An Exploratory Statistical Analysis of Some Socioeconomic Characteristics of Families in Hamilton, Ontario, 1871

by Frank T. DENTON and Peter J. GEORGE *

I. — INTRODUCTION.

The project discussed in this paper involves an analysis of historical Census data obtained from a sample of households drawn from the manuscript of the population Census of Hamilton in 1871. As far as we are aware, it represents the first attempt in Canada to apply modern techniques of cross-section regression analysis to historical Census returns for individual persons or families, as distinguished from published totals. [1] In particular, the analysis is concerned with the extent to which variations in family size and school attendance can be explained by reference to age, birthplace, religion, ethnic origin, occupation, and other variables for which the 1871 Census provided information. We view the project described here in part as of interest in its own right and in part as a pilot study of the problems and possibilities of using the historical Census microdata for larger-scale analysis in the future. [2]

* The authors are respectively Professor and Assistant Professor of Economics, McMaster University. They gratefully acknowledge some helpful discussions with Alan Green of Queen's University in the initial stages of the project. An earlier version of this paper was presented at the Fourth Conference on Quantitative Methods in Canadian Economic History held at the University of Victoria, March 20 and 21, 1970.

[1] However, statistical techniques of a descriptive nature have been employed by Michael B. Katz in his recent examination of the relationship between wealth, occupation, and age in Hamilton in 1851 and 1852. Data for his study were drawn from the manuscript of the 1851 Census and the assessment rolls of 1852 for Hamilton. See Michael B. KATZ, "Social Structure in Hamilton, Ontario", in Stephan THERNSTROM and Richard SENNETT, eds., *Nineteenth Century Cities: Essays in the New Urban History* (New Haven, 1969), pp. 209-244. We are indebted to Professor H. V. Nelles of the University of British Columbia for drawing our attention to Katz's paper.

A recent statistical study based on British historical demographic microdata is David J. LOSCHKY and Donald F. KRIER, "Income and Family Size in Three Eighteenth-Century Lancashire Parishes: A Reconstitution Study", *The Journal of Economic History*, Volume XXIX (September, 1969), pp. 429-448.

[2] A proposal to undertake a sampling of the Canadian Census originated at the second conference on Quantitative Research in Canadian Economic History held at Queen's University in February, 1967. The Executive Committee of the Conference was delegated to consider further the proposal and, after some preliminary discussions, met with W. D. Porter, Director of the Census Division, Dominion Bureau of Statistics, and members of his staff to explore the feasibility of a sample of the Census from 1871 to 1961. While the representatives of the Bureau discouraged an attempt to sample the entire Census immediately, they did encourage the members of the Committee or others to proceed with a pilot project involving the manuscript of the 1871 Census, which had just been made available on microfilm. In November, 1969, the study reported here was initiated, based on 1871 Census returns for Hamilton, Ontario.

II. — THE POPULATION CENSUS OF 1871.

Individual returns are available on microfilm for the 1871 and earlier Censuses.[3] However, there does not exist a complete record of all individual returns: in particular, the agricultural Census of 1871 has been lost, and there are several incomplete returns for earlier population and agricultural Censuses.[4]

The 1871 Census of Canada was published in five volumes.[5] Volumes I to III are based on the data collected during the contemporary Census, whereas Volume IV is essentially a summary of all previous Censuses taken at different intervals in the area comprising Canada in 1871, and Volume V compares the findings of the 1851, 1861 and 1871 Censuses, and surveys the historical vital statistics of Quebec and Halifax. Volumes I and II contain primarily data compiled from Schedules Nos. 1 and 2 entitled "Nominal Return of the Living" and "Nominal Return of the Deaths within the Last 12 Months", respectively.[6] In Volume I are presented detailed summary tables of the characteristics of the population, including sex, religion, ethnic origin, birthplace, and marital status, by enumeration district and sub-district. Volume II contains further population data — ages of population, ages of married and widowed, and the age, sex, month of death, marital status, birthplace, religion, and occupation of persons who died during the previous year, for each enumeration district — and occupational data of the living by enumeration district. Volume III presents the summary tables prepared from the data collected in the enumerations of property, agricultural output and equipment, forest products, fisheries, and mineral products. It also provides information on certain details of manufacturing establishments — value

[3] The provinces and Census years for which microfilms are available from the Public Archives are: Ontario — 1842, 1851, 1861, 1871; Quebec — 1825-31, 1842, 1851, 1861, 1871; Prince Edward Island — 1841, 1861; New Brunswick — 1851, 1861, 1871; Nova Scotia — 1861, 1871; Manitoba — 1870.

Canadian source materials for genealogical research, including Census records, records of births, marriages and deaths, land records and immigration records, are briefly summarized in PUBLIC ARCHIVES OF CANADA, *Tracing Your Ancestors in Canada* (Ottawa, 1968).

[4] For Ontario, see Public Archives of Canada, Manuscript Division, *Check-List of Ontario Census Returns, 1842-1871* (Ottawa, 1963), in which the microfilm reel numbers are catalogued by cities, towns, townships, and counties for the Censuses of 1842, 1851, 1861, and 1871, and any variations in or incompleteness of returns are noted.

[5] GOVERNMENT OF CANADA, *Census of Canada, 1870-71*, Volumes I-V (Ottawa, 1873-8).

[6] The schedules are presented in the Annual Report of the Minister of Agriculture for 1870 in GOVERNMENT OF CANADA, *Sessional Papers*, 1871, Volume VI, No. 64, pp. 111-2.

of raw materials, value of output, number of employees, and wages, by type of establishment and enumeration district. [7]

The published Census tables report summary data in some detail, especially the results compiled from the "Nominal Return of the Living" schedules. For example, it can be found that in the township of Morrison in the district of Muskoka, of 601 persons enumerated, 9 were Baptists and 49 were Lutherans, [8] that there were 14 persons of German origin in the township of Otonabee in Peterborough East but no Welshmen, [9] and that in St. Mary's Ward, Hamilton, there were two persons reported as having been born at sea. [10] However, with the individual returns for 1871 now available on microfilm, it is possible to look beneath the published data. [11] It is now possible to develop finer cross-classifications of data than those presented in published sources and to analyze in detail the relationships among variables such as age, family size, birthplace, religion, ethnic origin, and occupation. With this in mind, we have drawn a sample of families from the individual returns for the City of Hamilton contained in Reel No. C-616 of the Public Archives microfilms of the Ontario Census, 1871. Background information on Census procedure is contained in the Annual Report of the Minister of Agriculture for 1870. [12]

[7] Compiled from Schedules No. 3 — "Return of Public Institutions, Real Estate, Vehicles and Implements", No. 4 — "Return of Cultivated Land, Field Products, and of Plants and Fruits", No. 5 — "Livestock, Animal Products, Home-Made Fabrics and Furs", No. 6 — "Return of Industrial Establishments", No. 7 — "Return of Products of the Forest", No. 8 — "Return of Shipping and Fisheries", No. 9 — "Return of Mineral Products". See *ibid.*, pp. 113-9.

[8] *Census of Canada, 1870-71,* Volume I, pp. 142-3.

[9] *Ibid.*, pp. 270-1.

[10] *Ibid.*, pp. 342-3.

[11] It is possible, for example, to identify the two persons in St. Mary's Ward who were "born at sea" as Margaret S. Ford (29 years old, Irish origin, member of the Church of England, unmarried, no occupation) and Richard Madge (61 years old, English origin, member of the Church of England, married, bricklayer). See Ontario Census 1871, Public Archives Microfilm Reel No. C-616. Ford was enumerated in dwelling No. 124, division No. 2 of St. Mary's Ward, and Madge in dwelling No. 188 in the same division.

Note: All examples of individual returns presented in the remainder of the paper are drawn from Reel No. C-616.

