AN EXAMPLE OF A TIME SERIES APPROACH TO SHORT TERM POPULATION PROJECTIONS FOR SMALL AREAS

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

H. Krebs and B. Kale
State of Wisconsin

At St. Louis a year ago, we reported on the development and testing of an estimation procedure for the population of all Wisconsin municipalities. That development was necessary to implement the 1971 Wisconsin legislation which changed the state's method of sharing its revenue with its 1,800 odd municipalities. Previously, state-collected revenues had been returned to the municipality whence it came, the 1971 legislation directed that the tax sharing be on the basis of annual population estimates of the municipalities. The same legislation mandated that short-term population projections be provided to the municipalities for budget revenue anticipation. This work, like the earlier, was done by an informal seminar involving University and State people.

In the nature of the problem, a projection of next year's estimate would be preferable to a projection of population per se. Many of the municipalities are very small, two-thirds have a thousand or less population. Most lack sophisticated budget planning capability, thus the need for projections. Very little demographically useful data is generated by or about these communities, and this constrains the methodological options of those who would estimate or project population.

Derivation of the Model

The movement through time of the population size for an area, or of the estimate of population size, would seem to be a phenomenon well suited to time series analysis. It can be described by a sequence of values $P_0, P_1, \ldots, P_t \ldots$, where P_t is the population of the area at time t. However, to make a forecast using such an approach it is first necessary to identify a suitable model, and then to estimate the model's parameters from the available data. It is sometimes possible to use purely statistical tools to identify a suitable model for making forecasts; to let the data speak for themselves, so to speak. Such model identification procedures are well described in Box and Jenkins (1970), but these require a relatively long data series. These statistical methods have been applied to a series consisting of the Swedish population from 1780 to 1970 by Saboia (1974) with success. However, the statistical tool is not the only way to identify an appropriate time series model. Reasonable models may be constructed by examining the underlying mechanisms which give rise to the series in the first place. This latter approach is the only one available when the observed length of the series is short, that is, when there are only a few observation points. The Box and Jenkins procedures use estimates of the auto-correlations of the time series for which the standard error behaves like $1/\sqrt{n}$ where "n" is the number of observations in the series. With only, say, fourteen points in the series, the standard error would be approximately 0.27.

In this paper, we describe how we arrived at one such time series model starting from the familiar population accounting equation, and present an evaluation of the results for two applications. First for the annual forecasting of Wisconsin Municipal population estimates and the second for the forecasting of the Swedish population.

The algebra of a simplified derivation of the time series model from the demographic equation (1), is straightforward.

$$P_t = P_{t-1} + B_{t-1} - D_{t-1} + M_{t-1}; \tag{1}$$

To keep the presentation short we have eliminated some niceties which do not affect the conclusion of the argument presented here. In equation (1) the population at time t is the population at some earlier time plus births, B_{t-1}, and minus deaths, D_{t-1}, in the intervening period plus the net migrants, M_{t-1}, for that period. If we represent births as being the product of the population at the beginning of the time period, P_{t-1}, times some birth rate, b, plus some small error element, e, with an expectation of 0; and do the same for deaths with a death rate, d, then

$$B_{t-1} = P_{t-1}\, b + e, \tag{2}$$

$$D_{t-1} = P_{t-1}\, d + e. \tag{3}$$

The net migration is the difference between the in-migrants and the out-migrants, or,

$$M_{t-1} = I - O + e. \tag{4}$$

Collecting the first three terms on the right hand side of (1) and using the relationships in (2) and (3) we can write

$$P_t = A \cdot P_{t-1} + M_{t-1} + e \tag{5}$$

where A = (1 + b-d), which represents the natural increase in that population over the period.

