With or Without GIS? Evaluating Accuracy, Timeliness and Costs of Population Estimates for User-Defined Areas

Irina V. Sharkova
Population Research Center
Portland State University

Paper presented at the Population Association of America Annual Meeting, Los Angeles, California, March 23-25, 2000. *Address correspondence to Irina V. Sharkova, Population Research Center, Portland State University, P.O. Box 751, Portland OR 97207-0751. E-mail: sharkovai@pdx.edu. Phone (503) 725-5160.*

Abstract

With or Without GIS? Evaluating Accuracy, Timeliness and Costs of Population Estimates for User-Defined Areas

Population estimates for the majority of user-defined areas -- from traffic analysis zones to school attendance areas to watersheds -- are more and more often produced with the help of Geographic Information Systems (GIS). The use of GIS is warranted when spatial overlay or geocoding techniques have to be employed, i.e., when user-defined areas are not compatible with Census geographies, or when administrative records such as birth records or school enrollment are to be used. Yet implementation of GIS is pricey, a factor that most of the end users of population estimates such as school districts, police, fire protection and water districts, and other local governmental agencies must consider. Does it pay off to use GIS in population estimates for user-defined, or special, areas, or other, cheaper methods could produce acceptable results?

This paper provides empirical analysis of several commonly used population estimation techniques from the triple-constraints perspective, i.e. taking into considerations accuracy, timeliness, and costs of the estimates. It develops total population estimates for several user-defined areas in Multnomah County (Oregon): school districts, fire protection districts and water districts, and then tests these estimates against the numbers produced by the 1996 American Community Survey. The population estimates were developed using both the methods that do not require GIS applications (share in voter registration), and those that do (variants of the housing unit method). Time and resources spent on the database and estimates development were analyzed in relation to the differences in accuracy between methods.

The study demonstrates that application of GIS technology to population estimates is not warranted for *every* type of special or user-defined areas. GIS-based methods show a fair to considerable improvement over simpler methods as population of a special area reaches and exceeds the threshold of about 20,000 persons; however, this improvement comes at a very high cost (an increase of 5 to 10 times). Accuracy and utility of population estimates for very small areas (under 2,500 persons) cannot be improved enough to justify the use of GIS-based techniques over simpler methods such as voter registration. In data-poor research environments, a simple voter registration method, or once-a-decade investment in the GIS-assisted Census-year total population count would provide acceptable results from the triple-constraints prospective of accuracy, costs and timeliness of the estimates.

OVERVIEW

As Geographic Information Systems have become more common in applied demographic research, there seem to be a lack of empirical studies verifying the superiority of GIS-based population estimation methods comparing to simpler, cheaper and less involved techniques. The present study addresses this shortcoming by conducting an empirical analysis of several commonly used population estimation techniques for small, user-defined areas from the triple-constraints perspective, i.e. taking into considerations accuracy, timeliness, and costs of the estimates. It develops 1996 population estimates for eleven user-defined areas in Multnomah County (Oregon): school districts, fire protection and water districts, and then compares these estimates with the numbers produced by the 1996 American Community Survey. The population estimates are developed using both the methods that require little or no applications of Geographic Information Systems (such as a share in voter registration or application of county growth rates to the GIS-derived Census-year total population), and those that do (variants of the housing unit method). Time spent on the database and estimates development and resulting costs were recorded and are analyzed in relation to the differences in accuracy (precision and directional bias) and utility of the methods.

The study demonstrates that GIS-based methods show a fair to considerable improvement over simpler methods as population of a special area reaches and exceeds the threshold of about 20,000 persons; however, this improvement comes at a very high cost (an increase of 5 to 10 times). Accuracy and utility of population estimates for very small areas (under 2,500 persons) cannot be improved enough to justify the use of GIS-based techniques over simpler methods such as voter registration. The results of the study indicate that more involved and costlier methods do not offer a proportionate increase in precision of the estimates, although they do tend to reduce a directional bias of the estimate and its total absolute error.

BACKGROUND

The needs of planning and policy agencies, non-profit organizations, service districts and local governments in accurate, timely and affordable population estimates do not require an elaborate justification; for most of them, a current estimate of a population in their jurisdiction is necessary at least once a year, during preparation of the new fiscal year's budget (Davis 1995, Lee and Goldsmith 1982). While population estimates at a county and city level are available free of charge, significant effort is involved when a population estimate is required for areas whose boundaries do not correspond to any standard regions of data collection and dissemination, especially when the population size of these areas is less that 15,000-30,000 persons (Tayman, Schafer and Carter 1996). As a rule, small areas rarely collect new housing information or other symptomatic indicators that would allow them to relatively easily update the Census-year population counts (Morrison 1982). The situation is further complicated for user-defined, or special, areas since they often do not even have an accurate population count provided by the most recent decennial census. Transportation districts, water districts, special election districts, fire protection and emergency districts, school districts are the examples of the areas which boundaries do not follow census geographies.

