THE ACCURACY OF STATE POPULATION PROJECTIONS

Michael A. Stoto and Alicia P. Schrier, Harvard University

I. Introduction

State and local population projections are widely used for planning and policy purposes. But forecasting is a difficult task. Often only scanty or incomplete data are available for local areas. More importantly, even trends that have remained in force for a long time are not guaranteed to continue next year.

One facet of population or other forecasting models is their scope and complexity. At one extreme, population can be modeled as part of a larger, more complex, economic system. Clearly, some economic factors affect population movements, and to some extent, population size and composition affect the local economy. The systems analysis or econometric approach attempts to model all of the interrelating factors in the interest of better accuracy on each. In an intermediate position are complex demographic models that take into account age, sex, and perhaps other factors, but do not explicitly involve economic activity. Finally, at the other extreme, population can be forecast by simple extrapolation of the immediate past total population figures. These approaches differ both in required inputs and expected outputs. The systems analysis end of the spectrum requires a large amount of data, and a substantial modeling effort. One payoff is that a more complex and complete forecast includes economic as well as demographic results. The simple extrapolation end of the spectrum, on the other hand, requires fewer inputs, but also promises fewer outputs. One aim of this paper is to compare these approaches in terms of the accuracy of the resulting total population forecasts, the one thing they have in common.

Sensible policymakers realize that, by their nature, human populations are not perfectly predictable, and take this variability into account. Our second goal, then, is to measure the amount of uncertainty about population growth given a forecast, and to provide realistic confidence intervals for policymakers and planners who use the forecasts.

In this paper we examine the accuracy of population projections for U.S. states as one level of sub-national forecasts. Our approach is historical. Judgement is a key factor in complex projection methods, in terms of choosing a model, appropriate data, and assumptions for the future, so we want to measure its effect. Our way of doing this is to compare projections actually made by government agencies to the eventual outcome. In this way we hope to measure the amount of uncertainty associated with projections as they are actually used in practice, and thus provide a realistic measure of the uncertainty inherent in the projections we make today.

Our data consists of three series of projections for 48 U.S. states from 1950 to 1980. Alaska and Hawaii were excluded because the complete series is not available. We examine a set of economic projections made by the Bureau of Economic Affairs of the U.S. Department of Commerce called the OBERS series, and a set of age- and sex-specific demographic projections produced by the U.S. Bureau of the Census. The base year, target year, and year of publication of these projections are given in Table 1. We examined projections ranging from five to ten years into the future. In some cases, forecasts based on different fertility assumptions were available. Since many users have no basis for choosing one set of assumptions over another, we averaged these forecasts. In the last two Census Bureau forecasts, two sets of migration assumptions were used, one thought to be realistic, and one not. These were treated as two independent sets of projections. In other years, separate forecasts were presented based on five- and ten-year migration patterns. Following our reasoning with the fertility assumptions, we simply averaged these two sets of projections. In addition, for comparison purposes we have constructed a set of simple geometric extrapolations of total state projections, which simply assume that the total growth rate will remain constant.

Our measure of error or discrepancy is based on differences between the projected and actual annual growth rates. Let P_0 be the population at the beginning of the projection period, and P_T be the actual population T years later. Thus the actual average growth rate is

$$r = \frac{1}{T}\,\ln\left(\frac{P_T}{P_0}\right).$$

Similarly, let $\hat{P}_0$ be the assumed population at the beginning of the period (final estimates are often not available when the calculations are made--we use the most recent estimates at the time of publication) and $\hat{P}_T$ be the projected population. Thus the projected average growth rate is

$$\hat{r} = \frac{1}{T}\,\ln\left(\frac{\hat{P}_T}{\hat{P}_0}\right).$$

Multiplying by 100 to put results in terms of percent, our measure of error is

$$\Delta r = 100(\hat{r}-r) = \frac{100}{T}\,\ln\left(\frac{\hat{P}_T}{\hat{P}_0}\cdot\frac{P_0}{P_T}\right).$$

Stoto (1982) has shown that, for national population projections, this measure has a reasonably stable distribution over time. Furthermore, it adjusts for the length of the projection period in that average yearly errors in the growth rate are about the same for five-year, ten-year, or longer projections.

