Survey Methodology, June 1992
Vol. 18, No. 1, pp. 155–163
Statistics Canada

A Variation of the Housing Unit Method for Estimating the Population of Small, Rural Areas: A Case Study of the Local Expert Procedure

LINDA K. ROE, JOHN F. CARLSON and DAVID A. SWANSON[1]

ABSTRACT

This paper examines the suitability of a survey-based procedure for estimating populations in small, rural areas. The procedure is a variation of the Housing Unit Method. It employs the use of local experts enlisted to provide information about the demographic characteristics of households randomly selected from residential unit sample frames developed from utility records. The procedure is nonintrusive and less costly than traditional survey data collection efforts. Because the procedure is based on random sampling, confidence intervals can be constructed around the population estimated by the technique. The results of a case study are provided in which the total population is estimated for three unincorporated communities in rural, southern Nevada.

KEY WORDS: Survey-based; Utility records; Confidence intervals; Nevada.

1. INTRODUCTION

In its most recent survey of state and local agencies preparing population and housing estimates, the U.S. Bureau of the Census found that about 89 percent of the agencies surveyed use the Housing Unit Method (HUM) (Byerly 1990). This method was also found to be widely used in an earlier survey (U.S. Bureau of the Census 1978). The method has been found to provide accurate estimates of the total population (Lowe, Pittenger and Walker 1977; Lowe, Weisser and Myers 1984; Smith and Lewis 1980, 1983; Smith and Mandell 1984) as well as a strong conceptual and practical foundation for a municipal estimation system (Martin and Serow 1979; Rives and Serow 1984; Swanson, Baker and Van Patten 1983).

One of the strong features of the HUM is that it can be implemented in a variety of forms, which allows it to be adapted to a range of data environments (Swanson, Baker and Van Patten 1983). This adaptability has been exploited primarily by subnational demographic centers for purposes of revenue sharing and related programs (Martin and Serow 1978; Swanson, Baker and Van Patten 1983). However, as pointed out by Rives (1982), the method has potential uses in other arenas.

As an example, consider the case of environmental impact statements. Concerns over legal and environmental issues have resulted in decisions to locate unpopular facilities in sparsely populated rural areas for which census and other socioeconomic data are usually not available (Freudenburg 1982; Brown, Geertsen and Krannich 1989; Munsell 1988). As a consequence, it has become necessary to develop methods of inquiry, particularly suited for small, rural areas, that fully exploit available data, are less costly and, in many cases, less intrusive, than area, telephone, and mail surveys. We believe that the variation of the HUM that we propose in this paper contributes to this type of methodological development.

[1] Linda K. Roe and John F. Carlson, Science Applications International Corporation, 101 Convention Center Drive, Las Vegas, Nevada 89109; David A. Swanson, Center for Social Research and Department of Sociology, Pacific Lutheran University, Tacoma, Washington 98447-0003, USA.

The HUM variation that we describe in this paper combines two methods that are, in themselves, well known. However, they have largely been developed in isolation from each other, as well as from the HUM. These are: (1) random sampling; and (2) "local expert" interviews. As discussed later, these methods, combined with the HUM may lead to a means of obtaining the population size and, eventually, composition data required to meet the information needs of impact assessment projects and other activities affecting small, rural areas.

2. CONSIDERATIONS IN ASSESSING IMPACTS IN SMALL, RURAL AREAS

The location of new plants or industries in rural areas generally requires a work force exceeding that which is available in the local area. Population growth in the communities that are in close proximity to the site can be expected to vary according to the size of the project and the number of employees that will be hired to build, then to operate and maintain the completed facility. Whether rapid increases in the overall number of individuals are expected, or significant changes in the age and sex distribution, the altered population structure will have an effect on the type and amount of public services needed (Summers 1982). Thus, impact assessments require information regarding anticipated increases in school enrollment, housing requirements, health care needs, and other services. Before such projections can be made, however, information on the current population in the impacted area must be determined in order to have a "jump-off" or "launch" population for forecasting purposes (Carlson, Williams and Swanson 1990; Pittenger 1976; U.S. Department of Energy 1988).