[12] *Sessional Papers,* 1871, Volume VI, No. 64, pp. 108-60. See especially the "Instructions to Officers", pp. 125-42.

Although the instructions to the individual enumerators are available, it has not been possible to locate at the Public Archives, the Dominion Bureau of Statistics, or in published sources the instructions to Government clerks who collated the individual returns and prepared summary tables for publication. It would be useful, for example, to find the "manual" which instructed clerks where to place particular occupations or trades within the 135 categories and 6 summary classes in the published Census.

W. D. Porter has advised us that "tabulating equipment was not used until the 1911 Census, and the compilations of the 1871 Census must have been performed manually by such means as the use of spread-sheets, etc". (Letter, January 9, 1970.) This suggests that a worthwhile project might be simply to employ the manuscripts of individual returns to check the accuracy of the published Census data. Insofar as allocating individuals'

The data on microfilm are presented in a form which is well-ordered for sampling and for locating "households" or "dwellings" and "families" in sequence.[13] For example, district No. 24 (Hamilton) is divided into five sub-districts (St. George's Ward, St. Mary's Ward, St. Andrew's Ward, St. Lawrence Ward, and St. Patrick's Ward). Each sub-district is divided into two or three divisions with a different enumerator in each division. The households and families in each division are numbered in order of visitation. For each person in each family or household, the enumerator completing the "Nominal Return of the Living" recorded:

1. Surname.
2. Given names.
3. Sex.
4. Age.
5. Month of Birth for Children Born within the Previous Twelve Months.
6. Country or Province of Birth.
7. Religion.
8. Origin.
9. Profession, Occupation or Trade.
10. Marital Status.
11. Whether or not Married within the Previous Twelve Months.
12. Whether or not Going to School.
13. Whether or not over 20 and Unable to Read.
14. Whether or not over 20 and Unable to Write.
15. Infirmities.

There was a column for remarks by the enumerator concerning unusual characteristics of the family or any of its members.[14]

occupations into categories is concerned, the lack of a "manual for clerks" implies that it would probably be necessary to group the detailed classes by trial and error in order to reconstruct the totals in the summary classes.

[13] The enumerator was instructed to go "personally from house to house", and to enumerate the family or families he found there, using the working definition that "A Family... may consist of one person living alone, or of any number of persons living together under one roof, and having their food provided together." *Sessional Papers*, 1871, Volume VI, No. 64, p. 128. Following this definition, enumerators frequently listed as one family for Census purposes two married couples and their children. They were directed to treat operators and residents of boarding-houses as one family. On the other hand, two or more families might be enumerated separately within one house, and receive different family numbers.

[14] For example, 16-year-old Annie Butler who was enumerated in household No. 121, division No. 1 of St. Mary's Ward, is noted as an "orphan", and Ann White (40 years old) in household No. 430, division No. 2 of St. Mary's Ward, as a "hypochondriac".

III. — A SAMPLE OF FAMILIES IN HAMILTON, ONTARIO.

The population of Hamilton in 1871 was reported as 26,716, the number of families as 5,084, and the number of dwellings as 4,830.[15] It was decided to draw a systematic ten-per-cent sample of the enumerated dwellings, in the order in which they appear on Reel C-616, starting with division No. 1 of sub-district "a" (St. George's Ward) and finishing with division No. 2 of sub-district "e" (St. Patrick's Ward). Beginning with a random number between one and ten, every tenth house was sampled regardless of division or sub-district boundaries. The location of dwellings for inclusion in the sample was facilitated by the sequential numbering of dwellings in the order of their visitation.

Not all of the families drawn into the sample were considered suitable for use in the study. In fact, since one of the major concerns was an attempt to explain variations in family size, it was decided at the outset to employ only data for "normal" families. For present purposes, a "normal" family is defined to be one in which (i) the husband (or head) and his wife are both present, and (ii) the husband is between 20 and 59 years of age, inclusive. Dwellings containing "non-normal" families only were discarded. Most of the dwellings remaining in the sample were single-family households, containing one "normal" family. However, dwellings which contained two or more families were encountered occasionally. The further decision was taken to include in the sample two or more "normal" families living in the same dwelling only if they were enumerated as separate families. On the other hand, two or more "normal" families enumerated as one Census family were discarded.

Once the decision rule to include only "normal" families enumerated separately had been invoked, there still remained some difficulties with the observations. First, there was a problem of partial illegibility of surnames and given names, especially in division No. 3 of St. Lawrence Ward. Where illegibility did not preclude the interpretation of data germane to the study, the family was retained in the sample. Secondly, there were four incomplete returns of the characteristics of a husband or wife in a family: for example, in one case the wife's origin was not given, and in another the husband's occupation was not indicated.[16]

[15] *Census of Canada, 1870-71*, Volume I, pp. 8-9.

[16] The enumerators were instructed to employ a dash (—) to indicate an answer of "nothing", "no", "not concerned", or "unknown". *Sessional Papers*, 1871, Volume VI, No. 64, p. 131.

Since it was not appropriate to make inferences from the answer "not given" (for example, that the man was either unemployed or outside the labour force), it was decided to drop families with incomplete returns from the sample. Thirdly, there was the occasional apparent transcription error in the reporting of answers by the enumerators.[17] For example, in one case, the occupation "labourer" was written beside the name of an eight-year-old daughter rather than beside the father. Such obvious errors were corrected, and the family retained in the sample. However, "normal" families whose returns could not be amended with confidence were discarded. Finally, there were 21 "normal" families in the sample which contained children of different surnames from the head and his wife. This might arise, for example, from a second marriage of the mother. These families were retained in the sample, and any children not over sixteen were entered in a separate category labelled as "other dependents 16 years and under".

We present in Tables I-XV some summary data relating to the sample. The ten-per-cent sample of dwellings yielded 483 households. After the definition of "normal" family and the decision rules described above had been implemented, there remained 350 dwellings and 357 "normal" families to constitute the observations upon which our study is based. Summary counts by birthplace, religion, and origin for the husbands and wives of the sample "normal" families are presented in Tables I, II, and III.

The development of a satisfactory occupational classification presented a somewhat different problem. Since information on income and level of education is not available from the 1871 Census, and since both might be expected, *a priori*, to influence both family size and school attendance, an attempt was made to employ occupational data in such a way as to represent income-educational effects, at least crudely. A socio-economic index of 320 occupations in the 1961 Census of Canada has

[17] The enumerators were charged "not to omit anything of importance", "not to record the same thing twice", "not to exaggerate anything", and "not to underrate anything". *Sessional Papers*, 1871, Volume VI, No. 64, p. 129.

An assumption basic to this study is, of course, that enumerators were scrupulous in the performance of their function; that is, enumerators did not make mistakes, except for "obvious" ones.

been developed by Bernard R. Blishen,[18] the ordering of occupations being represented as a function of income and education. In the absence of anything comparable for the latter part of the nineteenth century, this index was adapted[19] for use with the 1871 Census to produce a rough classification of reported 1871 occupations into 5 categories, ranging from high income-education occupations at one end (Category 1) to low income, unskilled occupations at the other (Category 5). The numbers of persons assigned to the 5 categories are reported in Table IV.

Table V summarizes the sample information on school attendance of children between 10 and 16 years old, inclusive, by age and sex of child. Tables VI-XV present a number of two-way cross-classifications: characteristics of husbands by characteristics of wives in VI-VIII; occupational category by other characteristics of husband in IX-XI; and area of residence by characteristics of husband in XII-XV.

