Socio-economic Characteristics of Families in Wentworth County, 1871: Some Further Results

by Frank T. DENTON and Peter J. GEORGE *

I. — INTRODUCTION

An earlier paper[1] described how techniques of cross-section regression analysis were applied to a sample of historical micro-data drawn from the manuscript of the 1871 population Census of Hamilton, Ontario. The analysis was concerned with age, birthplace, religion, ethnic origin, occupation, and other variables for which the 1871 Census provided information, as determinants of variations in family size and school attendance. The findings, although somewhat tentative, indicated an apparent and somewhat surprising weakness in the relationships between many of these socio-economic variables, on the one hand, and the number of children in a family unit and the propensity of older children to attend school, on the other.

As work on the project progressed, it became apparent that the study would benefit substantially from an expansion of the sample. It would be desirable to have a larger sample of 1871 Census data, but also one which included rural as well as urban families. Accordingly, the original ten-percent sample of enumerated dwellings in Hamilton was augmented by taking a systematic twenty-percent sample of dwellings in the districts of Wentworth North and South, using the same basic sampling procedures as in the earlier phase of the study.[2] The two samples together yielded 1,100 "normal" families in Wentworth County, and these constitute the observations on which the results reported here are based. Of these, 429 families were located in the urban areas of Hamilton and Dundas, and the remaining 671 in the rural areas of Wentworth North and South.[3] In addition, the sample of "normal" families produced 1,016 children between ten and sixteen years of age, for use in the analysis of school attendance.

Boys numbered 520 and girls 496, with 358 children living in urban areas and 658 in rural areas. The enlarged sample made possible an analysis of the influence of urban-rural and farm-nonfarm differences on family size and school attendance, as well as a more reliable analysis of the influence of the other variables which had been considered earlier.

II. — ANALYSIS OF FAMILY SIZE

The 1,100 observations on "normal" families were used first in an analysis of family size. Regression techniques[4] were employed to study the extent to which differences in the number of children (CHILD) in the family unit were related to differences in husband's occupation,[5] differences in the age, religion, birthplace, and ethnic origin of both husband and wife, and basic urban-rural and farm-nonfarm differences. Definitions and symbols for all relevant variables are listed at the end of the paper. Table I presents the results of F-tests for groups of variables.[6]

The family-size regression includes four independent variables which are labelled "demographic" (DEM) — wife's age and its square (AGEW and AGEW2), the difference between wife's age and husband's age (AGEDIF), and the number of children in the family other than those with the family surname (ODEP). As in the earlier paper, these variables are highly significant as a group (row 2). Variables representing husband's occupation continue to be insignificant as a group (row 7), as do the dummy variables representing religion and ethnic origin of both husband and wife, and birthplace of husband. On the other hand, variables representing birthplace of wife do appear to be significant at the 1 percent level (row 12). In sum, of all the variables representing occupation, religion, birthplace, and ethnic origin, birthplace of the wife alone can be regarded as being a significant source of influence on family size.

The twenty-eight variables relating to religion, birthplace, and origin can be separated into two groups — those relating to the husband and those relating to the wife. One may then ask whether the husband's or the wife's characteristics are more influential on family size. Whereas the F-tests indicate no significance for husband's characteristics (rows 17 and

---

[4] A general discussion of the use of "dummy" variables in these regressions and of the interpretation of the F-tests is presented in DENTON and GEORGE, "An Exploratory Statistical Analysis," pp. 23-24.

[5] Occupations reported in the 1871 Census were classified according to a rough "socio-economic" index scale. Procedures used are discussed in DENTON and GEORGE, "An Exploratory Statistical Analysis," pp. 21-22, and in "The Influence of Socio-Economic Variables on Family Size in Wentworth County, Ontario, 1871: A Statistical Analysis of Historical Micro-Data," in *The Canadian Review of Sociology and Anthropology* (November 1973), especially footnote 8.

[6] In order to conserve space, the actual regression equations for family size and school attendance are not included in this paper. They are available from the authors on request.

18), wife's characteristics (row 18) are significant at the 5 percent level, a result that appears to be attributable entirely to the effect of wife's birthplace.

