POPULATION ESTIMATING AND THE CENSUS OF 1970, THE CALIFORNIA EXPERIENCE

Walter P. Hollmann, State of California, Department of Finance

I. CITIES

After the years of working in darkness the California population estimators were at last able to confront their work systematically with the census. Perhaps it was not total darkness, since The State of California has a limited census program of its own. A series of test estimates were prepared for practically all of the cities participating in a program of local estimates designed for the redistribution of tax monies. The test estimates were compared with the results of the census in terms of a number of common elements:- total population, household population, housing units of several types, households and persons per household.

For the purposes of this paper, and by way of illustration, estimates for the cities within Orange and Santa Clara counties and the City of San Diego were scrutinized. One of the selected counties lies in the north, about forty miles south of San Francisco, the other in the south, adjacent to Los Angeles. Both have experienced unusual growth and they are served by different major power companies. One or more cities in each of the counties operates a municipal electric utility, while electric service is provided to the City of San Diego by a third major stockholder-owned utility.

The estimates were all prepared by a housing unit method, for two rather critical reasons, which, on reflection, are the same...no data options. Since the purpose of the program is the distribution of tax money, a minimum time must elapse between the reference date and the date at which the new estimate can be used for entitlement calculations. Utility records and construction statistics are available on a nearly current basis. Other types of estimates, which I will discuss later, require data that are far too slow in coming and some are all but impossible to obtain for incorporated cities. Secondly, the data usually available for cities do not permit estimates by other methods.

The housing unit method, for those unfamiliar with this type of effort, is very simple. Total housing units are estimated by adding to those recorded in the latest census new construction and annexations and subtracting demolitions and the rare disincorporations. Households or occupied housing units are estimated in a similar fashion by using residential electric customer increase since the census year. Group quarters population, handled separately, is based upon local data while average household size is artfully increased or decreased from the benchmark on the basis of observed trend, or type of construction or occasionally upon school enrollment statistics, if the latter are collected for the incorporated area. House trailer or mobilhome population is handled separately from other household population because of the popularity of electric master metering in mobilhome parks. For those who prefer formulae:

$$HU_t = HU_o + HUC_{o-t} - HUD_{o-t} + HUG_{o-t} - HUL_{o-t},$$

where HU_t is the estimate of total Housing Units (less mobilhomes) at the time of the estimate, and

HU_o = Units at the benchmark date
HUC_{o-t} = Housing Units constructed between o and t
HUD_{o-t} = Housing Units demolished between o and t
HUG_{o-t} = Housing Units annexed or gained between o and t
HUL_{o-t} = Housing Units de-annexed or lost between o and t

Construction includes units moved in; demolition includes those moved out.

$$HH_t = HH_o + \Delta REC_{o-t} + \Delta MMHH_{o-t}$$

where HH_t and HH_o represent households (less those in mobilhomes) at the time of the estimate and the benchmark respectively

ΔREC_{o-t} = change in Residential Electric Customers, suitably corrected for lags in recording annexations

$\Delta MMHH_{o-t}$ = change in master metered housing units (other than mobilhomes) to which an estimated vacancy has been applied.

Finally, $P_t = HH_t \times PP_t + TrP_t + GQP_t$

or, the total estimated population equals the product of the estimated households and the average household size plus the populations in mobilhomes and in group quarters.

A comparison of the total estimated populations with the census leads to the inescapable conclusion that the method contains a sharp upward bias. Table 1, following shows the percentage difference from the corresponding statistic in the Census of 1970 of each of five

elements related to housing unit estimates from their estimated values. In Orange County the average error of the population estimates was 2.33 percent high; eleven cities were estimated to have 17,900 more inhabitants than the census reported while another 11 were estimated to have 8,200 persons fewer than were reported. Among sixteen cities the estimates fell short by 6100 households while among the remaining six, 1900 too many were estimated. The method underestimated households but its failure (or rather the estimators' failure) to perceive the full effect of the drop in household size resulted in high estimates of population in spite of low numbers of households. Only five of the 22 were estimated low, the remaining 17 were high. In Santa Clara County the total population was high by an aggregate of 33,500 in eight cities, low by 6900 in seven cities. Households were estimated high by 4200 in five cities, and low by 4400 in ten; characteristically the average household size was underestimated in only two of the cities.

