BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-90/02

REPORT ON DEMOGRAPHIC ANALYSIS SYNTHETIC ESTIMATION FOR SMALL AREAS

by

Cary T. Isaki
Statistical Research Division
Bureau of the Census
Room 3132, F.O.B. #4
Washington, D.C. 20233 U.S.A.

Recommended: Nash J. Monsour

Report completed: January 2, 1990

Report issued: January 9, 1990

SMALL AREA ESTIMATION OF TOTAL RETAIL SALES USING PUBLISHED DATA

BY

CARY T. ISAKI

STATISTICAL RESEARCH DIVISION

BUREAU OF THE CENSUS

Table of Contents

Small Area Estimation of Total Retail Sales Using Published Data
by Cary T. Isaki

Executive Summary

The purpose of this report is to document a study of possible small area estimators of economic data at the state and county small area level where no such estimates currently exist. The estimators are particularly appealing because they require data inputs available to the public. The users can also augment the proposed methods by utilizing data available for their particular application.

The main idea is to model successive retail sales change from data published in two successive Retail Trade Censuses and construct regression equations based on explanatory variables available in County Business Patterns and the Bureau's population estimation reports. Then, using the Annual Retail Trade Survey U.S. figure, estimates of total retail trade sales for states and counties can be produced. The models and estimators used are detailed in section III of the report. Principally, regression and synthetic estimation is used. In this report, 1972 and 1977 census results are used for modelling and 1982 census results are used to assess the performance of the estimators.

Various measures of performance were used to assess how well the estimators performed. The measures of performance are defined in section V with the numerical work detailed in the Appendix and comparisons among estimators presented in section VI. Focussing on states and the absolute relative error (ARE) we have that the ARE's range from .03% for New Mexico to 9.13% for Alaska. The ARE's exceeded 5% for seven states with a mean ARE of 2.68%. Excluding Alaska, the largest ARE was 6.53% for West Virginia. The

mean ARE of counties by state ranged from a little over 2% for Connecticut to a little over 13% in Nevada. ARE's for individual counties were rather large, however. The three counties with the largest AREs were in Georgia, Nebraska and South Dakota with ARE's of 110.1%, 83.7% and 84.9%, respectively. These counties were rather small in terms of 1982 retail sales and when compared to other counties in their respective states. Hence, the error is not likely to affect the use of such estimates in allocation measures based on sales. Because, 1982 sales is known for all counties, it is possible to identify units with large ARE's in this study. In practice, we will not be able to identify such units.

We observe that for many states a hybrid estimator consisting of regression based estimates for large units and synthetic estimates for smaller units appears to improve on either of its component parts. We term such an estimator H. Finally, a strategy termed T, is proposed for estimating counties by state. Strategy T uses the synthetic, regression and H for various states depending on certain conditions based on number of counties and the regression fit. A summary of the performance of the various estimators is given in section VI.d, Table 3.

Since the estimators are derived from models of data from a previous time period, there is no guarantee that past performance will imply similar performances in the future. Also, we have no way of examining the performance of the methods for intervening years, e.g., years 1978 through 1981 in our application. The results of the 1987 censuses are soon to be completed. It will be possible, then, to repeat the work to observe the utility of the above approach of small area estimation of retail sales. Determination of usefulness of the estimates rests with the users of the CBP publication.

Small Area Estimation of Total
Retail Sales Using Published Data
by Cary T. Isaki

I. Introduction

The Bureau of the Census publishes annual economic data at the county level in each economic census (at five year intervals). The Bureau also publishes much annual economic data at the county level yearly (but not all economic data of interest). The Bureau publishes important economic data annually and monthly but not at the small area level such as the county. For example, annual retail sales at the state and county level are not regularly available for non-census years. In the wholesale and service industries areas, annual state estimates are not published regularly.

The purpose of this report is to document research results concerning the estimation of annual total retail sales for states and counties in non-census years. The proposed methods can be adapted for use in the wholesale and selected services areas (at least at the state level).

