where employer's records do not permit accurate count of workers, total wages of $7,000 or more in 1937. 1939 and thereafter: 8 or more in current year or 6 or more in preceding year, or, where employer's records do not permit accurate count of workers, total wages of $6,000 or more in preceding calendar year.

[20] And total wages of $150 or more in a quarter, effective Jan. 1, 1939.

of different workers earning wages in covered employment in those years.

Most States use a sampling procedure to determine the number of workers with wage credits in order to avoid the clerical and tabulating problems involved in an actual count of millions of wage items. In 1940 the average State had 1,250,000 separate wage items for 400,000 workers, and all but 7 of the State agencies had wage items for at least 100,000 workers.

Under the old-age and survivors insurance program, earnings of individual workers are summarized periodically from the permanent wage records; hence the total figure for workers with wage credits during a year is known exactly except for delinquent items. On the other hand, under a current benefit program like unemployment compensation, with no need for such permanent records, similar data on workers and annual earnings under State unemployment compensation laws are not automatically available.

Sampling Techniques

The normal wage-record file.—Most of the States use approximately the same filing system for individual wage items. Except for the States mentioned above as using reports on separations, the State agency receives from each reporting unit a quarterly report showing employees who earned some wages in covered employment during the preceding quarter. A complete report shows the workers' names, social security account numbers, and the amounts of wages earned during the quarter; some State agencies receive additional data, such as dates of hiring and separation. If State procedures permit the submission of data for several different workers on a single form, these data will be transferred to records of individual wage items for filing except that in some States the data are transferred to ledger cards so that all of a worker's wages are available on a single record. If a worker has been employed by more than one covered employer during the quarter, the file will contain more than one wage item for the worker for that quarter. If the employer has failed to submit the worker's social security account number with his wage and name data, the worker's wage item will be filed in the alphabetic section of the file. If the data are complete, the wage items will be filed in order of their social security account numbers in the numerical section of the

file except that, in some State agencies, wage items for workers with social security account numbers obtained outside the State are kept in the alphabetic file while in others the wage items in the numerical file are grouped by calendar quarters. In most States, the old wage items are removed at regular intervals, so that the file always contains wage items for only 4 consecutive calendar quarters. In a few States having 2-year base periods, the file contains wage items for 8 consecutive calendar quarters.

End-digit sampling.—Because of its simplicity and because the records selected may be used in analyses of various characteristics of workers, sampling of wage items is usually based on the end digits of workers' social security account numbers. This was the technique used by 19 State agencies in estimating the number of workers with wage credits for 1940. Social security account numbers have been allotted to workers in such a manner that for all practical purposes the final 3 and possibly the final 4 digits are distributed at random with respect to such worker characteristics as wages, recency of entry into the labor market, or State of employment.[1] In choosing an end-digit sample, a particular set of digits is selected arbitrarily, and the wage items for all workers with social security account numbers ending in those digits are included. For example, the sample may include the wage items for all workers with social security account numbers ending in 01. Since 01 is one of 100 possible 2-digit combinations in which a social security account number could end, the sample may be considered a 1-percent sample of the file, and any characteristics of the sample may be multiplied by 100 to obtain an estimate of the characteristics of the numerical file. Since the alphabetic file generally contains less than 2 percent of all the wage items, even a rough estimate of its contents, when added to the estimate of the contents of the numerical file, should provide an adequate final estimate.

In most State agencies using an end-digit sample, the simple expansion described above (multiplication of the number of workers with wage items in the sample by 100) was used to

[1] Social security numbers are currently being allotted in such a manner that the last 4 digits will be random with respect to all worker characteristics. It is expected that the diffusion of these randomly distributed numbers into the group of numbers which were distributed in numerical order to applicants and in some cases in blocks to workers for a particular firm will soon remove any lack of randomness in the last 4 digits which was inherent in the original method of assigning account numbers.

determine an estimate for checking purposes only. The final estimate was made by using one of the following expansion factors:

(1) The ratio of all wages in the combined alphabetic and numerical file to all wages in the sample; or

(2) the ratio of the number of wage items in the combined alphabetic and numerical file to the number of wage items in the sample. Either of these two expansion factors when multiplied by the number of workers with wage credits in the sample gives an adequate estimate of all workers with wage credits, and automatically allows for workers with wage items in the alphabetic file.

