PROCEDURES FOR ESTIMATING M.C.D. POPULATIONS FOR STATE REVENUE SHARING

C. D. Palit, H. Reinhardt, J. Sweet, and H. Winsborough
University of Wisconsin

H. Krebs, F. Backus, G. Ferwerda, D. Giovannini, and L. Golicz
State of Wisconsin

In November, 1971, the Wisconsin Legislature changed the basis for the sharing of tax revenues with its minor civil divisions to one which is based on estimates of the populations of the minor civil divisions made annually. There being 1800-plus minor civil divisions in 72 counties in Wisconsin, this presented a formidable task. The State Department of Administration was charged with making the estimates, and early in 1972 contracted with the University of Wisconsin for methodological development. At that time an informal "seminar" was organized which included some academic demographers, statisticians, and people from state government, who had been making county estimates for Wisconsin, and were familiar with some of the data sources that are useful in making population estimates.

Rather than rehearse the entire history of the enterprise, I'd like to discuss three aspects of our experience which may be of general interest:

1. What we came to view as the dimensions of the problem;
2. How we came to the solution we did;
3. How the method was tested.

One of the most important aspects of the problem, as we quickly realized, was the complementary problem of the large number and small size of the units to be estimated. In 1973 there were 1,872 cities, villages, and towns in Wisconsin, and what was remarkable about them is that 88 percent had populations of less than 2,500. Table 1 shows the distribution by type and size.

If that is contrasted with the estimation problem at the county level, where only 12.5 percent of the counties of Wisconsin have populations less than 10,000, and none less than 2,500, it is clear first, that a good-sized data set was necessarily involved, and that a set of that size is difficult to monitor year to year.

Second, it was imperative to use data that had a uniform collection and management system. Building permit data, each municipality applying its own rules; or utility meters, each of several hundred companies using its own standards; or school census data, several hundred districts responding with no State surveillance; all proved fatally flawed. Third, in many of our communities, the smallness of the numbers was going to lead to considerable random variability and therefore instability of estimates.

The second limitation on our enterprise was the legislative requirement that the estimates be current. This had been interpreted by the court to mean that the estimate effective date and year of production must be the same. Thus, we had to produce January 1, 1973 estimates by August 1, 1973.

Table 1
MUNICIPALITIES BY TYPE BY SIZE
Wisconsin, 1970

Municipality Size - 1970	Town		Village		City		Total		
	No.	%	No.	%	No.	%	No.	%	Cum. %
Less than 500	385	30.3	184	46.1	10	4.9	579	30.9	30.9
500 - 1,000	518	40.8	121	30.3	10	4.9	649	34.7	65.6
1,000 - 2,500	289	22.8	69	17.3	62	30.5	420	22.4	88.0
2,500 - 5,000	61	4.8	13	3.3	43	21.2	117	6.3	94.3
5,000 - 10,000	12	0.9	7	1.8	32	15.8	51	2.7	97.0
Over 10,000	5	0.4	5	1.2	46	22.7	56	3.0	100.0
Total	1,270	100.0	399	100.0	203	100.0	1,872	100.0	100.0

The implications of this limitation were that any data sources that could not be made available in time for August 1 estimates were in effect useless to us.

The third influence or dimension of this problem was clearly the fact that a considerable sum of money was riding on the outcome of these estimates. The revenue-sharing amounted to $35

per capita and that, for almost all communities, is a sizable part of their budget. This therefore subjected the whole process to close scrutiny and made for some very unacademic political pressures.

However, the greatest limitation on the methods that could be chosen was the availability of data. We set out to assess the problem by reviewing the conventional methods for small area population estimation and the data required for use of these methods. What we found was discouraging. Vital events, for instance, which are indispensable for the Vital Rates method, Component II and Composite, and are sometimes part of the Ratio Correlation method, were simply not available at the M.C.D. level at all, and they were usually not available by the August 1 deadline.