The report is divided into 10 sections. Section II describes the data used in model formulation, analysis and assessment. Section III contains a description of the models, resulting small area estimators, strategies that combine estimators and a literature review. Section IV describes the procedure used to transcribe data to diskette, i.e. the keying and editing and also the computing hardware and software. Section V contains the measures of performance used to evaluate the various small area estimators. Section VI contains the results of the research with comparisons of the estimators by measures of performance and separately by estimators over measures of performance. Sections VII and VIII contain a discussion and summary and suggestions for future work, respectively. Sections IX and X contain the list of references and an Appendix. The Appendix includes the data summary and measures of performance for the small area estimators by state.

I.a. Goals of the research

Small area estimation research aims at constructing and analyzing estimators of characteristics at the small area level. In the current context, various models are investigated for producing small area estimators given the availability of published data. By constructing small area estimators we hope to fill the current gap of the unavailability of annual total retail sales at the county level for all counties in the U.S. Using measures of performance we then evaluate the competing small area estimators so as to select the best overall performer.

I.b. Procedure

The procedure that was followed in this research was to consider the various sources of information available. Then methods of small area estimation were proposed. These methods were then applied in an operational setting. The resulting performance of the estimators were then measured against a known standard. The numerical county results are provided by state in the Appendix. The state results are also provided in the Appendix.

I.c. Results

The results of the research indicate that viable small area estimation methods are available and clearly implementable. Several promising estimators and a strategy appear to provide useful estimates (of course, the user of the data is the ultimate assessor of usefulness).

II. Background

We briefly describe the sources of information available for use in small area estimation, their limitations and document their source.

II.a. Sources of Information

As mentioned previously, the Bureau of the Census' Census of Retail Trade is the main source of data concerning county level information. For our work, we used county data from years 1972 and 1977 for model development and county data for 1982 to assess the model results. From the Retail Trade Census publications, county total retail sales, total retail payroll, number of retail establishments and number of retail employees were used. The data referred to retail employers only, i.e. all retail establishments with payroll.

The second data source is the Bureau's County Business Patterns publication. In this publication, the above mentioned data, excluding retail sales, are produced annually by county. The availability of this current information motivated the construction of the small area estimators that follow. As will become evident, the CBP data was not used in the current analysis because nearly equivalent information was easily accessible in the census publications. The CBP data are crucial for practical applications, however. The third data source is available through the Bureau's Population Division where annual population counts are estimated for counties by state. The fourth data source is the Bureau's current programs where the Business Division provides monthly retail sales estimates by varying degrees of kind of business and by geography. The most detailed kind of business sales estimates are provided at the total U.S. geographic level. The least detailed kind of business sales estimate (total retail) are provided for some large cities but rarely, if ever, at the county level. The Bureau's current programs include the Monthly Retail Trade Survey and the Annual Retail Trade Survey.

II.b. Limitations

Ideally, total retail sales (not just for employers) would be desirable. Since the CBP data is for employers only and because it is crucial for updating estimates, we are limited to providing retail sales for employers only. Occasionally, the CBP does not provide information for some counties in a state. When this happens, we may not be able to provide a retail sales estimate for those counties. These counties are invariably small in population and in retail sales.

The Retail Trade Census also omits publication of some county information. These omissions are due to disclosure, i.e. publication of such information would violate the Bureau's need to maintain confidentiality of information received. Invariably, the counties affected have a very small number of establishments. Out of 3115 counties considered for model development, 60 were not used for model development or assessment because of disclosure. Another seven counties were removed because they were deemed outliers. "Counties" (referred to as burroughs) in Alaska were not modelled because geographic identification of the "counties" differed between censuses. Hence it was not possible to link county census data over the three census years.

II.c. How the Data Are Used

In this section we briefly describe how the data from the various sources are used in this report. The data from the 1972 and 1977 Retail Trade Census are used to model data and construct small area estimators. The Bureau's annual estimates of population and of retail trade are also used for these purposes. Finally, separate portions of the 1982 Retail Trade Census data are used to both produce small area estimates and serve as a means of assessment of the methods.

Note that the CBP data are not used in this report. This is because 1) CBP data collected during economic census years have been found to be nearly identical with comparable data published in the census and 2) it was easier to access CBP equivalent data from published census volumes than by accessing CBP volumes. For example, the same source table that provided CBP equivalent data in the Retail Trade Census volume also contained retail sales. We avoided some clerical processing error by eliminating the step of combining recorded information via matching of counties (sales from census volumes matched with other CBP data.