The lack of any strong connection between family size and most of the socio-economic variables, which was a major finding in the earlier paper, was confirmed for the extended sample including both urban and rural households. However, the importance of basic urban-rural differences, which could not be considered in the earlier paper, stands out clearly now (row 3): in the regression equation, the coefficient of the urban-rural variable is significant at the 1 percent level and negative, reflecting a tendency for urban families to be smaller than rural families. On the other hand, the farm-nonfarm variable is not significant. These results together imply that urban-rural differences are important but that whether a rural family is a farm family or not is of little importance.

The tendency for rural families to be larger than urban families is well known, of course. But this tendency might have been thought to be associated largely with differences in the socio-economic characteristics of rural and urban families. That this appears not to be the case — that there appears to be a basic difference after allowance has been made for the effects of differences in parental birthplace, occupation, age, and so on — is perhaps one of the major contributions of the present analysis.

III. — ANALYSIS OF SCHOOL ATTENDANCE

The analysis of school attendance relates to children from ten to sixteen years of age, inclusive. It employs the 1,016 observations on children in this age group in the sample of "normal" families. As before, extensive use is made of dummy variables, and the dependent variable itself (GTS) is a dummy variable, having value one if a child is attending school and value zero otherwise. Many of the independent variables correspond closely to those in the family-size analysis. These include occupation of father; religion, birthplace, and ethnic origin of mother and father; and urban-rural and farm-nonfarm variables. However, the "demographic" (DEM) variables are different, being five in number, and consisting of age of child and its square (AGEC and AGEC2), sex of child (SEX), total number of children in the family who have the family surname (CHILD), and age of father (AGEF). The results of F-tests on groups of variables in the school-attendance analysis are reported in Table II.

Among the "demographic" variables, age of child is highly significant, as one would certainly expect: the probability that a child will be attending school is obviously much less for a sixteen-year-old than for a ten-year-old. Surprisingly enough, the sex of the child, which was significant in the earlier study, turns out not to be significant in the regression analysis at even the 10 percent level with the enlarged sample. Number of

children in the family and age of father are not significant. As a group, the "demographic" variables are significant at the 1 percent level (row 2).

Some of the non-demographic variables performed rather better with the extended sample. Religion and origin of either father or mother still give no evidence of influencing school attendance. Birthplace performs somewhat better: when the ten variables representing birthplace for mother and father are considered together, the F-test indicates significance at the 5 percent level. On the basis of the complete set of variables for religion, birthplace, and origin, the father's characteristics test as significant at only the 10 percent level, the mother's characteristics not at all (rows 17 and 18).

In the earlier paper, the equations were suggestive of a significant influence for father's occupation, and the present study confirms that occupation of father has a strong influence on school attendance. As a group, the four occupation variables are significant at the 1 percent level (row 7).

Urban-rural differences appear to be important, as in the case of family size. The variable URBAN itself tests as significant at the 5 percent level (row 3), and when it is paired with the variable FARM, the two together test as significant at the 1 percent level (row 5).

The general picture that emerges with respect to the determinants of school attendance is as follows: except for the obvious influence of the child's own age, "demographic" variables have no discernible effect. Parental birthplace does give indications of being significant in some degree and occupation of father appears highly significant. Basic urban-rural differences also appear significant, especially when combined with a farm-nonfarm categorization of the rural population.

IV. — SUMMARY AND CONCLUSIONS

Use of the extended sample for Wentworth County has provided new and more reliable evidence with regard to the factors influencing family size and school attendance a century ago. The additional evidence represents a substantial advance beyond the results of the authors' earlier paper. For one thing, the greater sample size makes more reliable the results of the statistical tests that have been employed. For another, the inclusion of rural areas in the sample, and the distinction between rural and urban and between farm and nonfarm, introduces an important new dimension into the analysis.

"Demographic" and urban-rural factors were found to be highly significant in the case of family size. However, "socio-economic" influences, as represented by occupation, religion, birthplace, and ethnic origin, were found generally to be weak. Only birthplace of wife appeared to be of much significance in this regard. This confirms the preliminary

finding in the earlier paper with regard to the somewhat surprising weakness of many of these influences.