Recent proliferation of Geographic Information Systems in applied demographic research has resulted in improved possibilities related to population estimates for small user-defined areas. Practical solutions for the modifiable areal unit problem (MAU) have been suggested by a number of researchers (Gosh and Rushton 1987, Fotheringham and Wong 1991, Openshaw 1984), and it became common to develop population estimates for areas incompatible with Census geographies. Yet for many administrators of great many special areas it is not feasible to invest in the development of in-house GIS capabilities, especially when a GIS-based population estimation system is needed only once a year or two, during a budget planning process. In such instances, a consultant - an applied demographer - is called into action, and a population estimate for a user-defined area is produced. It is often left to the applied demographer to choose the best - cheapest and the most accurate - method of population estimation.

Availability of a modern technology such as GIS allowing a very precise delineation of special areas and an easy incorporation of administrative records into population estimation process creates almost an urge to use it whenever possible. Is there enough evidence of the true superiority of the technique comparing to other, cheaper and simpler methods? Are the gains in precision and utility of a population estimate produced with help of GIS high enough to warrant the increased cost of such work? When advising a client on a method of population estimation, does an applied demographer have any hard evidence of the superiority of a complex GIS approach?

The idea for this paper evolved from a rich experience of the current author in small-area population estimates. Over the course of several years, the current author prepared population estimates for special and custom-defined areas, as a part of her duties in the Population Research

Center at PSU (Sharkova and Edmonston 1999a; 1999b). Being a "true believer" in the miracle of GIS, the author's first choice has always been a population estimation method based on applications of GIS - providing a customer could afford this approach. Is this choice always better?

RESEARCH DESIGN AND DATA CONSIDERATIONS

This paper provides empirical analysis of several commonly used local population estimation techniques from the triple-constraints perspective, i.e. taking into considerations accuracy, timeliness, and costs of population estimates (Swanson et al. 1997, Swanson and Tayman 1995). This prospective posits that in order to be a useful tool for an informed decision-making, population estimates and forecasts have to be accurate just *enough* rather than in absolute terms. In other words, rather than attempting to develop a *perfectly* precise forecast or estimate - an goal hardly achievable even with unlimited resources - an applied demographer should focus on *improving* certainty and precision of the forecast or estimate so that it can support a cost-effective decision-making process (Swanson et al. 1996).

We developed population estimates for a set of special areas in Multnomah County, OR, including fire protection districts, school districts, and water districts (see Fig. 1). To minimize errors associated with the MAU problem, only those districts that lay completely within Multnomah County were chosen for the study. To achieve diversity in districts' sizes and locations within the county, a total of thirty four areas were considered next: fifteen electoral districts (for the Oregon House of Representatives), four fire protection districts, seven water districts, two Metro Council districts, and six school districts. After close examination of the data, contacts with the Multnomah County Elections Office and consultations with historical maps, the sample was reduced 3-fold, mainly due to the fact that many districts changed - or were suspected to have changed - their boundaries between 1990 and 1996. As Table 1 shows, the eleven areas selected for the study range in population size from very small (1,200 persons in 1996) to very large (412,000 persons in 1996); if the largest area - Portland School District - is excluded, the average size of the remaining ten areas was about 20,000 persons in mid-1996.

A total of five estimation methods were tested, from the simplest one, based on the voter registration data, to more complex techniques, involving GIS and geocoded building permits information. Since the 1990 total population of most special areas is unknown (with an exception of the school district special tabulation data which we did not use for this analysis in order to keep conditions comparable), the **voter registration method** has a clear benefit comparing to other methods: it allows to avoid a time-consuming GIS-based estimate of the 1990 population of the district. It does so by assuming that the relationship between the total population of the district and the number of its registered voters is the same as for the county where the district is located.

Table 1. Selected Districts (Multnomah County, OR) and their 1990 and 1996 Population.

DISTRICTS	**Total Population April 1, 1990**	**Total Population July 1, 1996**
Rural Fire Protection #30	962	1,920
Rockwood PUD / Water	44,020	47,182
Powell Valley Road Water District	24,114	23,474
Lusted Water District	945	1,961
Valley View Water District	884	1,204
Corbett Water District	2,704	1,640
David Douglas School District #40	41,984	44,399
Portland School District #1	397,195	412,000
Reynolds School District #7	43,616	51,387
Riverdale School District #51	2,225	2,229
Parkrose School District #3	22,514	23,868

The remaining four methods involved an application of GIS to a certain degree, if only to arrive at the 1990 total population of each area. GIS (ArcView 3.2) was used to intersect the 1990 Census block boundaries and the districts' boundaries, and to come up with an allocation coefficient for each block. The allocation coefficient was used to recalculate block-level 1990 Census data to each district's boundaries.

Among the methods tested, two techniques involved a simple application of known county and/or city **population growth rates** to the base 1990 population derived with the help of GIS. First of them ("GIS & Rates 1") applied Multnomah County annual growth rate to each special area; the growth rate was calculated using the 1990 Census data and July 1, 1996, population estimates for Multnomah County produced by the Population Research Center as a part of its official estimation program. The second GIS- and rates-based method ("GIS & Rates 2") used annual growth rates of the cities of Multnomah County or its unincorporated area rather than the county rate. In this instance, a decision was made by the author as to what rate to apply depending on the location of a district in a particular city of the county and the author's knowledge of the type of growth trends that could be expected in a given district.