Suppose it is hypothesized that county ratios of successive census retail sales are linearly related to comparable ratios of payroll, number of establishments, etc. Then, the 1972 and 1977 Census of Retail Trade and population data are used to fit a regression. The resulting regression coefficients are used together with corresponding ratios based on 1977 and 1982 Retail Trade Census and 1982 Annual Retail Trade Survey data to estimate 1977 to 1982 ratios of retail sales and also 1982 county and state retail trade sales. Finally, 1982 Census of Retail Trade sales data are then used to assess the method.

Note that in the above, say for estimating sales in 1981, 1982 Retail Trade Census type information is not available. This is where, in practice, 1981 CBP data would be used. 1982 Census of Retail Trade data was used because we were assessing performance and 1982 retail sales was needed.

III. Models (Methods)

We considered two general types of estimators for small area estimation (county and state retail sales). The two types of estimators are synthetic and regression. Within the regression method several models are considered. Basically, the regression method is the ratio correlation method introduced by Snow (1911) in the demographic small area context. A combination of the synthetic and regression estimates, termed a hybrid estimate, is also examined. Finally, a strategy that combines all of the estimation procedures is proposed and evaluated. We begin with a description of the estimation methods.

III.a. Synthetic Estimator

Synthetic estimation in the small area context basically assumes that relationships existing for a larger area also hold for the individual smaller areas contained within the larger area. In our situation of estimating retail sales for counties, consider the larger area as being the state.

Let $Xstate_{1980}$ denote an estimate of the state's annual retail sales for 1980, the year of interest. Let X_{1977i} denote the i^{th} county's annual retail sales from the last census, 1977. Then, if we assume each county's 1980 sales has increased in the ratio as did the state's $(Xstate_{1980}/\sum_i^N X_{1977i})$, then a synthetic estimator of county i's 1980 annual retail sales is

$$SYN_i = \left[X_{1977i} / \sum_i^N X_{1977i} \right] Xstate_{1980} \qquad (1)$$

where

N is the number of counties in the state. Another way to view SYN_i is that county i's share of 1980 total retail sales is the same as its state share of 1977 retail sales.

Retail sales and retail payroll are highly correlated at the U.S. and state level. Another estimator considered in our work is labelled $SYN1_i$. $SYN1_i$ is not strictly speaking, a synthetic estimator but we introduce it here for brevity. The basic assumption underlying $SYN1_i$ is that the rates of change of retail sales equals the rates of change for retail payroll for the levels of analysis (states or counties).

For example, if X_{1977i}, Y_{1982i} and Y_{1977i} represent the 1977 retail sales, 1982 retail payroll and 1977 retail payroll of county i and we are interested in an estimate for 1982 retail sales, the synthetic estimator is defined by

$$SYN1_i = \left[Y_{1982i} \;/\; Y_{1977i}\right] X_{1977i}. \qquad (2)$$

In the following, we omit the word retail. For years between 1977 and 1982, Y_{1982i} in (2) would be replaced by the appropriate figure. The synthetic estimator is simple to construct and requires information from the County Business Patterns (for payroll) and from the most recent economic census. In a sense, $SYN1_i$ is a special case of a regression model below when ratios of level are used instead of ratios of proportions with no intercept and regression coefficient of unity. Mitch Trager of the Bureau of the Census suggested $SYN1_i$ as a possible small area estimator. The reader is referred to Purcell and Kish (1980) for an excellent review of small area estimation methods particularly synthetic estimation. Levy (1971) lays out a synthetic estimation model in a survey framework where national estimates of mortality are adequate but state estimates require synthetic methods as a means of estimation. Where no confusion can result, we also use N in place of SYN1.

III. b. Regression Estimators

The regression method is similar to the synthetic method but also allows for variables other than payroll in "explaining" the change in sales. Suppose that population change is also related to the change in sales. Let the variable Z denote population. The regression method assumes that the rate of change of sales is a linear combination of the rate of change of payroll and of population. In notation, we have

$$\left[X_{1977i}/X_{1972i}\right] = \beta_0 + \beta_1\left[Y_{1977i}/Y_{1972i}\right] + \beta_2\left[Z_{1977i}/Z_{1972i}\right] + e_i \qquad (3)$$

where $e_i \sim N(0,\sigma^2)$.