We now attack the migration component, M_{t-1}, by considering it as a series in its own right and assuming that a first order auto-regressive model would provide an approximation of its behavior over time, i.e.,

$$M_{t-1} = C \cdot M_{t-2} + e. \tag{6}$$

The migration for the time t-1 to t is the migration in the previous year times some coefficient C plus a small error factor. Equation (7) expresses the net migration in the year t-2 as being the population at the end of that year minus an expression which takes into account the population at the beginning of the year and the natural increase occurring during the year.

$$M_{t-2} = P_{t-1} - A \cdot P_{t-2} + e \tag{7}$$

When that expression is substituted into equation (6) we have equation (8),

$$M_{t-1} = C \cdot (P_{t-1} - A \cdot P_{t-2}) + e \tag{8}$$

and when the now complete representation of the migration component is substituted, along with the natural increase component, into the demographic equation (1), we get equation (9), which reduces to the equation (10).

$$P_t = A \cdot P_{t-1} + C \cdot (P_{t-1} - A \cdot P_{t-2}) + e \tag{9}$$

$$= (A + C) \cdot P_{t-1} - (A \cdot C) \cdot P_{t-2} + e \tag{10}$$

Now let

$$E = A + C \tag{11}$$

$$F = -A \cdot C \tag{12}$$

then,

$$P_t = E \cdot P_{t-1} + F \cdot P_{t-2} + e. \tag{13}$$

We let "E" stand for the coefficient of the population at time t-1 and "F" represent the coefficient of population at time t-2, producing equation (13) which is the second order auto-regressive model.

An Application to Wisconsin's Municipal Population

The time series, in this case, was not a series of population counts, but a series of population estimates. A few words of explanation about how those estimates were derived is important to an understanding of the whole process. The estimation methodology developed and reported to this body last year, is essentially a censal ratio method which uses a difference estimator to update the symptom to population ratio of the places estimated. The population, $P_{u,t}$, of the u^{th} place at time t is equated to a count of symptom, $S_{u,t}$, related to that place at time t divided by the symptom to population ratio, $r_{u,t}$, of that symptom to the place's population at time t.

$$P_{u,t} = \frac{S_{u,t}}{r_{u,t}} \tag{14}$$

To estimate $P_{u,t}$, the ratio in the denominator is the estimate defined below:

$$\hat{r}_{u,t} = r_{u,o} + (\hat{r}_{a,t} - r_{a,o}) \tag{15}$$

where the ratio, $\hat{r}_{a,t}$, is the symptom-population ratio for a larger area 'a' which includes the u^{th} place, and of course $\hat{r}_{a,t}$ is an estimate.

In constructing the time series, the population counts for 1960 and 1970 were used, and estimates for the intervening years were computed with the estimation model above. In order to eliminate the error of closure, the intercensal yearly estimates were adjusted by the interpolation of the symptom to population ratios. Had we interpolated the actual population estimates we would have destroyed the information about year-to-year change which appears in the symptom counts for each year. Thus, in equation (14), the ratio, $r_{u,t}$, is estimated by a weighted average of two other estimates rather than averaging population estimates as such. The intercensal interpolation used two estimates of the symptom-population ratio for each year. One used 1960 as the base period, and one was computed with 1970 as the base period, from which the estimate was made retrospectively. These two were then weighted in proportion to the propinquity to the base year, and the weights, of course, were constrained in each case, to sum to one, i.e.,

$$\hat{r}_{u,t+60} = 0.1\ (10-t)\ \hat{r}^{60}_{u,t+60} + t \cdot \hat{r}^{70}_{u,t+60} \tag{16}$$

where: t = 1, 2, ..., 9 and k in $r^k_{u,t}$ denotes base year.

Table 1.

Selected Error Measures for One, Two, and Three Year Ahead Population Projections of 1,835 Wisconsin Municipalities Base Years 1967-1970

Base Year	Measure	One Year Ahead	Two Year Ahead	Three Year Ahead
1967:	MSE	211,742	348,023	505,681
	Mean|%|Err	5.12	7.20	8.58
	% Misalloc	1.75	2.57	3.16
1968:	MSE	159,547	156,618	321,234
	Mean|%|Err	5.66	6.78	10.37
	% Misalloc	1.81	1.98	3.11
1969:	MSE	185,761	215,218	320,972
	Mean|%|Err	6.66	10.77	11.75
	% Misalloc	1.99	2.86	3.29
1970:	MSE	43,211	156,527	430,630
	Mean|%|Err	6.52	5.74	8.87
	% Misalloc	1.54	1.86	2.96