The understanding of major factors affecting the distribution of people in isolated rural areas is critical in constructing demographic profiles and projections. These communities are likely to have been affected by periods of both boom growth and decline (Krannich and Greider 1984). Historical patterns of population change, as well as current trends, may differ substantially from averages derived from that of the county as a whole or even other sub-county areas. This presents a special problem because accurate demographic information is usually available only for years in which the Federal Census is conducted. However, census data, including information on households, are not typically available for unincorporated places with small populations. Since cost is usually a major factor, the possibility of conducting special censuses or large sample surveys, particularly on a regular basis, is often precluded, even in small, rural areas. An additional problem associated with such counts is that they require interviewers to contact individual households, which imposes on time and privacy and adds to disruption burdens that may be already high for local residents (Brown, Geertsen and Krannich 1989; Krannich, Berry and Greider 1989; Schleifer 1986).

The estimation of the size of the current population of a small, rural area could, in principle, be accomplished through several techniques. However, data limitations and a desire for accuracy severely curtail the range of candidates and, realistically, point to a single technique: HUM (Smith 1986; Smith and Mandell 1984; Lowe, Weisser and Myers 1984; Swanson, Van Patten and Baker 1983; Smith and Lewis 1983, 1980).

3. THE HOUSING UNIT METHOD

The concept of the HUM relies on the fact that nearly everyone sleeps under some kind of shelter. The U.S. Bureau of the Census, for example, chooses to define two classes of shelters: group quarters; and housing units. All persons are assigned to one shelter class or the other. The HUM holds that these shelters can be identified, counted, and classified as occupied or vacant. Also, all occupied shelters must have a specific number of occupants. Therefore, the population of any given place must be equal to the sum of the housing units times the occupancy rate times the average number of persons per occupied housing unit (household) plus the number of persons in group quarters. The four elements of the HUM provide an exact demographic identity, with the population of a given place given by

$$P = [(H) * (O) * (PPH)] + GQ,$$

where

$$\begin{aligned} P &= \text{total population,} \\ H &= \text{total housing units,} \\ O &= \text{proportion of occupied units,} \\ PPH &= \text{mean number of persons per household,} \\ GQ &= \text{group quarters population.} \end{aligned}$$

The key accuracy issue in using the HUM is in the determination of each of the components. Moreover, as Smith (1986, pp. 245-246) observes:

> "The Housing Unit Method is a robust, comprehensive, and extremely flexible form of population estimation with a number of characteristics that make it useful for small-area analysis. It is not confined to a single technique or type of data; rather, it can utilize a number of different techniques and data sources, including those that may be applicable in one area but not another."

As also noted by Smith (1986), there are two major approaches used to generate the "number of households" element of the HUM. One relies on measures of construction activity and the estimation of an occupancy rate; the other uses utility data, such as residential electrical customers. A major advantage of the second approach is that it can directly provide the number of households, which eliminates or substantially reduces a number of potential data inaccuracies, including the need to estimate time lags between when permits are issued and units are completed, completion rates, demolitions, conversions, and occupancy rates. Starsinic and Zitter (1968) as well as Rives and Serow (1984) find that the "utility data" approach to the HUM is advantageous, although they also acknowledge certain limitations.

Another advantage of using utility data is that the same data used to obtain total households can also be used as a complete frame from which samples can be drawn in order to obtain an estimate of the average number of persons per household (*PPH*). There are three forms that traditional data collection usually take in obtaining this type of sample information: mail, telephone, and personal interview. We propose that in their place "local experts" be used to minimize both cost and disruption burdens.

4. LOCAL EXPERTS

The local expert procedure (also refered to as the key informant procedure) of obtaining information about a community is well-established in the field of cultural anthropology. It is generally acknowledged as a "reliance on a small number of knowledgeable participants, who observe and articulate social relationships for the researcher " (Seidler 1974, p. 816). Further, Poggie (1972) finds that when the questions asked in the field relate to noncontroversial, concrete, and directly observable public phenomena, local experts are a highly reliable and precise source of information.

There are two key issues in using the local expert procedure in conjunction with utility records and the HUM. The first is to identify and recruit people who are truly local experts on the composition of the households presented to them in the sample. The second is to be able to obtain household identifying information that is familiar to the local experts (*e.g.*, a street address and the name of the householder instead of a utility company billing code).

