# ADVANCING METHODOLOGICAL KNOWLEDGE WITHIN STATE AND LOCAL DEMOGRAPHY: A CASE STUDY*

David A. Swanson
Department of Sociology and Anthropology
University of Mississippi
P.O. Box 1848
University, MS 38677-1848
(email: dswanson@olemiss.edu)

*An earlier version was presented at the 2003 annual meeting of the Southern Demographic Association, Arlington, VA, October 23rd-25th. The author is grateful for comments and suggestions made at this session.

**ADVANCING METHODOLOGICAL KNOWLEDGE WITHIN STATE AND LOCAL DEMOGRAPHY: A CASE STUDY**


## ABSTRACT

Much of the academic literature dealing with state and local demography involves the development and evaluation of methods for estimating population. The focus on estimation methods is not surprising because they are used in many states to allocate resources. The quality control in regard to the validity and reliability afforded these methods by the traditional academic peer review process is important because, among other things, it serves to reduce the high potential for conflict that exists when resources are at stake. There are, however, methods being used by state and local demographers that have not been subject to peer review. While not necessarily unsound, these "fugitive" methods serve to keep the potential for conflict high because of the uncertainty regarding their validity and reliability. This paper examines just such a situation in the form of a case study. It is a discussion of a regression model developed in Nevada following the 2000 census that led to conflict over its use to estimate the population of Clark County, Nevada in 2002. The discussion reveals statistical and methodological shortcomings in this model that lead to an alternative model not subject to these shortcomings. This example illustrates how this type of analysis and discussion can lead to a wider understanding of methods on the part of practitioners through the corrective process of academic peer review. It also suggests that states in which estimates are used to allocate resources would be well-served by subjecting new methods being considered for use to academic peer review before they are adopted.

## ADVANCING METHODOLOGICAL KNOWLEDGE WITHIN STATE AND LOCAL DEMOGRAPHY: A CASE STUDY

## INTRODUCTION

Many of the methods commonly used by applied demographers to estimate subnational population have been subject to peer review through discussions and evaluations in the academic literature (Bogue, 1950; Bousfield 1977; Cannan, 1895, Eldridge, 1947; Erickson, 1974; Hamilton, 1964; Krotki, 1978; Land and Hough, 1986; Mandell and Tayman, 1982; McVey, 1974; Morrison, 1982; Namboodiri, 1972; Namboodiri and Lalu, 1971; O'Hare, 1976, 1980; Purcell and Kish, 1980; Rives, Serow, Lee, and Goldsmith, 1989; Roe, Swanson, and Carlson, 1992; Rosenberg, 1968; Schmitt, 1952; Schmitt and Crosetti, 1954, Schmitt and Grier, 1966; Shryock, 1938; Siegel, Shryock and Greenberg, 1954; Smith and Lewis, 1980; Smith and Cody, 1994; Smith, Nogle, and Cody, 2002; Snow, 1911, Spar and Martin, 1979; Starsinic and Zitter, 1968; Swanson, 1980; Swanson and Tedrow, 1984; and Zitter and Shryock, 1964, and U.S. Bureau of the Census, 1949).

Unfortunately, there have been and remain some methods used for subnational population estimation that have not been subject to the quality control process available though traditional academic peer review Some practitioners at the state and local level simply use, develop, and refine methods that are never evaluated in this manner. While these methods are not necessarily unsound, they are subject to a higher level of uncertainty in regard to their validity and reliability, especially if they are developed and used by analysts without much or any formal training in the methods of applied demography. The uncertainty surrounding the validity and reliability of these “fugitive” methods is not a trivial matter. In most states, annual postcensal population estimates are used to allocate resources and there often are questions about these estimates that lead to conflict in the form of appeals and even legal actions

As a recent case in point, take the method developed for use in Nevada subsequent to the 2000 Census to estimate the total population of Nevada’s counties (City of Las Vegas, 2002). It can be considered fugitive in that this method - known as the “Nevada Regression Model” (NRM) - represents a regression-based method used to allocate resources without the benefit of the quality control provided by academic peer review. This paper provides for the first time just such peer review. It evaluates the Nevada Regression Model (NRM) using as a case study of the use of the model’s estimate of the population of Clark County, Nevada in 2002. Not surprisingly this evaluation was initially generated by conflict over the use of this model that generated an appeal of the numbers it produced by the City of Las Vegas (2002).

The evaluation reveals statistical and methodological shortcomings in the model. This leads to the development of an alternative model, one not subject to these shortcomings and, importantly, with a direct link to the standard forms of regression models used for subnational population estimation. The case study serves to illustrate how the discussion of the use of such fugitive methods as the NRM can lead not only to a wider understanding of methods on the part of practitioners, but the advancement of methodological development and evaluation through the corrective process of peer review.

## BACKGROUND

In Nevada, the State Demographer produces annual estimates for the Department of Taxation, the agency responsible for the certification of annual estimates and the allocation of funds tied to them. In December of 2002, the City of Las Vegas initiated an appeal against the 2002 population estimate derived for Clark County because this estimate impacted the estimate for the City of Las Vegas, which is within Clark County (City of Las Vegas, 2002).