School enrollment data, which are essential for the Component II and the Composite methods and are frequently used in the Ratio Correlation method could not be used at the M.C.D. level because the geographic allocation is by school districts which respect no political boundaries.

A third standard data set that would have been useful in both Component and Composite is Medicare enrollments, but this too we found was unavailable at the M.C.D. level, and in any case was not available in time to be useful for our estimates.

So, what we were reduced to, was finding administrative data that could be used symptomatically. I'll spare you a listing of the dozens of sets that we investigated and rejected. What we came down to as reliable administrative data sets associated with population were motor vehicle registration data and data from state income tax returns.

Given these data limitations, the choice of the method was clearly a restricted one. What I'd like to examine now is the process by which we decided on the method chosen.

These results of our data search reduced our choice of method to a Ratio Correlation procedure, which has the appeal of being in wide use and being familiar in the literature, or a Censal Ratio type procedure. Unfortunately the only data series long enough for us to construct a ratio correlation regression equation was the motor vehicle data. The tax data had a serious discontinuity in the early 60's as a result of passage of a withholding provision into the revenue statute. The other choice, you will recall, was the Censal Ratio method, where a symptom population ratio in the census year is updated to the estimate year, and that estimated ratio, with the current symptom count, is used to produce a population estimate. The Censal Ratio procedure allowed the use of both the motor vehicle data and income tax data to produce estimates.

One would prefer more than one symptomatic data set and therefore one would prefer to use the Censal Ratio method, we nevertheless performed a test to determine the direction in which we should go.

The test consisted of making estimates of the 1970 populations for 71 of Wisconsin's counties from the 1960 base, and then comparing the errors of the estimates. This approach permitted us not only to compare our alternatives to each other, but also to evaluate our alternatives against the results of other county estimating procedures tested in the Federal-State Cooperative Program. For our test, we used three estimation methods: a Ratio Correlation procedure based on the automobile series, a conventional Censal Ratio procedure patterned after the Vital Rates procedure, but using motor vehicle data.

We referred to the third method as the Ratio Difference Estimator. This estimator was also a Censal Ratio procedure, but unlike conventional Censal Ratio procedures, uses a difference estimator to update the Symptom-population ratio of the census year to the time for which the estimate is desired.

Briefly the procedure is as follows: for each area for which the population is to be estimated, the ratio of symptom to population for the base year is calculated, viz.,

$$r_o(u) = \frac{\left\{\begin{array}{l}\text{Symptom at time zero, or}\\ \text{base year for the } u^{th} \text{ area}\end{array}\right\}}{\left\{\begin{array}{l}\text{Population count at time}\\ \text{zero for the } u^{th} \text{ area}\end{array}\right\}}$$

A similar ratio $R_o(S)$ = is computed for the area S; where S is a larger area containing all the "u" areas, i.e.,

$$R_o(S) = \frac{\text{Total symptom at time zero for S}}{\text{Total population in S at time zero}}$$

An independent estimate for the population at time t is obtained and used to estimate the ratio,

$$R_t(S) = \frac{\text{Total symptom for S at time t}}{\text{Total population for S at time t}}$$

We call this estimate $\hat{R}_t(S)$ and use it to estimate the ratio,

$$r_t(u) = \frac{\text{Symptom at time t for } u^{th} \text{ area}}{\text{Population at time t for } u^{th} \text{ area}}$$

with difference estimator,

$$\hat{r}_t(u) = r_o(u) + \hat{R}_t(S) - R_o(S).$$

The estimate $\hat{r}_t(u)$ is then used to generate an estimate of the population at time t in the u^{th} area by dividing the symptom for the u^{th} area at time t with the estimating ratio, i.e., the estimated population for the u^{th} area at time t is,

$$\hat{P}_t(u) = \frac{\text{Symptom at time t for } u^{th} \text{ area}}{\hat{r}_t(u)}$$

The test comparisons were made by reference to a series of five measures of accuracy. The first was the mean percentage absolute deviation from the census counts. This is conventional and gives a good overall measure of average performance. Its disadvantages are that it is:

1) unweighted, i.e., a 5% error on 50 weighs the same as a 5% error on 700,000; and that 2) since it deals in absolute values, it gives no indication as to the presence or absence of bias.