Let $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the resulting least squares estimates. If we can assume that the relationship in (3) holds for future changes (the β's do not change much), then we can utilize (3) to estimate sales change for years following 1977. For example, if 1981 sales is of interest, denoting the regression estimator of sales for unit i as $\hat{R}_i$ we have

$$\hat{R}_i = \left[\hat{\beta}_0 + \hat{\beta}_1\left\{Y_{1981i}/Y_{1977i}\right\} + \hat{\beta}_2\left\{Z_{1981i}/Z_{1977i}\right\}\right]X_{1977i}. \qquad (4)$$

To produce $\hat{R}_i$ one requires data from both the 1972 and 1977 economic censuses, the 1981 County Business Patterns and population estimates for 1972, 1977 and 1981. All of these data are published by the Bureau.

A variant in (3) above is to model ratio of sales proportions versus ratios of payroll and population proportions. If the unit of analysis i is the state, then the variables X, Y and Z are replaced by proportions of state to total U.S. characteristics. For example, in place of X_{1977i} in (3) we would use $P^X_{1977i} = X_{1977i} / \sum_{i=1}^{51} X_{1977i}$. The resulting regression estimator of retail sales, denoted $\hat{S}_i$, is then

$$\hat{S}_i = \left[\hat{\alpha}_0 + \hat{\alpha}_1 \left\{ P^Y_{1981i}/P^Y_{1977i} \right\} + \hat{\alpha}_2 \left\{ P^Z_{1981i}/P^Z_{1977i} \right\} \right] P^X_{1977i} \text{ X } U.S._{1981} \quad (5)$$

where $U.S._{1981}$ denotes the 1981 estimated sales for the U.S. This latter figure is obtainable from the Bureau's Annual Retail Trade Survey. $\hat{S}$ requires this additional bit of information over that required by $\hat{R}$.

When the unit of analysis is the county, then the proportions in (5) represent that over the relevant state and the U.S. figure is then replaced by the state's 1981 sales total. Since this state figure is not always available nor useful, we use the estimated state figure available from (5).

Schmitt and Crosetti (1964) modelled county population changes between the 1930 and 1940 censuses using 1) to 4) and assessed accuracy by comparing their results against the 1950 census. Their data set consisted of 39 counties in Washington state. Rosenberg (1968) proposed stratifying Ohio counties and constructing separate ratio-correlation models by strata. He reported gains in his procedure over the unstratified ratio-correlation method. Namboodiri and Lalu (1971) and Namboodiri (1972) considered several ways of averaging a set of simple regressions of ratios of population change from 1940 to 1950 censuses for estimating 1960 population in North Carolina counties. Their method consistently outperformed the ratio-correlation method. They pointed to a change in the variance covariance matrices (1940 to 1950 versus 1950 to 1960) as the principal cause. Such a change in the covariance matrices is termed temporal instability.

O'Hare (1976) proposed using differences of proportions (termed difference correlation) instead of ratios of proportions as the dependent variable in multiple regression. Applied to Michigan counties, using differences of proportions yielded smaller mean absolute relative errors than several other population estimation methods. Mandell and Tayman (1982) using Florida counties demonstrated that the improved performance of the difference-correlation method over the ratio-correlation method is dependent on the choice of explanatory variables. Martin and Serow (1978) compared several methods of estimating total population and population of subgroups (age and race) for counties in Virginia. They considered such methods as stratification, dummy variables, nonstratified multiple regression and averages of simple regression. They concluded that the nonstratified multiple regression (ratio-correlation) performed consistently better over all types of dependent variables.

In our work we chose to use the ratio-correlation method as presented above and also with a slight modification to be discussed later. Because of temporal instability there is no guarantee that a model based on previous census relationships will hold in the future. What is hoped is that the model based on previous census data produces adequate estimates in the applied setting.