The remaining two methods employed GIS for more than the initial aggregation of the 1990 data. Using geocoded data on building permits issued for each district between January of 1990 and January of 1996, the number of authorized additional housing units was derived for each district. In the first variant of the **housing units method** ("GIS & HU 1"), only the total number of permits and housing units was used. It was then multiplied by the district-specific 1990 housing vacancy rate to obtain the number of occupied housing units, and then multiplied by the district-specific 1990 persons per household rate. The resulting figure for population added between April 1, 1990, and July 1, 1996, was then combined with the 1990 total population for each district to arrive at the 1996 population estimate. The second GIS- and housing units-based method ("GIS & HU 2") followed the same logic, but separated building permits and housing units by type into single-family residential, multi-family residential, and mobile homes. It then utilized 1990 district-specific PPH rates disaggregated by type of the housing units.

These population estimates were then compared with the 1996 American Community Survey population totals for the districts (see Table 2). Since the 1996 ACS only provided information about household population, population in group quarters reported by the 1990 Census of Population was added to the 1996 household population in order to come up with an estimate of the 1996 total population. As it was the case with deriving the 1990 total population of each district, GIS was used to allocate the 1996 total block group population to the districts. As with the 1990 data, the 1996 population was assumed to be distributed evenly across each block group, and a proportion of each block group's area in the district to the total area of the block group was used to derive an allocation coefficient[1]. This method inevitably introduced an unknown error in our calculations; it is our hope that the errors balanced themselves across the areas.

---

[1] We are aware of more precise methods for allocation of spatial data; for example, it was possible to use 1990 block-level population, tax lot data, or area proportions of residential areas only to derive allocation coefficients for the 1996 ACS data. However, the increase in precision in allocating 1996 data would have increased the time spent on the project beyond what was feasible.

Table 2. 1996 Population Estimates by District and Method.

DISTRICTS	American Community Survey	Voter Registration (Method 1)	GIS & Rates 1 (Method 2)	GIS & Rates 2 (Method 3)	GIS & HU 1 (Method 4)	GIS & HU 2 (Method 5)
Rural Fire Protection #30	1,920	1,246	1,048	1,102	1,003	1,004
Rockwood PUD / Water	47,182	42,099	47,949	50,460	46,503	46,307
Powell Valley Road Water District	23,474	23,267	26,266	27,642	25,350	25,189
Lusted Water District	1,961	1,196	1,029	1,684	982	983
Valley View Water District	1,204	1,216	963	1,209	935	936
Corbett Water District	1,640	3,564	2,946	1,386	2,879	2,877
David Douglas School District #40	44,399	41,310	45,731	48,127	44,720	44,458
Portland School District #1	412,000	432,173	432,646	455,306	411,637	410,251
Reynolds School District #7	51,387	49,096	47,509	50,711	52,714	52,451
Riverdale School District #51	2,229	2,725	2,424	2,424	2,296	2,297
Parkrose School District #3	23,868	23,559	24,523	24,326	22,888	22,854
Total	611,266	621,453	633,033	664,376	611,907	609,607

RESULTS

To analyze whether an increase in time and resources spent on the development of population estimates with the help of GIS resulted in an increased utility of the estimates, we relied on the tools developed by other demographers. Rees (1994: 1680) suggested that the reduction in the total absolute error of an estimate is an appropriate measure of its accuracy. The mean (algebraic) percentage error (MALPE) and the mean absolute percentage error (MAPE) of a forecast are also used to evaluate forecast precision and directional bias (Perry and Voss 1996, Smith 1987, Tayman 1996). Following Costner (1965), Swanson and Tayman (1995) advocated the use of the PRE, or a proportionate reduction in error, in evaluating the utility of a forecast. In their analysis of the utility of population forecasts Tayman and Swanson (1996: 524-527) used MAPE- and MALPE-based calculations of PRE to identify whether forecasts for census tracts and counties performed better than the no-cost, “naïve” alternative of assuming that the 1990 population of the areas under consideration remained the same as in 1980.

We utilized these tools in order to test whether applications of GIS improved the estimates’ precision, directional bias and utility. Table 3 lists the results of comparison of each estimation technique with the 1996 ACS-based estimate, as well as a proportionate reduction of error (PRE) achieved by each method in comparison with the least time-consuming voter registration-based technique.

Table 3. Population Estimates Bias, Precision, and Utility, by Method.

Measures of Accuracy and Utility	**Voter Registration 1**	**GIS & Rates 1 2**	**GIS & Rates 2 3**	**GIS & HU 1 4**	**GIS & HU 2 5**
Total Absolute Error	44,856	33,617	57,164	9,017	9,943
Total Absolute Error PRE		25.1%	-27.4%	79.9%	77.8%
MALPE	5.9%	0.7%	1.7%	3.3%	3.5%
MALPE PRE		87.7%	70.6%	44.3%	40.4%
MAPE	20.4%	21.2%	11.7%	19.6%	19.5%
MAPE PRE		-3.7%	42.9%	4.1%	4.7%

Accuracy of the Estimates. Total absolute error of the estimate decreases significantly - from almost 45,000 persons to just under 10,000 persons - as the complexity of the estimation technique increases. We suspect that this improvement is largely due to the improved estimate for the largest district in the group, the “monster” of Portland Public Schools. The MAPE of the estimates is larger than their MALPE suggesting that the directional bias of each estimate on average is lower than the errors associated with precision of the estimates. As a category, estimates derived with the aid of GIS seem to perform somewhat better than the voter-based method, producing a notably lower total absolute error, lower MALPE and somewhat lower MAPE parameters (see also Charts 1-3). The method that applied area-specific growth rates selected based on local knowledge to GIS-derived 1990 total population produced the best combination of MALPE and MAPE errors.