Aside from age of child, "demographic" characteristics were not found to be significant in the determination of school attendance; even sex of child was not found to be significant. On the other hand, urban-rural combined with farm-nonfarm differences were found to be important. Occupation was also found to be important, with the probability of a child attending school generally greater the higher on the index scale the father's occupation. There was some moderately strong evidence to indicate that parents' birthplace may be important but none to suggest any significant influence on the part of religion or ethnic origin.

DEFINITIONS OF VARIABLES USED IN ANALYSIS OF FAMILY SIZE

CHILD   number of children in family with family surname

AGEW   age of wife expressed as difference in years from sample mean

AGEW2   square of AGEW

AGEDIF   difference in years between husband's age and wife's age

ODEP   number of children in family other than those with family surname

OCCH   occupation of husband: dummy variables based on index scale
           OCCH3 = 1 if index 50.00-59.99; 0 otherwise
           OCCH4 = 1 if index 40.00-49.99; 0 otherwise
           OCCH5 = 1 if index 30.00-39.99; 0 otherwise
           OCCH6 = 1 if index less than 30.00; 0 otherwise
           OCCH7 = 1 if index 60.00 or greater; 0 otherwise

RH   religion of husband: dummy variables

           RH1 = 1 if Church of England; 0 otherwise
           RH2 = 1 if Roman Catholic; 0 otherwise
           RH3 = 1 if Presbyterian; 0 otherwise
           RH4 = 1 if Baptist; 0 otherwise
           RH5 = 1 if Methodist; 0 otherwise
           RH6 = 1 if any other religious denomination; 0 otherwise

RW   religion of wife: dummy variables defined in same way as corresponding variables for husband - RW1, RW2, RW3, RW4, RW5, RW6

BH   birthplace of husband: dummy variables

           BH1 = 1 if England; 0 otherwise
           BH2 = 1 if Ireland; 0 otherwise
           BH3 = 1 if Scotland; 0 otherwise
           BH4 = 1 if U.S.A.; 0 otherwise
           BH5 = 1 if Canada; 0 otherwise
           BH8 = 1 if any other birthplace; 0 otherwise

BW   birthplace of wife: dummy variables defined in same way as corresponding variables for husband - BW1, BW2, BW3, BW4, BW5, BW8

OH
: ethnic origin of husband: dummy variables
OH1 = 1 if English; 0 otherwise
OH2 = 1 if Irish; 0 otherwise
OH3 = 1 if Scottish; 0 otherwise
OH4 = 1 if German; 0 otherwise
OH9 = 1 if any other ethnic origin; 0 otherwise

OW
: ethnic origin of wife: dummy variables defined in same way as corresponding variables for husband - OW1, OW2, OW3, OW4, OW9

URBAN
: dummy variable = 1 if family lives in an urban area; 0 otherwise

FARM
: dummy variable = 1 if family head is a farmer; 0 otherwise

DEFINITIONS OF VARIABLES USED IN ANALYSIS OF SCHOOL ATTENDANCE

GTS
: dummy variable = 1 if child goes to school; 0 otherwise

AGEC
: age of child expressed as difference in years from sample mean

AGEC2
: square of AGEC

SEX
: dummy variable = 1 if child is male; 0 otherwise

CHILD
: number of children in family with family surname

AGEF
: age of father in years (actual age, not difference from mean)

OCCF
: occupation of father: dummy variables defined in same way as corresponding variables for husband in analysis of family size - OCCF3, OCCF4, OCCF5, OCCF6, OCCF7

RF, RM
: religion of father, religion of mother: dummy variables defined in same way as corresponding variables for husband and wife in analysis of family size - RF1, RF2, RF3, RF4, RF5, RF6; RM1, RM2, RM3, RM4, RM5, RM6

BF, BM
: birthplace of father, birthplace of mother: dummy variables defined in same way as corresponding variables for husband and wife in analysis of family size - BF1, BF2, BF3, BF4, BF5, BF8; BM1, BM2, BM3, BM4, BM5, BM8

OF, OM
: ethnic origin of father, ethnic origin of mother: dummy variables defined in same way as corresponding variables for husband and wife in analysis of family size - OF1, OF2, OF3, OF4, OF9; OM1, OM2, OM3, OM4, OM9

