### A METHOD OF PROJECTING THE NUMBER OF HOUSEHOLDS IN SMALL AREAS

## By: William Hodgkinson, Jr., American Telephone and Telegraph Company

The purpose of this paper is to outline a methodology for breaking down a given projection of households for an area into individual household projections for subdivisions of that area. It assumes that population projections are available for each subdivision. Household projections are needed in marketing studies and in various fields of long-term planning. Households, of course, are the natural market for long lists of commodities and services. The procedures to be described do not constitute the basic type of research which is involved, say, for estimates of the future number of households in the United States. Rather, they are such as to make use of such estimates.

A household is defined in the Population Census as one person or a group of not necessarily related persons occupying a dwelling unit. In the Housing Census there are data on occupied dwelling units, and since the same definition of a dwelling unit is used in both Censuses, there is an identity of concept. This identity unfortunately did not carry over into the tabulations for 1950 because of necessarily separate processing, but the differences between the two are relatively small. The critical part of the definition is bound up in the concept of "dwelling unit." In the main a dwelling unit is a house, an apartment, trailer, etc.

The population census in this country has always been taken on a <u>de jure</u> basis, so that wherever they might be enumerated, persons are attributed to their permanent residences. One sees a reflection of this fact in the counts of "non-resident dwelling units" in the housing census.

The Bureau of the Census uses a dichotomy for classifying the population: everybody is assumed to live either in a household or in a quasi-household. The essential point is that a quasi-household is <u>not</u> a household. The population in quasi-households is made up of the institutional population (hospitals, prisons, homes for the aged, etc.) and "other." This latter group covers a wide variety of abodes those who live in all but the smallest lodging houses, in hotels, those quartered in lumber and mining camps, and those members of the resident armed forces who live in barracks. The following are the figures as of April 1, 1950:

Type of Residence	Population	of Total
Total	150,697,361	100.00
Households	145,030,888	96.24
Quasi-households	ر جور,666,473	3.76
Institutions	1,566,846	1.04
Other	4,099,627	2.72

If estimates are available of the future population in an area, and if in addition one knows for that area the future ratios of population to households (often referred to as "population per household," or "persons per household"), the problem of this paper is obviously solved: these ratios when divided into the population give the number of households. Any procedure which attempts to derive household from population estimates must concern itself with this relation.

Any error in such ratios is of course transmitted to the household estimates as a reciprocal. Thus a 10% underestimate of the ratio produces an 11.1+% overestimate of households, etc. It is significant, therefore, to have some idea of possible variations in the ratio. To this end a distribution was made of all 3,102 counties (or equivalents) in 1950 according to the magnitude of the ratio of total population living in households to the number of households. (These ratios were published to two places of decimals in Table 42 of the individual State parts of Volume II of the 1950 Census of Population.) Data derived from this compilation are presented in Table 1. It is to be noted that the distribution is of counties and not of households, since the emphasis here is on variations in the averages between counties. The mean of the averages for all counties is 3.54 (compared to the mean size of all households in the nation of 3.38). The range in the distribution is from 2.19 persons for Esmeralda County, Nevada, to 5.12 persons for Leslie County, Kentucky, or 2.93 persons. This range is over 80% of the mean and 7.7 times the standard deviation of 0.38. The coefficient of variation is 10.7% for the nation but is naturally less in all divisions except the extensive Mountain division. Time does not permit more than mere mention of the well-known regional variations in the ratio. It is evident, if proof were needed, that account must be taken of area variations in average household size if projections of the number of households are to do more than indicate a gross order of magnitude.

Even if projections are given of average household size for the nation or other large area, it remains a formidable task to make direct projections of average household size for a small area. Population and households, the numerator and denominator of the measure, are influenced by different factors or by the same factors in varying degree.

The problem is simplified by projecting what is termed herein, "relative household size." This is a ratio of two ratios: the numerator is the ratio of total population to the number of households for the small area, while the denominator is the same ratio for the large area for which projections of households are available. Two characteristics of this ratio are to be noted (see equation (1) below). In the first place, extensive cancellations take place upon its formation; secondly, the ratio is also identical with that obtained by dividing the proportion of the large area's population contained in the small area by the proportion of the large area's households contained in the small area.

