

DISCUSSION

Paul Burke, U.S. Department of Housing and Urban Development

This discussion covers the Canadian and United States papers on estimates of migration between censuses, and the Wisconsin paper on estimates of population for small areas between Censuses [1].

The Canadian and U.S. papers have complementary strengths and complementary weaknesses, which are discussed first. The discussion of errors in the Wisconsin paper is illuminating in its own right, and has implications for the Canadian and U.S. papers as well.

I. CANADIAN AND U.S. ESTIMATES

Norris-Britton. This paper compares empirically five ways of estimating Canadian migration, using two different administrative data bases. It reminds us that there are usually several good approaches to a problem, and the best test is an empirical one. The paper would benefit from a discussion of why some of these methods in some provinces are better than others. Presumably different kinds of people migrate to some provinces, so some of the data bases pick up these migration flows better than others. For example estimates based on child allowances would be less reliable where migrants are usually childless, such as possibly Alberta and the Northwest Territories. A theoretical discussion should explain these patterns, and might also suggest other data bases that should be explored to compensate for weaknesses in the data used so far.

It also is possible that different methodologies would be appropriate for different provinces. The paper notes that different methodologies are already used for population estimates within some provinces. There seems no technical reason not to use different methodologies in estimating migration too, as long as the differences are carefully justified in terms of producing more accurate data for each province. One would require the total domestic migration leaving all the provinces to equal the total domestic migration arriving in all provinces (netting out international migrants). However, this constraint could be satisfied by estimating each one-way flow, say from Quebec to Alberta, separately, using the methodology that best fits that particular flow. Then all the totals would add up correctly.

Word - Zitter. This paper is complementary to the Canadian one in offering a careful discussion of why the U.S. estimates work out the way they do. The authors discuss in detail errors induced by census undercounts, possible overcounts, shifting migration patterns, problems of racial identification, and file-matching problems. In this discussion the United States paper gives valuable information, which would have been similarly useful in the Canadian paper.

The estimates the paper gives of black population are good, especially considering the caveats the authors point out. Even better are the estimates of blacks as a percentage of total popula-

tion. This percentage is fairly accurately measured, and is of wide interest. The paper sets itself an unnecessarily high standard by trying to measure changes in this percentage, which are less important than the percentage itself, and make the methodology appear less useful than it is.

As the authors point out, once the methodology is developed, it can be applied to other groups, such as the elderly, the young, whites, perhaps the poor. Some of these groups may be even more important to track than blacks.

On the other hand, the United States paper uses only one basic approach to population estimates, only one data source, and lacks the ability of the Canadian paper to compare different approaches. Social Security earnings records could be one alternative; private marketing data such as the Donnelly data described in another paper in this session would be another possible data source. Much of the interest in the Canadian paper comes from its comparison of different data bases, and the same approach would be equally valuable in the U.S. work.

Different techniques are particularly important in any measurement of the migration of hispanics, an issue which will become more and more important in this decade. The data on blacks depend on codes assigned when people applied for their social security numbers, often decades ago. These data probably would not measure hispanic status accurately [2].

Common Weaknesses in the Word-Zitter and Norris-Britton Papers. Both papers evaluate migration estimates by adding net migration to other components of population change, comparing the resulting total population to a census count, and attributing any discrepancy to the migration estimates. This is wrong. They are asking migration data to estimate net migration accurately and also to estimate all the other errors in the other components. Since we know that other errors are large, such as the change in the census undercount in the U.S., and perhaps international migration in both countries, it is misleading to judge estimates of domestic migration by how well they make the total population come out.

Migration estimates can and should be compared directly to census measures of migration. This test is important enough that it should wait, if necessary, for the census migration data to become available. Other small area data bases can also be used for tests in the meantime, such as the SMSA samples of the Annual Housing Survey in the U.S.

If one wants to eliminate other errors in estimating total population, a much more reliable approach would be to estimate each error separately and study its trend, rather than assuming that all these other errors will for some reason follow the trends of domestic migration.

When comparing migration estimates one also should be careful not to assume the census is always right. Censuses ask people to remember

when they moved and where they moved from. Administrative data show where people actually are when the forms are filled out, and should inherently be more reliable. Part of the testing process should be to compare groups which are identical on both files (such as children in the Canadian Census and family allowance records), to test the accuracy of the census.

A common aspect of the Canadian and United States papers is that the multipliers, to obtain total migration from the subset of migrants covered by administrative data, are not shown explicitly. The stability of these multipliers is crucial, since each country plans to use past multipliers to estimate future trends. It would be very helpful to show these multipliers explicitly for various areas at various times. This idea is expanded in my comments on the Palit, et al. paper below.