As stated earlier, appeals by local governments of estimates generated by a State Demographer are not uncommon within the world of state and local demography because, as is the case in Nevada, many states use current population estimates in formulas that allocated funds to local governments. This is precisely why Las Vegas appealed the 2002 numbers produced by the NRM (City of Las Vegas, 2002). The methodological heart of the conflict between the City of Las Vegas and the Nevada Department of Taxation centered on the particular regression model that the State Demographer used to generate an estimate for Clark County and for which the City had deep reservations, based in part on empirical evidence (City of Las Vegas, 2002). The essence of the NRM as applied to Clark County, is found in the following regression equation (City of Las Vegas, 2002: 14):

$$\overset{?}{P} = -145379 + 8500.326*GOVT + 578.312*RESADJLF + 599.404*LABOR \quad (1)$$

$$r^2 = 0.999$$

$$s.e.e = 13{,}210.166$$

The model described in Equation (1) and presented by the Nevada State Demographer to the city of Las Vegas was developed using the SPSS "stepwise regression" procedure using the data shown in Table 1. The variable names shown in Table 1 are pretty much self-evident (e.g., "POPULATI" = p = total population of the Clark County, "GOVT" = government employment in Clark County; "RESADJLF" = residentially adjusted Labor Force for Clark County, and "LABOR" = Labor Force in Clark County) as are the coefficients and other model characteristics.

(TABLE 1 ABOUT HERE)

Before proceeding with the evaluation, it is important to note that Nevada regulations largely define the NRM. This is found in section 360.340 of the Nevada Administrative Code (NAC). Additionally, in section 360.365, the NAC also directs the Department of Taxation to use the NRM on an equal basis with the housing unit method for estimating the population of the state as a whole and the counties in Nevada. Further, the NAC states in section 360.370 that, in essence, the estimate for an incorporated place (e.g., the City of Las Vegas) must be "controlled" to the estimate of the county in which it is located. This means that the NRM is an important component in determining the population of an incorporated place such as Las Vegas.

In terms of the NRM, NAC 360.340 states: "Nevada Regression Model means a method, as determined by the department and the demographer employed by the department, by which the population of an area is estimated using population as a dependent variable and employment, labor force, school enrollment, and any other relevant data as independent variables." We will return to this description during the course of the evaluation because of the restrictions it places on the methods that can be used in Nevada to estimate county populations (and indirectly, the population of subcounty areas).

The first step in evaluating the model presented in equation (1) was to replicate it. However, even using the same SPSS "stepwise regression" procedure and the data provided by the Nevada State Demographer to the City of Las Vegas, the resulting model was slightly different, as is shown in Equation (2):

$$\overset{?}{p} = -145{,}216 + 8{,}490.820*GOVT + 578.698*RESADJLF + 599.787*LABOR \quad (2)$$

$$r^2 = 0.999$$

$$s.e.e = 13{,}209.232$$

There are some minor differences in the corresponding coefficients found in equations (1) and (2). They may represent rounding differences or some minor adjustments to data. However, the two models were deemed sufficiently close to continue with the analysis. As such, the model given in Equation (2) is referred to throughout the remainder of this paper as the "Nevada State Demographer's model."

Using the model described in Equation (2), and 2002 values for the three independent variables[1], one finds the following 2002 population estimate for Clark County:

$$1{,}502{,}881 = -145{,}216 + 8{,}490.82 * 80.392 + 578.698 * 856.822 + 599.787 * 783.049 \qquad (3)$$

(GOVT) (RESADJLF) (LABOR)

## CRITICAL ANALYSIS

The second step in the evaluation was to examine the regression model itself. A major component of the evaluation of any regression model is to determine how well it meets the major assumptions underlying multiple regression, which, if violated, can lead to various problems, some of which bias the regression coefficients, some of which affect statistical inference, and some of which do both (Schroeder, Sjoquist, and Stephan, 1986; Lewis-Beck, 1980). In regard to these major assumptions, the Nevada State Demographer's model was found to have problems in regard to high collinearity: Tolerance values are close to zero (from 0.005 to 0.008) for each of the three independent variables, which means not only that the standard errors estimated for the coefficients are artificially inflated, but that the computations can lose numerical accuracy and, further, that the estimate of a variable's regression coefficient is "unstable" (SPSS, 1999: 220-221). This evaluation suggests that the model is a member of the "inadequate set" of possible regression equations that could be generated using the available data and the regulations governing the development of the NRM.

The statistical evaluation of the model must also be complemented by a logical evaluation and although there are several additional statistical grounds on which the regression model can be criticized,[2] the logical evaluation led to the identification of a fundamental methodological shortcoming. This shortcoming lies in the fact that the model's unit of observation is essentially ordered by time, yet the model, as implemented, reveals a failure to understand (and take advantage of) the temporal data that were used to construct it. This methodological shortcoming is manifested in two major ways. The first is that the standard "goodness of fit" measure ($r^2$) used to construct an optimum regression model using these data is inappropriate. The second is that the model uses cross-sectional observations inherently ordered by time in a manner that disregards important temporal information and can, consequently, result in a mis-specified model (Hanke, Wichern, and Reitsch, 2001: 294-298).

## The "Goodness of Fit" Manifestation of The Methodological Shortcoming

In examining the "goodness of fit" issue it is important to understand what $r^2$ does and the nature of the "benchmark" estimator from which $r^2$ is calculated. Almost everybody who uses a regression model understands that "$r^2$" measures the proportionate reduction in error provided by the regression model, such that 0 ? $r^2$ ? 1. However, many regression users do not fundamentally understand that the benchmark against which the regression model is judged is in fact the mean of the dependent variable. That is, $r^2$ measures how much proportionate reduction in error one gets by using the regression model as opposed to using the mean of the dependent variable as the estimator

(Draper and Smith, 1981: 17-22). This is can be quickly illustrated using the data in Table 1 and portrayed in Figure 1.

As found from the data shown in Table 1, over the years from 1980 to 2001, the mean population (POPULATI) of Clark County is 861,184.55, with a standard deviation of 327,430.83. It is the mean value of 861,184.55 (rounded here for discussion purposes to 861,185) that is the "estimator" against which the regression model's precision is being judged (i.e., $r^2$ ). Using the mean as the estimator for "POPULATI" yields an average error of 327,430 (i.e., the standard deviation) over the 22 observations from 1980 to 2001. The average error in using the mean as an estimator is, not surprisingly, quite high. This is because in using the mean, one is not taking into account the fact that population values exhibit a monotonic, nearly linear increase from 1980 to 2001. This can be seen in Figure 1, which shows annual population values, the (linear) trend and the mean value (the horizontal line). This variable exhibits a clear temporal trend. Although not shown here, the three independent variables also exhibit temporal trends to varying degrees.