II. Comparison of Methods

We begin with an overall comparison of the distribution of Δr for each of the three projection methodologies, the difference between projected and actual annual growth rates. The first Census projection uses 1950 as a jump-off year, but as we see later, the distribution of Δr is very different for this year, probably due to the effects of demobilization after World War II in the base data for 1945 through 1950. Therefore our comparisons will be for the five- and ten-year projections made in '55, '65, '70 and '75 that can be evaluated now using 1980 census data. For comparison purposes, we can calculate geometric extrapolations for the years '55 through '75. The OBERS data only start in '70, so we make a separate comparison.

Table 2 gives summary measures of Δr for five- and ten-year census and geometric projections. First, over the period from 1955 to 1980, the Census and geometric projections were relatively unbiased. The major component of the root mean square error (RMSE) is variance. Second, the variance, and hence the RMSE of each group, are relatively constant. Three groups have essentially the same standard deviation -- F-tests for pairwise comparison of the five-year Census, ten-year Census, and ten-year geometric variances are not significant. The five-year geometric group has a slightly higher standard deviation, but this is largely due to two large outliers (Nevada for jump-off years 1960 and 1965). With these points removed there is no noticeable difference in variability. These results have two ramifications. First, we see that in terms of Δr, the average error in the annual growth rate, five- and ten-year projections are about equally accurate. Second, also in terms of Δr, simple geometric extrapolation is almost as accurate as the more complex Census methodology.

Table 3 shows summary measures of Δr for two five-year OBERS projections, and for the corresponding Census and geometric projections. First, we see that the OBERS projections were more biased, on average, than the other two sets of projections. Second, and more importantly, the standard deviation of the Δr for the OBERS projections is more than twice as high as the other sets. The F-tests for equality of variance are significant at $p < .001$ for each comparison. This means that the added economic detail in the OBERS model, while perhaps enhancing the value of the output in other ways, did not increase, and in fact substantially decreased, the accuracy of the model for population projections.

III. Policy Implications

The federal government has recently considered the use of standardized population projections for the allocation of federal funds (Griffith, 1980). Its rationale was that future rather than current population is the relevant need factor for the long term projects, but individual states making their own projections could manipulate the assumptions to unfairly increase their share of the funds. The proposed solution was that the federal government make state projections, and the states break these down for smaller geographic areas. A model combining features of the Census and OBERS methodology would provide the state projections. Our results indicate two weaknesses of this plan.

First, although the OBERS method might give important economic detail, it seems decidedly inferior to the other methods in terms of projections of total population. Additional attention to detail does not seem to lead to accuracy. Thus there is no reason to believe that the combined model, which will be more complex than either alone, will be more accurate.

Second, simple geometric extrapolations are about as accurate as anything we have examined. In addition, they do not require judgement, and thus are not subject to political pressure. Furthermore, they can be applied simply and easily to states, and also to smaller geographic areas.

Since simple geometric extrapolations avoid political disputes about projection methods and assumptions, and do not suffer in accuracy, they are more appropriate for allocating federal funds for long term projects.

IV. Region and Jump-Off Year Effects

Besides differences in projection methodologies, there are reasons to believe that other factors would be associated with bias and variance in population projections. Stoto (1982) found a substantial jump-off-year bias in national projections -- all of the projections made at a certain time tended to be off in the same direction. Since the geographical regions of the United States are so varied, and have very different amounts and patterns of migration and growth, we might expect regional bias in the projections, and perhaps more variability in faster growing areas. In this section we explore the effects and interactions of these factors.

Table 4 gives summary measures by jump-off year for each of the sets of projections. There are substantial differences. In particular, we see that the 1975 OBERS projections were substantially more error prone than the 1970 series. Thus our previous observation about the poor accuracy of the more complex econometric model may need to be tempered. Future projections may be more like the earlier set, no worse, but no better, than the Census or geometric methods. Or they could be as bad, or perhaps even worse, than the more recent set.