III.c. Other Estimators

In the course of developing and analyzing synthetic (denoted N in what follows) and regression estimators of county sales it was observed that the synthetic estimators performed better for measures of performance that weighted errors equally while the regression estimators appeared to perform better under measures of performance that weighted errors by the size (sales) of the county. This suggested an hybrid estimator, denoted H, that estimated low sales counties using N and high sales counties using $\hat{R}$ or $\hat{S}$. Using X_{1977i} as the indicator variable and its 75 percentile point (X_0) as the "break point" the hybrid estimator using $\hat{S}$ was defined as

$$H_i = \begin{cases} N_i & \text{if } X_{1977i} < X_0 \\ \hat{S}_i & \text{otherwise} \end{cases} \quad . \tag{6}$$

The choice of X_0 was completely arbitrary. The measures of performance will be discussed in a later section.

Finally, a strategy is proposed for providing county estimates of retail sales. This strategy, denoted T, is based on the assumption that when the regression model fit is relatively high (we arbitrarily chose an R^2 of .80 as the cutoff) then only the regression $\hat{S}$ would be used. Several states consisted of a small number of counties. For those states a synthetic estimator N (we used $SYN1_i$) was used. The regression estimator $\hat{R}$ performed better than $\hat{S}$ for models with number of counties less than or equal to 20. For the remaining situation, H was used. We then have an estimation strategy T where

$$T = \begin{cases} \hat{S} \text{ if } R^2 > .80 \text{ and } n > 20 & (17) \\ H \text{ if } R^2 < .80 \text{ and } n > 20 & (27) \\ \hat{R} \text{ if } R^2 > .40 \text{ and } 14 < n < 20 & (3) \\ N \quad \text{otherwise} & (2). \end{cases} \qquad (7)$$

The numbers in parentheses represent the number of states where counties are estimated using the designated method.

The strategy T can be produced only upon completion of individual modelling of county data by state. It is easier to present it here in its final form than to await development of its component parts. Discussion of these estimators follow in Section VI. Note that in the following, SYN1 is sometimes referred to as N.

IV. Data Preparation

Prior to model fitting and analysis, data from the various sources required transcription, keying, editing and computation.

IV.a. Data transcription

Most of the data used in this study required transcription from published Census Bureau volumes or reports. The 1972, 1977 and 1982 Retail Census information for counties by state and for states themselves were clerically transcribed onto sheets. Each row of data represented those for the geographic unit of interest. The data - number of establishments, sales, payroll, number of employees and population were transcribed in blocks of census years. The geographic units were also described by name. For eg., a county's name and a number were written down. When a disclosure was specified, a dash was placed in the appropriate data field.

IV.b. Data keying

SRD's clerical staff both transcribed and keyed the data onto diskettes for use on the IBM PC. They left "disclosure items" as blanks in the file. Keying was 100% verified although data errors were still evident. Some errors were the result of transposition of figures while others were due to recording erroneous figures. Editing of the keyed figures must be done in this type of activity.

IV.c. Data editing

The clerks were told to also record the state totals for characteristics transcribed. This step was crucial for the data editing phase. The keyed data was read into MINITAB 5.1.1 for the PC and simple tabulations of the data revealed possible data errors. In addition, creation of ratios of the data items (required for some regression modelling) also helped in detecting data errors.

While the clerks were instrumental in matching data for geographical units over time, it was necessary to review the overall data sources for the occasional situation where counties were split or combined with other counties between censuses. Besides detecting the definitional changes, the economic data had to be edited to reflect the changes. As previously mentioned, for the units in Alaska, this was not possible and so county estimation for Alaska was not undertaken.

IV.d Computing

The county raw data was combined into a single file, MI827772.DAT (we use Michigan (MI) counties as an example for the discussion in this section). A similar process was used for state data so we concentrate on the county data in the exposition. Computer macros were written to execute commands within MINITAB. All commands, files, etc. that varied by state were accomodated by using the two letter alphabetic abbreviation for the state. The macros together with their input and output files are listed below. The notation C16-C20 is used to denote columns of data 16 through 20.

i) Mich.fil = Reads census data in columns C1-C15, eg., the first five columns are census data for 1982, using MI827772.dat

- Writes out to MIPROP.REG a 25 column data file in which C1-C15 are the input data, C16-C20 are the 77/72 ratios of proportion and C21-C25 are the 82/77 ratios of proportion

- Writes out to MIRAT.REG a 25 column data file in which C1-C15 are the input data, C16-C20 are the 77/72 ratio of levels and C21-C25 are the 82/77 ratios of level

ii) MIPROP1.REG = Reads data in MIPROP.REG and computes measures of performance of the regression estimator based on proportions

- Writes out to MIHYB2.REG and MIHYB1.REG columns C1-C5, information needed to compute and evaluate hybrid estimators. It also computes and evaluates the hybrid estimator.

iii) MIRAT1.REG = Reads data in MIRAT.REG and computes measures of performance of the regression estimator based on ratios and the synthetic estimator, SYN1.