The cancellations assist in the analysis of a given situation and hence permit more intelligent forecasting by clearing away extraneous detail which masks the essential point. A notable example of cancellation is the historic decline in average household size which has persisted off and on from the first decennial census in 1790. It is of course cancelled out since it tended to affect all areas. A further point not to be overlooked is that any breaks in the continuity of the Census record due to definitional or procedural changes also tend to cancel, since both large and small areas are similarly affected. The varying treatment of quasi-households is an example.

What is left after the cancellations is a record over time of average household size in the small area as a proportion of that in the large area. This of course is precisely what is needed to estimate average household size in the small area, since by assumption the average size of households in the large area is known. The ratios of relative household size tend to fluctuate in time about unity, since any residual trend represents a divergence in average household size between the two areas. This fact suggests that the <u>widening</u> of such divergencies (the ratio moving further from unity) can be of only limited duration. Plotted as a time series on a chart, the line for unity will, in effect, operate as a magnet. The following are three examples of exceptional deviation from unity which has persisted for a long time. The reasons, however, seem fairly evident:

	Aroostook County, Me.	Dukes County, Mass.	Nantucket County, Mass
1890	1.24	.73	.72
1900	1.24	.70	.70
1910	1.23	.71	.71
1920	1.23	.78	.72
1930	1.27	.81	.83
1940	1.23	.87	.89
1950	1.19	.85	.85

The measure, relative household size, has served its purpose as such, when, from its past performance and what other considerations come into play, it is projected into the future.

Simple graphical projections seem indicated, with readings made from the charts. It is at this point that the second property of this ratio which was mentioned above, becomes of service. For if we divide the proportion of large area population in each of the component small areas by the figure for relative household size, we obtain the proportion of households contained in each small area. The sum of these proportions will be precisely unity at any census date. For projected values, however, this sum will deviate from unity because each projected relative household size was determined independently. Hence a small adjustment is required. Application of the proportions, adjusted to add to unity, is the final step, yielding a projection of households for each component of the large area. The algebraic relations will now be set out, together with a practical computing procedure.

For simplicity of notation a subscript i is suppressed in writing each lower case letter. Thus "q" stands for " $q_1$ ". The letter i denotes a particular county, the counties always to be taken in the same sequence, i = 1, 2, . . ., N. N is the number of counties in the state. Summations, denoted by S(), are all from i = 1 to i = N. In general a different set of the relations below will hold for each point in time.

p = population,	h = households,
P = S(p),	H = S(h),
q = p/P,	r = h/H,
S(q) = 1;	S(r) = 1.

It is assumed that H and the N values of q are given. Relative household size, denoted by s, is defined as:

(1) 
$$s = \frac{p/h}{P/H} = \frac{p/P}{h/H} = \frac{q}{r}$$
, so that  
(2)  $r = q/s = h/H$ .

Summing equation (2) we have:

(3) S(r) = S(q/s) = (1/H)S(h) = 1.

At any census equation (3) must hold exactly. If for the projection period the values of s are determined independently of one another, however, equation (3) will not hold exactly. Writing as primed characters those quentities subject to adjustment and assuming that the adjustment can be spread proportionally over all counties, we have, with K a constant:

(4)  $(1/K)S(r^{*}) = (1/K)S(q/s^{*}) = (1/KH)S(h^{*}) = 1.$ 

Multiplying each member by K we see that

(5) 
$$K = S(q/s^{*})$$
.

Equating the third member of (3) to the second member of (4), and also equating the second members of (3) and (4), we obtain:

### 222

- (6)  $h = (H/K)(q/s^{t})$ , and
- (7)  $s = Ks^{*}$ .

Practical computation:

- a) Read values of s' from charts (or otherwise) and
- b) divide them into corresponding (given) values of q.
- c) The sum of these quotients is K, the adjustment factor.
- d) Divide H by K and multiply this constant quotient into each quotient obtained in b), thus deriving the adjusted number of households in each county.
- e) Compute adjusted values of s by multiplying each s' by K. Effect of adjustment can then be gauged.