Table 5 gives summary measures of Δr for the nine Census regions of the U.S. for each of the three projection methods. We have pooled five- and ten-year projections, and sorted the regions by their average growth rate over the 1950 to 1980 period. Two features deserve comment.

First, errors are substantially larger in some regions than in others. Bartlett's chi-square test for homogenity of variance (Snedecor and Cochran, 1967) indicates significant differences across the nine regional groups for all three projection methods.

Second, large errors are associated with fast growing regions. The correlations between the average growth rate and the RMSE for geometric, Census, and OBERS projections are 0.83, 0.75, and 0.67 respectively.

To explore the effects of jump-off year, region, and growth rate, and their interactions more fully, we have made analysis of variance and analysis of covariance calculations. Let Δr_{ijk} be the error term for state k in region j of the projection made in jump-off year i. We first fit the model

$$\Delta r_{ijk} = \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

for the five- and ten-year Census and geometric projections, and the five-year OBERS projections. The summary statistics appear in Table 6.

The results indicate that there are substantial differences from region to region in terms of average Δr or bias, and also from year to year. The interaction mean square is significantly larger than the error mean square. This suggests that, even after adjustment for regional and jump-off year effects, projection errors for states within a region tend to be correlated. Every component of the error is larger for the OBERS projections.

Further calculations bring in the effect of the projected growth rate, $\hat{r}$, on the Δr term. Adding this to the analysis as a covariate results in a significant reduction of the residual mean square error in every case.

The projected growth rate also has an effect on the variability of Δr. We find that the standard deviation of Δr tends to be correlated with $\hat{r}$.

V. Confidence Intervals for Future Projections

If one is willing to assume that current projections are no better or worse than those we have examined, our results provide us with the information we need to construct confidence intervals for future projections. The essence of our approach is the realization that population growth is inherently variable. Rather than fruitlessly searching for the "perfect" forecasting method, we concentrate on describing the inherent variability in population trends. Thus a policymaker can decide whether a forecast's accuracy is sufficient for a given purpose, and not be misled by seemingly precise numerical results.

To construct a confidence interval for a new population projection, we can assume that the true, but currently unknown average growth rate, r, has a Normal distribution with mean $\hat{r}$, the projected average growth rate, and standard deviation σ. If z_α is the value such that a standard normal random variable exceeds z_α in absolute value with probability α, a 100(1-α) percent subjective confidence interval for r is

$$(\hat{r} - z_\alpha \sigma,\ \hat{r} + z_\alpha \sigma)$$

and for P_T is

$$(\hat{P}_0 \exp((\hat{r}-z_\alpha\sigma)T),\ \hat{P}_0 \exp((\hat{r}+z_\alpha\sigma)T))\ .$$

Even in complex projection methods where the growth rate is not constant, the average can be computed as

$$\hat{r} = \frac{1}{T} \ln\left(\frac{\hat{P}_T}{\hat{P}_0}\right).$$

The important question is the value of σ.

Figures 1 and 2 summarize graphically the information necessary to examine historical trends in the accuracy of population projections. The OBERS projections are a special case, and we will discuss them below. For the geometric and Census, we see in Figure 1 that there is no discernable trend in the bias of the projections. Figure 2 indicates a slight negative trend in the standard deviation of Δr, although the 1970 projections were uncharacteristically variable. To assess this negative trend, we have calculated a regression of $\ln \sigma$ versus a time index with the value 1 for 1950, 2 for 1955, and so on. The results are:

Geometric - $\ln \sigma = 0.172 - .099t$, $R^2 = .20$, n= 11
Census - $\ln \sigma = 0.078 - .065t$, $R^2 = .11$, n= 10

These figures suggest a drop in the standard deviation of Δr for geometric projections between 1950 and 1980 from 1.075 to 0.590, and for Census projections from 1.013 to 0.686. Of course, these calculations based on only 10 or 11 observations are not very reliable. Thus we have weak evidence that projections are getting more accurate, either because modern techniques are better, or, more likely, because the underlying population dynamics are more stable. The fact that the accuracy of the geometric projections improved, and improved more than the Census projections, suggests the latter explanation.