Random wage item sampling technique.—In 1940, six State agencies used a sample consisting of single wage items selected at approximately equally spaced intervals throughout numerical files. A wage item received a weight of one if it was the only wage item for a worker in the sample; a weight of one-half if it was one of two wage items for that worker; a weight of one-third if it was one of three wage items for that worker, and so on. These weights were assigned so that the sum of the weights of all wage items in the numerical file would be equal to the number of workers with wage credits in that file, because the weights of each worker's wage items when summed equaled one. After the sample was chosen, an equation was set up from the assumption that the sum of the weights of the wage items in the file (number of workers with wage credits) would be in the same proportion to the total number of wage items in the file as the sum of the weights of the wage items in the sample was to the number of wage items in the sample. Since all these quantities except the number of workers with wage credits were known, that number could be readily derived. This estimating technique makes adequate allowance for workers in the alphabetic file.

Block sampling.—Five State agencies used a sample consisting of a number of blocks of wage items selected at regular intervals throughout the files. The blocks were combined to form the total sample, and the final estimate was made by assuming that the number of workers with wage credits in the complete file bore the same relationship to the number of wage items—or, in some States, total wages—in the file as the number of workers in the sample bore to the number of wage items—or wages—in the sample. Since all these

quantities except the number of workers with wage credits were known, that number could readily be obtained. This estimating technique also makes adequate allowance for workers in the alphabetic file.

A block sample does not give an estimate with as small a standard error as equal-sized samples based on end-digits or random wage items; this error also increases with the size of the blocks. However, a sample with blocks so small that at least one is taken from every tray of wage items will approximate, in estimating efficiency, an equal-sized sample of the other types. Care must be taken that the blocks contain approximately an equal number of wage items rather than the wage items for an equal number of workers, because the latter selection oversamples those sections of the file which contain an above-average number of wage items per worker.

Ledger-card sampling.—Two State agencies used a ledger-card sample. In these agencies, all wage items were posted to individual ledger cards so that each worker had in effect a single wage record. The samples were selected by choosing single ledger cards at equally spaced intervals throughout the files. In one agency, the estimate was made by assuming that the percent of ledger cards with 1940 postings in the file would equal the percent of ledger cards with 1940 postings in the sample. This percentage as determined from the sample was multiplied by the known number of ledger cards in the file to estimate the number of workers with 1940 wage credits. In the other agency, the estimate was made by assuming that the number of postings of 1940 wage records for each worker with any 1940 postings was the same in the entire file as in the sample. The known number of 1940 postings was divided by the number of 1940 postings per worker who had 1940 postings (as determined from the sample) to estimate the number of workers with 1940 wage credits. The chief requirement for accurate sampling of ledger cards is that the sample cards should be selected at approximately equally spaced intervals throughout the files. Block sampling of ledger cards should be avoided unless there is some objective assurance that the blocks are small enough to give an efficient sample.

First-wage-item sampling.—One State agency in 1940 used a first-wage-item sample. For each worker, the 1940 wage item filed nearest the front

of the numerical file was designated as the worker's first wage item. The numerical file was broken at approximately equal intervals, and the 10 wage items found immediately following each break were included in the sample. A count was made of the number of first wage items in the sample. The final estimate was made by assuming that the percentage of 1940 wage items that were first wage items would be the same in the combined alphabetic and numerical files as in the sample. This percentage as determined from the sample was multiplied by the known number of wage items in the entire file to estimate the number of first wage items included and therefore the number of workers with wage credits in the file. This technique makes adequate allowance for workers in the alphabetic file.