Reference has been made herein to two different ways of computing average household size in an area: (1) to relate total population to households, and (2) to relate only the population living in households to households. For any one date it is obviously imperative that average household size should be computed on the same basis for both the small areas, for which household estimates are to be made, as well as for the large area for which household estimates are already available. The use of either base (preserving consistency at any date) is permissable if it does not produce distortion in relative household size.

Distortion can be caused by variations in the relative balance between population living in households as opposed to population in quasihouseholds. The danger is of a disproportionate change in quasi-household population in the small area.

The following table indicates the magnitude of average household size using both bases for different types of area.

Ratios of Population by Type of Residence to Total Households in Area, 1950

	In House-	In Quasi-Households				
Area	holds	tutions	<u>Other</u>	Total	TOTAL	
Urbanized Areas:						
Central cities	3.186	.025	.138	.163	3.349	
Urban fringes	3.400	.021	.070	.091	3.491	
Total	3.248	.024	.118	.142	3.390	
Other urban	3.224	.036	.107	.142	3.367	
Total urban	3.242	.027	.115	.142	3.384	
Rural nonfarm	3.454	.092	.088	.180	3.634	
Rural farm	3.984	-	.012	.012	3.996	
Total rural	3.667	.055	.057	.112	3.780	
U. S.	3.384	.037	.096	.132	3.516	

Table 2 gives an illustration of the first part of the procedure for estimating the number of households in each geographic division. The data through the next to the last column should be clear. The last column gives relative household size. The United States average in the next to last column becomes the denominator for the corresponding date for each of the nine divisions.

The accompanying charts show fluctuations from 1870 in relative household size, with possible projections indicated for the period after 1950.

Readings from these charts provide all the data needed to produce estimates of future households in these areas except for the required population projections and the national household projection. These charts with their accompanying tables are included to give some idea of the behavior of these ratios. It is also worth noting that the measures of relative household size need not be changed if a different household projection should be substituted for the large area, or if different general population projections should be used.

# RELATIVE HOUSEHOLD SIZE



## VARIATIONS IN MEAN HOUSEHOLD SIZE BETWEEN COUNTIES

ъ.	n	5	c
т	7	2	U

	Number of Counties		Mean*		Standard Deviation		Coefficient of Variation					
Area	Total	Metro.	Non- Metro.	Total	Metro.	Non- Metro.	Total	Metro.	Non- <u>Metro.</u>	<u>Total</u>	<u>Metro.</u>	Non- <u>Metro.</u>
United States	3,102	284	2,818	3.54	3.51	3.56	.38	.22	.39	10.7	6.3	1 <b>0.9</b>
Regions												
Northeast North Central South West <u>Divisions</u>	217 1,056 1,416 413	77 82 95 30	140 974 1,321 383	3.41 3.36 3.76 3.33	3.51 3.50 3.50 3.49	3.40 3.36 3.78 3.35	.17 .25 .38 .36	.13 .16 .25 .20	.19 .25 .38 .36	5.1 7.3 10.0 10.7	3.8 4.6 7.3 5.8	5.6 7.5 10.0 10.8
New England Middle Atlantic East North Central West North Central South Atlantic East South Central West South Central Mountain Pacific	67 150 436 620 582 364 470 280 133	20 57 56 26 54 17 24 9 21	47 93 380 594 528 347 446 271 112	3.39 3.41 3.34 3.37 3.89 3.87 3.51 3.43 3.13	3.51 3.51 3.55 3.53 3.57 3.39 3.36 3.08	3.39 3.41 3.34 3.38 3.92 3.89 3.52 3.44 3.14	.19 .16 .17 .29 .37 .34 .29 .38 .18	.20 .15 .14 .17 .25 .21 .27 .16 .16	.22 .17 .17 .29 .36 .34 .29 .38 .19	5.6 4.8 5.0 8.5 9.4 8.8 8.2 11.0 5.8	5.7 4.2 4.1 5.3 7.0 5.8 7.9 4.7 5.1	6.6 5.0 5.1 8.6 9.1 8.7 8.1 11.1 5.9

\* Unweighted means of county averages of persons per household.

