

*A MEASURE OF ANNUAL WORK EXPERIENCE USING INDIVIDUAL INCOME TAX RECORDS
- SOME INITIAL RESULTS -*

Douglas Norris and Anne Mussely, Statistics Canada

INTRODUCTION

In recent years, there has been an increasing demand for small area labour market data on both employment and unemployment. Currently, the main source of small area data on the labour force is the quinquennial Census of Population. In addition, the Canadian monthly Labour Force Survey (LFS) provides detailed labour market data for Canada and the provinces and more limited data for larger subprovincial regions (census metropolitan areas and economic regions). During the last few years, work has been ongoing on a number of fronts to develop more frequent small area labour market data. One approach has been to develop small area labour market data from the Labour Force Survey (Drew et al., 1982).

Other potential sources of small area data are administrative records. In particular, the annual individual income tax records and the monthly unemployment insurance records both contain a vast amount of information on the labour force. Bobet et al. (1982) reported on some exploratory work on using the unemployment insurance records to develop small area indicators of unemployment.

The focus of this paper is on the development of a measure of the size and characteristics of the "annual" work force as derived from individual income tax records. Although much of the work with administrative records has been directed at the development of small area data, until recently it has not been possible to compare empirically the administrative records data to alternative data sources. In recent months, however, data from the 1981 Census of Population have become available and this provides an opportunity for an assessment of the administrative data.

This paper, then, is the first in a series to evaluate small area labour market indicators of employment, unemployment and income by comparing data derived from administrative records to similar data available from the Census of Population.

THE CONCEPT OF ANNUAL WORK EXPERIENCE

The concept of annual work experience is well established. The 1981 Census of Population collected data on the work experience, including the number of weeks worked and whether these were mostly full time or part time. These data were collected for the calendar year 1980, the year preceding the census date. An important use of these data is in interpreting and analysing the income data collected in the Census. A second source of work history data is the Annual Work Patterns Survey (AWPS) that has been conducted since 1977. This survey is a supplement to the monthly Labour Force Survey, generally in January of each year. The survey is used to collect data on labour force experience during the preceding calendar year. Respondents are classified as to whether they were employed, unemployed, not in the labour force or some combination of these for each of the 12 months. Individuals are then classified as to their "annual" labour force status (for a complete description of the survey, see Statistics Canada, 1982).

Individual income tax records provide a third source of annual labour force data. The tax records provide detailed data on income by source, in particular employment income including wages and salaries and self-employment income. In addition, the records include basic demographic data on age, sex and marital status. The tax records can therefore provide a measure of the size and characteristics of the annual work force, by considering those persons who report employment income, i.e., either wages and salaries and/or self-employment income. Although the data derived from tax records are limited in that there is no indication of the duration of work, a strength of the tax records is that data can be produced for small geo-

graphic areas. A similar measure of the work force based on earnings can also be derived from the annual Survey of Consumer Finances, however the Survey provides only limited sub-provincial data.

The measure of annual work experience derived from reported earnings on tax records differs in a number of ways from similar measures derived directly from census or survey questions on work experience. On the one hand, the tax data miss certain persons, in particular those who worked and received employment income but did not file a tax return and those who worked but received no income during the year (e.g., unpaid family workers). On the other hand, tax records include some persons who received employment income but did not work during the year, perhaps having worked in the previous year.

Note that the tax records provide an indication of annual work experience, that is the number of persons who worked anytime during the year. This is in contrast to the more common measure of employment derived from the monthly Labour Force Survey that refers to persons working during the reference week. The Survey also provides an annual average employment level that is an average of the monthly levels. This is different from the annual measure as discussed here. For example, the AWPS for 1980 reported that 12,738,000 persons were employed sometime during 1980 but the average number employed each month was 16% lower (10,751,000). This difference reflects the movement of individuals into and out of the labour force during the year.

SOME EMPIRICAL COMPARISONS

As indicated above, a prime motivation for using the tax records is the ability to produce data at a small area level of geographic detail. Experimental data on the annual work force have been developed for each of the years 1976-81 for all Census Divisions (counties) in Canada. The recent availability of data from the Census of Population allows for a comparison of the administrative data for 1980 to data derived from the Census.

Table 1 shows a comparison of the number of persons in the work force as indicated by reported earnings on the tax records compared to the census data on persons reporting some work experience during 1980, and those that both worked and received earnings. Overall, the tax records are 6.1% lower than the Census counts of persons who worked in 1980 with the differences being higher for females than for males. Part of this difference is explained by persons who worked but received no income. If such persons are taken out, the difference is reduced to 3.6%. This remaining difference may be explained in part by marginal workers with low income that may not necessitate tax filing. (For example, students working for a few months during the summer).