Utility of the Estimates. Almost every PRE measure demonstrated an increased utility of a GIS-based population estimate technique in comparison with the voters-registration method. The most of the reduction in *total absolute error* occurred when a GIS was used to aggregate geocoded building permits data (Methods 4 and 5). PRE for *directional bias* showed a notable improvement associated with Methods 2 and 3 (88% and 71% improvement over the voter-registration method), and a decent improvement by Methods 4 and 5 (44% and 40% improvement). MAPE-based utility of the estimates was highest for the “informed” GIS and Rates method (Method 3).

Chart 1. Total Absolute Error and Utility (PRE) by Estimation Method

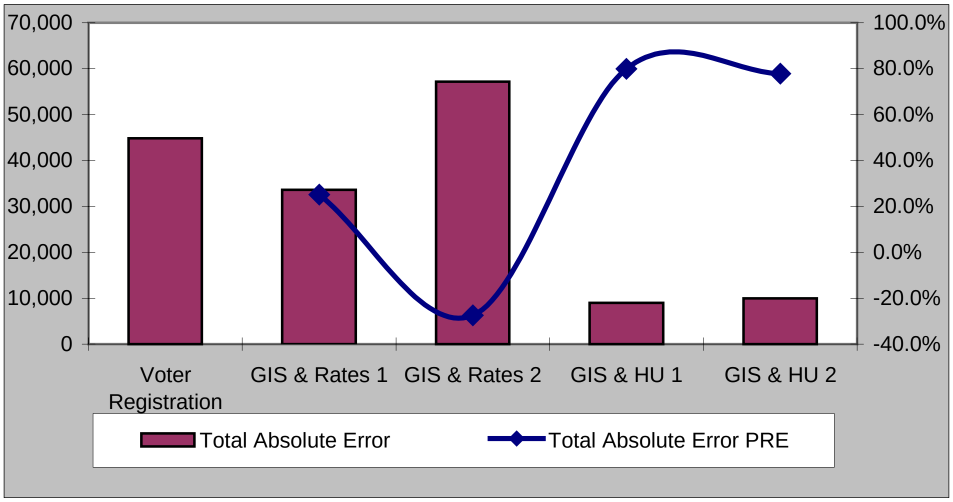
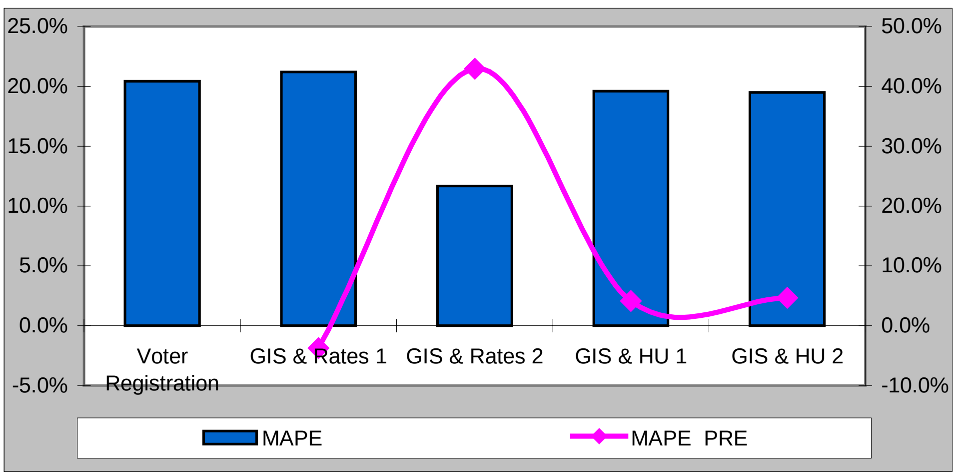


Chart 2. Precision Error and Utility (PRE) by Estimation Method



As it was suggested in the literature, **population size** of the area under consideration has a direct effect on the precision of population forecasts: the smaller the area, the higher the magnitude of an error (Tayman, Schafer and Carter 1998). Table 4 provides the comparison of the accuracy and utility of the population estimation methods disaggregated by the population size of the area: for small districts (1,200 to 2,250 persons), medium-size districts (23,000 to 51,500 persons), and large districts (just one case - Portland Public Schools - is present in this category).