The second major assumption is that the previously observed Δr's are something like a random sample of what one would expect in the future. In a number of ways, they are. For one, the same set of 48 states was used throughout. One projection method, geometric extrapolation, will not change, and the other is relatively constant in general terms although details change from time to time. But we have found substantial regional differences in bias and variability, so it probably makes sense to construct different confidence intervals for each region. We have also found significant jump-off year and interaction effects, meaning that all of the projections for one geographical region made at one time tend to be off in the same direction. One implication of this is that our estimates of σ are not as precise as we might imagine. Observations from five jump-off-years and six states, yield only five observations of the common bias and thirty observations of the residual variance, rather than thirty independent observations of Δr.

As we have discussed earlier, we have been able to evaluate only two OBERS projections, and the second was substantially more variable than the first. This is most likely due to a change in methodology -- the corresponding Census and geometric projections improved over the same interval. It is thus difficult to guess the accuracy of the 1980 OBERS projections.

Since all of the projections we have examined exhibit some overall bias, but the bias is not in a predictable direction, the root mean square error (RMSE) is a good estimate of σ for constructing confidence intervals (Keyfitz, 1982). For the Census and geometric projections, Tables 2 to 5 indicate that the bias component of RMSE is generally very small.

Without being specific regarding region or growth rate, Table 2 suggests values of around 0.9 percent for Census and 1.0 percent for geometric projections. The trend analysis of Figure 2, and the more recent data in Table 2, suggest lower values of 0.8 percent or even lower for both. For OBERS projections 1.8 percent may be the best guess, but as we have said before, this is highly uncertain.

With more information, one can be more specific about the choice of σ. As we have seen in Table 5, the RMSE varies from 0.5 to 1.7 percent. To construct a confidence interval for, say, Utah, one would use the Mountain States RMSE as σ. For Census projections, this is 1.4 percent, and for geometric projections, it is 1.7 percent. Given increased accuracy for more recent projections, perhaps these should be lower.

Another approach uses the relationship between the error in Δr and the underlying growth rate of the population. A regression of ℓn RMSE versus the average growth rate of Table 5 for geometric and Census projections gives

$$\text{Geometric} - \ell n\ \text{RMSE} = -0.74 + 0.40r,\ R^2 = .50$$
$$\text{Census} - \ell n\ \text{RMSE} = -0.75 + 0.37r,\ R^2 = .74\ .$$

Thus for a Census projection with a growth rate of 1 percent, we would use

$$\sigma = \exp(-0.75 + .37(1)) = 0.68 \text{ percent},$$

and for a growth rate of 3 percent we would use

$$\sigma = \exp(-0.75 + 0.37(3)) = 1.43 \text{ percent}.$$

Average growth rates have been declining over time, a fact which may account for the increasing accuracy we have discussed above. Estimating σ based on the projected growth rate may be essential for future projections that will involve slower growth rates.

Some projections users tend to think of the range of estimates generated by different assumptions and presented in a single report as a sort of confidence interval. The maximum ranges presented in Census Bureau publications, in terms of Δr, range from about 0.3 to 1.0 percent for the average state. Even if σ is as low as 0.8 percent, the probability of the eventual population being contained in the maximum range is less than 15 percent if the range is 0.3 percent, and 46 percent if the range is 1.0 percent. Published ranges of projections are thus substantially narrower than we would want for realistic confidence intervals.

Judgement, of course, should play a role in choosing σ for confidence intervals. Each state is different, and special circumstances were not considered here. The figures presented here provide a starting point. Local knowledge should be used to adjust them to fit the circumstances.

VI. Conclusion

An historical examination of the accuracy of U.S. state population projections reveals two interesting facts. First, simple geometric extrapolation provides forecasts that are as accurate as much more complex demographic models, in terms of estimates of the total population. The even more complex and inclusive OBERS econometric model performs substantially worse in terms of total population. Complexity simply does not pay off. One policy implication is that for allocation of federal funds, simple geometric extrapolation is preferable because it requires no judgement, yet does not suffer in terms of accuracy. Of course, if demographic or economic detail are required, the simple methods are no substitute for the more complex models.