The second measure of accuracy used was the simple percent of deviations that were positive. This enables one to evaluate the estimates for the presence of bias.

The third was the mean square error.

And finally, two measures were used that are not well-represented in the literature of population estimation, the number and percent of misallocation. Misallocation is calculated as the sum of the absolute values of the deviations, that sum divided by two. This has the advantage of being a weighted measure, of being intuitively appealing inasmuch as it actually measures the number of people who wound up in the wrong jurisdiction as a result of estimation errors. For purposes of tax sharing in particular, this was the most direct measure of performance we could devise, for by multiplying the misallocation number by $35 you have an exact amount of money misallocated among the jurisdictions estimated. The percent misallocation is the percent of the total population which was misallocated.

Table 2
ERROR MEASURES OF VARIOUS ESTIMATION METHODS
Wisconsin Counties
1960 Based 1970 Estimates

Method	% Error	M.S.E.	Mis-allocated	% Mis-allocated
Vital Rates	5.49	19836576	90320	2.05
Regression*	4.52	14896030	78312	1.77
Component*	6.43	79084064	118312	2.68
Composite*	3.85	7980251	59422	1.35
Ratio Difference**	1.96	5258575	35902	0.81
Censal Ratio**	2.25	15965140	47834	1.08
Ratio Correlation**	2.20	17552576	48929	1.11

* These are the results from the Federal State Co-op Program adjusted so as to sum to the 1970 Census total for the 71 counties used. (Wisconsin now has 72 counties.)

** Only one Symptomatic data set, viz., number of passenger automobiles registered to each county was used to make these estimates.

Table 2 represents the various test measures as produced in the Federal-State Cooperative Program in the test for Wisconsin counties, and then those of the ratio correlation, conventional censal ratio and censal ratio with a difference estimator. While the ratio correlation method and the ratio difference method have similar percent errors, the mean square error seems to indicate that the distribution of errors is more favorable for the ratio difference method. This is seemingly confirmed by the percent misallocation figures.

While these results were gratifying at the county level, it told us little about how this method would perform at the M.C.D. level. For an M.C.D. test, we used 1960 data and 1970 data and compared the estimate with the 1970 census counts. This could only be done with the motor vehicle registration data, which one could expect would be less accurate alone than in combination with other data series. The mean deviation of the 1,800-plus minor civil division estimates was about 10% and the misallocation about 3.2% after 10 years. Graph 1 shows the distribution of errors. Table 3 shows the error measures of this test.

Graph 1
DISTRIBUTION OF PERCENT ERROR OF 1960 BASED ESTIMATES
FOR 1970 MCD POPULATIONS (AUTOS ONLY)