Chart 3. Directional Bias and Utility (PRE) by Estimation Method

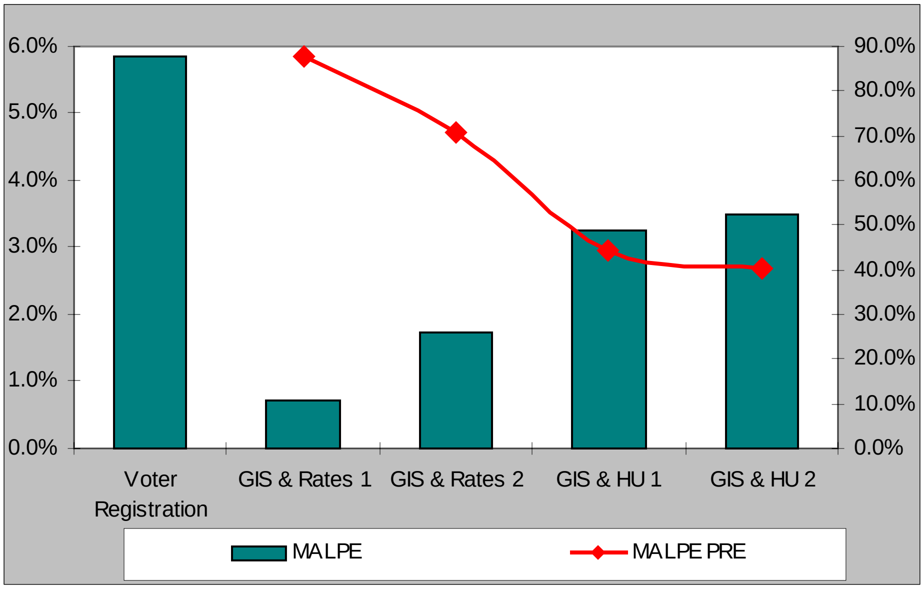


Table 4. Population Estimates Bias, Precision, and Utility, by Method, by District Size.

	Evaluation Tools	Voter Registration 1	GIS & Rates 1 2	GIS & Rates 2 3	GIS & HU 1 4	GIS & HU 2 5
Small districts	Total Absolute Error	3,873	3,547	1,550	3,470	3,466
N=5	Total Absolute Error PRE		**8.4%**	**60.0%**	**10.4%**	**10.5%**
	MALPE	9.0%	4.9%	12.6%	8.3%	8.3%
	MALPE PRE		**45.1%**	**-40.6%**	**7.5%**	**7.8%**
	MAPE	38.3%	40.3%	16.3%	39.7%	39.7%
	MAPE PRE		**-5.2%**	**57.5%**	**-3.7%**	**-3.6%**
Medium districts	Total Absolute Error	10,979	9,424	12,309	5,183	4,727
N=5	Total Absolute Error PRE		**14.2%**	**-12.1%**	**52.8%**	**56.9%**
	MALPE	5.3%	-2.3%	-6.7%	-1.2%	-0.7%
	MALPE PRE		**55.6%**	**-27.6%**	**78.2%**	**87.1%**
	MAPE	5.3%	5.4%	7.3%	3.4%	3.1%
	MAPE PRE		**-1.5%**	**-37.5%**	**36.2%**	**40.9%**
Large district	Total Absolute Error	30,004	20,646	43,306	363	1,750
N=1	Total Absolute Error PRE		**31.2%**	**-44.3%**	**98.8%**	**94.2%**
	MALPE	-6.9%	-5.0%	-10.5%	0.1%	0.4%
	MALPE PRE		**27.8%**	**-51.4%**	**98.7%**	**93.9%**
	MAPE	6.9%	5.0%	10.5%	0.1%	0.4%
	MAPE PRE		**27.8%**	**-51.4%**	**98.7%**	**93.9%**

Clearly, the importance of GIS in increasing precision and utility becomes more obvious as the population size of the area under consideration increases. However, even for small-size districts the *total absolute error* decreases by 10% and more when GIS is involved in preparation of the estimates. Simple GIS-based methods gain about 50% in utility, while complex GIS-based methods add up to 99% in utility comparing to the voters registration method if one considers the *total absolute error* of the estimate. The GIS-related gains are not as convincing if one considers the *accuracy* of the estimates for small districts as measured by MAPE and MALPE. Indeed, while the *directional bias* improves by almost 8% when complex GIS-based estimation methods are involved, the *precision* of the estimates decreases by almost 4 percent, comparing to the simplest, voters registration-based method.

Among all estimation methods used **for small districts** the simple GIS-based technique used in conjunction with the “informed” growth rates (Method 3) turns out to be most useful as judged by *total absolute error* and MAPE; it furnishes an improvement of 60% comparing to Method 1. However, this technique errs notably with regard to the directional bias. Other techniques do not seem to provide a significant enough improvement in either precision or directional bias to warrant their use for small areas comparing to the voters registration method.

Accuracy and utility of population estimates for **medium-size districts** improve the most when the most complex GIS method (Method 5) is involved. The utility of the estimates derived this way increase by 41 to 87 percent comparing to the voter registration method, while the accuracy ranges from -0.7% for MALPE to 3.1% for MAPE. Method 4 provides similarly encouraging results. Method 3 decreased the utility of population estimates for both medium-size and large districts; it works worse than its “close relative” - Method 2. The latter shows a slightly to moderately improved utility for most types of areas as measured by most evaluation tools comparing to the cheapest method.

There is only one **large district** in the study due to data constraints discussed above. The district estimates’ accuracy and utility improve significantly when complex GIS methods are used: the errors are at a level of 0.1 to 0.4 percent for methods 4 and 5, respectively, while accuracy increases by 94 to 99 percent comparing to method 1.

The table reveals several unexpected outcomes. It is surprising that, for most of the areas, Method 2 that uses county growth rates shows a better fit to the ACS population counts then city-specific, or “informed” growth rates of Method 3. It is also unexpected that utilizing aggregated (Method

4) rather than housing type-specific (Method 5) PPH rates produces a more accurate estimate for the large district.