(FIGURE 1 ABOUT HERE)

In estimating the population of Clark County in any given year (such as 2002), this temporal ordering should be taken into account. For example, as a "current year" estimate it would be easy to show that last year's population would yield a more accurate number than the mean population. For 2001, using the population estimated for 2000, the error is -80,756 (where –80,756 = 1,405,099 – 1,485,855); using the mean population as an estimator of the population in 1001, the error is –624,670 (where –624,670 = 861,185 – 1,485,855). Similar error differences are found for the remaining years.

It is should be clear that the mean population value is not an estimator that would be realistically considered as a viable possibility. Yet, this is precisely what the model implicitly does. The incredibly high $r^2$ (.999) is largely obtained because the regression model's precision is benchmarked against using this very same mean as an estimator. This extremely high $r^2$ is both unrealistic and deceptive. It is unrealistic because this level of "goodness of fit" is rarely, if ever, seen in a regression model used to estimate population; it is deceptive because it "overstates" the actual precision of the model. This is the first manifestation of the fundamental methodological shortcoming in the model. Unfortunately, this shortcoming was viewed as a strength by the Nevada State Demographer (City of Las Vegas, 2002).

## The "Temporal Ordering" Manifestation of the Methodological Shortcoming

The second manifestation of the methodological shortcoming is based on the manner in which the symptomatic indicators and population data were structured to construct a regression model. It fails to take into account the temporal ordering of the entire set of data, as exhibited in Table 1. Instead, the model uses cross sectional observations as if the observations were independent of time, which disregards information to an adequately specified regression model.

This oversight can be seen in several ways. The most obvious is from an evaluation of the presence of “autocorrelation.” This condition is common in time series data, where successive values of a variable are correlated with previous values (Hanke, Wichern, and Reitsch, 2001: 294-298). While the Durbin-Watson (d) statistic calculated for the Nevada Regression Model was inconclusive at *p*=.05 (Hanke, Wichern, and Reitsch, 2001: 298-300) in that it only indicated that positive autocorrelation may be present (d = 1.4142, with n=22, and k=3, where $d_l$ = 1.05 and $d_u$ = 1.66), it is clear from the plot of the residuals (lagged one time period) shown in Figure 3 that positive autocorrelation is present in this model and would be even more pronounced in the absence of a single outlying observation. Figure 1 strongly suggests that the model may be adversely affected by the fact it is not allowing for the inherent time series nature of the data. This again points to the inadequacy of the model in that there are several problems that autocorrelation can cause, including spurious regressions and invalid statistical inference (Hanke, Wichern, and Reitsch, 2001: 296-297).

(FIGURE 2 ABOUT HERE)

Instead of taking each year as a “case” and the data for each year as the values of the variables for each “case,” it would have been much more logical to construct variables that take into account annual changes such as a ratio or a difference. Doing so would have taken advantage of the temporal ordering of the data. And would, moreover, be consistent with the autocorrelation present using a lag of one year. Let us take “ratios” as an example. The use of ratios as a measure of change has a long history in the use of regression models for subnational population estimation and variations on it are widely used (Erickson, 1974; Namboodiri and Lalu, 1971; O'Hare, 1976; Schmitt and Crosetti, 1954; Schmitt and Grier;1966; Swanson, 1978, 1980; Swanson and Tedrow, 1984; Swanson, Tayman and Barr, 2000). It has an advantage over the "difference" method in that as a measure of change it is more independent of absolute population size (Swanson, 1978). It is also suitable to use where there are not extremely small or zero-values observations (Swanson, 1978). These are all characteristics important for Clark County, given the growth in its population between 1980 and 2001 and the absolute size of its population and the symptomatic indicators.

In the ratio method, the variables for a given year are formed by finding the ratio between the current year and the preceding year. In the case of 2001, instead of taking the information given in Table 1 for 2001, one would divide each 2001 value by the corresponding value for 2000. Table 2 shows how the data from Table 1 would appear if annual ratios were used to take advantage of the inherent temporal ordering in the data.

(TABLE 2 ABOUT HERE)

Using the data in Table 2, one could construct a regression model using annual ratios in the symptomatic indicators to estimate annual changes in population. Of the symptomatic indicators available, a reasonable starting selection would include school enrollment, government employment and residentially adjusted labor force. While they are correlated both with population change and with one another, they each capture slightly different aspects of population change and as such represent a better choice than the variables selected by the “stepwise” procedure.

An example of such a regression model to estimate “RPOP” using ratios of the same three symptomatic indicators (RSCHOOL, RGOVT, and RRESADJL) found in equation 2 (as well as in equations 1 and 3) is provided in Equation (4)

$$\overset{?}{\text{RPOP}} = 0.415 + (0.328*\text{RSCHOOL}) + (0.17*\text{RGOVT}) + (0.114*\text{RRESADJLF}) \qquad (4)$$

$r^2 = 0.673$

s.e.e = 0.0104

Note that $r^2$ is much lower in the model shown by equation (4) than that found in the model developed by the Nevada State Demographer: 0.673 vs. 0.999. This is largely because the average annual ratio of population change (1.05638) is far more accurate as an estimator of the ratio of change in a given year than the average population (861,184.55) is as an estimator of the population of Clark County in a given year. This can be seen by comparing Figure 1 and Figure and 3 and, also, by examining the coefficient of variation for each of these two estimators, respectively.[3] In the case of “POPULATI”, the coefficient of variation is 0.38 = (327,430.83/861,184.55); in the case of “RPOP,” it is much lower, 0.02 = (.016756/1.05638). The coefficient of variation for POPULATI is 19 times the size of the coefficient of variation for RPOP. This means that the average of POPULATI has 19 times more error in it as an estimator than does the average of RPOP. Because the average annual ratio of change is much more accurate, there is far less “variance” (i.e., error) to be “explained” in the regression model using ratios. Hence, it has an $r^2$ of 0.673 as opposed to the $r^2$ of 0.999 found in the Nevada State Demographer’s stepwise model. This is a case where “less” is better – a subtle point, but an important one to understand in this case study. Again, a comparison of Figure 1 with Figure 3 illustrates this point.