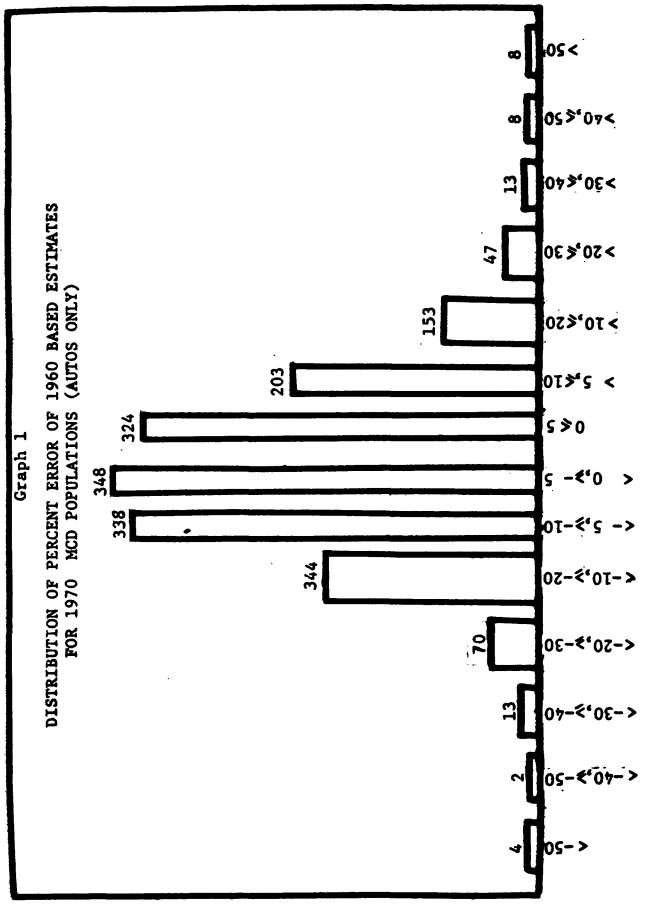

Table 3

Summary Statistics on the Estimation Errors of 1960 Based Ratio-difference Estimates for 1970 Wisconsin M.C.D. Populations

Average Percent Error	10.03
Mean Square Error	440,262
Misallocation	140,753
Percent Misallocation	3.19

The only hard testing available for the income tax data was a test of Dane County, and its 60 M.C.D.'s. This was fortuitous in that Dane County had been one of the sites of a dress rehearsal census in 1968. While the tax data by itself did not perform any better than the motor vehicle data, what we found was that the averaging of the tax estimates with those made from motor vehicle data reduced the observed errors. Mean percent absolute error between 8 and 9%, misallocation percent 2.4. The percent misallocation of the various tests are summarized in Table 4.

Based on our detailed evaluation of all these tests, we made the following choices in constructing the estimation model for minor civil divisions:

First, that it should be a two-step process that proceeds from the state control total to county estimates; these county estimates then serve as county control totals to the M.C.D. estimates within each county.

Our empirical results confirmed the suggestion made earlier to us by Peter Morrison that this would improve the estimates at the M.C.D. estimates level.

The second decision was that the Censal Ratio method, with the difference estimator as the updating mechanism, be used on three separate series of data and the unweighted average of these be the final estimate for the M.C.D. The three series were passenger automobiles, income tax filers, and dollar value of exemptions claimed.

The third decision concerning the actual estimation model was to apply the Censal Ratio method estimating model to the uncountable population only, that is, removing the institutional population from the base population and getting an independent count of that institutional population at the estimate year to add to the estimate generated by the Censal Ratio method.

Table 4

MISALLOCATION RESULTS FOR SOME ESTIMATES OF WISCONSIN POPULATIONS USING THE RATIO-DIFFERENCE ESTIMATOR

		% of Total Population Misallocated
1	1960 Based 1970 estimates for Wisconsin Counties (autos only)	0.81
2	1960 based 1970 estimates for Wisconsin Municipalities (autos only)	3.19
3	1968 based 1970 estimates for Dane County's municipalities (autos only)	1.6
4	1968 based 1970 estimates for Dane County's municipalities (Average of auto, files, and $ exemption estimates)	1.5
5	1973 Estimates for Wisconsin municipalities with special census in 1973	2.0

The difficulties were surmounted and the estimates were made on time, to the surprise of some of us.

As a result of the preliminary estimates, some 80 challenging cities submitted claims of underestimation. Of these, we found about 20 to be substantial, generally meaning some data error had occurred. The data in these cases were adjusted before the final estimates were made. Final estimates did not satisfy everybody; some 14 municipalities brought suit for judicial review of the estimation method. After an evidentiary hearing in the circuit court, which consumed some fifteen days, the judge strongly upheld the reasonableness of this method.

There have been some 50 special censuses conducted in Wisconsin M.C.D.'s around the time of the estimates, and these generally confirm our expectations as to the probable precision of the estimates. The average error is about 6% and misallocation is about 2%.

Reference: U. S. Bureau of the Census, Current Population Reports, Series P-26, No. 21, "Federal-State Cooperative Program for Local Population Estimates: Test Results--April 1, 1970."