It is clear that at least within our hands, the method carries a serious upward bias and much more study is required not only of the quality of the statistics used but also of the determinants of changing household size. Were the Bureau to provide persons per household by units in structure one might be able to estimate household size with more precision by use of the change in housing composition. Which gives rise to another concern with the Census. The column headed "Singles" in Table 1 presents the percent error in our estimates of single housing units in the housing inventories of the cities in Orange and Santa Clara counties. The results are dismal; 20 of the 25 cities on the list were underestimated, and seriously. These were cities within the mail-out area and a study of the wording of the question on units-in-structure suggests overreporting of singles. If this is so, substantial doubt is cast upon the units-in-structure data.

In a recent set of provisional estimates for California counties, the housing unit method was used but with a more vigorous attempt to establish an average household size on the basis of past trends in the variable itself and the partial indicator, school-enrolled children per household. Although we can not know if we were more accurate, the fact that the results of the population calculation by this method for the 58 counties of California seemed to agree more closely with other methods than in the past was encouraging.

II. COUNTIES

Each summer, estimates of county populations in California are prepared for the current July 1 on a provisional basis and are revised for the preceding year. The methods used recently were based upon experimental work performed during the decade 1960-70; four were used for provisional estimates and six for revisions.

Two of the methods, designated Department of Finance Regression I and II are component techniques. The equations which appear in the handout are

$$Z = 1.103 + .734X + .374Y$$

$$Z = .172 + .215X + .921Y$$

and were derived from 1950-60 experience. The independent variables are, X equal to the percent change in occupied housing units and percent change in residential electric customers, respectively, in the two equations and Y equal to the percent change in school enrollment in grades 3 through 8 over the enrollment in grades 2 through 7 in the prior year in both equations, for each of the ten years, multiplied geometrically. The dependent variable Z represents net civilian migration as a percent of civilian population at the beginning of the year. It is apparent that the second equation weights school enrollment change more heavily than the first. In the post-censal estimating period, both equations perforce use the change in residential electric customers since estimates of household change are based upon that available statistic.

The housing unit method, described earlier, is the third method, while Census Bureau Component Method II is the fourth.[1] Two methods which present problems in timely data acquisition are the Composite Method of Bogue and Duncan[2] and the Ratio Correlation Method described by Crosetti and Schmitt.[3] The former method generates broad age groups of the population to be estimated from symptomatic indicators for which rates can be assumed. The latter is used to divide the total population of the State among the counties on the basis of the relationship of shares of seven independent variables to population shares--births, deaths, elementary school enrollment, fee-paid auto registrations, income tax returns, covered employment and taxable sales.

For the tests, the first four methods were used to calculate the populations of the state and its 58 counties as of April 1, 1970, and the averages of the four as well as the averages of the

four adjusted for the assumed effects of the two additional methods were calculated. The standards of comparison were the state and county final total census populations less estimated military, a civilian figure subject to later change when sample data are ready for use. Two methods for which current data could not be obtained--Ratio Correlation and Composite--were calculated for July 1, 1969 and compared with the latest intercensal estimates for that date, estimates which were based upon as much of the detail of the Census of 1970 as was available. Thus, there are eight comparisons possible--each of four methods, their arithmetic mean and an adjusted mean, six with the 1970 census, and two additional methods with a 1969 standard based upon the 1970 census.

Statewide, the Housing Unit Method estimated 4.8 percent high while Census Bureau Method II estimated 3.0 percent low. The average of the four methods was .16 percent high when adjusted for the presumed effects of the two missing methods and only .09 when left unadjusted. Apparently the four methods for which reasonably current data can be used, when averaged, yield the best estimate of the State's population, but this is not necessarily the most satisfactory technique if minimum error in county populations is the criterion.

California has 58 counties of which 24 are metropolitan, i.e. they are Standard Metropolitan Statistical Areas or parts thereof, and 34 are nonmetropolitan. Of the 24, those of 500,000 inhabitants or more are 10 in number; of the 34 those with 20,000 or more inhabitants number 18. On this basis, four strata were identified, large and small metropolitan counties and large and small nonmetropolitan.