(FIGURE 3 ABOUT HERE)

Unlike the Nevada State Demographer’s model, the model described by equation (4) does not suffer from high collinearity: for each of the three independent variables, the tolerance values range from 0.660 to 0.798, a range well above zero. This means not only that the standard errors estimated around its coefficients are not artificially inflated, but that the model’s computational precision is reliable and that the estimates of its regression coefficients are stable (SPSS, 1999: 220-221). Moreover, the model meets the other major assumptions underlying multiple regression analysis, including: (1) the residuals do not exhibit any patterns that would suggest model mis-specification; and (2) in the form of ratios, the data are much more “stationary” than in the form used by the State Demographer - the Durbin-Watson statistic (d = 1.851) clearly shows that no autocorrelation is likely to affect the model (at $p$ =.05, with d = 1.851 where $d_l$ = 1.05 and $d_u$ = 1.66, n= 22, and k =3, neither positive nor negative autocorrelation is likely; see, e.g., Hanke, Wichern, and Reitsch, 2001: 298-300). This evaluation (pretty much standard in the development of adequate regression models) suggests that the model shown in Equation (4) is adequate.

Using the model in equation (4), the Clark County population in 2002 would be estimated in two steps:

Step 1. Estimate the relative change in population from 2001 to 2002 using model (4)

1.04281 ? (0.415*1.00) + (0.328*1.035) + (0.17*1.006) + (0.114*1.029)

Step 2. Estimate the 2002 population, by multiplying the estimated relative change by the 2001 population:

1,549,459 ? (1.04281)*(1,485,855)

The model given in equation (4) provides a far more realistic picture of precision by taking into account the inherent temporal ordering of the data not taken into account by the Nevada State Demographer's model. It also uses symptomatic indicators that in the professional judgment of many demographers represent a better choice than those selected by the Nevada State Demographer's stepwise procedure.[4] Not only is it a member of the "adequate" set of potential regression models, it also could be used in conjunction with the other methods allowed under Nevada's statues and regulations to generate an alternative estimate of Clark County's 2002 population.[5]

Another highly desirable characteristic of the model given by equation (4) not found in the Nevada State Demographer's model is that each of the coefficients has a "substantive" interpretation: each gives an approximate "weight" to the effect of change in the symptomatic indicator with which it is associated on changes in population. Note that the sum of coefficients is approximately 1.00 (in actual fact, the sum is approximately 1.03, where 1.03 ? 0.415 + 0.328 + 0.17 + 0.114). Thus, one can view the estimated change in population as a weighted average of the changes in the symptomatic indicators. This is a typical (and highly desirable) characteristic of a regression-based model that uses ratios to estimate population (Prevost and Swanson, 1985). Even the intercept term has this interpretation when you realize that it can be interpreted as the weight given to "no change in population from last year," which is represented as 0.415 = (0.415*1.00) in "Step 1" above. There is no corresponding substantive interpretation for any of the coefficients in the Nevada State Demographer's model. As such, it is not as adequate as the model given by equation (4) in terms of several areas important for the evaluation of population models, including face validity, plausibility and explanatory ease.[6]

Clearly, additional refinements could be made to the Clark County regression-based model illustrated by equation (4). However, these refinements are constrained not only by the statutes and regulations governing the development of estimates in Nevada, but by commonly accepted practice within the state. In the case of the latter, it appears that it is commonly accepted practice to develop regression models that are specific to each county. This is, perhaps, unique among states in that this practice does not allow the use oft the standard "ratio-correlation model, which takes counties, not time points within a given county, as the observations.

However, given the constraints imposed by statues, regulations, and practice, additional refinements are possible. They could include: (1) looking at lagged observations (Swanson and Beck, 1994); (2) considering the use of rates instead of ratios

(Swanson and Tedrow, 1984); (3) considering the development of confidence intervals around the estimates (Swanson, 1989); (4) paying even closer attention to regression diagnostics, to include autocorrelation (Swanson and Tedrow, 1984); (5) developing a better understanding of the relationship between values of the dependent variable and earlier versions of the regression model; and (6) examining historical changes in symptomatic indicators and population levels from a substantive perspective (McKibben and Swanson, 1997). However, without belaboring the point, the three fundamental corrections to the existing model are clear: (1) explicitly incorporate the inherent temporal ordering exhibited in the data; (2) forego the use of a stepwise procedure that is not itself subject to standard evaluations of regression adequacy beyond the most basic criteria such as $r^2$ and statistical significance; and (3) consider conventional approaches to developing regression-based population estimates (e.g., ratio-correlation and its variations) instead of applying multiple regression to a single county using a historical time series based on annual numbers, which, in part, are based on estimated values of the dependent variable, which, in turn, are at least partly derived from the regression model itself. These suggestions appear to be consistent with the statues and regulations governing the use of the NRM, if not what appears to be common practice in Nevada.

