mind some model, of the form

$$\text{undercount rate} = f(\text{SUB}, \text{ALL}, \text{MULT}, \text{NM}, \text{MOB}, \text{POV}, \text{Error}).$$

The analysis of DIFF in their Table 9 suggests something of this form with f as a linear function with invented coefficients! What evidence do we have for believing such models and how much credence should we give to the analysis of the component parts?

Freedman and Wachter conclude that the 357 poststrata have too much residual variability on the proxy variables within state. This comes as no surprise to us. In its original analyses following the 1990 census, the bureau used 1,392 poststrata and then smoothed the resulting adjustment factors to remove variability. The later approach, which dropped the number of poststrata from 1,392 to 357, was sure to introduce greater heterogeneity and the "random" features in it are no longer kept under control through smoothing. As Fay and Thompson (1993) note: "Although [census analysts] eliminated smoothing from consideration in 1992, there may have been hidden costs to this decision. . . ."

The statistical literature clearly suggests that, even if FW were successful in showing substratum heterogeneity, by doing so they may have strengthened rather than weakened the case for the use of adjusted data.

Finally, we wonder what FW's analysis really has to say about the wisdom of using adjusted counts for various intercensal purposes. As we move further into the decade and away from the April 1, 1990, census date, there must be more and more error in census counts, both adjusted and unadjusted. Does there come a point when the cumulative errors due to the passage of time swamp the undercount problem? Or does the differential undercount between the white majority and various minority groups that we have observed for over 50 years in decennial census data only become worse?

Comment

Lars Lyberg and Sixten Lundström

The American census adjustment debate must represent the pinnacle of statistical methodological controversy. Usually, statistical discourse is conducted by laconic academics who address technical issues of obscure merit to nonstatisticians. Indeed, it is an anomaly for the profession that an essentially technical issue, such as census adjustment, would attract such widespread and vocal attention.

Our comments should be prefaced with the fact the Swedish censuses are not affected by the type of undercoverage that characterizes the U.S. census undercount. Our approach to census taking is vastly different from that used in America. Sweden is known for its high-quality population registers and uses a register-based approach for the actual count. Since it is extremely difficult to function in Swedish society without a personal identity number (PIN) (and many of the benefits and amenities offered by Swedish society require a PIN), every legal resident is included in the population register. Any undercount is very small (a few hundred) and is linked to lags or delays in the reporting of vital statistics. These delays usually do not last more than 10–13 weeks; so, both in principle and practice, Sweden can conduct an accurate population census any week of the year. A word of caution, though, over the last few years, we have had an increasing problem with overcoverage due to immigrants who repatriate without notifying the authorities.

The United States, on the other hand, lacks Swedish-style population registers and bases its census on a master sample, that is, tracking everyone down by figuring out the number and location of dwellings and then ascertaining who and how many live in a given dwelling. This is obviously a daunting task when multiplied by an entire nation of geographically and ethnically diverse individuals.

When studying the articles by Freedman and Wachter, by Breiman, by Belin and Rolph and by numerous others dealing with the U.S. census adjustment, we have made a number of general observations. First, it should be kept in mind that statistics is the theory and practice of dealing with uncertainty. Second, surveys never produce "true" numbers. What surveys do produce are estimates and every source available should be used to make

Lars Lyberg is Head of Research and Development at Statistics Sweden and is Chief Editor of the Journal of Official Statistics. Sixten Lundström conducts work at the Statistical Research Unit at Statistics Sweden.

these estimates more accurate. Modelling is used at every stage of survey design. When new things are learned, these models are revised to reflect the insights provided by the new information. It can be seen as an iterative process moving toward certainty, but never actually getting there. Third, applied work is very different from theoretical work. Theoretical work is very neat and tidy in that the most important elements are defined by the scientist him/herself. Theoreticians state the premises under which their findings are valid, thereby conducting their research under highly controlled conditions. The practitioner, on the other hand, must do battle with the vicissitudes of daily life. Applied work has the added element that almost anything, regardless of how important, can spin out of control with no regard for the resulting inconvenience or damage.

Fourth, all sciences are subject to controversies; such debate is necessary, and examples from statistics include Fisher versus Neyman, frequentists versus neo-Bayesians, model-based samplers versus design-based samplers and now census adjusters versus census nonadjusters. Furthermore, scientific debates can contain nonscientific ingredients. Perhaps in cases where science cannot reach a consensus, it is best to turn the discussion over to politicians.

The statistical information collected in the U.S. census is used in many different areas, and many different political decisions are based on it. The undercount problem, however, is described only in connection with its use for congress apportionment and the allocation of tax funds. We have not found any description of the "models" used for these purposes, but they are probably not particularly complicated. On the other hand, equitable and fair allocation of funds is another issue with its own difficulties. Perhaps the main problem is not the census undercount, but rather the oversimplified allocation model? If this is the case, the problem really belongs in the political sphere.

In our opinion, the U.S. census must be the most complex survey in the world. There are other surveys that are larger (e.g., the Chinese census), but none comes close to the level of ambition of the U.S. census. The combination of problems associated with population mobility, lack of reliable and complete registers, illegal immigration, homeless people, enumerator security and demands for high-quality data makes the U.S. census an enormous undertaking. The Census Bureau must be commended for its skill and competence in dealing with this task. From the start of the modern survey era, the bureau has been on the forefront, and preeminent statisticians like Morris Hansen, Bill Hurwitz, Bill Madow and Barbara Bailar are among those who have managed the bureau's continuous development. The Census Bureau is the agency where the most extensive and complicated evaluation studies have been conducted. Due to the agency's efforts, the entire survey community has learned a great deal about survey errors. The bureau has led research on, for instance, survey models, processing errors, coverage errors, response variance, estimation and the use of new technology.