IV. — A REGRESSION ANALYSIS OF FAMILY SIZE.

One of the uses that we have been making of the sample of "normal" families is in the analysis of family size. More specifically, we employ regression techniques to explore the extent to which variations in number of children in the family unit are related to variations in husband's occupation, variations in the age, religious denomination, and birthplace of both husband and wife, and differences associated with area of residence within the City of Hamilton, as it was in 1871. Although the analysis is not yet complete, a number of alternative regression equations have been estimated and some of these are presented in Tables XVI

[18] Bernard R. Blishen, "A Socio-Economic Index for Occupations in Canada", *Canadian Review of Sociology and Anthropology*, Volume IV (1967), pp. 41-53.

Also see Bernard R. Blishen, "The Construction and Use of an Occupational Class Scale", *Canadian Journal of Economics and Political Science*, Volume XXIV (November, 1958), pp. 519-31.

[19] Whereas Blishen calculated an index of socioeconomic rank of occupations to two decimal points, we are mainly concerned with defining broad groups and have employed summary categories with arbitrary bounds as specified in the note accompanying Table IV.

We hasten to add that we do not think that the 1961 occupational categorization is strictly appropriate for 1871; rather, it represents merely a rough guide for classifying 1871 occupations, the application of which has been tempered by our "judgement". Thus, where the meaning of occupational titles has undoubtedly changed over the 90-year period, the 1871 occupation was moved into a more appropriate category.

Moreover, it is worth noting that Katz rejected the use of a scale based on the 1961 Census occupations, preferring instead to devise a scale which he believed to be more representative of mid-nineteenth century conditions. Katz grouped occupations by function into seven major categories — Artisans, Businessmen, Professionals, Labourers, Public Employees, Gentlemen, and Unemployed — and cross-classified occupational groups by wealth measured as total assessed value of property. See Katz, "Social Structure in Hamilton", p. 215.

and XVII. Table XX displays the results of some tests of significance for various combinations of variables, based on the equations of Tables XVI and XVII and on other equations not shown in the tables. Definitions and symbols for the relevant variables are as listed following Table XVII. The variable CHILD is the dependent variable in all of the regressions pertaining to family size. (Note that, because all families in the sample are "normal", total family size is equal to number of children plus two parents. All of the variation in family size thus arises from variation in number of children.)

The analysis makes extensive use of "dummy variables", variables that are set equal to one if particular characteristics are present and to zero if they are not. For example, the variable BH3 is defined to have value one if the husband was born in Scotland, value zero otherwise; RW4 is defined to have value one if the wife's religious denomination is Baptist, zero otherwise. With the exception of variables pertaining to age and number of children, all variables used in the analysis are of the dummy-variable type.[20]

The regressions discussed here thus involve sets of such variables — a set for birthplace, a set for origin, and so on. A well-known problem which arises immediately in the specification of the regressions has to do with the fact that the dummy variables for each set always sum to unity and that some sort of constraints are necessary in order to permit estimation of the equations. Following common practice, we have satisfied this requirement by omitting one dummy variable from each set in the specification of the regression equations. The category represented by the excluded variable in a particular set is then to be interpreted as a reference category and the coefficient of any other variable in the same set is to be interpreted as the difference between the original coefficient of that variable and the original coefficient of the excluded variable. For example, the variable OCCH6 has been omitted from the set of

[20] A general discussion of the use of dummy variables in regression analysis can be found in any one of a number of places. For example, see Arthur S. GOLDBERGER, *Econometric Theory* (John Wiley and Sons, Inc., New York, 1964) and also his *Topics in Regression Analysis* (The Macmillan Company, New York, 1968); J. JOHNSTON, *Econometric Methods* (McGraw-Hill, New York, 1960); Daniel B. SUITS, "Use of Dummy Variables in Regression Equations", *Journal of the American Statistical Association*, Volume 52 (December, 1957), pp. 548-551. A treatment that is particularly relevant to the present analysis is given in Emanuel MELICHAR, "Least-Squares Analysis of Economic Survey Data", *Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 1965*, pp. 373-385.

occupational variables in the regression equations presented in Tables XVI and XVII; if β_6 is the original coefficient of this variable and β_3 is the original coefficient of the variable OCCH3, then the regression coefficients reported in the tables for OCCH3 would be interpreted as estimates of $\beta_3 - \beta_6$.

The usual t-test of regression analysis can be applied to individual variables in the equations to determine whether their estimated coefficients differ significantly from zero. However, in the sort of regression analysis with which we are concerned here, it is often of more interest to consider the variables in groups rather than one at a time and to test the total contribution of each group as a whole. A standard form of the F-test is available for this purpose [21] and has been applied to obtain the results reported in Table XX. It is possible to carry out the test on a large number of groups, both individually and in various combinations. The procedure for doing this involves estimating the regression equations with all variables included and again with all variables except the group being tested. However, thus far, there are available the necessary regression equations for tests on only those groups specified in the table.

Without attempting detailed comment here on the regression results, a few conspicuous points may be noted. First, the high t-ratios in Tables XVI and XVII for the variables relating to wife's age (AGEW and AGEW2) reflect the obviously strong influence of these variables, the quadratic form of the relationship indicating a decline in family size after a certain age of wife as children grow up and leave home. Also important is the apparent negative influence of the presence of other young dependents in the family (ODEP), an influence which may be the result in part of a substitution effect (foster children being a substitute for natural children) and in part a result of the erroneous inclusion of natural children in the "other-dependent" category in households for which the correct interpretation of the Census returns is uncertain.

Husband's and wife's ages are highly-correlated variables and rather than include husband's age in the regression equations, with the attendant difficulty of identifying its separate influence, a variable representing

21 See Melichar, "Least-Squares Analysis of Economic Survey Data", for a discussion of this test.

the difference between husband's and wife's ages (AGEDIF) was introduced. However, this variable made no significant contribution.

The contribution to the explanation of family size of variables pertaining to occupation, religious denomination, birthplace, origin, and area of residence is surprisingly weak, at least on the basis of the early results reported here. The individual t-ratios for these variables are consistently low and the F-tests on various combinations of them failed to produce evidence of any significant relationships, with the possible exception of a relatively weak one between family size and area of residence. (See Table XX.) The lack of evidence of any strong connection between family size and the various socioeconomic variables considered is perhaps one of the most interesting results of the analysis so far.

The overall explanatory power of the regression equations is not high, as evidenced by the relatively low coefficients of determination ($\bar{R}^2$, after adjustment for degrees of freedom). The highest value of $\bar{R}^2$ for any of the equations in Tables XVI and XVII is .2247 and the highest value attainable with any subset of the variables, allowing for possible inclusion of both husband and wife variables in the same equation, was found to be only .2569.[22] This suggests that random elements and family characteristics other than those allowed for in the analysis were the dominant factors in determining family size.

V. — A REGRESSION ANALYSIS OF SCHOOL ATTENDANCE.

A second major use to which the sample is being put is in the analysis of variations in school attendance of persons 10-16 years of age. (In total, there were 298 persons in this age group in the 357 "normal" families of the sample.) As in the case of family size, extensive use is made of dummy variables and, indeed, the dependent variable itself (GTS — see the list of variables following Table XIX) is a dummy variable, with a value of one for a person recorded as attending school

[22] This result and the similar one reported later in connection with the analysis of school attendance were obtained by employing a stepwise regression program involving a selection algorithm which chooses new variables in such a way as to maximize the coefficient of determination at each step subject to the restriction that once a variable has been brought into the regression equation it remains in for all subsequent steps.

and a value of zero otherwise.[23] In addition to area of residence, occupation of father, and religious denomination, origin, and birthplace of both mother and father, the list of explanatory variables includes age of child and its square, a dummy variable for sex of child, a variable for age of father, and another variable representing the total number of children in the family to which the particular 10-16-year-old belongs. Selected regression equations are presented in Tables XVIII and XIX and the results of F-tests based on these and other equations are presented in Table XXI.[24]

The variables representing age, sex, and total number of children in the family appear significant, in varying degree. In some equations there is evidence to suggest a significant relationship between school attendance and occupational category of father and between school attendance and birthplace of one or other of the parents; in others, though, this evidence is lacking. The F-tests that it has been possible to carry out so far suggest that the overall influence of occupation may be significant but additional tests are required in order to establish the presence of the influence more firmly. The values of $\overline{R}^2$ in Tables XVIII and XIX range as high as .3606 and the highest value for any subset of all of the variables considered in the analysis was found to be .3877.