Comparisons between the tax data and the census data by age and sex are shown in Table 2. As can be seen, the largest differences are at the younger ages. The Census data on persons who worked during 1980 are 14% higher for ages 15-24. The differences are minimal for older males and are about 4% for both males and females in the prime working ages 25-44. If only earners are considered the differences are minimal in the prime working ages and in fact for males 45 and over, the tax data are 3% higher than the Census.

The tax and census data can also be compared by employment income class. Table 3 shows the results for both males and females. As expected, the tax data for persons earning less than \$3,000 are less than the census counts. For males, the census count is 14% higher than the tax count while for females the census count is 17% higher. This, together with the comparisons by age in Table 2, supports the conclusion that the tax data miss young marginal workers, many of

whom may not be required to file income tax returns. For incomes in excess of \$3,000, the counts of the employed from the tax data are 99% of the census counts for both males and females. However, within the group reporting more than \$3,000 employment income, there are substantial variations by income class. The tax data are higher than the census for the income class \$3,000-\$9,999 but lower for higher income classes particularly incomes in excess of \$25,000. These differences will be investigated further in a future study. The data from the tax records can also be compared to the data derived from the Annual Work Patterns Surveys for the four years 1977-80. Table 4 shows the comparisons for males and females. The Survey data show the annual work force to be 6 to 7% higher than the tax records and again, the differ-

ences are higher for females than for males. For males, the differences showed an increase over time from 3.2 % in 1977 to 5.3% in 1980. For females, the differences fluctuated around 10%. The survey universe is slightly different from that of the Census and the tax records in that it excludes residents living in the Northwest Territories and Yukon, persons living on Indian reservations, inmates of institutions and full time members of the armed forces. A comparison of the Census and Annual Work Patterns Survey data adjusting for the differences in universes showed that the Census count of the annual work force was 2.3% lower than the survey count. This may be explained by the fact that the Census was held five months after the end of 1980, the reporting year, compared to only one month for the survey and there may be

TABLE 1
A Comparison of the Annual Work Force from Census and Tax Records,
Canada, 1980.

Sex	Tax	Census Worked		Census Worked and Received Earnings	
		Number	Percent Difference	Number	Percent Difference
Males	7,020,717	7,378,840	4.85	7,207,620	2.59
Females	4,811,371	5,230,360	8.01	5,065,635	5.02
Both Sexes	11,832,088	12,609,190	6.16	12,273,255	3.59

NOTE: Percent Difference = (Census - Tax)/Census * 100

TABLE 2
A Comparison of the Annual Work Force from 1981 Census and Tax
Records by Age Group and Sex, Canada 1980.

Sex/Age	Tax	Census Worked in 1980		Census Worked and Received Earnings	
		Number	Percent Difference	Number	Percent Difference
Males					
15-24	1,519,109	1,726,390	12.01	1,691,615	10.20
25-44	3,300,094	3,438,755	4.03	3,381,600	2.41
45 and over	2,201,514	2,213,695	0.55	2,134,405	-3.14
All Ages	7,020,717	7,378,840	4.85	7,207,620	2.59
Females					
15-24	1,266,271	1,506,005	15.92	1,476,930	14.26
25-44	2,363,019	2,459,745	3.93	2,395,705	1.36
45 and over	1,182,081	1,264,360	6.53	1,193,000	0.92
All Ages	4,811,371	5,230,360	8.01	5,065,635	5.02
Both Sexes					
15-24	2,785,380	3,232,390	13.83	3,168,545	12.09
25-44	5,663,113	5,898,500	3.99	5,777,305	1.98
45 and over	3,383,595	3,478,300	2.72	3,327,405	-1.69
All Ages	11,832,088	12,609,190	6.16	12,273,255	3.59

NOTE: Percent Difference = (Census - Tax)/Census * 100

TABLE 3
A Comparison of Number of Persons Who Worked and Received Employment
Income in 1980 from the 1981 Census and Persons Who Reported
Employment Income on 1980 Tax Returns, by Employment Income Class
and Sex, Canada.