Note: Except for New England, metropolitan counties are those included in <u>Census of Population: 1950</u>, Vol. II, Part 1, Table 26. For the New England States, they are the counties (rather than towns and cities) given in <u>County and City Data Book: 1952</u>, Appendix Table D-1. Berkshire County, Mass., was also regarded as metropolitan.

	<u>Illustrat</u>	ive Computation	for Geographic	<u>Divisior</u>	15
	<b>Com</b>			Pop./Ho	buseholds
Area	Date	<b>Population</b>	<u>Households</u>	Area	of U.S.
<u>United</u> <u>States</u>	1870 1880 1990 1910 1920 1920 1930 1940	38,558,371 50,155,783 62,622,250 75,994,575 91,972,266 105,710,620 122,775,046 131,669,275 150,697,361	7,579,363 9,945,916 12,690,152 15,963,965 20,255,555 24,351,676 29,904,663 34,948,666 42,857,335	5.087 5.043 4.935 4.760 4.341 4.341 4.106 3.768 3.516	100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0
<u>New</u> England	1870 1880 1990 1900 1910 1920 1930 1940 1950	3,487,924 4,010,529 4,700,745 5,592,017 6,552,681 7,400,909 8,166,341 8,437,290 9,314,453	740,271 872,075 1,034,262 1,236,929 1,464,942 1,703,812 1,981,499 2,208,351 2,616,797	4.712 4.599 4.545 4.521 4.473 4.344 4.121 3.821 3.559	92.6 91.2 92.1 95.0 98.5 100.1 100.4 101.4 101.2
<u>Middle</u> <u>Atlanti</u>	1870 <b>c</b> 1880 1890 1900 1910 1920 1930 1940 1950	8,810,806 10,496,878 12,700,800 15,454,678 19,315,892 22,261,144 26,260,750 27,539,487 30,163,533	1,757,223 2,151,666 2,677,980 3,320,337 4,235,675 5,085,080 6,374,380 7,294,488 8,622,808	5.014 4.878 4.743 4.655 4.560 4.378 4.120 3.775 3.498	98.6 96.7 96.1 97.8 100.4 100.9 100.3 100.2 99.5
<u>East</u> <u>North</u> <u>Central</u>	1870 1880 1890 1900 1910 1920 1930 1940 1950	9,124,517 11,206,668 13,471,840 15,985,581 18,250,621 21,475,543 25,297,185 26,626,342 30,399,368	1,757,835 2,213,547 2,820,912 3,488,620 4,214,820 5,143,913 6,362,823 7,290,676 8,829,542	5.191 5.063 4.776 4.582 4.330 4.175 3.976 3.652 3.443	102.0 100.4 96.8 96.3 95.4 96.2 96.8 96.9 97.9
<u>West</u> <u>North</u> <u>Central</u>	1870 1880 1890 1900 1910 1920 1930 1940 1950	3,856,594 6,157,443 8,890,439 10,347,423 11,637,921 12,544,249 13,296,915 13,516,990 14,061,394	722,476 1,175,470 1,777,693 2,143,928 2,592,069 2,957,849 3,317,881 3,698,161 4,153,167	5.338 5.238 5.201 4.826 4.490 4.241 4.008 3.655 3.386	104.9 103.9 101.3 101.4 98.9 97.7 97.6 97.0 96.3