Worker sampling.—Although a number of State agencies used a worker sample in their 1939 estimates, only one State used it in 1940. The technique employed in 1939 was to break the numerical file at approximately equal intervals and to choose for the sample the wage items of the worker following the wage items of the worker at whose records the break occurred. The wage items of the worker at whose records the break occurred were not suitable, since file breaks are more likely to fall among a worker's wage items if he has an above-average number of such items. Assuming that the number of 1939 wage records per worker with some 1939 wage records was the same in the entire file as in the sample, the final estimate was made by dividing the known number of wage records in the entire file by the number of 1939 wage records per worker as determined from the sample.

This estimating technique has been proved to give an estimate averaging about 10 percent below the actual number of workers with wage credits, because the number of workers whose wage items were sampled from each tray is approximately proportional to the number of wage items in the tray, whereas it should be proportional to the number of workers with wage items in the tray.

Chart 2.—*Regression of workers with wage credits on workers employed in high month of employment*

18   *Social Security*
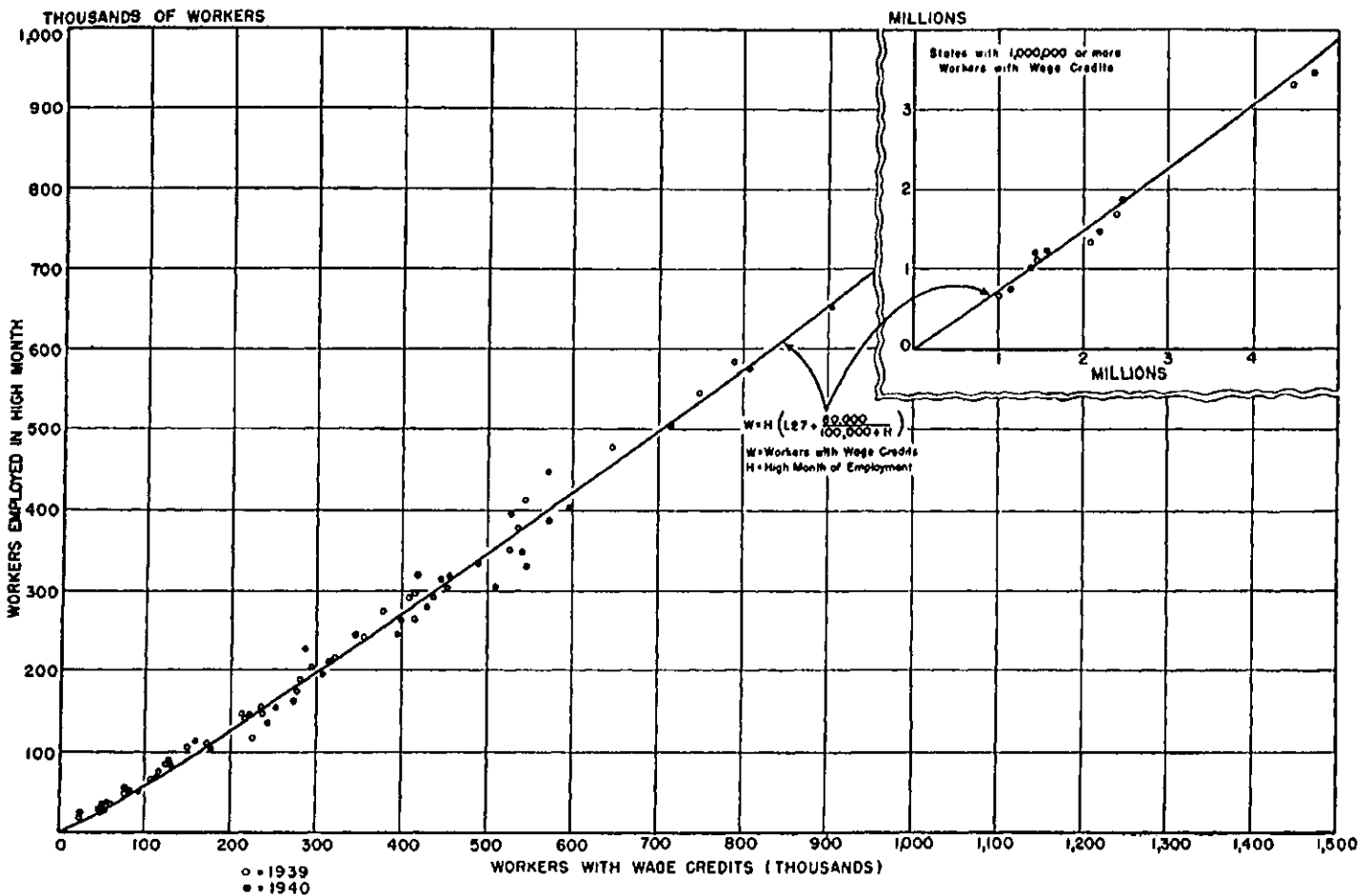
Thus, the trays with few wage items per worker were undersampled and those with above-average wage items per worker were oversampled.