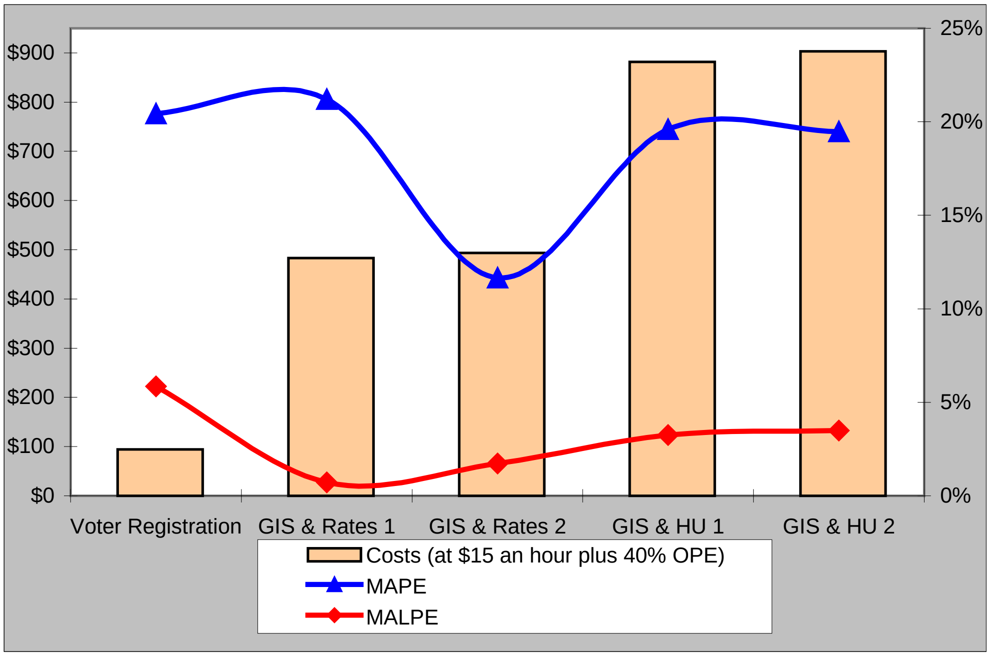
Timeliness and Costs. Having established that application of GIS increases the utility and accuracy of population estimates for medium and large areas, we need to compare these results with the time and costs involved in developing the estimates. Table 5 reports empirical data regarding the time spent on developing population estimates by a graduate research assistant with some familiarity with population estimation techniques. Where the current author's time was involved, it was reflected in the totals after being adjusted for the level of experience.

Table 5. Population Estimates Method by Time and Costs Involved.

METHODS STEPS	Voter Registration 1	GIS & Rates 1 2	GIS & Rates 2 3	GIS & HU 1 4	GIS & HU 2 5
Database development:					
Obtaining voter registration data (phone, fax)	3.0	0.0	0.0	0.0	0.0
District boundaries: find and download	0.0	7.5	7.5	7.5	7.5
1990 Census Blocks boundaries: download	0.0	4.0	4.0	4.0	4.0
Intersecting districts and Census Blocks	0.0	3.5	3.5	3.5	3.5
Changing projections	0.0	1.0	1.0	1.0	1.0
Aggregating building permits data by district	0.0	0.0	0.0	9.0	9.0
Obtaining 1990 STF1b data for Census Blocks	0.0	3.0	3.0	3.0	3.0
Recalculating 1990 STF1b data for districts	0.0	2.0	2.0	10.0	10.0
Population estimates development:					
Instruction, explanation	0.5	1.0	1.0	2.0	2.0
Calculating 1996 population estimates	1.0	1.0	1.5	2.0	3.0
Total (hours)	**4.5**	**23.0**	**23.5**	**42.0**	**43.0**
Approximate costs (at $15 an hour)	$67.50	$345.00	$352.50	$630.00	$645.00
Plus 40% OPE	$94.50	$483.00	$493.50	$882.00	$903.00

As Table 5 and Chart 4 (below) show, the time and associated costs increase almost 10 times between the simplest and the most complex methods. The increase is not gradual or even: Method 2 is about 5 times more time- and resources-consuming than Method 1, Method 4 is twice as expensive as Method 2, while differences between Methods 2 and 3 as well as Methods 4 and 5 are insignificant.

Chart 4. Estimate Cost and Accuracy by Estimation Method



Although we only tested five methods and have too few observations, we attempted to calculate Pearson coefficients of correlation between the money spent on developing the estimates and the amount of error associated with each method (see Table 6). The coefficients indicate that for the districts under consideration, the increased costs tend to dramatically decrease the total absolute error of the estimate, notably decrease its mean algebraic percentage error, yet have no influence on the mean absolute percentage error.

Table 6. Correlation between Costs and Accuracy of Estimates.