## CONCLUDING REMARKS

As stated at the outset of this paper, much of the academic literature dealing with state and local demography involves the development and evaluation of methods for estimating population.  The quality control in regard to validity and reliability afforded these methods by the peer review process embedded in the academic literature is important because, among other things, it serves to reduce the high potential for conflict that exists when resources are at stake. There are, however, methods being used by state and local demographers that have not been subject to peer review.  While not necessarily unsound, these "fugitive" methods serve to keep the potential for conflict high because of the uncertainty regarding their validity and reliability. The Nevada Regression Model examined in this paper illustrates these points. Moreover, its examination illustrates the importance of subjecting fugitive methods to peer review in that in this case the method in question has been found to suffer from serious shortcomings. The shortcomings, once identified, led directly to a regression model that does not suffer from them. This case study is designed to illustrate how this type of analysis and discussion can lead to a wider understanding of methods on the part of practitioners through the corrective process of peer review. It also suggests that states in which estimates are used to allocate resources would be well-served by subjecting new methods being considered for use to peer review before they are adopted.

This case study also serves as an affirmation of the benefits provided by the Schmitt and Crosetti (1954) ratio-correlation method that, among other positive features, serve to free it from the presence of autocorrelation, an adverse condition to which any regression model using time-ordered data is particularly at risk. Whether or not Schmitt and Crosetti were consciously aware of this and other positive features, it is a testament to their judgment and insights that led them to develop a regression-based estimation method that avoids the autocorrelation problem inherent in the Nevada Regression Model. The illustration of the very real problems that may affect in a regression model that strays too far from the ratio (and difference)- correlation logic should serve as a clear example of the benefits to be gained by understanding a method that has nearly 50 years of use, evaluation, and refinements, much of which is in the academic literature. Clearly, the feedback and communication provided by this exposure has provided a large measure of quality control in regard to the use of ratio-correlation model and its variants and has contributed to its widespread use and satisfactory performance. The Nevada Regression Model would be equally well-served by the same exposure, as would other fugitive estimation methods employed by state and local demographers.

**ENDNOTES**

1. The full set of values for all of the independent variables (in thousands) in 2002 is as follows:

Employment, 889.651;

Labor Force,783.049;

Private Non-Farm, 808.933;

Government, 80.392;

Farm, 0.326;

Residence Adjusted Employment, 856.822; and

School Enrollment ,254.912.

2. Other shortcomings include: (1) The model was not constructed using a random sample, (hence why use statistical inference as a guide to variable selection?); (2) stepwise regression is an automated procedure in which the analyst yields model construction to criteria that may not be appropriate - it is a controversial approach (Tabachnick and Fidell, 1996:150-156); and (3) an inadequate use of regression diagnostics – there is no discussion of residuals (e.g., hetereoskedasticity, model mis-specification), no discussion of the low tolerance values, no discussion of autocorrelation, and so on.

3. The coefficient of variation is formed by dividing the standard deviation by the mean. It is independent of the unit of measurement and for this reason is useful in comparing distributions where the units of measurement are different (Spiegel, 1961: 73), as is the case with POPULATI and RPOP.

4. It is important to note that the use of judgment in selecting independent variables for use in a regression model is not an inherently arbitrary process. Ultimately, all regression models should be evaluated for adequacy using a set of well-defined conventions (Belsey, Kuh, and Welsch, 1980; Berry and Feldman, 1985; Tabachnick and Fidell, 1996: 131-139). These evaluations not only reveal the efficacy of potential independent variables (and transformations thereof), but also the adequacy of the judgments used in constructing a given model. It is exactly this type of evaluation that determines if a model is a member of the inadequate set or the adequate set that potentially could be generated in given situations.

5. One other method that can be used under the Nevada Regulations in informing estimates of Clark County and the statistical and administrative units within it is the Housing Unit Method (Roe, Swanson, and Carlson, 1992; Smith and Lewis, 1980). Estimates resulting fro this method formed the empirical framework within which the City of Las Vegas decided to appeal the initial estimate generated by use of the NRM.

6. These criteria are listed as important for the evaluation of population projections by Smith, Tayman, and Swanson (2001: 280-299), but they apply equally to the evaluation of population estimates.

7. The definitions for these variables are found in the Appeal Letter. (City of Las Vegas, 2002).

8. These were formed by taking annual ratios of the data in Table 1.

## REFERENCES

Belsey, D., E. Kuh, and R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York, NY: John Wiley and Sons.

Berry, W. and S. Feldman. 1985. *Multiple Regression in Practice*. Sage University Papers, Quantitative Applications in the Social Sciences. No. 50. Beverly Hills, CA: Sage Publications.

Bogue, D. 1950. A Technique for making extensive population estimates. *Journal of the American Statistical Association* (45): 191-193

Bousfield, M. 1977. Intercensal estimation using a current sample and census data. *Review of Public Data Use* 5: 6-15.

Cannan, E. 1895. The probability of the cessation of the growth of population in England and Wales during the next century. *The Economic Journal* 5: 506-515.

City of Las Vegas. 2002. *City of Las Vegas Appeal Letter* (December 12th, to Charles F. Chinnock, Executive Director, Department of Taxation, State of Nevada, 1500 East College Parkway, Suite 115, Carson City, NV 89706; from Douglas A. Selby, City Manager, City of Las Vegas, City Hall, 400 Stewart, Las Vegas, NV 89101).

Draper, N. and H. Smith. 1981. *Applied Regression Analysis, Second Edition.* New York, NY: Wiley-Interscience.

Eldridge, 1947. Problems and methods of estimating postcensal population. *Social Forces* 24: 41-46.

Erickson, E. 1974. A regression method for estimating population changes of local areas. *Journal of the American Statistical Association* 69: 867-875.

Hamilton, C. H. 1964. Comparison of two formulas in making population estimates. *Rural Sociology* 29: 426-431.

Hanke, J., D. Wichern, and A. Reitsch. 2001. *Business Forecasting, 7th Edition*. Upper Saddle River, NJ: Prentice Hall

Krotki, K. 1978. (Ed.). *Developments in Dual System Estimation of Population size and Growth*. Edmonton, Alberta, Canada: The University of Alberta Press.