Other programs and steps were used in the computing phase. A program Set1. col was used to define the intercept column used in regression. Also, regression software was used prior to MIPROP1.REG, etc., to determine the regression coefficients to be used and to determine the model. Partial residual plots, residual plots , normal plots and t-statistics were examined. Individual plots of the raw data points assisted in this phase of the analysis. With a few exceptions, the analysis for Michigan counties was repeated for each of the remaining states. The exceptions will be discussed in Section VI.

V. **Measures of Performance**

As a means of assessing the performances of the various methods for estimating total retail sales for states and for counties within states, we elected to use data from the 1972 and 1977 economic censuses for modelling purposes and to use 1982 retail sales economic census data as target values. In addition, population estimates for the aforementioned years were also used. In practice, we would be interested in estimating for years 1978 through 1981 but 1982 data is the only available data to directly assess the performance of the methods.

Several measures of performance were used to indicate the accuracy of the methods. The measures were computed for all states and the District of Columbia as a group (51) and separately for all counties within each state. Let E_i denote the estimate of 1982 retail sales for the $i^{\underline{th}}$ unit (state or county) and let A_i denote the actual 1982 retail sales for the $i^{\underline{th}}$ unit. We define the following measures -

a. $$\text{MARE}(E) = n^{-1} \sum_{i=1}^{n} |(E_i - A_i)/A_i|$$

b. $$\text{WARE}(E) = \sum_{i=1}^{n} |(E_i - A_i)/A_i| \, P_i^A$$ where

$$P_i^A = A_i / \sum_{i=1}^{n} A_i$$

c. $$\alpha(E) = \sum_{i=1}^{n} A_i \, [(E_i - A_i)/A_i]^2$$

d. $\beta(E) = \sum_{i=1}^{n} (P_i^E - P_i^A)^2 / P_i^A$ where

$$P_i^E = E_i / \sum_{i=1}^{n} E_i \quad . \tag{8}$$

The first three measures in (8) indicate how close the estimate E is to its target value A. They differ in the manner of emphasis placed on the absolute relative error. The mean absolute relative error (MARE) treats each unit i equally. The weighted absolute relative error (WARE) places a premium on the units with larger sales. The α measure further magnifies the importance of the large absolute relative errors. The fourth measure, β, provides an assessment of how well the unit sales proportions are estimated. This latter measure would be of interest if unit sales are used for allocation.

To some extent the particular measure of performance dictates the final choice among the small area estimators and strategies provided in section III. In particular, when assessing the performance of the estimators, the measure MARE favors a synthetic while β favors a regression. We now provide a closer look at the data and the measures of performance of the estimators in the next section.

VI. Results

Sales data were modelled at the U.S. and county by state levels. The results are presented in three parts. The first part examines the estimation of state sales. The second examines the county sales estimates for four selected states. The final part provides an overall summary of the county by state performance of the estimators and strategies.

For simplification of notation, we denote the predicted sales ratio of levels by $\hat{R}$ and the predicted sales ratio of proportions by $\hat{S}$. The relevant unit of analysis will be evident from the discussion. $\hat{R}$ and $\hat{S}$ are also referred to as estimators of total retail sales. Similarly, instead of recording an explanatory variable as for e.g., Y_{1977i}/Y_{1972i}, we merely refer to it as Pay (Payroll).

VI.a. State sales estimates

The Bureau's Current Retail Monthly program provides monthly estimates of retail sales for about 20 states. It is possible, then, to obtain annual estimates for such states as well. We looked at four of the states - California, Florida, Michigan, and New Jersey and compared the 1982 Retail Census figures with the survey estimate. Table 1 below provides the relative errors of the survey estimate.