Second, the accuracy of the projections, by whatever method, is influenced by a number of factors, including the year the projection is made, the geographical region, and the underlying growth rate. The last two are related and are important for estimating future accuracy, and constructing confidence intervals.

The figures presented here give a rough indication of our ability to predict future population for states. Errors of 1 and sometimes even 2 percent in annual growth rates of the the same order of magnitude are simply not good. Future population growth and movements are simply very variable, and even our best efforts are not very precise. On the other hand, they are clearly better than the alternatives of assuming no change, or even the same proportional change in every state. Projections may not be perfect, but they are not worthless. The figures we present here will help policymakers decide when and if the accuracy of population forecasts is good enough for their purposes, and help prevent a false sense of security in seemingly precise computer output.

Acknowledgments

The authors are grateful to Ken Feldman and Barry Scribner for help in data collection and preliminary analysis, and to Jane Durch for comments.

References

Griffith, Jeanne E. (1980), "Standardizing population projections required in federal fund allocations," Statistical Reporter, 80-4: 57-63.

Keyfitz, Nathan (1982), "The limits of population forecasting," Population and Development Review, 7:579-593.

Long, John F. (1977), "Prospects for a composite demographic-economic model of migration for subnational population projections," Report of the Conference on Economic and Demographic Methods for Projecting Population, Washington: American Statistical Association.

Snedocor, George W. and Cochran, William G. (1967), Statistical Methods, Ames, Iowa: Iowa State Press.

Stoto, Michael A. (1982), "The accuracy of population projections," forthcoming in Journal of the American Statistical Association.

Tukey, John W. (1977), Exploratory Data Analysis, Reading, MA: Addison-Wesley.

U.S. Bureau of the Census (February 1981), 1980 Census of the United States.

U.S. Bureau of the Census, Current Population Reports, Series P-25, Numbers 56, 70, 110, 126, 227, 229, 375, 380, 460, 477, 488, 727.

U.S. Bureau of Economic Analysis (1974), 1972 OBERS Projections Series E, States, Vol. 4.

U.S. Bureau of Economic Analysis (1977), "Population, personal income and earnings by states projections to 2000," (unpublished).

Table 1: Projection Series

Series	Jump-Off Year	Target Years	Assumptions and Source
Geometric	'50	'55,'60	'40,'45 population from CPR P-25 #72; '50 population from CPR P-25 #56
Geometric	'55	'60,'65	'45 population from CPR P-25 #72; '50,'55 population from CPR P-25 #160
Geometric	'60	'65,'70	'50,'55 population from CPR P-25 #160;'60 population from CPR P-25 #375
Geometric	'65	'70,'75	'55 population from CPR P-25 #160;'60,'65 population from CPR P-25 #375
Geometric	'70	'75,'80	'60,'65 population from CPR P-25 #375;'70 population from CPR P-25 #796
Geometric	'75	'80	'65 population from CPR P-25 #375;'70,'75 population from CPR P-25 #796
Census	'50	'55,'60	Medium fertility assumption from CPR P-25, #56, 1952
Census	'55	'60,'75	Average of high and low fertility assumptions (Series B and C) from CPR P-25 #126, 1955
Census	'65	'70,'75	Average of high and low fertility assumptions (Series I-B and I-D) from CPR P-25 #375, 1967
Census I	'70	'75,'80	'60-'70 migration patterns average of high and low fertility assumptions (Series I-C and I-E) from CPR P-25 #477, 1972
Census II	'70	'75,'80	No interstate migration, average of high and low fertility assumptions (Series III-C and III-E) from CPR P-25 #477, 1972
Census I	'75	'80	Average of '65-'75 and '70-'75 migration patterns (Series II-A and II-B) from CPR P-25 #796, 1979
Census II	'75	'80	No interstate migration (Series II-C) from CPR P-25 #796, 1979
OBERS	'70	'75,'80	1972 OBERS Projections Series E (published in 1974) interpolated to '75 from Long (1977)
OBERS	'75	'80	1977 OBERS Projections, not adjusted for undercount

NOTE: CPR refers to the Bureau of the Census Current Population Reports series.