Employment Income	Males			Females		
	Tax Number ('000)	Census Number ('000)	Percent Differ- ence	Tax Number ('000)	Census Number ('000)	Percent Differ- ence
Less than \$3,000	727	850	14.4	1,039	1,251	16.9
\$3,000 +	6,294	6,358	1.0	3,773	3,815	1.1
\$3000-\$9999	1,550	1,363	-13.7	1,877	1,796	-4.5
\$10000-\$24999	3,425	3,528	2.9	1,770	1,865	5.1
\$25,000 +	1,319	1,468	10.1	125	154	18.1
TOTAL	7,021	7,208	2.6	4,811	5,066	5.0

NOTE: Percent difference = (Census - Tax)/Census * 100

TABLE 4
Percentage Differences Between Annual Work Force from Tax Records and
the Annual Work Patterns Surveys by Sex, Canada 1977-1980
(excludes Yukon and Northwest Territories)

Sex	1977	1978	1979	1980
Males	3.17	3.98	4.66	5.32
Females	10.24	9.88	10.55	9.61
Both Sexes	5.85	6.28	7.03	7.05

NOTE: Percent Difference = (Survey - Tax)/Survey * 100

TABLE 5
Percentage Differences Between Measures of Annual Work Force Derived
From Tax Records and 1981 Census by Sex and Provinces, 1980.

Provinces	Males	Females	Both Sexes
Newfoundland	1.47	4.44	2.53
Prince Edward Island	7.27	10.49	8.60
Nova Scotia	2.52	7.08	4.32
New Brunswick	1.97	5.52	3.34
Quebec	4.63	7.55	5.77
Ontario	4.73	7.86	5.98
Manitoba	3.47	7.80	5.26
Saskatchewan	4.10	10.77	6.72
Alberta	7.43	9.82	8.35
British Columbia	5.99	8.24	6.90
Yukon	5.87	8.38	6.87
Northwest Territories	6.35	7.97	6.76
Canada	4.85	8.01	6.16

NOTE: Percent difference = (Census - Tax)/Census * 100

a higher recall lapse in the Census.

The comparison of the data from the Census and tax records at the provincial level is summarized in Table 5. Differences range from 2.5% in Newfoundland to 8.6% in Prince Edward Island. Differences are generally lowest in the Atlantic provinces with the exception of Prince Edward Island and highest in Prince Edward Island, the western provinces and the Territories. Differences for females are particularly high for Prince Edward Island and Saskatchewan. These provinces have the highest concentration of workers without earnings with 3.7% in Saskatchewan and 3.4% in Prince Edward Island.

As indicated earlier, the main strength of the tax records is the ability to derive data for small areas. Estimates of the annual work force for 1980 from the tax records have been tabulated for the 265 census divisions (or counties) in Canada. Compared to the Census data, the counts from the tax records were generally lower and the absolute average difference between the two sources was about 7%. This average deviation is much higher than the 4% difference observed for Canada as a whole. In fact, as shown in Table 6, 40 census divisions (15% of all divisions) showed deviations greater than 10% and 5 showed deviations greater than 20%. The average deviations by province are shown in Table 7. Average deviations range from a low of 2.3% in Newfoundland to a high of 12.2% in Manitoba.

In addition to differential tax filing rates due to variations in unpaid family workers and low income earners, the differences across census divisions are also partly due to problems in geocoding the tax records to a small area level of detail. There are problems, for example, in coding the tax records to census divisions, especially in smaller rural areas. The use of the postal code as a geographic indicator is discussed in detail elsewhere (Kopustas et al, 1983).

The net coverage bias caused by using the postal code to identify census divisions has been indirectly estimated. The estimates were made by comparing aggregate census division counts of children aged 1-14 from family allowance records in June 1981 to census counts. Since the family allowance data are derived using the postal code as an indicator of census division, and since the family allowance records are nearly universal in coverage, differences between the two sources, especially larger differences, are mainly due to problems of using the postal code as a geographic identifier. At the national level and, in fact, for most provinces the two sources give nearly identical counts of children. Therefore the ratio of the counts from the Census to counts from the Family Allowance records can be taken as a measure of net geographic coverage bias. These were therefore used to adjust the estimates of annual work force derived from the tax records. The adjusted counts were then again compared to the Census counts of annual work experience. The results are shown in Tables 6 and 7. The adjustment reduced the deviations between the tax data and Census data although a number of large differences remain. An initial analysis of the larger differences indicates that many of them occur in small census divisions (work force less than 10,000) and in northern

parts of the country. The problems in northern areas have surfaced in other work with the tax records and may be due to the high turnover of workers in the North, many of whom may be in the North only temporarily. In some cases, the addresses used for filing income tax returns (during March and April) may differ from the "permanent place of residence" in the June Census. The role of these factors, such as unpaid family workers and low income earners, in explaining the differences between the tax and census data will be assessed as data become available.