# RELATIVE HOUSEHOLD SIZE

TABLE 2

- 2 -

	Area	Census Date	Population	Households	Pop./Ho in Area	useholds Percent of U.S.
•	<u>South</u> Atlantic	1870 1880 1890 1900 1910 1920 1930 1940 1950	5,853,610 7,597,197 8,857,920 10,443,480 12,194,895 13,990,272 15,793,589 17,823,151 21,182,335	1,132,621 1,463,361 1,687,767 2,078,603 2,539,270 2,991,628 3,511,860 4,291,395 5,540,342	5.168 5.192 5.248 5.024 4.803 4.676 4.497 4.153 3.823	101.6 103.0 106.3 105.5 105.8 107.7 109.5 110.2 108.7
	<u>East</u> <u>South</u> <u>Central</u>	1870 1880 1900 1910 1920 1930 1940 1950	4,404,445 5,585,151 6,428,770 7,547,757 8,409,901 8,893,307 9,887,214 10,778,225 11,477,181	833,694 1,053,186 1,217,097 1,520,339 1,796,832 1,977,381 2,273,359 2,626,791 2,991,927	5.283 5.303 5.282 4.965 4.498 4.498 4.349 4.103 3.836	103.9 105.2 107.0 104.3 103.1 103.6 105.9 108.9 109.1
	<u>West</u> <u>South</u> <u>Central</u>	1870 1880 1900 1910 1920 1930 1940 1950	2,029,965 3,334,220 4,544,123 6,532,290 8,784,534 10,242,224 12,176,830 13,064,525 14,537,572	408,717 644,364 854,023 1,287,871 1,827,105 2,242,810 2,868,262 3,386,552 4,103,354	4.967 5.322 5.072 4.5228 4.5245 4.5245 3.54 3.54 3.54 3.54 3.54 3.54 3.54 3.	97.6 102.6 107.8 106.6 105.9 105.2 103.4 102.4 102.8
	<u>Mountain</u>	1870 1880 1890 1900 1910 1920 1930 1940 1950	315,385 653,119 1,156,326 1,674,657 2,633,517 3,336,101 3,701,789 4,150,003 5,074,998	73,597 144,891 239,940 367,932 614,656 803,853 914,408 1,126,190 1,446,725	4.285 4.508 4.552 4.285 4.150 4.048 3.685 3.508	84.2 897.6 975.6 95.6 98.6 98.8 97.8 99.8
	<u>Pacific</u>	1870 1880 1990 1900 1910 1920 1930 1940 1950	675,125 1,114,578 1,871,287 2,416,692 4,192,304 5,566,871 8,194,433 9,733,262 14,486,527	152,929 227,356 380,478 519,406 970,186 1,445,350 2,300,191 3,026,062 4,552,673	4.415 4.902 4.918 4.653 4.321 3.852 3.563 3.216 3.182	86.8 97.7 99.8 97.2 88.7 86.8 85.4 90.5

TABLE 2

For a discussion of comparability of data see U.S. Census of Population: 1950, Vol. IV, Special Reports, Part 2, Chapter A, General Characteristics of Families, pp. 8 - 9. The data herein follow Table B, p. 8, of this reference. Table B, however, is for the U.S. as a whole and covers only the period 1890 - 1950.

Household data taken from the following Population volumes:

	vorune	
Data for	for	Reference
1870	1890	Part I, Table 87, p. 914;
1880 - 1900	1900	Vol. II, Part II, Table LXXXVIII,p.clx;
1910 - 1920	1920	Vol. II, cap. XIV, Table 2, p. 1267;
1930	1930	Vol. VI, Table 40, p. 33;
1940	1940	Vol. IV, Part 1, Table 51, pp. 162 - 163;
1950	1950	Vol. II, Part 1, Table 47 for U.S.
		and Table 22 of each State Part.

Prior to 1950, except for 1900 and 1930, the data labeled "households" are for "families" and necessarily include the relatively small number of quasi families. In 1900 and 1930, however, the data are for "private families", (excluding quasi families) and are generally comparable with "households" in the 1950 Census.

Population data taken from the following volumes:

	VOLUNC	
Data for	for	Reference
<b>1870 - 1</b> 950		
(except 1890)	1950	Vol. II, Part 1, Table 6;
1890	1900	Vol. II, Part II, Table LXXXVIII, p.clx.

Persons in Indian territory and on Indian reservations were enumerated for the first time in 1390. No "family" data, how-ever, are available and hence the 325,464 persons are excluded.