URBAN
: dummy variable = 1 if family lives in an urban area; 0 otherwise

FARM
: dummy variable = 1 if family head is a farmer; 0 otherwise

Table 1: RESULTS OF SIGNIFICANCE TESTS (F-TESTS) FOR SELECTED GROUPS OF INDEPENDENT VARIABLES IN REGRESSION ANALYSIS OF FAMILY SIZE (DEPENDENT VARIABLE CHILD; 1,100 OBSERVATIONS)

Group	Number of variables in group	F-ratio	Is contribution of group significant at the —		
			1% level?	5% level?	10% level?
1. All variables	38	13.61	YES	YES	YES
2. DEM	4	94.14	YES	YES	YES
3. URBAN	1	11.46	YES	YES	YES
4. FARM	1	1.06	NO	NO	NO
5. URBAN+FARM	2	10.24	YES	YES	YES
6. DEM+URBAN+FARM	6	68.54	YES	YES	YES
7. OCCH	4	.27	NO	NO	NO
8. RH	5	.72	NO	NO	NO
9. RW	5	1.23	NO	NO	NO
10. RH+RW	10	.85	NO	NO	NO
11. BH	5	1.25	NO	NO	NO
12. BW	5	3.40	YES	YES	YES
13. BH+BW	10	2.11	NO	YES	YES
14. OH	4	.13	NO	NO	NO
15. OW	4	.21	NO	NO	NO
16. OH+OW	8	.19	NO	NO	NO
17. RH+BH+OH	14	.74	NO	NO	NO
18. RW+BW+OW	14	1.74	NO	YES	YES
19. RH+BH+OH+RW+BW+OW	28	1.10	NO	NO	NO
20. OCCH+RH+BH+OH+RW +BW+OW	32	1.02	NO	NO	NO

Note: For purposes of this table, the group DEM ("demographic" variables) includes AGEW, AGEW2, AGEDIF, and ODEP; F-ratios are computed from

$$F = (R_A^2 - R_B^2)(1 - R_A^2)^{-1}(n - k_1 - k_2 - 1)(k_1)^{-1},\ \text{where}\ R_A^2\ \text{and}\ R_B^2$$

are the coefficients of determination in regressions which include and exclude, respectively, the group of independent variables being tested, n is the number of observations, k_1 the number of independent variables in the group, and k_2 the number of independent variables not in the group.

Table II : RESULTS OF SIGNIFICANCE TESTS (F-TESTS) FOR SELECTED GROUPS OF INDEPENDENT
VARIABLES IN REGRESSION ANALYSIS OF SCHOOL ATTENDANCE (DEPENDENT
VARIABLE GTS; 1016 OBSERVATIONS)

Group	Number of variables in group	F-ratio	Is contribution of group significant at the —		
			1% level ?	5% level ?	10% level ?
1. All variables	39	8.27	YES	YES	YES
2. DEM	5	48.84	YES	YES	YES
3. URBAN	1	4.59	NO	YES	YES
4. FARM	1	1.85	NO	NO	NO
5. URBAN+FARM	2	6.30	YES	YES	YES
6. DEM+URBAN+FARM	7	36.27	YES	YES	YES
7. OCCF	4	3.68	YES	YES	YES
8. RF	5	.74	NO	NO	NO
9. RM	5	.51	NO	NO	NO
10. RF+RM	10	.95	NO	NO	NO
11. BF	5	2.15	NO	NO	YES
12. BM	5	2.21	NO	NO	YES
13. BF+BM	10	2.01	NO	YES	YES
14. OF	4	1.54	NO	NO	NO
15. OM	4	.37	NO	NO	NO
16. OF+OM	8	1.43	NO	NO	NO
17. RF+BF+OF	14	1.64	NO	NO	YES
18. RM+BM+OM	14	1.08	NO	NO	NO
19. RF+BF+OF+RM+BM+OM	28	1.44	NO	NO	YES
20. OCCF+RF+BF+OF+RM +BM+OM	32	2.06	YES	YES	YES

Note: For purposes of this table, the group DEM ("demographic" variables) includes
AGEC, AGEC2, SEX, CHILD, and AGEF; F-ratios are computed as described in the
note to Table I.