The regression equations with which we have been working so far apply to all 298 persons 10-16 in the sample, males and females together, with a single dummy variable for sex included among the explanatory variables. The implicit assumption is that all parameters of the equation other than the coefficient of the sex variable are identical for males and females. This assumption is of doubtful validity and present plans call for estimating separate equations for males and females in future

[23] The true value of GTS must be zero or one. However, the value of GTS calculated from a regression equation will probably not be one of these values. Following common practice, we may interpret the calculated value as the *probability* that a person with given characteristics will be going to school, on the assumption that the value lies between zero and one. Although there is no guarantee that the calculated value will lie between these bounds, in practice the assumption should be valid for most cases.

[24] When the dependent variable is a dummy variable the assumption of normality of errors in the regression equation is violated and the F and t-tests, which are based on this assumption, must be regarded as approximate only. This is true also in the case of number of children in the family-size regressions since the dependent variable there must take on non-negative discrete values. However, it is presumably more of a problem in the case of a variable restricted to values of zero or one. For a discussion of the problem in the dummy-variable case, see the note by Orley ASHENFELTER in Appendix A of William G. BOWEN and T. Aldrich FINEGAN, *The Economics of Labour Force Participation* (Princeton University Press, 1969), pp. 644-648.

analyses of school attendance. (As shown in Table V, the 10-16-year-olds in the sample are almost evenly divided — 148 boys, 150 girls.) Since it seems quite plausible to argue that the influence of socioeconomic variables on parental attitudes towards the education of children may vary as between male children and female children, this respecification of the regression relationships would seem to be a promising possibility.

VI. — CONCLUDING REMARKS.

As noted in the introduction, we view the project on which this paper reports partly as a pilot study of the research potential of the historical Census microdata and partly as an analytical study of interest in its own right. As a pilot study of the uses of the Census data, one of the principal conclusions would seem to be that the existing microfilm files provide a rich source of information for future work. The files, at least those for 1871, are well organized geographically, and by households within geographic areas; as such, they lend themselves readily to sampling. The techniques applied to two particular problems in this paper, the analysis of variations in family size and of variations in school attendance, are readily applicable to other problems. In particular, regression analysis employing dummy variables is a technique well suited to the analysis of data the largest part of which is qualitative rather than quantitative in nature, as is the case with the historical Census population microdata.

As for the substantive findings with respect to family size and school attendance, it should be stressed that the work to date is incomplete and that the findings are somewhat tentative at this point. The apparent weakness of the relationships between many of the socioeconomic variables tested, on the one hand, and the number of children in a family unit and the propensity of the older children to attend school, on the other, is perhaps surprising.[25] Additional tests, respecification of some of the regression equations, and possibly augmentation of the existing sample may help to sharpen the analysis.

25 Our results, although preliminary, would appear not to provide much support for Katz's hypothesis that family size and school attendance are explained by religion and ethnic origin: "Promoting or retarding chances of individuals in this period [1851-2] were cultural patterns (family size and school attendance) associated with their social, ethnic, or religious affiliation..." KATZ, "Social Structure in Hamilton", p. 211.

There are many possibilities for future analyses of the historical Census data. A matching of data from the agricultural and population Censuses would juxtapose information on family characteristics and information on size and type of farm and type of farming activity. (An experimental matching of 1861 data for Ancaster Township in Wentworth County is now underway; this may be extended to 1851 in order to permit analysis of changes over time.) The analysis of individual farm yields in relation to farm area, labour inputs, family characteristics, and other variables would seem to be feasible.

An analysis of age differences between husband and wife in relation to their ethnic origin, religious denomination, and birthplace is another subject which we propose to investigate with the 1871 Hamilton sample data. (Parenthetically, it may be noted that the overall average age for husbands in the sample is 38.3 years; for wives it is 34.1 years.) A comparison of age of mother (for some well defined age or age range) with age of oldest child could be used to indicate average age of women at first childbirth, for women with various characteristics, subject to proper allowance for the effects of mortality on the probabilities of both the child and the mother surviving so as to be enumerated in subsequent Censuses. We intend to explore this possibility as well.

A matching of Census records from one Census to another poses some tricky technical problems associated with computer linkage of records on the basis of limited information which is subject, in some degree, to errors and inconsistencies. (For example, a man's name might have been recorded as Thomson and his age as 30 in one Census, while ten years later his name might have been recorded as Thompson and his age as 41.) However, the potential for analysis that such a matching would afford would be substantial. Matched data might be used for studies of migration and for the formation of family histories. They might also be used to study the influence on male children who grow up and form their own families of the size of family from which they came. (Do male children from large families tend to have large families or small families ?) For this latter sort of analysis, a linkage of Census records two or more decades apart would probably be required. To be a little more specific, one might take males 10-15 years of age in the 1851 Census and try to locate them in the 1871 Census, where they should be reported as 30-35 years of age. A regression analysis could then be carried out in which family size of 30-35-year-old fathers in 1871 would be related to their personal characteristics (e.g. religious denomination), to the characteristics of their wives and, especially, to the size of their parents' families as recorded in 1851. An area in which migration in and out was relatively small might be chosen so as to minimize the matching problem.

A suggestion that some checks on the accuracy of published Census tabulations might be made by verifying them from the original records has been offered earlier (footnote 12). We mention also the possibility of an analysis of the composition of various occupational categories in terms of religion, origin, birthplace, and age, using an occupational classification procedure based on a socioeconomic index of the sort employed in the present study.

All of the foregoing has to do with the analysis of microdata derived from the historical Census microfilm files. While the analysis of microdata is the principal concern of this paper, it may be worth noting the possibility of using published data as a basis for cross-section regression analysis. Instead of the individual person or family, the unit of observation would be the enumeration district — there were 206 such districts in Canada in the 1871 Census — and instead of dummy variables, variables in the form of proportions would typically be used (proportion of population in a given age group, proportion of population reported as Roman Catholic, and so on). Using this approach, and deriving data solely from published sources, one might be able to analyze variations in net migration rates, average family size, birth rates (based on numbers of persons under 10 years of age, adjusted for mortality, if necessary), and other variables of interest, relating them to various relevant variables representing the population composition and characteristics of the areas. [26] Working with published area data in this way would have the advantage of permitting analyses involving Censuses later than that of 1871, which is the earliest one for which individual enumeration records are available to the public at this time.

[26] For an example of cross-section regression analysis based on Canadian Census areas, see Appendix D of Frank T. DENTON, *An Analysis of Interregional Differences in Manpower Utilization and Earnings* (Economic Council of Canada Staff Study No. 15, April, 1966).

Table I: Birthplaces of Husbands and Wives, Sample of 357 "Normal" Families, Hamilton, 1871

Country of Birth	No. of Husbands	No. of Wives
England	115	90
Ireland	77	80
Scotland	66	55
U.S.A.	23	17
Canada	61	103
Germany	11	9
Other	4	3
	357	357

Note: Canada includes Newfoundland; Germany includes Prussia.
Source: Reel C-616.

Table II: Religious Denominations of Husbands and Wives, Sample of 357 "Normal" Families, Hamilton, 1871

Religious Denomination	No. of Husbands	No. of Wives
Church of England	93	88
Roman Catholic	66	71
Presbyterian	91	88
Baptist	18	20
Methodist	72	73
Other	17	17
	357	357

Note: Presbyterians of all kinds including Church of Scotland, Baptists of all kinds, and Methodists of all kinds.
Source: Reel C-616.