Table 2: Summary Measures of Δr by Projection Method and Duration (1955-1975)

Projection Method	Duration	N	Bias	Std. Dev.	RMSE
Geometric	5 years	240	-0.067	1.030	1.032
Geometric	10 years	192	0.252	0.870	0.906
Census	5 years	288	-0.200	0.874	0.897
Census	10 years	192	-0.143	0.849	0.861

Table 3: Summary Measures of Δr by Projection Method and Duration (1970-1975)

Projection Method	Duration	N	Bias	Std. Dev.	RMSE
Geometric	5 years	96	-0.251	0.777	0.816
Census	5 years	192	-0.318	0.769	0.832
OBERS	5 years	96	-0.465	1.737	1.798

Table 4: Summary Measures of Δr by Projection Method and Jump-Off Year

Projection Method	Years	N	Bias	Std. Dev.	RMSE
Geometric	'50-'55	48	2.153	1.181	2.456
Geometric	'55-'60	48	-0.553	0.921	1.074
Geometric	'60-'65	48	0.132	1.068	1.076
Geometric	'65-'70	48	0.590	1.182	1.321
Geometric	'70-'75	48	-0.360	1.040	1.101
Geometric	'75-'80	48	-0.141	0.340	0.368
Census	'50-'55	48	-0.338	1.348	1.390
Census	'55-'60	48	-0.064	1.308	1.310
Census	'65-'70	48	0.138	0.595	0.611
Census I	'70-'75	48	-0.142	0.893	0.904
Census II	'70-'75	48	-0.336	0.760	0.831
Census I	'75-'80	48	-0.347	0.532	0.635
Census II	'75-'80	48	-0.448	0.835	0.948
OBERS	'70-'75	48	0.230	0.679	0.717
OBERS	'75-'80	48	-1.230	2.154	2.480
Geometric	'50-'60	48	0.347	1.006	1.064
Geometric	'55-'65	48	0.590	0.813	1.005
Geometric	'60-'70	48	0.150	0.590	0.609
Geometric	'65-'75	48	0.474	0.793	0.924
Geometric	'70-'80	48	-0.207	1.022	1.043
Census	'50-'60	48	-0.553	0.955	1.104
Census	'55-'65	48	-0.181	0.696	0.719
Census	'65-'75	48	0.008	0.455	0.455
Census I	'70-'80	48	-0.261	1.020	1.053
Census II	'70-'80	48	-0.137	1.077	1.086
OBERS	'70-'80	48	-1.108	1.902	2.201

Table 5: Summary Measures of Δr by Projection Method and Region

Projection Method and Region	r	N	Bias	Std. Dev.	RMSE
GEOMETRIC					
Mountain	2.51	72	-0.013	1.736	1.736
Pacific	2.18	27	0.062	1.011	1.013
South Atlantic	1.77	72	0.050	0.876	0.877
W. South Central	1.47	36	-0.283	0.644	0.704
E. North Central	0.96	45	0.414	0.452	0.613
New England	0.94	54	0.069	0.702	0.705
E. South Central	0.80	36	-0.235	0.646	0.687
W. North Central	0.63	63	0.196	0.587	0.619
Middle Atlantic	0.58	27	0.445	0.611	0.756
CENSUS					
Mountain	2.51	80	-0.897	1.070	1.396
Pacific	2.18	30	-0.249	0.830	0.867
South Atlantic	1.77	80	-0.200	0.748	0.774
W. South Central	1.47	40	-0.395	0.869	0.955
E. North Central	0.96	50	0.333	0.874	0.935
New England	0.94	60	0.019	0.728	0.728
E. South Central	0.80	40	-0.090	0.507	0.515
W. North Central	0.63	70	0.030	0.388	0.389
Middle Atlantic	0.58	30	0.350	0.686	0.770
OBERS					
Mountain	2.51	24	-1.655	0.999	1.933
Pacific	2.18	9	-1.057	0.651	1.241
South Atlantic	1.77	24	-0.660	2.316	2.408
W. South Central	1.47	12	-1.223	0.427	1.295
E. North Central	0.96	15	0.657	0.370	0.754
New England	0.94	18	-0.088	1.436	1.439
E. South Central	0.80	12	-0.291	0.473	0.555
W. North Central	0.63	21	-0.381	0.775	0.864
Middle Atlantic	0.58	9	1.016	0.218	1.039