5. CASE STUDY

The data collection activity on which our population estimates rely is part of a program to assess the socioeconomic characteristics of communities located near Yucca Mountain, Nevada, the proposed site of a geologic nuclear waste repository (U.S. Department of Energy 1988). The data will comprise part of the set used in a comprehensive impact analysis of the proposed repository.

Yucca Mountain is located in Nye county, approximately 90 miles northwest of Las Vegas in a sparsely populated, desert area. The impact analysis is focused on the communities that are within a fifty mile radius of the Yucca Mountain site. The study areas includes the unincorporated communities of Amargosa Valley, Beatty, and Pahrump in southern Nye county and Indian Springs in Clark county. Tax boundaries specified by the county commissioners are used to deliniate community boundaries for purposes of the impact analysis.

6. DATA AND METHODS

During a preliminary phase of the research, contacts were made with community leaders and residents. These contacts resulted in a network that later facilitated the collection of data. Field notes were taken describing the general layout of each community in the study area. These included the types and locations of businesses and residential areas. Four separate housing types were defined using the guidelines developed by the U.S. Bureau of the Census.

Following the preliminary investigation, the road system and other features were mapped for each community. Using these maps and utility records, representatives of the electrical company servicing southern Nye county identified the location and type of housing, if any, associated with all current electrical connections. This information was added to the housing unit file constructed from the utility records for each community. Because of the lack of adequate utility records for Indian Springs, housing information for this area was collected by a "windshield survey," a systematic, block-by-block canvassing of housing units by teams operating from automobiles (Lowe, Pittenger and Walker 1977). As a consequence, Indian Springs is not included in the test results reported in this paper.

The preliminary fieldwork indicated that substantial differences in *PPH* could be expected across the communities in the study area. Thus, a random selection of units from the housing unit file was drawn separately for each community, based on the number of housing units in each community. A conservative approach was used to determine the size of each community's sample. It assumed a 5% margin of error, a significance level of .05 and interest in a dichotomous variable with a 50-50 distribution (Cochran 1977). Once the initial size was determined, an additional 15% was added to allow for missing cases. The final sample size for Amargosa Valley was 175 housing units, for Beatty it was 222, and for Pahrump, 355.

Local experts were initially identified through the contact network on the basis of their experience in community activities and their familiarity with local residents. Each potential expert was interviewed and asked to complete a form designed to assess their qualifications. A written explanation of the project and specific instructions regarding the data collection process were provided and discussed. The persons selected as local experts were given instructions regarding confidentiality. For this project, we found that the "meter readers" employed by the local utilities constituted a good source of local experts. The local experts were provided with the sample set of housing units for their respective communities. In most cases, two local experts worked together, which made it possible to verify the accuracy of information as it was recorded. For each unit, the local experts communicated to the researcher only the number of persons in the household as of July 15, 1990, the age (using eight age groups) and gender of each household member, and the retirement status of any member fifty years of age and over. If either of the two local experts was unsure about the composition of a given household, another member of the community was contacted to confirm the data. In the case where the composition of any part of the household could not be confirmed, "data unknown" was recorded for the entire unit. The data were recorded on a form that listed and identified the sample units by an attribute number (designated according to location on the housing unit map), the electrical meter number assigned to the unit, and the type of housing unit. All residential units, including those identified as "burned down" or otherwise destroyed, unoccupied or "removed from pad" (in the case of mobile homes and trailers) are considered part of the final sample. Units identified as "not a residence" were eliminated from the frame and not included in the sample. There were a few units for which data were unknown. These units are not included in the final sample, which may cause some slight bias.

7. RESULTS

The first data product is the number of households, which is derived directly from the active meter records, screened and classified by utility company personnel. Table 1 displays these figures by community along with other results that are discussed later.

Table 1 also provides the estimated *PPH*, which is taken from aggregate number of persons identified in the occupied sample units by the local experts. Also found in this table is the estimated household population of each community, which was found by applying the HUM formula to the household and *PPH* components. There were no group quarters identified in any of the communities.