It does not make sense to criticize the Census Bureau for collecting bad data. The Post Enumeration Survey (PES), like all surveys, suffers from errors. However, these errors are smaller than those generated in the census. The absolute best a survey practitioner can do is use the preferred procedures. In the case of the U.S. census, the PES is a preferred procedure which provides information that can be used to increase the accuracy of the census count. We believe it is important to appreciate that the PES is not meant to generate true numbers, but rather numbers closer to the true numbers. When discussing the U.S. undercount problem, or the quality of any survey for that matter, it is vital to recognize the limitations inherent in survey practice and view the survey results in this light. It is also germane to this discussion to point out that data of lesser quality than the PES are used every day with excellent results and no complaints.

The article by Freedman and Wachter provides an example of shortcomings associated with evaluation surveys. The PES provides acceptable or perhaps even unbiased estimates for the nation and for some other very large domains. For small domains, model-dependent estimators have to be used. Such estimators include some bias because of model error. For census adjustment, the bureau uses a type of synthetic count estimator. The model error associated with this estimator is due to heterogeneity in poststrata. The heterogeneity, measured by proxy variables, has been shown to be quite large within some poststrata.

At Statistics Sweden, we use a synthetic count estimator regularly when estimating the number of employed persons by different categories of hours worked in each municipality. Variables defining the poststrata are sex, age, industry and income. From a methodology study where the parameter values were already known, we learned that the estimator worked well for the majority of municipalities, but performed poorly for a few municipalities. Despite these results, this method has not received much criticism. We believe that users are aware that this estimate is the best we can do and it is better than no estimate at all.

This leads us to another comment. A census should result in a single published count. Publishing two counts gives the impression that users have a "choice" even though the producer of the cen-

sus never intended such an interpretation. In the 1970 Swedish census, Statistics Sweden presented two numbers: one regular set of estimates with missing data and one with imputed values added. Surprisingly, many users (but perhaps not so surprising after all) knew exactly which estimate to use. In the 1975 census, imputation was not performed, which made comparison to the 1970 census awkward.

We find it both reasonable and natural to use auxiliary information to improve an estimate. After all, this is what survey design is all about. Various model assumptions are made in every design step, but the final result should be expressed as a single count or estimate. We sympathize with Belin and Rolph regarding their general conclusion about the protracted controversy on the undercount problem. An impressive amount of work has been done, but it appears as if we have reached the point where further methodological resources, time and money would be a waste.

Most U.S. statistical agencies have committed themselves to modern quality thinking, that is, various forms of total quality management. It seems as if it would be better to use the "debate resources" to improve the regular census count procedures, thus decreasing the need for extensive and expensive evaluation procedures. This is especially true for the U.S. undercount, where the discussion fails to result in a consensus.

Comment

David Steel

Evaluating and possibly adjusting the census for undercount raises a lot of difficult statistical and general issues. The papers here consider several of these and add to the already large literature on the subject. While the basic questions are now clear, the answers are not. To enable readers to make a judgement about any prejudices I might have on these issues, I should point out that as a former officer of the Australian Bureau of Statistics (ABS) I was involved in the evaluation of the 1981 and 1986 censuses. While the views I have are entirely my own, they are influenced by this past involvement. In terms of my prejudices this could work either way: having been involved in adjustment, I may have a bias to that view to justify my past work; alternatively, detailed knowledge of the many problems involved could lead me to be against adjustment.

In Australia, population estimates based on census counts adjusted for undercount have been released as the official population estimates since 1976. The estimates are produced for states and territories and local government areas. Population estimates are used to determine the number of seats each state has in the federal House of Representatives and the allocation of funds to states and local government areas. The decision to adjust was prompted by the high undercount rate showed by the 1976 Post Enumeration Survey (PES) and the fact that the 1976 census count fell considerably below the population estimates for 1976, which were based on updated 1971 census results. There has been general acceptance and remarkably little controversy surrounding the adjustment. A clear distinction is made between census counts and population estimates. Census counts are produced without any adjustment. There are similarities to the situation in the United States. The level of undercount is basically estimated from a PES which involves an independent household survey and matching between the census and the survey to determine missed people and some categories of erroneous enumerations. Dual system estimates (DSE's) are calculated. The results of the PES are compared with demographic analysis and other population indicators such as school enrolments and Medicare enrolments (Medicare applies to all age groups), primarily at the national level, but with some analysis below this. Synthetic estimation is used to obtain population estimates for local government areas. The procedures for the PES and census evaluation are decided in advance. The view is that quality must be designed into the census and the PES. The estimated level of net undercount is remarkably similar to the United States: 1.9% in 1986 and 1.8% in 1991. In 1991 the state undercount rates ranged from 1.2 to 4.1%. The ranking of the states in terms of undercount has been consistent over time. Further details are given in Choi, Steel and Skinner (1988), Trickett (1992) and Australian Bureau of Statistics (1990).

David Steel is Senior Lecturer, Department of Applied Statistics, University of Wollongong, Northfields Avenue, Wollongong, New South Wales 2522, Australia.