Land, K. and G. Hough. 1986. Improving the accuracy of intercensal estimates and postcensal projections of the civilian noninstitutional population: a parameterization of institutional prevalence rates. *Journal of the American Statistical Association* 81: 62-74.

Lewis-Beck, M. 1980. *Applied Regression: An Introduction.* Sage University Papers, Quantitative Applications in the Social Sciences No. 22. Beverly Hills, CA. Sage Publications

Mandell, M. and J. Tayman, 1982. Measuring temporal stability in regression models of population estimation. *Demography* 19: 135-146.

McKibben, J. and D. Swanson. 1997. Linking substance and practice: a case study of the relationship between socio-economic structure and population estimation. *Journal of Economic and Social Measurement* 24: 135-147.

McVey, W. 1974. *An empirical assessment of a modified censal ratio estimation technique*. Unpublished Ph.D. dissertation, The University of Alberta, Edmonton, Alberta, Canada.

Morrison, P. 1982. Different approaches to monitoring local demographic change. *RAND paper P-6743*. RAND Corporation, Santa Monica, CA.

Namboodiri, N. K. 1972. On the ratio-correlation and related methods of subnational population estimation. *Demography* 9: 443-453.

Namboodiri, N. K. and N. M. Lalu, 1971. The average of several simple regression estimates as an alternative to the multiple regression estimate in postcensal and intercensal population estimation: a case study. *Rural Sociology* 36: 187-194.

O'Hare, W. 1976. Report on a multiple regression method for making population estimates. *Demography* 13: 369-379.

O'Hare, W. 1980. A note on the use of regression models for making population estimates. *Demography* 17: 341-343.

Prevost, R. and D. Swanson. 1986. A new technique for assessing error in ratio-correlation estimates of population: a preliminary note. *Applied Demography* 1 (November): 1-4.

Purcell, N. and L. Kish. 1980. Postcensal estimates for local areas(or domains). *International Statistical Review* 48: 3-18.

Rives, N. W. Serow, A. Lee, and H. Goldsmith. 1989. *Small Area Analysis: Estimating Total Population*. Rockville, MD: U.S. Department of Health and Human Services.

Roe, L., D. Swanson, and J. Carlson. 1992. A variation of the housing unit method for estimating the population of small, rural areas: a case study of the local expert procedure. *Survey Methodology* 18: 155-163.

Rosenberg, H. 1968. Improving current population estimates through stratification. *Land Economics* 44: 331-338

Schmitt, R. 1952. Estimating current populations of census tracts. *Sociology and Social Research* 37: 12-15.

Schmitt, R. and A. Crosetti. 1954. Accuracy of the ratio-correlation method for estimating postcensal population. *Land Economics* 30: 279-281.

Schmitt, R. and J. Grier. 1966. A method of estimating the population of minor civil divisions. *Rural Sociology* 31: 355-361

Schroeder, L., D. Sjoquist, and P. Stephan. 1986. *Understanding Regression: An Introductory Guide.* Sage University Papers, Quantiative Applications in the Social Sciences No. 57. Beverly Hills, CA. Sage Publications

Shryock, H. 1938. Methods of estimating postcensal population. *American Journal of Public Health* 28: 1042-1047.

Siegel, J., H. Shryock and B. Greenberg. 1954. Accuracy of postcensal estimates of population for states and cities. *American Sociological Review* 19: 440-446.

Smith, S. and S. Cody. 1994. Evaluating the housing unit method: a case study of 1990 population estimates in Florida. *Journal of the American Planning Association* 60: 209-221.

Smith, S. and B. Lewis. 1980. Some new techniques for applying the housing unit method of local population estimation. *Demography* 17: 323-339.

Smith, S., J. Nogle, and S. Cody. 2002. A regression approach to estimating the average number of persons per household. *Demography* 39 (4): 697-712.

Smith, S., J. Tayman, and D. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York, NY: Kluwer Academic/Plenum Publishers

Snow, E. C. 1911. The application of the method of multiple correlation to the estimation of post-censal populations. *Journal of the Royal Statistical Society* 74 (part 6): 575-620.

Spar, M.  and J. Martin. 1979.  refinements to regression-based estimates of postcensal population characteristics. *Review of Public Data Use* 7: 16-22.

Spiegel, M. 1961. *Theory and Problems of Statistics*. New York, NY: McGraw Hill (Schaum's Outline Series).

SPSS, Inc. 1999. *SPSS Basic 10.0 Applications Guide*. Chicago, IL: SPSS, Inc.

Starsinic, D.  and M. Zitter. 1968. Accuracy of the housing unit method in preparing population estimates for cities. *Demography* 5: 475-484

Swanson, D. 1989. Confidence intervals for postcensal population estimates: a case study for local areas. *Survey Methodology* 15: 271-280.

Swanson, D. 1980. Improving accuracy in multiple regression estimates of population using principles from causal modeling. *Demography* 17: 413-428.

Swanson, D. 1978. An evaluation of ratio and difference regression methods for estimating small, highly concentrated populations: the case of ethnic groups. *Review of Public Data Use* 6: 18-27.

Swanson, D. and D. Beck. 1994. A new short-term county population projection method. *Journal of Economic and Social Measurement* 20: 25-50.

Swanson, D. and L. Tedrow. 1984. Improving the measurement of temporal change in regression models used for county population estimates. *Demography* 21: 361-372.

Swanson, D., J. Tayman, and C. Barr. 2000. A note on the measurement of accuracy for subnational demographic estimates. *Demography* 37: 193-201.

Tabachnick, B. and L. Fidell. 1996. *Using Multivariate Statistics, 3rd Edition*. New York, NY: HarperCollins.