The fourth stratum contained one county (Alpine) of 484 and one of 2365 inhabitants (Sierra); mean errors were calculated with and without these two. Table 2 displays the mean errors from the different methods and combinations of methods for the strata. Except in the state totals, absolute values of errors were averaged, hence no effect of upward or downward bias is presented. No Ratio Correlation value is shown for the State since use of the method is restricted to allocation by county of a predetermined whole. The best performance in estimating the large metropolitan counties shown in the table is that of Ratio Correlation, with a mean error of one percent. The best method for all counties was the Adjusted Average with a mean error of five percent and a very creditable two percent for the largest counties; it estimated Los Angeles County within .06 percent, but this is not shown.

Table 2 compares the success of the methods in another way in comparing for each method, separately for the metropolitan and nonmetropolitan, the number of counties estimated high with the number estimated low. In the same table a comparison of the methods with respect to the magnitude of errors of estimating is shown. Four ranges were selected which might be subjectively characterized as "excellent", "good", "fair to poor" and "unacceptable", or in percent error ranges, 0 to 1.99, 2.00 to 4.99, 5.00 to 9.99 and 10 percent or greater.

It is apparent that for California counties the composite method seems to estimate low as often as high for metropolitan and nonmetropolitan counties, taken together, although it tends to be slightly high in nonmetropolitan counties. The results show that it was "excellent" in half the metropolitan counties but less than one-sixth of the nonmetropolitan; in fact this otherwise successful method was 10 or more percent off the mark in nearly a third of the nonmetropolitan counties. Ratio Correlation, with a slight tendency to estimate high seems to have yielded the best results but the fact that it is controlled to an accepted total renders this less than miraculous. Unfortunately, neither of these methods are satisfactorily timely, i.e., they can not be used for developing current-year estimates.

Turning to the four methods that can be used for current-year estimates, the tendency of the Housing Unit Method to estimate high is obvious in both metropolitan and nonmetropolitan counties. Its accuracy, when measured against the census is only fair for metropolitan counties, (14 of 24 metropolitan counties within five percent) and poor for nonmetropolitan (seven of 34). Census Bureau Method II, one of the poorer methods for estimating the State, was about as accurate (13 of 24 metropolitan counties within five percent, nine of 34 nonmetropolitan).

Although the first of the two Department of Finance regression methods performed rather well with 19 of 24 metropolitan counties and 15 of 34 nonmetropolitan counties (within five percent), the adjusted average was slightly better. The case might be made, indeed has been made, that averages of methods are preferable to single methods in this type of estimating.

What remains to be done in California is a standardized test of Method II,

the Composite and Ratio Correlation to the specifications of the Bureau to enable those in the Cooperative Program[5] to evaluate methods for counties across the nation.

Inquiries were made of several other western states in order to compare our experience with that of others. Although it was determined that no two states were engaged in the same estimating activities with respect to their counties, the efforts of the several states were not without common elements and concerns. Utah, for example, reported that its estimates of major counties were satisfactory but trouble was encountered with the smaller ones. Its errors were between five thousandths of a percent to 20 percent with a median of 3.7 percent and their methods a modified Method II and another component method as well as membership statistics from the L. D. S. Church.

The State of Washington is in the process of preparing tests of methods. Census Bureau Method II underestimated the State by eight percent and yielded the poorest distribution by county, especially in the larger ones. Ratio Correlation gives the best basis for distributing the States' population. Of the 39 counties of the State, it came closest in 20, the Composite Method came closest in 14 and Method II in 10.

Hawaii, with four counties, used a still different approach. The State total was estimated by the Bureau, using a weighted average; 25 percent Ratio Correlation, 25 percent Method II and 50 percent a special Hawaii component method using arrival and departure data for net migration. As the decade progressed, the two traditional methods diverged increasingly from the special component method and in 1970 were substantially closer to the Census. It was concluded that departures were less scrupulously recorded than arrivals by the steamship companies and airlines; the result was an overestimate of the population by the Hawaii method. Table 4 shows the extent of the error of estimation and the percent distribution among Hawaii's four counties.

Hawaii analysts feel that a part of the error is the result of a census undercount and they are encouraged in this belief by surveys of selected areas and covered worker statistics.

In closing my remarks to this panel on "Why Did Intercensal Estimates Go Wrong in the 1960s?" I should allude to some of the problems we have had in our local estimates work with the Census benchmark. In the material from the first count which was hungrily consumed by local planners, a substantial number of misallocations were detected, errors attributable to the address coding guide or to other sources of inaccurate coding. Such errors, which may be negligible when dealing with a unit as large as a county, loom very large when attempting to evaluate an estimating method for a city. We are indeed fortunate in having the summary tape information for without it we would have been unable to understand, and sometimes to detect census errors. With it we can occasionally reconstruct what probably happened. Is it not fair to ask that in local estimates, at least, on the basis of an active program of data collection and estimates of population and housing, "Where did the Census go wrong?". It is also fair to ask whether small area data from the second and subsequent counts can be very useful to local people unless a substantial effort toward their improvement is undertaken.