As a simplified example of this bias, consider a 2-tray file. Tray 1 represents 600 workers with 2,400 wage items; tray 2 represents 1,200 workers with 2,400 wage items; thus the file represents 1,800 workers with 4,800 wage items. A 15-worker representative sample from tray 1 would have, on the average, 60 wage items; a 15-worker representative sample from tray 2 would have, on the average, 30 wage items; thus the combined worker sample would represent, on the average, 30 workers with 90 wage items. A truly representative sample from the file, however, would represent, on the average, 30 workers with 80 wage items; the combined sample would therefore understate the actual number of workers with wage credits by 12.5 percent.

The one State agency which used the worker sample in 1940 files its wage items by quarters instead of years. As a result, equally filled trays of wage items contain wage items for approximately equal numbers of workers, and the bias is not serious.

Actual count.—For 1940 seven State agencies found it possible to make an actual count of the number of workers with wage credits.

Regression estimates.—For the nine State agencies receiving some wage reports on a wage and separation basis in 1940 and for one State agency whose files were being reorganized so that accurate sampling was not possible, special estimates were made by the Bureau of Employment Security on the basis of the relationship between 1940 workers with wage credits and the high month of 1940 employment for each of the 41 States for which both types of data were available. It was discovered that there was a high correlation between the number of workers with wage credits in 1940 in a given State and that State's high month of employment in 1940. The number of 1940 workers with wage credits was plotted for each of the 41 States against the State's high month of employment. A curve of regression fitted to these 41 points was found to be hyperbolic (see chart 2). Since the equation for this curve involved only two variables, one representing workers with wage credits and the other repre-

senting high month of employment, an estimate of workers with wage credits could be readily made for any State by substituting in the hyperbolic equation the value for that State's 1940 high month of employment and solving for the remaining unknown variable.

Summary

Of the sampling methods used, the social security number end-digit method meets the requirements of more States than any other method. Moreover, a sample chosen by this technique may be readily used for estimating the distribution of annual wages per covered worker and other data available in the file. In one State agency the wage items are filed in social security account serial number order. That is, the order of filing is determined primarily by the last four digits of the social security account number. This system makes it possible to select an end-digit sample by separating out large blocks of the file. This filing procedure introduces no complications in the agency procedures and is worth considering for its sampling advantages.

The random wage item sample has approximately the same efficiency as the first wage item sample but probably cannot be drawn with as little work, since it requires more than twice as many file breaks to obtain the same-sized sample.

The ledger-card and workers samples are valuable in a limited number of States which have unusual filing situations.

The added accuracy from an actual count is probably not sufficient to justify the added labor in most States. However, it is practical in States having few covered workers or in States in which it can be obtained as a byproduct of other operations.

Block samples lose precision rapidly as the blocks are enlarged; hence, care must be taken not to expand the blocks to a point at which the sample chosen is less efficient than smaller samples of other types.

For estimating the number of workers with wage credits, the most appropriate sample in any State is one that will provide the required accuracy with a minimum amount of labor. In using any of the available techniques, the standard error of estimate probably should not be allowed to exceed 2 percent.