U.S. Bureau of the Census, 1949. Illustrative examples of two methods of estimating the current population of small areas. Current Population Reports, Series P-25, No. 20. Washington, DC: U.S. Bureau of the Census.

Zitter, M. and H. Shryock. 1964. Accuracy of methods of preparing postcensal population estimates for state and local areas. Demography 1(1): 227-241.

**Table 1. Original Regression Data[7]**

| YEAR | CENPOP | POPULATI | SCHOOl | PRIVATE | GOVT | FARM | TOTAL | RESADJLF | LABOR |
|---|---|---|---|---|---|---|---|---|---|
| 2002 | , | , | 254912 | 808,651 | 80,392 | 0,326 | 889,651 | 856,822 | 783,049 |
| 2001 | 1478788 | 1485855 | 246289 | 784,15 | 79,93 | 0,33 | 864,41 | 832,99 | 763,01 |
| 2000 | 1405099 | 1405099 | 231655 | 777,14 | 78,87 | 0,33 | 856,35 | 825,93 | 726,68 |
| 1999 | 1336506 | 1336506 | 217139 | 739,68 | 75,70 | 0,34 | 815,72 | 787,05 | 646,75 |
| 1998 | 1269276 | 1269276 | 203616 | 689,40 | 72,48 | 0,34 | 762,21 | 735,87 | 615,29 |
| 1997 | 1200286 | 1200286 | 190822 | 658,18 | 68,29 | 0,34 | 726,81 | 701,88 | 588,93 |
| 1996 | 1124654 | 1124654 | 178896 | 609,51 | 64,86 | 0,32 | 674,69 | 651,26 | 552,99 |
| 1995 | 1059946 | 1059946 | 166788 | 554,76 | 62,15 | 0,30 | 617,22 | 597,71 | 523,87 |
| 1994 | 994007 | 994007 | 156348 | 517,76 | 59,84 | 0,36 | 577,96 | 560,41 | 495,22 |
| 1993 | 919138 | 919138 | 145327 | 454,74 | 57,47 | 0,37 | 512,57 | 498,40 | 460,53 |
| 1992 | 875308 | 875308 | 136188 | 423,89 | 56,36 | 0,33 | 480,58 | 468,17 | 445,99 |
| 1991 | 836003 | 836003 | 129233 | 418,33 | 54,44 | 0,34 | 473,11 | 460,92 | 431,33 |
| 1990 | 770280 | 770280 | 121984 | 408,44 | 50,69 | 0,40 | 459,54 | 447,30 | 402,50 |
| 1989 | 708750 | 708750 | 111460 | 369,94 | 48,88 | 0,41 | 419,23 | 407,71 | 375,42 |
| 1988 | 661690 | 661690 | 105151 | 335,17 | 45,10 | 0,45 | 380,72 | 371,99 | 347,78 |
| 1987 | 616650 | 616650 | 100027 | 305,57 | 44,42 | 0,44 | 350,43 | 340,98 | 327,32 |
| 1986 | 587760 | 587760 | 95412 | 277,71 | 42,95 | 0,46 | 321,12 | 312,75 | 307,76 |
| 1985 | 562280 | 562280 | 91446 | 261,82 | 41,08 | 0,44 | 303,34 | 295,36 | 294,47 |
| 1984 | 539030 | 539030 | 89742 | 248,02 | 40,37 | 0,41 | 288,79 | 279,92 | 284,14 |
| 1983 | 525050 | 525050 | 89258 | 234,34 | 41,31 | 0,43 | 276,08 | 265,33 | 275,48 |
| 1982 | 507510 | 507510 | 89646 | 230,14 | 41,23 | 0,41 | 271,78 | 261,38 | 267,35 |
| 1981 | 491620 | 491620 | 89547 | 234,38 | 39,15 | 0,43 | 273,96 | 263,68 | 256,09 |
| 1980 | 469362 | 469362 | 88543 | 227,77 | 37,73 | 0,42 | 265,92 | 255,13 | 242,54 |