SUMMARY AND CONCLUSIONS

This paper has presented a measure of the annual work force derived from individual income tax records. One strength of the tax records is the ability to derive annual data for small geographic areas. To evaluate the tax data, comparisons were made to the 1981 Census data. The results showed that, at the national level, the tax records provided estimates that were about 6 to 7% lower than the Census. The differences were larger for females and for younger age groups. Most of the differences are probably attributable to undercoverage of workers without earnings or with very low earnings, particularly young workers. There is close agreement between the tax and census data for the prime working ages. Compared to measures of the annual work force derived from the Annual Work Patterns Survey, the tax data were again lower but the differences did not vary greatly over time, although there was a gradual increase in the difference between the two sources for males.

Evaluation at the census division level produced mixed results. Initial comparisons showed some large differences between the tax data and Census counts. Some of these differences are due to the known problems of using the mailing address, in particular the postal code, as a geographic indicator. An attempt was made to adjust for the postal code problems and this produced somewhat more favourable comparisons between the Census and tax data. However, large differences remain and therefore either the adjustment for geographic coding bias was unsatisfactory or more likely the remaining differences may be due to differential coverage of the tax records, possibly due to differences in the incidence of marginal workers. A next step will be to do more detailed comparisons between the tax and Census data. Further work will also be required to investigate those geographic areas where there are large differences between the data sources.

The problems of geographic coding and differential coverage are clearly of importance if one is interested in measures of the size of the work force or, in fact, the level of any factor. On the other hand, these coverage problems may be less important if one is primarily interested in change over time. If one is interested in measures of size, one option that could be considered is to use the Census counts as a benchmark and "update" the Census using indicators of change derived from the tax records.

TABLE 6
Distribution of Unadjusted and Adjusted* Absolute Percentage Differences** Between Estimates of Annual Work Force Derived from Tax Records and 1981 Census, Canada.

Absolute Percentage Differences	Unadjusted Differences		Adjusted Differences	
	Number	%	Number	%
< 4.9	108	40.75	143	53.96
5.0 - 9.9	117	44.15	99	37.36
10.0 - 14.9	29	10.94	14	5.28
15.0 - 19.9	6	2.26	5	1.89
> 20	5	1.90	4	1.51
Total	265	100.00	265	100.00
Average Percentage Difference	7.08		5.52	

*The adjusted data are an attempt to adjust for coverage problems caused by geographic coding. See text for details.

**The percentage differences are calculated as $(\text{Census} - \text{Tax}) / \text{Census} * 100$.

TABLE 7
Absolute Average Percentage Differences* for Census Divisions Between Estimates of Annual Work Force Derived from Tax Records and 1981 Census, by Provinces (Average Percentage Differences for Census Divisions).

Provinces	Unadjusted	Adjusted**
Newfoundland	2.32	2.38
Prince Edward Island	8.96	9.69
Nova Scotia	5.88	3.77
New Brunswick	6.32	2.92
Quebec	4.97	4.08
Ontario	6.38	5.08
Manitoba	12.21	9.64
Saskatchewan	7.25	5.45
Alberta	8.33	6.26
British Columbia	8.43	8.14
Yukon	6.87	5.96
Northwest Territories	8.05	8.41
Canada	7.08	5.52

*The percentage differences are calculated as $(\text{Census} - \text{Tax}) / \text{Census} * 100$.

**The adjusted data are an attempt to adjust for coverage problems caused by geographic coding. See text for details.

REFERENCES

- Bobet, E., Leyes, J., Radley, L., 1982. "The Use of Unemployment Insurance Records to Derive an Unemployment Indicator". In *Statistics of Income and Related Administrative Records Research: 1982*, pp.21-24. Department of the Treasury, Internal Revenue Service, Statistics of Income Division, October 1982.
- Choudhry, G.H., Drew, J.D., Singh, M.P., 1982. "Evaluation of Small Area Estimation Techniques for the Canadian Labour Force Survey". In *Survey Methodology*, Vol.8, No.1 and 2, 1982, pp.17-47. Ottawa: Methodology Staff, Statistics Canada.
- Kopustas, N. et al., 1983. "Geocoding Administrative Records". Presented at the 1983 Annual Meetings of the American Statistical Association, Toronto, Ontario, August 16-18, 1983.
- Statistics Canada, 1982. "Patterns of Full and Part Year Employment and Unemployment: Results of the Annual Work Patterns Surveys 1977-1980". In *The Labour Force*, Catalogue No.71-001 (March) pp.79-87. Ottawa: Supply and Services Canada.