Costs by method	**MAPE**	**MALPE**	**Total Absolute Error**
$94.50	20.4%	5.9%	44,856
$483.00	21.2%	0.7%	33,617
$493.50	11.7%	1.7%	57,164
$882.00	19.6%	3.3%	9,017
$903.00	19.5%	3.5%	9,943
Pearson correlation coefficient	**0.007**	**-0.312**	**-0.797**

DISCUSSION

The results of this exploratory study seem to confirm our original suspicion that application of powerful yet costly GIS technology to population estimates is not warranted for *every* type of special or user-defined areas. Accuracy and utility of population estimates for very small areas (in our study those with under 2,500 persons) cannot be improved enough to justify the use of GIS-based techniques over simpler methods such as voter registration. GIS-based methods show a fair to considerable improvement over simpler methods just as population of a district under consideration reaches and exceeds the threshold of about 20,000 persons; however, this improvement comes at a very high cost (an increase of 5 to 10 times). If we consider the results of the correlation analysis to be of any indication of possible relationship between time and costs involved in population estimates and changes in utility and accuracy of the estimates, we have to admit that more involved techniques do not seem to increase precision of the estimates (as measured by MAPE) while somewhat improving their directional bias and significantly decreasing the total absolute error.

The results of this study should be treated with caution for a number of reasons. Due to considerable data limitations, we could only study eleven districts, a very small sample for any statistical analysis. Due to time and costs constraints, we limited our study to five estimation methods. Several possible sampling and non-sampling errors were introduced into the analysis, although their magnitude is unknown. Estimates were made of 1996 total population since the American Community Survey only provided household population counts. The 1996 ACS data was collected in a 15% sample and available by block group rather than by Census block. Clearly, more extensive research has to be done in order to confirm regularities suggested by this study.

Each method under examination is likely to suffer from inherent errors as well. Voter registration technique is likely to produce higher errors in the areas with many recent immigrants and higher proportion of children. Applications of county and city growth rates (Methods 2 and 3) are particularly unreliable in the areas of unusually fast growth or decline. Housing unit methods presented here depend on a positional accuracy of geocoded building permit data[2] as well as consistency of vacancy and person per household rates over time.

[2] An extreme example of an error introduced by the geocoding process was encountered by the author during this study: in the geocoded building permits database, almost 3,000 permits were found in one single location in one of the districts. These permits - an false 200 percent increase in housing units for the district under consideration during a 5-year period - were simply assigned to this location by the database developer, apparently in an attempt to reach a 100% match rate for the whole county.

The study was conducted in a region blessed with rich geographic and demographic data. Oregon is known for a high-quality digital spatial database available on-line for every part of the state. Since Multnomah County is a part of the Metropolitan Regional Government (Metro), it benefits from an access to RLIS database developed by Metro, an extensive collection of fairly accurate spatial data down to tax lot level, including building permits data. Finally, it was one of the four test sites for the 1996 ACS, which provided Census-like information down to block group level. As a result, an applied demographer working in such data-rich region has a variety of choices in methods and symptomatic indicators to use in his or her population estimates and other demographic research, without a need of an up-front investment in data collection. Such luxuries are not available for applied demographers in data-poor regions, who may find it prohibitively time-consuming and expensive to invest in development of suitable database in addition to a GIS. In such instances, a simple voter registration method, or once-a decade investment in the GIS-assisted Census-year total population count would provide acceptable results from the triple-constraints prospective of accuracy, costs and timeliness of the estimates.

CONCLUSION

The study demonstrates that application of GIS technology to population estimates is not warranted for every type of special or user-defined areas. GIS-based methods show a fair to considerable improvement over simpler methods as population of a special area reaches and exceeds the threshold of about 20,000 persons; however, this improvement comes at a very high cost (an increase of 5 to 10 times). Accuracy and utility of population estimates for very small areas (under 2,500 persons) cannot be improved enough to justify the use of GIS-based techniques over simpler methods such as voter registration. In data-poor research environments, a simple voter registration method, or once-a-decade investment in the GIS-assisted Census-year total population count would provide acceptable results from the triple-constraints prospective of accuracy, costs and timeliness of the estimates.

Despite its limitations, we hope that the study provides important empirical evidence on the accuracy, timeliness and costs involved in developing population estimates with and without applications of GIS. We hope that this information can be used by applied demographers in real-life situations, when they have to make decisions about particular techniques of population estimation, and, more broadly, methods of demographic analysis.

BIBLIOGRAPHY

Costner, H. (1965) “Criteria for Measures of Association”. *American Sociological Review* 30: 341-353.

Davis, Craig H. (1995) *Demographic Projection Techniques for Regions and Small Areas: A Primer*. Vancouver, BC: UBC Press.

Fotheringham, Stewart A., and D. W. S. Wong (1991)”The Modifiable Areal Unit Problem in Multivariate Statistical Analysis.” *Environment and Planning A* 23: 1025-1044.

Gosh, A., and G. Rushton (1987) *Spatial Analysis and Location-Allocation Models*. New York: Van Nostrand Rheinhold.

Lee, Everett S., and Harold F. Goldsmith (1982) *Population Estimates: Methods for Small-Area Analysis* Beverly Hills, CA: SAGE Publications, Inc.

Morrison, Peter A. (1982) “Alternatives for Monitoring Local Demographic Change” In Lee and Goldsmith (Eds.) *Population Estimates: Methods for Small-Area Analysis* Beverly Hills, CA: SAGE Publications. 97-112.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem: Concepts and Techniques in Modern Geography*. Norwich: Geo Books.

Perry, Marc, and Paul R. Voss (1996) “Using Geo-Demographic Methods for Improving Small-Area Population Estimates.” Presented at the Annual Meeting of the Population Association of America, New Orleans.

Rees, P. (1994) “Estimating and Projecting the Populations of Urban Communities”. *Environment and Planning A* 26: 1671-1697.