**Table 2. Regression Data Using Annual Ratios[8]**

| YEAR | RCENPOP | RPOPULATI | RSCHOOl | RPRIVATE | RGOVT | RFARM | RTOTAL | RRESADJLF | RLABOR |
|---|---|---|---|---|---|---|---|---|---|
| 2002 | | | 1,035 | 1,031 | 1,006 | 0,982 | 1,029 | 1,029 | 1,026 |
| 2001 | 1,052 | 1,057 | 1,063 | 1,009 | 1,013 | 1,000 | 1,009 | 1,009 | 1,050 |
| 2000 | 1,051 | 1,051 | 1,067 | 1,051 | 1,042 | 0,982 | 1,050 | 1,049 | 1,124 |
| 1999 | 1,053 | 1,053 | 1,066 | 1,073 | 1,045 | 1,009 | 1,070 | 1,070 | 1,051 |
| 1998 | 1,057 | 1,057 | 1,067 | 1,047 | 1,061 | 0,982 | 1,049 | 1,048 | 1,045 |
| 1997 | 1,067 | 1,067 | 1,067 | 1,080 | 1,053 | 1,079 | 1,077 | 1,078 | 1,065 |
| 1996 | 1,061 | 1,061 | 1,073 | 1,099 | 1,044 | 1,039 | 1,093 | 1,090 | 1,056 |
| 1995 | 1,066 | 1,066 | 1,067 | 1,071 | 1,039 | 0,854 | 1,068 | 1,067 | 1,058 |
| 1994 | 1,081 | 1,081 | 1,076 | 1,139 | 1,041 | 0,975 | 1,128 | 1,124 | 1,075 |
| 1993 | 1,050 | 1,050 | 1,067 | 1,073 | 1,020 | 1,093 | 1,067 | 1,065 | 1,033 |
| 1992 | 1,047 | 1,047 | 1,054 | 1,013 | 1,035 | 0,974 | 1,016 | 1,016 | 1,034 |
| 1991 | 1,085 | 1,085 | 1,059 | 1,024 | 1,074 | 0,853 | 1,030 | 1,030 | 1,072 |
| 1990 | 1,087 | 1,087 | 1,094 | 1,104 | 1,037 | 0,976 | 1,096 | 1,097 | 1,072 |
| 1989 | 1,071 | 1,071 | 1,060 | 1,104 | 1,084 | 0,914 | 1,101 | 1,096 | 1,079 |
| 1988 | 1,073 | 1,073 | 1,051 | 1,097 | 1,015 | 1,025 | 1,086 | 1,091 | 1,062 |
| 1987 | 1,049 | 1,049 | 1,048 | 1,100 | 1,034 | 0,963 | 1,091 | 1,090 | 1,064 |
| 1986 | 1,045 | 1,045 | 1,043 | 1,061 | 1,045 | 1,046 | 1,059 | 1,059 | 1,045 |
| 1985 | 1,043 | 1,043 | 1,019 | 1,056 | 1,018 | 1,074 | 1,050 | 1,055 | 1,036 |
| 1984 | 1,027 | 1,027 | 1,005 | 1,058 | 0,977 | 0,942 | 1,046 | 1,055 | 1,031 |
| 1983 | 1,035 | 1,035 | 0,996 | 1,018 | 1,002 | 1,046 | 1,016 | 1,015 | 1,030 |
| 1982 | 1,032 | 1,032 | 1,001 | 0,982 | 1,053 | 0,952 | 0,992 | 0,991 | 1,044 |
| 1981 | 1,047 | 1,047 | 1,011 | 1,029 | 1,038 | 1,036 | 1,030 | 1,034 | 1,056 |

**Figure 1. Clark County Population, 1980 to 2001**

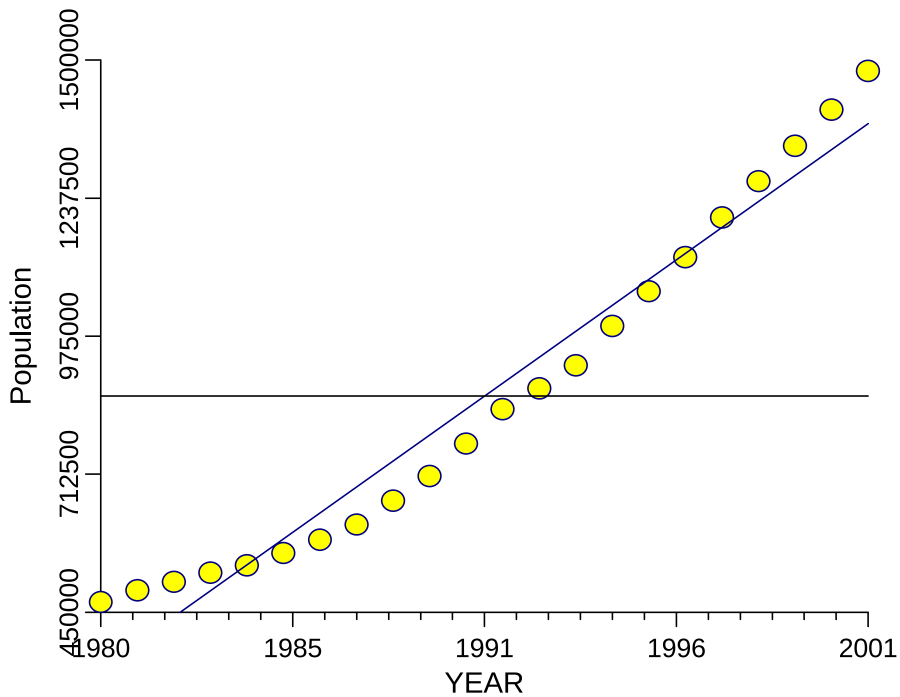

**Figure 2**

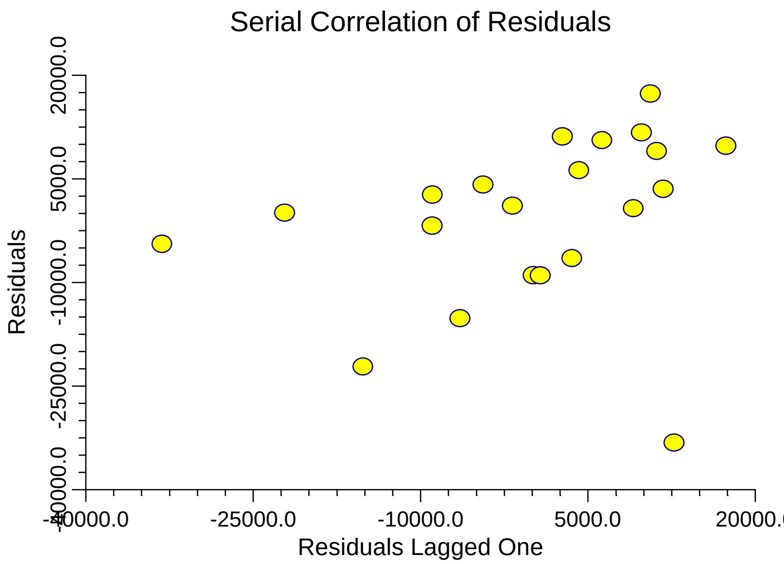
Serial Correlation of Residuals
Residuals
20000.0
5000.0
-10000.0
-25000.0
-40000.0
-40000.0
-25000.0
-10000.0
5000.0
20000.0
Residuals Lagged One

**Figure 3.  Ratio of Year-to-Year Population, Clark County, 1981 to 2001**