Table 1

Sample Characteristics and Results of the Accuracy Test*

Community	Households	Estimated 1990			1990 Census Count	95% Confidence Interval	
		PPH	*SE*	Population		Low	High
Amargosa Valley	326	2.58	.11	841	853	771	911
Beatty	672	2.43	.10	1,633	1,623	1,501	1,765
Pahrump	3,224	2.23	.06	7,190	7,425	6,810	7,569

* The Estimated data and confidence intervals are produced by the procedures described in the text. The 1990 census counts are taken from Table 3 in the "1990 Census Extract, Nevada, Public Law 94-171 Data," dated February 11, 1991 and distributed by Betty McNeal, Nevada State Data Center Librarian, Nevada State Library and Archives, Capitol Complex, Carson City, Nevada 89710. The count for the area "Amargosa Valley is made up of the 1990 population reported for Nye county's Amargosa Valley Division (761) and Crystal Division (92). The count for the area "Beatty" is taken from the Beatty Census Designated Place and the count for the area "Pahrump" is taken from the Pahrump Division of Nye county.

8. MEASURING UNCERTAINTY IN THE ESTIMATES

One major advantage of estimates based on random sampling is that confidence intervals can be generated. Rives (1982) advocates this approach. However, he did not consider the use of local experts and believed that his suggestion would only be followed in exceptional circumstances because of the high expense associated with traditional surveys. This was also noted by Morrison (1982) and Lee and Goldsmith (1982) in their critical review of Rives' suggestion.

In the case of the local expert procedure, the "statistic" is the *PPH* value, which in practice would vary from sample to sample depending on the variation in *PPH* values. Our interest is less in the *PPH* values than in the estimate of population, however, so we use a simple transformation introduced by Espenshade and Tayman (1982) and used more recently by Swanson (1989) to place the confidence interval originally generated for a given community's *PPH* value around each of the community population estimates.

Let

$$P = \text{estimated household population},$$
$$N = \text{number of households},$$
$$PPH = \text{estimated persons per household}.$$

Then

$$\text{lower limit } (P) = (N) * (PPH - (t_{n-2}, \alpha/2) * (se)),$$
$$\text{upper limit } (P) = (N) * (PPH + (t_{n-2}, \alpha/2) * (se)),$$

where

n = number of households sampled,
α = level of significance desired,
se = standard error of the estimated *PPH*,
t_{n-2} = $(\alpha/2)$100th percentile of the t distribution,with $(n - 2)$ degrees of freedom.

As an example, using a significance level of .05, the corresponding 95% confidence interval for the estimated 1990 population of Pahrump (7,190) is

$$\text{lower limit} = 6{,}810 = (3{,}224) * (2.23 - (1.96 * .06)),$$
$$\text{upper limit} = 7{,}569 = (3{,}224) * (2.23 + (1.96 * .06)).$$

9. TEST OF ACCURACY

Before turning to the test results, which are also included in Table 1, some data qualifications require discussion. The single most problematic issue in terms of comparing the estimates with the 1990 census results lies in the fact that the Bureau of the Census does not recognize the "tax districts" as administrative boundaries for the communities in the study area. This means that the Bureau's "statistical" geography must be used, which requires some adjustments so that the geography used for purposes of the impact analysis matches that used by the Bureau.

In terms of these adjustments, the area identified as Amargosa Valley for purposes of the impact study is known to vary from the Amargosa Valley Census Division of Nye county used by the Bureau in that the study's definition includes the Crystal Census Division of Nye county. Fortunately, this is a case where two pieces of statistical geography used by the Bureau can be combined to virtually match that used in the impact study. Thus, the 1990 census population counts shown in Table 1 for the Amargosa Valley include the Crystal Division along with the Amargosa Valley Division. "Beatty" is another area that is known to vary in terms of geography. It is identified as both a Census Designated Place and as the Beatty Census Division of Nye county by the Bureau. In this situation, it is the Census Designated Place that corresponds very closely to the definition of Beatty used in the impact study. Thus, the 1990 census population count for Beatty shown in Table 1 is for the Beatty CDP.

The third community, Pahrump, is identified as a Census Division of Nye county. This piece of statistical geography used by the Bureau is virtually identical to that used in the impact study. Consequently, the 1990 census population found in Table 1 for Pahrump is that given for this division of Nye county.