Table III: Origins of Husbands and Wives, Sample of 357 "Normal" Families, Hamilton, 1871

Origin	No. of Husbands	No. of Wives
English	143	129
Irish	100	118
Scottish	86	87
German	13	12
Other European	8	4
African	7	7
Other	—	—
	357	357

Note: The category "Other" is empty in the sample.
Source: Reel C-616.

Table IV: OCCUPATIONS OF HUSBANDS,
SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

Occupational Category	No. of Husbands
1	14
2	24
3	50
4	126
5	143
	357

Note: The professions, occupations, and trades listed in the 1871 Census were located in the appropriate range of Blishen's socioeconomic index of occupations in Canada, as adapted here.

Category 1 includes all occupations lying at 60.00 or over on the modified Blishen scale, category 2 those from 50.00 to 59.99, category 3 those from 40.00 to 49.99, category 4 those from 30.00 to 39.99, and category 5 those under 30.00.

Sources: Reel C-616; *Census of Canada, 1870-71,* Volume II, pp. 250-61; Bernard R. BLISHEN, "A Socio-Economic Index for Occupations in Canada", *Canadian Review of Sociology and Anthropology,* Volume IV (1967), pp. 44-51.

Table V: SCHOOL ATTENDANCE OF 10-16-YEAR-OLD CHILDREN,
SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

	Going to School		Not Going to School		
Age in Years	Boys	Girls	Boys	Girls	Total
10	30	20	2	—	52
11	23	20	1	2	46
12	15	27	4	2	48
13	20	12	7	5	44
14	11	14	12	8	45
15	4	9	11	12	36
16	0	5	8	14	27
Total	103	107	45	43	298

Source: Reel C-616.

Table VI: Cross-Classification of Husbands and Wives by Birthplace, Sample of 357 "Normal" Families, Hamilton, 1871

	Birthplace of Wife							
	England	Ireland	Scotland	U.S.A.	Canada	Germany	Other	Total
Birthplace of Husband								
England	68	17	5	3	21	—	1	115
Ireland	4	48	6	—	19	—	—	77
Scotland	8	2	40	1	15	—	—	66
U.S.A.	3	4	1	8	7	—	—	23
Canada	5	9	2	4	40	—	1	61
Germany	—	—	1	1	—	9	—	11
Other	2	—	—	—	1	—	1	4
Total	90	80	55	17	103	9	3	357

Note and Source: See Table I.

Table VII: Cross-Classification of Husbands and Wives by Religious Denomination, Sample of 357 "Normal" Families, Hamilton, 1871

	Religious Denomination of Wife						
	Church of England	Roman Catholic	Presbyterian	Baptist	Methodist	Other	Total
Religious Denomination of Husband							
Church of England	82	7	1	1	2	—	93
Roman Catholic	1	63	1	1	—	—	66
Presbyterian	2	1	86	—	2	—	91
Baptist	—	—	—	17	—	1	18
Methodist	2	—	—	1	69	—	72
Other	1	—	—	—	—	16	17
Total	88	71	88	20	73	17	357

Note and Source: See Table II.

Table VIII: Cross-Classification of Husbands and Wives by Origin, Sample of 357 "Normal" Families, Hamilton, 1871

	Origin of Wife							
	English	Irish	Scottish	German	Other European	African	Other	Total
Origin of Husband								
English	106	22	14	1	—	—	—	143
Irish	9	85	6	—	—	—	—	100
Scottish	11	9	66	—	—	—	—	86
German	—	—	1	11	1	—	—	13
Other European	3	2	—	—	3	—	—	8
African	—	—	—	—	—	7	—	7
Other	—	—	—	—	—	—	—	—
Total	129	118	87	12	4	7	—	357

Note and Source: See Table III.

Table IX: Cross-Classification of Husbands by Birthplace and Occupational Category, Sample of 357 "Normal" Families, Hamilton, 1871

	Occupational Category					
	1	2	3	4	5	Total
Birthplace						
England	3	9	15	42	46	115
Ireland	3	5	6	20	43	77
Scotland	2	6	10	21	27	66
U.S.A.	1	—	6	10	6	23
Canada	5	4	11	24	17	61
Germany	—	—	1	8	2	11
Other	—	—	1	1	2	4
Total	14	24	50	126	143	357

Note and Source: See Tables I and IV.

Table X: CROSS-CLASSIFICATION OF HUSBANDS BY RELIGIOUS DENOMINATION AND OCCUPATIONAL CATEGORY, SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

	Occupational Category					
	1	2	3	4	5	Total
Religious Denomination						
Church of England	5	9	12	29	38	93
Roman Catholic	1	1	6	23	35	66
Presbyterian	3	8	13	31	36	91
Baptist	—	—	4	8	6	18
Methodist	3	5	11	28	25	72
Other	2	1	4	7	3	17
Total	14	24	50	126	143	357

Note and Source: See Tables II and IV.

Table XI: CROSS-CLASSIFICATION OF HUSBANDS BY ORIGIN AND OCCUPATIONAL CATEGORY, SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

	Occupational Category					
	1	2	3	4	5	Total
Origin						
English	7	10	20	54	52	143
Irish	2	5	8	30	55	100
Scottish	5	9	17	26	29	86
German	—	—	3	8	2	13
Other European	—	—	—	5	3	8
African	—	—	2	3	2	7
Other	—	—	—	—	—	—
Total	14	24	50	126	143	357

Note and Source: See Tables III and IV.

Table XII: CROSS-CLASSIFICATION OF HUSBANDS BY BIRTHPLACE AND AREA OF RESIDENCE, SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

	Area of Residence					
	1	2	3	4	5	Total
Birthplace						
England	19	34	14	32	16	115
Ireland	14	18	17	19	9	77
Scotland	10	17	14	13	12	66
U.S.A.	7	6	5	5	—	23
Canada	7	19	11	14	10	61
Germany	2	2	2	3	2	11
Other	—	3	—	1	—	4
Total	59	99	63	87	49	357

Note and Source: See Table I.

Table XIII: CROSS-CLASSIFICATION OF HUSBANDS BY RELIGIOUS DENOMINATION AND AREA OF RESIDENCE, SAMPLE OF 357 "NORMAL" FAMILIES, HAMILTON, 1871

	Area of Residence					
	1	2	3	4	5	Total
Religious Denomination						
Church of England	18	29	14	21	11	93
Roman Catholic	8	19	15	16	8	66
Presbyterian	11	31	13	19	17	91
Baptist	5	3	5	2	3	18
Methodist	15	14	14	21	8	72
Other	2	3	2	8	2	17
Total	59	99	63	87	49	357

Note and Source: See Table II.

Table XIV: Cross-Classification of Husbands by Origin and Area of Residence, Sample of 357 "Normal" Families, Hamilton, 1871

	Area of Residence					
	1	2	3	4	5	Total
Origin						
English	29	36	19	39	20	143
Irish	14	30	20	25	11	100
Scottish	14	25	14	17	16	86
German	2	4	2	3	2	13
Other European	—	4	3	1	—	8
African	—	—	5	2	—	7
Other	—	—	—	—	—	—
Total	59	99	63	87	49	357

Note and Source: See Table III.

Table XV: Cross-Classification of Husbands by Occupational Category and Area of Residence, Sample of 357 "Normal" Families, Hamilton, 1871

	Area of Residence					
	1	2	3	4	5	Total
Occupational Category						
1	5	2	3	2	2	14
2	4	3	5	6	6	24
3	12	17	8	5	8	50
4	16	40	21	34	15	126
5	22	37	26	40	18	143
Total	59	99	63	87	49	357

Note and Source: See Table IV.