Table 1 shows the results of an evaluation of the model's performance for one, two, and three year ahead forecasts from base years 1967 through 1970. Forecasts were made for 1,835 municipalities. Three summary measures of performance are reported: the mean square error, the mean percent absolute error, and the percent misallocation. If you will look at the diagonal from lower left, that is 1969, up to the top right, these are the 1970, 1-year, 2-year and 3-year ahead projections based in 1969, 1968 and 1967. The mean percent absolute error and the percent misallocation increase, as one would expect, from the first to the second to the third. The mean square error in year two is not consistent with that pattern, however, one must remember that this measure is far more sensitive to outliers. If you will look at the verticals, we had considered the possibility that the incremental increase in the length of the series might improve the estimates as time went on but, the evidence over this four year period is inconclusive. If you look at the horizontals you will discover, as expected, that the projections degrade as you get further and further from the last estimate year.

Table 2 displays the error distribution as a percent of the estimate for a one year ahead projection. The error seems to be reasonably unbiased with 20.8% less than -5% and 17.4% greater than 5%. The range, however, is large. Some of these large errors represent the effect of perturbations in the data series.

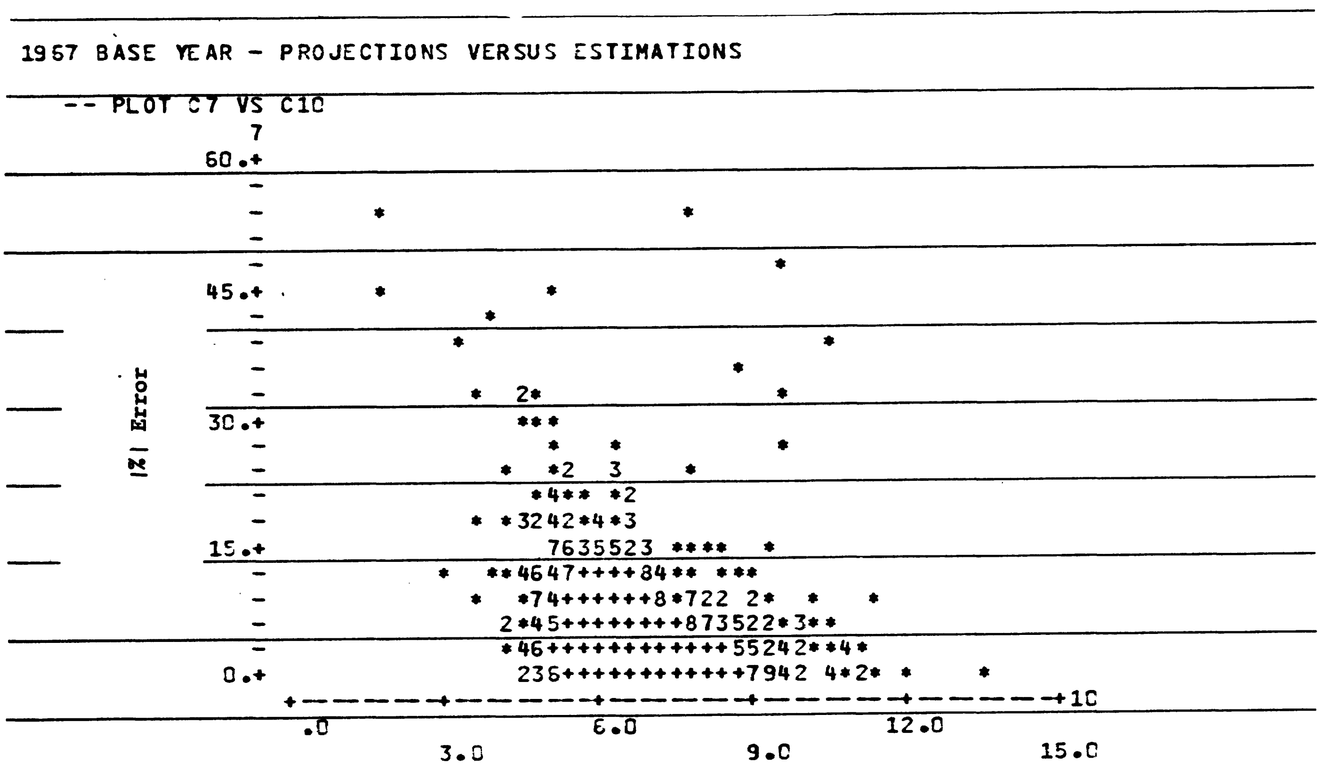
Table 2

Per Cent Deviations
One Year Ahead Population Projection of
1,835 Wisconsin Municipalities
Base Year 1967