II. WISCONSIN METHODOLOGICAL PAPER

Palit, et al. The paper makes a very important contribution by giving a formal discussion of the types of errors involved in using administrative data.

One of these potential errors is lag in administrative data. Lag is reasonably straightforward. Economists deal with lagging indicators all the time, and there is no particular problem, provided the lag is respected in making estimates. Economists also deal with leading indicators, and in administrative data we also have leading indicators. Purchases of maternity clothes or attendance at pre-natal clinics should be leading indicators of births.

Two other potential errors identified by the authors are random variation and systematic change. Systematic change happens when a law suddenly makes 16 year olds eligible for drivers' licenses. Random variation happens when slightly more (or fewer) people happen to apply for licenses in a year. The authors use a variance calculation to show that errors due to random variation

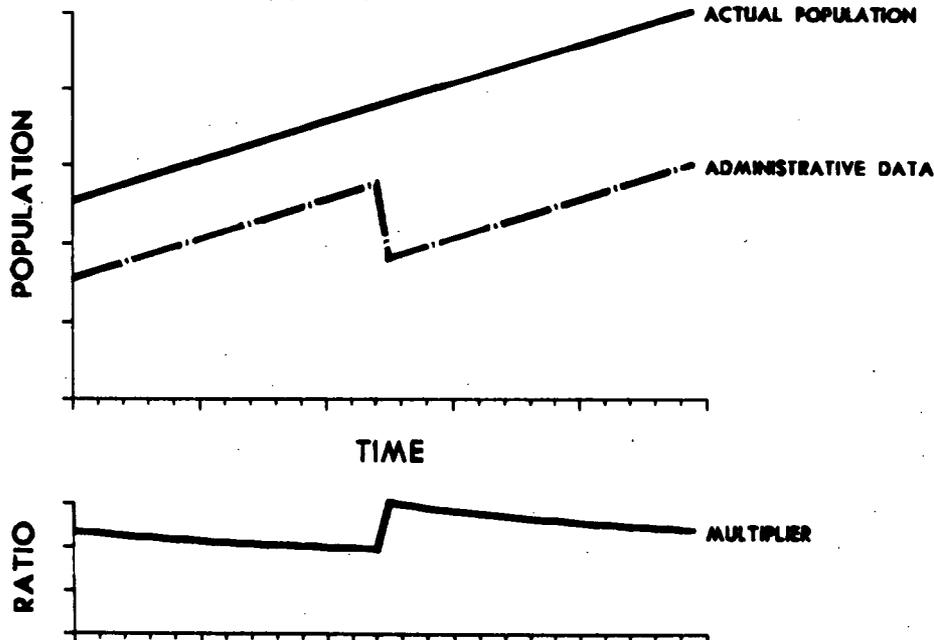
can be minimized by using as large an administrative data base as possible, relative to the population. This rule of thumb seems plausible, but I am not convinced by the variance calculation, since the administrative data are not in fact a random sample: people get drivers' licenses, and enter other administrative systems, for particular personal reasons, which may or may not act like a random sample.

The authors are right, however, in distinguishing small ripples in administrative data from major systematic shifts. Small ripples can be adjusted by any good smoothing function, as the authors point out. The multiplier to estimate total population need not be affected. Systematic shifts, however, require a counterbalancing shift in the multiplier, which is hard to do accurately.

The analysis in the Palit, et al. paper has two important methodological implications. Since the multiplier may be stable, may trend up or down, or may shift sharply, the first implication is that one should measure this multiplier as often as possible, from censuses, partial censuses, and other surveys, to see how stable it is. If it is stable or has a steady trend, it can be used confidently as long as the analyst projects any trend forward. The second methodological implication is that the analyst should constantly watch for events which can shift the multiplier suddenly, such as new laws, new enforcement, or social change. Sometimes a new multiplier cannot be calculated after such a shift (say, until the next census). More often, however, one can measure the shift, and estimate a new multiplier reasonably well.

Figure 1 shows some of the patterns in hypothetical administrative data and a hypothetical multiplier. It shows the importance of watching both the administrative data and the multiplier. It is for this reason that formal presentation of the multipliers found in the Canadian and United States studies would be useful, to show how much they differ from area to area and from one time period to another.

FIGURE 1. EXAMPLE



EDITORS' NOTES

[1] The remaining two papers at this session: those by Mary Kay Healy and by Michael A. Stoto and Alicia P. Schrier were discussed by Edward J. Spar. He has not written up his comments for the Proceedings; hence, they are not included in this report.

[2] Some recent work on hispanic origin is reported on by Jeffrey S. Passel and David L. Word, U.S. Bureau of the Census, in "Constructing the List of Spanish Surnames for the 1980 Census: An Application of Base Theorem," presented at the 1980 annual meetings of the Population Association of America in Denver, Colorado.