Table XVI: Regression Analysis of Family Size: Selected Equations (Dependent Variable CHILD; 357 Observations)

Independent Variable	Equation FS-1 Coefficient	Equation FS-1 t-ratio	Equation FS-2 Coefficient	Equation FS-2 t-ratio	Equation FS-3 Coefficient	Equation FS-3 t-ratio	Equation FS-4 Coefficient	Equation FS-4 t-ratio	Equation FS-5 Coefficient	Equation FS-5 t-ratio	Equation FS-6 Coefficient	Equation FS-6 t-ratio
Constant Term	3.3040	—	3.1807	—	2.9277	—	3.0493	—	3.3201	—	3.2193	—
AGEW	.1147	9.21	.1151	9.20	.1166	9.26	.1146	8.91	.1134	8.77	.1100	8.51
AGEW2	−.0062	5.21	−.0062	5.20	−.0063	5.29	−.0063	5.25	−.0062	5.07	−.0064	5.25
AGEDIF	.0197	.95	.0225	1.07	.0232	1.10	.0172	.80	.0140	.65	.0125	.58
ODEP	−.5663	2.78	−.6104	2.98	−.6106	2.96	−.5900	2.82	−.5468	2.56	−.5654	2.66
OCCH3	—	—	−.1882	.44	−.1651	.38	−.1432	.33	−.1371	.32	−.0919	.21
OCCH4	—	—	.5027	1.59	.5006	1.56	.4981	1.54	.4620	1.40	.5338	1.61
OCCH5	—	—	.2057	.88	.1921	.81	.2189	.91	.2085	.87	.2247	.94
OCCH7	—	—	−.3539	.66	−.2970	.54	−.3037	.55	−.3307	.59	−.2814	.50
RH6	—	—	—	—	−.0677	.13	−.2557	.48	−.3778	.69	−.5031	.91
RH2	—	—	—	—	.2868	.92	.0606	.18	.0661	.18	.0669	.18
RH3	—	—	—	—	.3211	1.13	−.1296	.34	−.1538	.39	−.1704	.43
RH4	—	—	—	—	.3376	.68	.2457	.46	.2571	.47	.3176	.58
RH5	—	—	—	—	.5582	1.85	.4939	1.60	.4646	1.49	.4294	1.38
BH1	—	—	—	—	—	—	−.2583	.78	−.5024	1.21	−.4954	1.19
BH2	—	—	—	—	—	—	.1925	.54	.1785	.38	.1459	.31
BH3	—	—	—	—	—	—	.5146	1.26	.6636	1.26	.7741	1.46
BH4	—	—	—	—	—	—	.1955	.39	.1748	.34	.0867	.17
BH8	—	—	—	—	—	—	.1309	.22	−.6944	.81	−.7524	.88
OH9	—	—	—	—	—	—	—	—	−.7364	1.22	−.4484	.73
OH2	—	—	—	—	—	—	—	—	−.2420	.46	−.1554	.30
OH3	—	—	—	—	—	—	—	—	−.3800	.66	−.4049	.70
OH4	—	—	—	—	—	—	—	—	.9120	.96	1.0356	1.10
AREA1	—	—	—	—	—	—	—	—	—	—	.3460	.91
AREA2	—	—	—	—	—	—	—	—	—	—	−.0524	.15
AREA3	—	—	—	—	—	—	—	—	—	—	−.3984	1.06
AREA4	—	—	—	—	—	—	—	—	—	—	.4275	1.22
R^2	.2291		.2383		.2472		.2568		.2641		.2813	
$\bar{R}^2$	.2203		.2208		.2187		.2172		.2156		.2247	

Note: t-ratio is ratio of estimated coefficient to its estimated standard error; R^2 is coefficient of determination, not adjusted for degrees of freedom; $\bar{R}^2$ is R^2, adjusted for degrees of freedom; omitted dummy variables are OCCH6, RH1, BH5, OH1, and AREA5.

Table XVII: Regression Analysis of Family Size: Selected Equations
(Dependent Variable CHILD; 357 Observations)

Independent Variable	Equation FS-1		Equation FS-2		Equation FS-7		Equation FS-8		Equation FS-9		Equation FS-10	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Constant Term	3.3040	—	3.1807	—	2.8111	—	2.4663	—	2.4042	—	2.2827	—
AGEW	.1147	9.21	.1151	9.20	.1160	9.26	.1078	8.01	.1076	7.92	.1036	7.63
AGEW2	−.0062	5.21	−.0062	5.20	−.0063	5.27	−.0061	5.02	−.0060	4.91	−.0062	5.10
AGEDIF	.0197	.95	.0225	1.07	.0215	1.02	.0269	1.26	.0271	1.25	.0261	1.21
ODEP	−.5663	2.78	−.6104	2.98	−.5913	2.86	−.5879	2.82	−.5923	2.81	−.6179	2.95
OCCH3	—	—	−.1882	.44	−.1452	.34	−.0057	.01	.0079	.02	.0297	.07
OCCH4	—	—	.5027	1.59	.4879	1.52	.5097	1.59	.4872	1.50	.5481	1.68
OCCH5	—	—	.2057	.88	.1935	.82	.2474	1.04	.2460	1.02	.2700	1.13
OCCH7	—	—	−.3539	.66	−.3063	.56	−.1610	.29	−.1068	.19	−.1244	.22
RW6	—	—	—	—	.4020	.79	.5275	.99	.4691	.86	.3831	.70
RW2	—	—	—	—	.4413	1.42	.4812	1.39	.4269	1.17	.4762	1.31
RW3	—	—	—	—	.5136	1.77	.5034	1.45	.4774	1.20	.5764	1.45
RW4	—	—	—	—	.3471	.72	.6763	1.32	.6251	1.19	.6240	1.19
RW5	—	—	—	—	.6253	2.06	.7034	2.27	.6800	2.16	.6603	2.10
BW1	—	—	—	—	—	—	.4082	1.34	.4788	1.36	.4771	1.36
BW2	—	—	—	—	—	—	.3327	1.01	.2911	.73	.2839	.71
BW3	—	—	—	—	—	—	.4106	1.08	.4576	1.01	.4191	.93
BW4	—	—	—	—	—	—	−.5815	1.07	−.6352	1.09	−.7133	1.23
BW8	—	—	—	—	—	—	.1785	.29	−.2211	.25	−.2229	.25
OW9	—	—	—	—	—	—	—	—	.2755	.40	.4537	.65
OW2	—	—	—	—	—	—	—	—	.1391	.33	.1650	.39
OW3	—	—	—	—	—	—	—	—	.0398	.08	.0134	.03
OW4	—	—	—	—	—	—	—	—	.6243	.69	.6482	.72
AREA1	—	—	—	—	—	—	—	—	—	—	.4773	1.26
AREA2	—	—	—	—	—	—	—	—	—	—	−.1133	.33
AREA3	—	—	—	—	—	—	—	—	—	—	−.2863	.76
AREA4	—	—	—	—	—	—	—	—	—	—	.4391	1.25
R^2	.2291		.2383		.2496		.2599		.2612		.2796	
$\bar{R}^2$	.2203		.2208		.2212		.2205		.2125		.2229	

Note: t-ratio is ratio of estimated coefficient to its estimated standard error; R^2 is coefficient of determination, not adjusted for degrees of freedom; $\bar{R}^2$ is R^2, adjusted for degrees of freedom; omitted dummy variables are OCCH6, RW1, BW5, OW1, and AREA5.