Deviation %	Number	%
<-45	3	0.16
-45<-35	4	0.22
-35<-25	10	0.55
-25<-15	35	1.91
-15<-5	330	17.98
-5<+5	1,133	61.74
5<15	296	16.13
15<25	20	1.09
25<35	2	0.11
35<45	1	0.05
45+	1	0.05

Figure 1 is a plot of the absolute value of the relative error against the logarithm of the place size. The astericks represent one observation each, any digit from two to nine represents that number of observations and the plus sign indicates more than nine observations. You will notice a distinct drift from the lower right to

Figure 1



Log Population Size

the upper left, indicating that the error is inversely related to size. But again, there are outliers that represent problems.

It was necessary before these could be used for budget anticipation purposes, to investigate the possible causes for the outliers. We found that, in general, they represented one of the following categories. Either there was a new incorporation and there were no past observations for the model to work with, or there were significant one-time annexations. Third, there were some other one-time events such as significant size construction crews moving in one year and out the next. Finally, there were some obvious symptom data errors, e.g. zeroes in the data. One of the useful things about time series models is that they yield an estimate of the variance of the projection. The variance estimate can therefore be used to flag unreliable projections which can then be dropped, and in this case we reverted to the last annual estimate as the best estimate of the next year's estimate.

The second application used the Keyfitz and Flieger data on the Swedish population quenquennially represented from 1780 to 1970, that had been used by Saboia (1974). In that work, Saboia derived, using the Box and Jenkins statistical devices, two time series models, for the Swedish data.

Table 3

Comparison of Three Time Series Models Fitted to Swedish Population Data Series

Source	Fitted Model	Residual Variance
Our Model ARIMA (2,0,0)	$\hat{P}_t = 1.347P_{t-1} - 0.323P_{t-2}$	0.00396
Saboia (1974) ARIMA (1,1,0)	$\hat{P}_t = 0.069 + 1.540P_{t-1} - 0.540P_{t-2}$	0.00442
Saboia (1974) ARIMA (0,2,1)	$\hat{P}_t = 2P_{t-1} - P_{t-2} + a_t - 0.663a_{t-2}$	0.00463

For ease of comparison we have summarized the models and their parameters in Table 3. Our model seems to have slightly smaller residuals, about 12% in one case and 17% in the other. Our auto-regressive model has a zero intercept, whereas Saboia found a small positive intercept. Our model is not constrained to stationarity. When these three models are applied to the task of making 1965 and 1970 projections from 1960 as the final data point, we can compare them with the actual populations from the Swedish series. As can be seen in Table 4, all three of the models perform well on the one step ahead 1965 projection, and none look very well for the two step ahead 1970 projection.

By way of conclusion, three things seem to be clear. One, in the absence of a long data series, one can derive time series models from the substantive mechanisms of the data series themselves. Second, the shorter data series, though inappropriate for deriving a model, can be used to estimate the parameters, and yield useful results. Finally, the advantage of the variance estimate of a time series model allows the evaluation of the projections in terms of confidence regions.

Table 4

Forecasts of the 1965 and 1970 Population of Sweden for Each Model

		Forecast		
Year	Actual Population*	Our Model ARIMA (2,0,0)	Saboia ARIMA (1,1,0)	Saboia ARIMA (0,2,1)
1965	7,734	7,728	7,666	7,716
1970	8,346	7,992	7,835	7,935

*in millions

REFERENCES:

Box, George E. P., and Gwilym M. Jenkins: Time series analysis 1970, Holden-Day, Inc., San Francisco.

Saboia, Joao L. M.: "Modeling and forecasting populations by time series: the Swedish case" in Demography V.11 no. 3 August, 1974.

Palit, C.D. et al: "Procedures for estimating M.C.D. population for state revenue sharing," Proceedings of the American Statistical Association, Social Statistics Section, 1974.