Sharkova, Irina V., and Barry Edmonston (1999a) “Beaverton School District 48 Population and Enrollment Forecast, 2000 to 2020: Results and Documentation”. Report to the Beaverton (OR) School District.

Sharkova, Irina V., and Barry Edmonston (1999b) “Hockinson School District 98 Population and Enrollment Forecast, 2000 to 2010”. Report to the Hockinson (WA) School District.

Smith, Stanley (1987) “Tests of Accuracy and Bias for County Population Projections” *Journal of the American Statistical Association* 82: 991-1003.

Swanson, David A., D. Richard Lycan, George Hough, Jr., and Irina V. Sharkova (1996) “Hillsboro School District 1J Population and Enrollment Forecast, 1995 to 2010: Results and Documentation”. Report to the Hillsboro (OR) School District.

Swanson, David A., Irina V. Sharkova, Richard Lycan and George Hough, Jr. (1997)

“Forecasting For School Attendance Zone Changes And Capital Needs: Merging GIS And Demographic Methods With Local Expert Judgment” *Market and Demography* 3(6): 23-31.

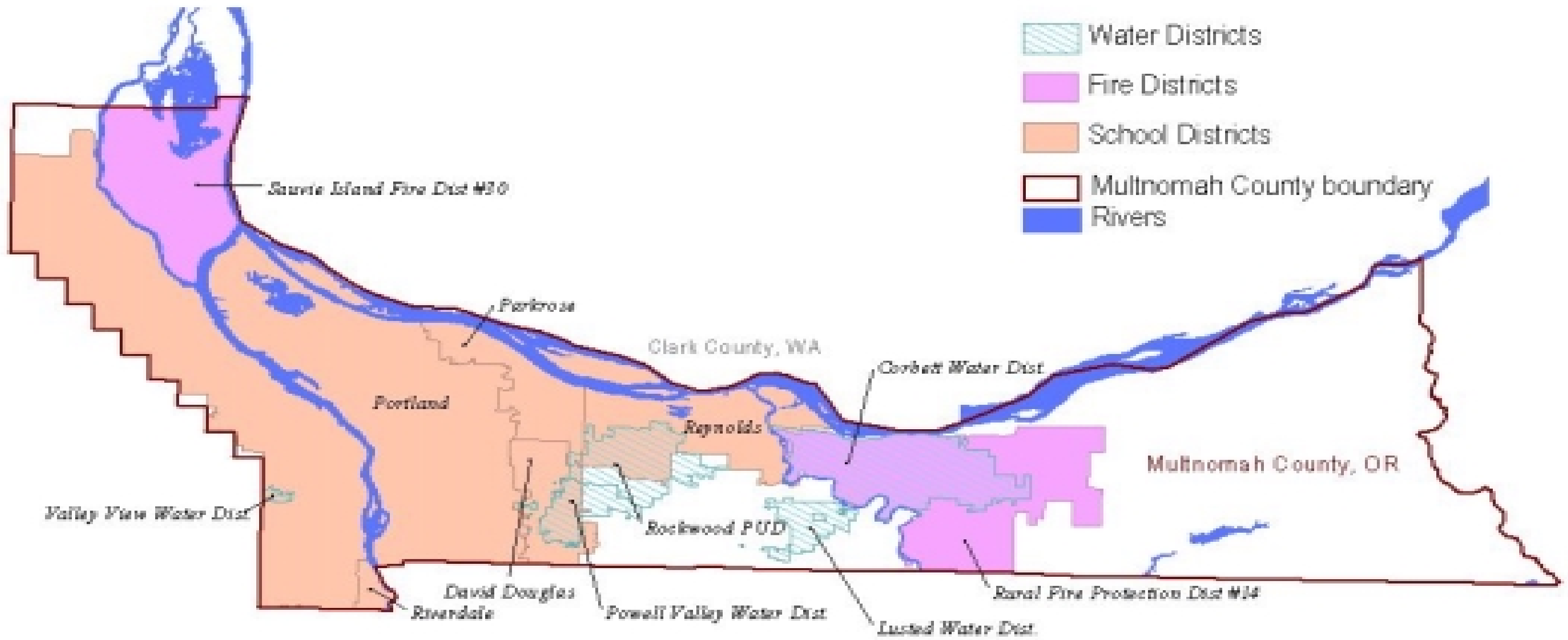
Swanson, David A., and Jeff Tayman (1995) “Between a Rock and a Hard Place: The Evaluation of Demographic Forecasts”. *Population Research and Policy Review* 14: 233-249.

Tayman, Jeff (1996) “The Accuracy of Small Area Population Forecasts Based on Spatial Interaction Land Use Modeling System.” *Journal of the American Planning Association* 62: 85-98.

Tayman, Jeff, Edward Schafer and Lawrence Carter (1998) “The Role of Population Size in the Determination and Prediction of Population Forecast Errors: An Evaluation Using Confidence Intervals For Subcounty Areas.”. *Population Research and Policy Review* 17: 1-20.

Tayman, Jeff, and David A. Swanson (1996) “On the Utility of Population Forecasts” *Demography* 33(4): 523-528.

Fig. 1 Selected Districts, Multnomah County, OR.



Source: RLIS (November 1999, Metro Regional Government)