DEFINITIONS OF VARIABLES EMPLOYED IN REGRESSION 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 (sample mean = 34.10).
AGEW2	—	square of AGEW.
AGEDIF	—	difference in years between husband's age and wife's age (husband's age minus wife's age *without* adjustment for means).
ODEP	—	number of children in family other than those with family surname.
OCCH3	—	dummy variable with value 1 if husband's occupation between 50.00 and 59.99, inclusive, in socioeconomic index scale; value 0 otherwise.
OCCH4	—	dummy variable with value 1 if husband's occupation between 40.00 and 49.99, inclusive, in socioeconomic index scale; value 0 otherwise.
OCCH5	—	dummy variable with value 1 if husband's occupation between 30.00 and 39.99, inclusive, in socioeconomic index scale; value 0 otherwise.
OCCH6	—	dummy variable with value 1 if husband's occupation less than 30.00 in socioeconomic index scale; value 0 otherwise.
OCCH7	—	dummy variable with value 1 if husband's occupation 60.00 or greater in socioeconomic index scale; value 0 otherwise.
RH1	—	dummy variable with value 1 if husband's religious denomination is Church of England; value 0 otherwise.
RW1	—	same for wife.
RH2	—	dummy variable with value 1 if husband's religious denomination is Roman Catholic; value 0 otherwise.
RW2	—	same for wife.
RH3	—	dummy variable with value 1 if husband's religious denomination is Presbyterian; value 0 otherwise.
RW3	—	same for wife.
RH4	—	dummy variable with value 1 if husband's religious denomination is Baptist; value 0 otherwise.
RW4	—	same for wife.
RH5	—	dummy variable with value 1 if husband's religious denomination is Methodist; value 0 otherwise.
RW5	—	same for wife.
RH6	—	dummy variable with value 1 if husband's religious denomination is other than specified above; value 0 otherwise.
RW6	—	same for wife.
BH1	—	dummy variable with value 1 if husband born in England; value 0 otherwise.
BW1	—	same for wife.
BH2	—	dummy variable with value 1 if husband born in Ireland; value 0 otherwise.
BW2	—	same for wife.
BH3	—	dummy variable with value 1 if husband born in Scotland; value 0 otherwise.
BW3	—	same for wife.
BH4	—	dummy variable with value 1 if husband born in U.S.A.; value 0 otherwise.
BW4	—	same for wife.

BH5	—	dummy variable with value 1 if husband born in Canada; value 0 otherwise.
BW5	—	same for wife.
BH8	—	dummy variable with value 1 if husband born elsewhere than in countries specified above; value 0 otherwise.
BW8	—	same for wife.
OH1	—	dummy variable with value 1 if husband's origin is English; value 0 otherwise.
OW1	—	same for wife.
OH2	—	dummy variable with value 1 if husband's origin is Irish; value 0 otherwise.
OW2	—	same for wife.
OH3	—	dummy variable with value 1 if husband's origin is Scottish; value 0 otherwise.
OW3	—	same for wife.
OH4	—	dummy variable with value 1 if husband's origin is German; value 0 otherwise.
OW4	—	same for wife.
OH9	—	dummy variable with value 1 if husband's origin is other than specified above; value 0 otherwise.
OW9	—	same for wife.
AREA1	—	dummy variable with value 1 if family lives in St. George's Ward; value 0 otherwise.
AREA2	—	dummy variable with value 1 if family lives in St. Mary's Ward; value 0 otherwise.
AREA3	—	dummy variable with value 1 if family lives in St. Andrew's Ward; value 0 otherwise.
AREA4	—	dummy variable with value 1 if family lives in St. Lawrence Ward; value 0 otherwise.
AREA5	—	dummy variable with value 1 if family lives in St. Patrick's Ward; value 0 otherwise.

Table XVIII: Regression Analysis of School Attendance: Selected Equations (Dependent Variable GTS; 298 Observations)

Independent Variable	Equation SA-1		Equation SA-2		Equation SA-3		Equation SA-4		Equation SA-5		Equation SA-6	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Constant Term	.9808	—	.9276	—	.9228	—	1.0169	—	1.0422	—	1.0784	—
AGEC	—.1250	10.66	—.1261	10.90	—.1264	10.75	—.1264	10.75	—.1255	10.62	—.1244	10.43
AGEC2	—.0188	2.90	—.0178	2.79	—.0177	2.76	—.0170	2.63	—.0172	2.63	—.0178	2.71
SEX	—.0918	2.08	—.0975	2.24	—.0920	2.08	—.0867	1.94	—.0896	1.99	—.0919	2.03
CHILD	.0111	.94	.0186	1.58	.0210	1.75	.0270	2.17	.0272	2.18	.0282	2.24
AGEF	—.0050	1.44	—.0053	1.53	—.0054	1.54	—.0052	1.47	—.0047	1.32	—.0052	1.42
OCCF3	—	—	.2207	2.62	.2214	2.57	.2289	2.66	.2378	2.71	.2432	2.74
OCCF4	—	—	.0418	.63	.0348	.51	.0297	.42	.0327	.46	.0368	.50
OCCF5	—	—	—.0174	.35	—.0205	.40	—.0341	.63	—.0292	.53	—.0206	.37
OCCF7	—	—	.3958	2.92	.4026	2.93	.3657	2.61	.3912	2.62	.4097	2.69
RF6	—	—	—	—	.0072	.06	.0211	.16	.0190	.14	.0504	.37
RF2	—	—	—	—	—.0141	.20	—.0093	.12	—.0074	.08	—.0102	.11
RF3	—	—	—	—	.0203	.32	.0393	.44	.0420	.45	.0501	.51
RF4	—	—	—	—	.0853	.87	.0956	.78	.0957	.75	.0907	.69
RF5	—	—	—	—	—.0461	.71	—.0535	.82	—.0449	.68	—.0341	.51
BF1	—	—	—	—	—	—	—.1615	2.08	—.2147	2.27	—.2253	2.37
BF2	—	—	—	—	—	—	—.1579	1.92	—.2300	1.60	—.2078	1.40
BF3	—	—	—	—	—	—	—.1647	1.73	—.0990	.68	—.0905	.61
BF4	—	—	—	—	—	—	—.1522	1.19	—.1772	1.32	—.1743	1.27
BF8	—	—	—	—	—	—	—.0693	.47	—.0404	.00	—.0413	.00
OF9	—	—	—	—	—	—	—	—	—.1981	1.30	—.2010	1.30
OF2	—	—	—	—	—	—	—	—	.0216	.14	—.0090	.05
OF3	—	—	—	—	—	—	—	—	—.1174	.72	—.1448	.88
OF4	—	—	—	—	—	—	—	—	—.0805	.00	—.1017	.00
AREA1	—	—	—	—	—	—	—	—	—	—	—.0239	.30
AREA2	—	—	—	—	—	—	—	—	—	—	.0364	.47
AREA3	—	—	—	—	—	—	—	—	—	—	—.0080	.08
AREA4	—	—	—	—	—	—	—	—	—	—	—.0557	.76
R^2	.3337		.3696		.3750		.3875		.3926		.3973	
$\bar{R}^2$	.3223		.3499		.3441		.3457		.3416		.3371	

Note: t-ratio is ratio of estimated coefficient to its estimated standard error; R^2 is coefficient of determination, not adjusted for degrees of freedom; $\bar{R}^2$ is R^2, adjusted for degrees of freedom; omitted dummy variables are OCCH6, RF1, BF5, OF1, and AREA5.