---

1 Bureau of the Census, Current Population Reports, Series P-25 No. 339, 1966.

2 Bogue, Donald J. and Duncan, Beverly "A Composite Method for Estimating Postcensal Population of Small Areas by Age, Sex and Color". U. S. Department of Health, Education and Welfare, Selected Studies, Vol. 47 No. 6, 1969.

3 Crosetti, Albert H. and Schmitt, Robert C., "A Method of Estimating the Intercensal Population of Counties", Journal of the American Statistical Association, December 1956, pp 587-590.

4 Table 4 furnished by Robert C. Schmitt, Department of Planning and Economic Development, State of Hawaii.

5 Meyer Zitter, "Federal-State Cooperative Program for Local Population Estimates," The Federal Registrar and Statistician, U. S. Department of Health, Education and Welfare, January 1968.

Table 1. COMPARISON OF ELEMENTS OF HOUSING UNIT METHOD, CENSUS AND TEST ESTIMATES, SELECTED CALIFORNIA CITIES, PERCENT DEVIATION

ORANGE COUNTY	HOUSEHOLDS	HOUSEHOLD POP.	HOUSING UNITS	SINGLES	PERSONS PER.
Anaheim	2.04	4.79	- .36	- 4.56	22.33
Brea	-1.49	-6.35	5.60	3.42	- 5.07
Buena Park	- .36	2.82	.01	-12.47	3.29
Costa Mesa	-1.20	3.39	1.38	--	4.66
Cypress	-2.05	1.79	- 4.54	--	4.02
Fountain Valley	-3.20	- .91	- 8.00	--	2.37
Fullerton	-3.97	-2.39	- 1.34	- 3.29	.16
Garden Grove	-1.28	.78	- 1.23	- 2.99	2.09
Laguna Beach	-6.46	-4.62	- 1.32	1.17	2.00
La Habra	-6.00	3.01	- 3.54	--	9.57
Los Alamitos	-8.31	-6.77	- 6.52	--	1.67
Newport Beach	-2.77	-2.34	4.76	- 3.30	.47
Orange	-2.67	-2.03	- 2.45	- 6.90	.64
Placentia	1.67	.86	4.39	- .44	- .80
San Clemente	-1.35	.44	- 7.29	--	1.90
San Juan Capistrano	-5.11	-6.37	4.66	--	- 1.36
Santa Ana	-1.93	- .23	2.01	- 3.70	1.73
Seal Beach	.27	-1.64	.09	--	- 1.90
Stanton	6.19	5.72	5.75	- 7.53	- .45
Tustin	3.39	8.82	4.17	3.61	6.09
Westminster	- .51	.06	- .80	- 3.32	.56
Yorba Linda	2.28	2.51	.26	--	.27

SANTA CLARA COUNTY	HOUSEHOLDS	HOUSEHOLD POP.	HOUSING UNITS	SINGLES	PERSONS PER.
Campbell	3.04	- 1.27	4.9	-16.8	- 4.13
Cupertino	1.73	8.41	- 2.8	-10.1	6.32
Gilroy	- 9.44	- 8.44	- 4.8	- 5.6	1.16
Los Altos	- .25	4.16	1.4	2.6	4.36
Los Altos Hills	- 3.92	.46	- 3.89	--	4.65
Los Gatos	- 9.64	- 8.31	- 9.3	-15.8	1.69
Milpitas	- .91	3.18	- 0.3	- .5	4.24
Monte Sereno	-15.13	-20.05	12.2	--	5.85
Morgan Hill	-14.34	-11.69	-16.60	--	2.87
Mountain View	3.85	9.07	10.60	- 8.40	5.16
Palo Alto	- 5.73	1.36	- 4.79	- 3.43	7.87
San Jose	1.88	4.46	3.57	- 1.17	2.39
Santa Clara	- 4.77	- 2.27	- 5.56	- 8.48	2.52
Saratoga	- 3.96	- 2.07	.18	- 1.56	1.86
Sunnyvale	2.33	6.08	7.83	1.56	3.50
City of San Diego	2.26	2.91	2.54	--	.89