Table 6: Analysis of Variance by Projection Method and Duration

Projection Method	Duration	N	Mean Square Error			
			Region	Year	Interaction	Error
Geometric	5 years	240	0.782	9.550	1.357	0.850
Geometric	10 years	192	1.773	6.161	1.027	0.560
Census	5 years	288	4.318	2.287	1.147	0.544
Census	10 years	192	4.733	0.613	1.305	0.427
OBERS	5 years	96	14.052	46.469	6.703	0.952

Figure 1. Bias of Geometric (A) and Census (B) Projections by Jump-Off Year

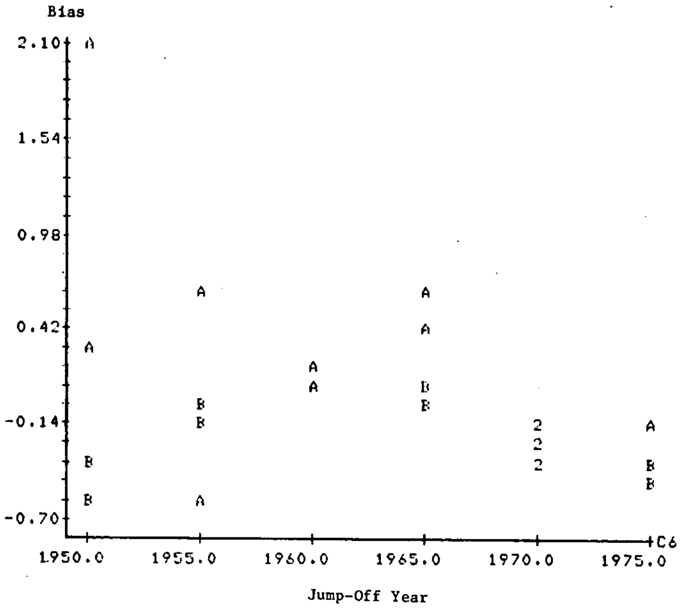
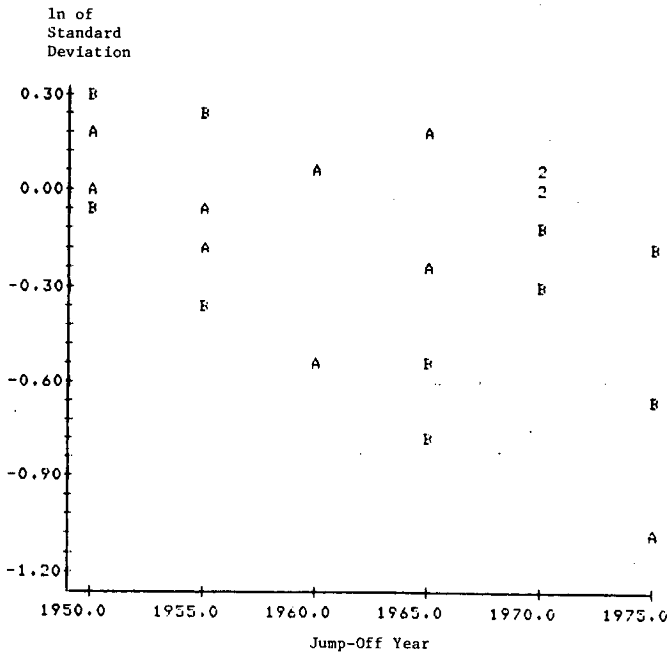


Figure 2. Standard Deviation of Geometric (A) and Census (B) Projections by Jump-Off Year



NOTE: A "2" indicates that two points appear at essentially the same location.