There are other differences between the estimates and the 1990 census figures. The official date of the census count is April 1st; the estimates are for July 15th. In terms of this difference, seasonal effects are believed to be very slight for the communities in question. With the exception of the outflow of some "snowbirds," who may have been counted in the study area because they had no usual residence elsewhere, there were no known migration streams of any consequence between April and July. Similarly, the other components of population change were slight.

Had the Bureau found transient persons with no usual residence elsewhere, the estimation procedure is likely to have missed them. These differences would also impact housing unit counts. If a transient person, identified as a resident for purposes of the decennial census, is found in a recreational vehicle it would be included in the community's "other" housing stock by the Bureau. Such accommodations would not be included in the data derived from the residential electrical meter records.

We believe, however, that such instances are rare and, further, that the test results are not confounded by comparing a household population with a population that resides mainly in households but also, to some extent, in group quarters.

The results of the test of accuracy are also summarized in Table 1, along with the "low" and "high" estimates corresponding to the 95 percent confidence interval placed around each community's estimated population. The estimated population is very close to the population reported by the Bureau. Overall, the mean absolute difference is 86 persons and the mean absolute percent difference is 1.7.

The three confidence intervals contain the 1990 census population in each of the three communities, respectively. This finding is of special interest because the intervals are relatively narrow for a 95 percent level of confidence. On average, the width, as measured from the estimated population to either boundary is 7.2 percent of the estimated population. This suggests that confidence intervals constructed around the estimates derived from this variation of the HUM are meaningful, even in the presence of some unknown level of nonsampling error.

Two of the three communities are underestimated. In the case of Pahrump, it appears that the estimation technique was not able to capture all of the recent growth that appears to be spilling from the Las Vegas Valley into the Pahrump area. It is not known to what extent this was due to missing households on the frame and what was due to underestimating Pahrump's *PPH* value.

10. SUMMARY

While the local expert procedure may not provide satisfactory population estimates in all small, rural areas (*e.g.*, vacation spots, with a high incidence of seasonal housing units and privately owned rental units), it appears to hold promise based on the data for the area included in this study. As with any estimation technique, the key criteria for determining if it could be implemented elsewhere revolve around the possibility of obtaining the required data and implementing the procedure within available funding. In the case of the local expert procedure, this would mean that utility data can provide the number of households and be used as a sample frame. Once a sample was selected, the procedure's effectiveness would depend on the recruitment and knowledge of local experts. If these criteria can be met, the procedure would seem to be feasible. The next step would be to determine how accurate it is in a given application.

We were not able to evaluate the accuracy of the age and other composition data estimated through the procedure at the time of this writing because these data were not yet available from the 1990 decennial census. However, we are encouraged by the test results for the total population, which indicate that the procedure has the potential for highly accurate estimates, even in small, rural areas experiencing rapid change.

ACKNOWLEDGMENT

The authors are grateful for comments made by Mark Flotow and anonymous reviewers on earlier drafts of this paper.

REFERENCES

BROWN, R., GEERTSEN, H., and KRANNICH, R. (1989). Community satisfaction and social integration in a boomtown: A longitudinal analysis. *Rural Sociology*, 54, 568-586.

BYERLY, E. (1990). State and local agencies preparing population and housing estimates. *Current Population Reports* Series P-25, 1063. Washington, DC: U.S. Bureau of the Census.

CARLSON, J., SWANSON, D., and WILLIAMS, C. (1990). The development of small area socioeconomic data to be utilized for impact analysis: Rural, Southern Nevada. In *High Level Radioactive Waste Management Proceedings of the 1990 International Conference*. LaGrange Park, Illinois: American Nuclear Society, 985-990.

COCHRAN, W. (1977). *Sampling Techniques*, (3rd Edition). New York: Wiley.

ESPENSHADE, T., and TAYMAN, J. (1982). Confidence intervals for postcensal state population estimates. *Demography*, 19, 191-210.

FREUDENBURG, W. (1982). Social impact assessment. In *Rural Society in the U.S.: Issues for the 1980s*, (Eds. D. Dillman and D. Hobbs). Boulder, Colorado: Westview Press, 296-303.