Table 2. COMPARISON OF ERRORS[1] OF SIX ESTIMATING METHODS AND AVERAGES OF METHODS, COUNTY GROUPINGS, CALIFORNIA

METHOD	1	2	3	4	5	6	7	8
	State as a whole	58 counties	10 Metro counties 500,000 and Over	14 Metro counties of less than 500,000	18 Non-Metro counties 20,000 and over	16 Non-Metro counties under 20,000	15 small counties less Alpine	14 small counties less Alpine, Sierra
Four-Method Average, 1970, Adjusted to Six	+0.161	4.977	2.021	3.707	5.277	7.599	6.439	5.974
Four Method Average, 1970, Unadjusted	-0.089	5.761	2.804	3.915	6.258	8.667	7.867	8.042
HHR D of F Regression I, 1970[2]	-0.217	5.043	2.689	3.203	5.114	8.046	6.489	6.877
REC D of F Regression II, 1970[3]	-2.031	5.667	3.973	5.229	5.404	7.406	7.651	8.150
Housing Unit, 1970	+4.844	8.215	4.702	4.649	8.347	13.382	11.188	11.184
Census Bureau Method II, 1970	-2.951	7.079	5.882	6.090	8.183	7.450	7.395	7.154
Composite, 1969	1.18	6.96	2.60	3.55	6.28	13.42	10.85	7.38
Ratio Correlation, 1969		4.87	.99	3.54	4.59	8.78	7.27	7.72

1. Excepting for State as a whole, errors are averaged without regard for sign.
2. Z = -1.103 + .734X + .374Y; X = Household ratio, HHR; Y = School enrollment ratio Grades 3-8/Gr.2-7
3. Z = .172 + .215X + .921Y; X = Residential electric customer ratio,REC;Y = School enrollment ratio

Population Research Unit
1623 10th Street
Sacramento, California 95814

Table 3. DIRECTIONS AND MAGNITUDES OF ERRORS OF SIX ESTIMATING METHODS AND AVERAGES OF METHODS, METROPOLITAN AND NONMETROPOLITAN COUNTIES, CALIFORNIA
Frequency distributions of 24 Metropolitan and 34 Nonmetropolitan counties.

	DIRECTIONS					MAGNITUDE							
	METRO		NONMETRO			METROPOLITAN				NONMETROPOLITAN			
	High	Low	High	Low		0-1.9	2-4.9	5-9.9	10%& over	0-1.9	2-4.9	5-9.9	10%& over
Four-Method Average, 1970, Adjusted to Six	14	10	30	4		11	9	3	1	6	9	13	6
Four-Method Average, 1970, Unadjusted	14	10	31	3		8	10	5	1	4	6	16	8
HHR D of F Regression I, 1970	15	9	26	8		11	8	4	1	7	8	13	6
REC D of F Regression II, 1970	9	15	28	6		7	9	5	3	6	14	5	9
Housing Unit, 1970	23	1	33	1		7	7	9	1	1	6	9	18
Census Bureau Method II, 1970	10	14	26	8		3	10	8	3	8	1	16	9
Composite, 1969	11	13	19	15		12	8	3	1	5	8	11	10
Ratio Correlation, 1969	15	9	17	17		15	7	1	1	8	7	13	6

TABLE 4

STATE OF HAWAII

County	1969 estimate		1970 census	
	Number	Percent	Number	Percent
State total.....	793,747	100.0	769,913	100.0
Hawaii.....	67,229	8.5	63,468	8.2
Honolulu.....	645,319	81.3	630,528	81.9
Kauai.....	31,666	4.0	29,761	3.9
Maui.....	49,533	6.2	46,156	6.0

APPENDIX

HOUSING UNIT METHOD CALCULATIONS

$$HU_t = HU_o + HUC_{o-t} - HUD_{o-t} + HUG_{o-t} - HUL_{o-t}$$

$$HH_t = HH_o + \Delta REC_{o-t} + \Delta MMHH_{o-t}$$

$$P_t = HH_t \times PP_t + TrP_t + GQP_t$$

DEPARTMENT OF FINANCE REGRESSION EQUATIONS

I. $Z = 1.103 + .734X + .374Y$

II. $Z = .172 + .215X + .921Y$