Table XIX: Regression Analysis of School Attendance: Selected Equations
(Dependent Variable GTS; 298 Observations)

Independent Variable	Equation SA-1		Equation SA-2		Equation SA-7		Equation SA-8		Equation SA-9		Equation SA-10	
	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio	Coefficient	t-ratio
Constant Term	.9808	—	.9276	—	.9645	—	1.0043	—	.8667	—	.9050	—
AGEC	—.1250	10.66	—.1261	10.90	—.1263	10.78	—.1286	11.01	—.1297	11.14	—.1289	10.94
AGEC2	—.0188	2.90	—.0178	2.79	—.0177	2.76	—.0186	2.86	—.0188	2.89	—.0192	2.94
SEX	—.0918	2.08	—.0975	2.24	—.0967	2.18	—.0896	2.01	—.0896	2.01	—.0922	2.04
CHILD	.0111	.94	.0186	1.58	.0210	1.75	.0271	2.16	.0328	2.57	.0334	2.59
AGEF	—.0050	1.44	—.0053	1.53	—.0057	1.64	—.0039	1.11	—.0029	.82	—.0030	.84
OCCF3	—	—	.2207	2.62	.2039	2.34	.1645	1.86	.1371	1.53	.1362	1.50
OCCF4	—	—	.0418	.63	.0195	.28	.0381	.54	.0394	.56	.0376	.52
OCCF5	—	—	—.0174	.35	—.0246	.48	—.0359	.68	—.0203	.38	—.0138	.26
OCCF7	—	—	.3958	2.92	.3914	2.84	.3017	2.10	.2668	1.62	.2689	1.62
RM6			—	—	—.0086	.07	.0008	.01	—.0500	.38	—.0315	.24
RM2	—	—	—	—	—.0649	.88	—.0744	.94	—.1155	1.37	—.1336	1.54
RM3	—	—	—	—	.0147	.22	—.0098	.12	—.0528	.62	—.0732	.82
RM4	—	—	—	—	.0455	.45	.0179	.16	.0193	.18	.0128	.12
RM5	—	—	—	—	—.0671	.98	—.0940	1.36	—.1147	1.61	—.1062	1.47
BM1	—	—	—	—	—	—	—.1976	2.80	—.1392	1.63	—.1385	1.60
BM2	—	—	—	—	—	—	—.1584	2.12	—.2093	2.18	—.1928	1.92
BM3	—	—	—	—	—	—	—.1204	1.38	—.2153	2.00	—.2063	1.90
BM4	—	—	—	—	—	—	—.0672	.44	.0944	.56	.1042	.61
BM8	—	—	—	—	—	—	—.0632	.47	—.1113	.34	—.0918	.28
OM9	—	—	—	—	—	—	—	—	—.2323	1.34	—.2176	1.24
OM2	—	—	—	—	—	—	—	—	.1455	1.36	.1427	1.30
OM3	—	—	—	—	—	—	—	—	.1887	1.66	.1970	1.71
OM4	—	—	—	—	—	—	—	—	.1380	.37	.1100	.30
AREA1	—	—	—	—	—	—	—	—	—	—	—.0505	.65
AREA2	—	—	—	—	—	—	—	—	—	—	—.0024	.03
AREA3	—	—	—	—	—	—	—	—	—	—	—.0227	.26
AREA4	—	—	—	—	—	—	—	—	—	—	—.0786	1.10
R^2	.3337		.3696		.3771		.3964		.4101		.4144	
$\bar{R}^2$	.3223		.3499		.3462		.3552		.3606		.3559	

Note: t-ratio is ratio of estimated coefficient to its estimated standard error; R^2 is coefficient of determination, not adjusted for degrees of freedom; $\bar{R}^2$ is R^2, adjusted for degrees of freedom; omitted dummy variables are OCCH6, RM1, BM5, OM1, and AREA5.

Table XX: F-TEST SIGNIFICANCE LEVELS FOR SELECTED GROUPS OF VARIABLES IN REGRESSION EQUATIONS FOR FAMILY SIZE

Group of Variables Tested	Equations Involving O, B, R for Husband but not Wife (Table XVI)			Equations Involving O, B, R for Wife but not Husband (Table XVII)			Equations Involving O, B, R for both Husband and Wife (Not Shown in Tables)		
	1% Level	5% Level	10% Level	1% Level	5% Level	10% Level	1% Level	5% Level	10% Level
All Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variables relating to AREA, O, B, R, OCC only	No	No	No	No	No	No	No	No	No
Variables relating to AREA, O, B, R only	No	No	No	No	No	No	No	No	No
Variables relating to AREA, O, B only	No	No	No	No	No	No	—	n.a.	—
Variables relating to AREA, O only	No	No	No	No	No	No	—	n.a.	—
Variables relating to AREA only	No	No	Yes	No	No	Yes	—	n.a.	—
Variables relating to O, B, R of *husband* only	—	n.a.	—	—	n.a.	—	No	No	No
Variables relating to O, B, R of *wife* only	—	n.a.	—	—	n.a.	—	No	No	No

Note: Definitions of symbols: AREA — area of residence; O — origin; B — birthplace; R — religious denomination; OCC — occupation (of husband); n.a. — "not available" or "not applicable".

Table XXI: F-TEST SIGNIFICANCE LEVELS FOR SELECTED GROUPS OF VARIABLES IN REGRESSION EQUATIONS FOR SCHOOL ATTENDANCE

Group of Variables Tested	Equations Involving O, B, R for Father but not Mother (Table XVIII)			Equations Involving O, B, R for Mother but not Father (Table XIX)			Equations Involving O, B, R for both Father and Mother (Not Shown in Tables)		
	1% Level	5% Level	10% Level	1% Level	5% Level	10% Level	1% Level	5% Level	10% Level
All variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variables relating to AREA, O, B, R, OCC only	No	No	No	No	Yes	Yes	No	No	Yes
Variables relating to AREA, O, B, R only	No	No	No	No	No	No	No	No	No
Variables relating to AREA, O, B only	No	No	No	No	No	No	—	n.a.	—
Variables relating to AREA, O only	No	No	No	No	No	No	—	n.a.	—
Variables relating to AREA only	No	No	No	No	No	No	—	n.a.	—
Variables relating to O, B, R of *father* only	—	n.a.	—	—	n.a.	—	No	No	No
Variables relating to O, B, R of *mother* only	—	n.a.	—	—	n.a.	—	No	No	No

Note: Definitions of symbols: AREA — area of residence; O — origin; B — birthplace; R — religious denomination; OCC — occupation (of father); n.a. — "not available" or "not applicable".

DEFINITIONS OF VARIABLES EMPLOYED IN REGRESSION ANALYSIS OF SCHOOL ATTENDANCE

GTS — dummy variable with value 1 if child attending school; value 0 otherwise.

AGEC — age of child expressed as difference in years from sample mean (sample mean = 12.67).

AGEC2 — square of AGEC.

SEX — dummy variable with value 1 if child is male; value 0 if female.

CHILD — number of children in family with family surname.

AGEF — age of father (*without* adjustment for mean).

OCCF3, OCCF4, OCCF5, OCCF6, OCCF7 — dummy variables representing occupation of father defined in same way as corresponding dummy variables for husband in regression analysis of family size.

RF1, RF2, RF3, RF4, RF5, RF6 — dummy variables representing religious denomination of father defined in same way as corresponding dummy variables for husband in regression analysis of family size.

RM1, RM2, RM3, RM4, RM5, RM6 — dummy variables representing religious denomination of mother defined in same way as corresponding dummy variables for wife in regression analysis of family size.

BF1, BF2, BF3, BF4, BF5, BF8 — dummy variables representing birthplace of father defined in same way as corresponding dummy variables for husband in regression analysis of family size.

BM1, BM2, BM3, BM4, BM5, BM8 — dummy variables representing birthplace of mother defined in same way as corresponding dummy variables for wife in regression analysis of family size.

OF1, OF2, OF3, OF4, OF9 — dummy variables representing origin of father defined in same way as corresponding dummy variables for husband in regression analysis of family size.

OM1, OM2, OM3, OM4, OM9 — dummy variables representing origin of mother defined in same way as corresponding dummy variables for wife in regression analysis of family size.

AREA1, AREA2, AREA3, AREA4, AREA5 — dummy variables representing area of residence defined in same way as in regression analysis of family size.