KRANNICH, R., BERRY, E., and GREDIER, T. (1989). Fear of crime in rapidly changing rural communities: A longitudinal analysis. *Rural Sociology*, 54, 195-212.

KRANNICH, R., and GREDIER, T. (1984). Personal well-being in rapid growth and stable communities: Multiple indicators and contrasting results. *Rural Sociology*, 49, 541-552.

LEE, E., and GOLDSMITH, H. (1982). Evaluation and synopsis. In *Population Estimates: Methods for Small Area Analysis*, (Eds. E. Lee and H. Goldsmith). Beverly Hills, California: Sage, 113-118.

LOWE, T., PITTENGER, D., and WALKER, J. (1977). Making the housing unit method work: A progress report. Presented at the Annual Meeting of the Population Association of America.

LOWE, T., WEISSER, L., and MYERS, B. (1984). A special consideration in improving housing unit method estimates: The interaction effect. Presented at the Annual Meeting of the Population Association of America.

MARTIN, J., and SEROW, W. (1979). Virginia's state-local cooperative program: conflict and cooperation in producing population estimates. State Government, 182-186.

MORRISON, P. (1982). Alternatives for monitoring local demographic change. In *Population Estimates: Methods for Small Area Analysis*, (Eds. E. Lee and H. Goldsmith). Beverly Hills, California: Sage, 97-111.

MUNSELL, K. (1988). Towns cope with hazardous wastes. *Small Town*, 18, 30.

PITTENGER, D. (1976). *Projecting State and Local Populations*. Cambridge, Massachusetts: Ballinger Press.

POGGIE, J. (1972). Toward quality control in key informant data. *Human Organization*, 31, 23-30.

RIVES, N. (1982). Assessment of a survey approach. In *Population Estimates: Methods for Small Area Analysis*, (Eds. E. Lee and H. Goldsmith). Beverly Hills, California: Sage, 79-96.

RIVES, N., and SEROW, W. (1984). *Introduction to Applied Demography Data Sources and Estimation Techniques*. Beverly Hills, California: Sage.

SCHLEIFER, S. (1986). Trends in attitudes toward and participation in survey research. *Public Opinion Quarterly*, 50, 17-26.

SEIDLER, J. (1974). On using informants: A technique for collecting quantitative data and controlling measurement error in organizational analysis. *American Sociological Review*, 39, 816-831.

SMITH, S. (1986). A review and evaluation of the housing unit method of population estimation. *Journal of the American Statistical Association*, 81, 287-296.

SMITH, S., and LEWIS, B. (1983). Some new techniques for applying the housing unit method of local population estimation: Further evidence. *Demography*, 20, 407-413.

SMITH, S., and LEWIS, B. (1980). Some new techniques for applying the housing unit method of local population estimation. *Demography*, 17, 323-339.

SMITH, S., and MANDELL, M. (1984). A comparison of local population estimates: The housing unit method versus component II, regression, and administrative records. *Journal of the American Statistical Association*, 79, 282-289.

STARSINIC, D., and ZITTER, M. (1968). Accuracy of the housing unit method in preparing population estimates for cities. *Demography*, 5, 475-484.

SUMMERS, G. (1982). *Industrialization. In Rural Society in the U.S.: Issues for the 1980s*, (Eds. D.Dillman and D. Hobbs). Boulder, Colorado: Westview Press, 164-174.

SWANSON, D. (1989). Confidence intervals for postcensal population estimates: A case study for local areas. *Survey Methodology*, 15, 271-280.

SWANSON, D., BAKER, B., and VAN PATTEN, J. (1983). Conceptual and practical features of the housing unit method. Presented at the Annual Meeting of the Population Association of America.

U.S. BUREAU OF THE CENSUS (1978). State and local agencies preparing population estimates and projections: Survey of 1975-76. *Current Population Reports* P-25, 725. Washington, DC: U.S. Bureau of the Census.

U.S. DEPARTMENT OF ENERGY (1988). Section 175 report. *Secretary of Energy's Report to the Congress Pursuant to Section 175 of the Nuclear Waste Policy Act, As Amended*. Washington, DC: Office of Civilian Radioactive Waste Management, U.S. Department of Energy.