Table 1. Estimates of 1982 Annual Retail Sales for Four States - Survey estimate, Regression estimate, 1982 Retail Census (in millions)

	State	Survey estimate	Regression estimate	1982 Census
1.	California			
	Total	132485	122028	120755
	Relative Error	.0971	-.0105	
2.	Florida			
	Total	54952	53172	54539
	Relative Error	.0076	-.0250	
3.	Michigan			
	Total	41042	39384	38454
	Relative Error	.0673	.0241	
4.	New Jersey			
	Total	38809	33807	35503
	Relative Error	.0931	-.0477	

The relative error of the survey estimate for Florida was small but those of the remaining states tended toward the high side. Because we needed sales estimates for all states in some of our county sales estimators, we considered developing synthetic and regression estimators for states. The results are displayed in the Appendix. The regression methods exhibited better measures of performance than the synthetic and among the regression methods $\hat{R}$ appeared the best performer. We chose to use $\hat{S}$ instead, however. Using $\hat{S}$ was in keeping with our formulation of strategy T. As can be seen in Table 1 and in the Appendix, $\hat{S}$ provided reasonable state sales estimates.

The regression model underlying $\hat{S}$ was estimated to be

$$\hat{S} = .124 + \underset{(31.28)}{.870}\ \text{pay} \qquad \hat{\sigma} = .0256$$

with an $R^2 = .952$. When utilizing $\hat{S}$ to estimate 1982 sales for states, the absolute relative errors ranged from .03% for New Mexico to 9.13% for Alaska. The absolute relative errors exceeded 5% for seven states with an overall mean absolute relative error of 2.68%. Excluding Alaska, the largest absolute relative error was 6.53% for West Virginia. See the Appendix, page a6.

VI.b. County estimates in four states

The analysis for counties by state began with examining the above mentioned four states. The aim was purely exploratory and if the measures of performance were found acceptable, then additional states were to be analyzed. The measures of performance are presented in Table 2 below.

Table 2. Measures of Performance of Several Regression and Synthetic Estimators of County Retail Sales by Each of Four States.

State Measure	$\hat{S}$	$\hat{SN}$	$\hat{R}$	SYN1	SYN
A. California					
N = 58 counties					
MARE	.0479	.0365	.0570	.0423	.0874
med (ARE)	.0347	.0228	.0472	.0253	.0644
max (ARE)	.2058	.1776	.2184	.2204	.3391
α	1004	1207	1575	1571	4844
$\beta \times 10^3$	.0078	.0066	.0078	.0093	.0374
SUM $\times 10^{-6}$	121350	119020	122818	118926	122028
B. Florida					
N = 67 counties					
MARE	.0588	.0588	.0721	.0668	.1214
med (ARE)	.0374	.0432	.0567	.0425	.1006
max (ARE)	.2070	.1870	.2039	.3111	.4386
α	1245	1632	3341	1702	7951
$\beta \times 10^3$	.0123	.0151	.0121	.0138	.1435
SUM $\times 10^{-6}$	53044	52766	51366	52628	53172
C. Michigan					
N = 83 counties					
MARE	.0621	.0618	.0678	.0611	.0962
med (ARE)	.0545	.0551	.0548	.0529	.0789
max (ARE)	.2989	.2727	.3087	.2413	.3796
α	909	846	1083	1437	5301
$\beta \times 10^3$	.0233	.0220	.0234	.0262	.1247
SUM $\times 10^{-6}$	38588	38342	39143	37197	39384
D. New Jersey					
N = 21 counties					
MARE	.0544	.0541	.0512	.0818	.0772
med (ARE)	.0508	.0567	.0473	.0827	.0909
max (ARE)	.1127	.0949	.1094	.1467	.1565
α	5757	5098	5165	12415	10314
$\beta \times 10^3$	.0362	.0262	.0362	.0425	.2005
SUM $\times 10^{-6}$	33650	33721	33775	32620	33807

Expressions for $\hat{S}$, $\hat{R}$ and $\hat{SN}$ are provided in the Appendix. Apart from $\hat{SN}$, all of the remaining estimators do not use 1982 sales. $\hat{SN}$, on the other hand, is similar to $\hat{S}$ except that the ratio of 1982 sales to 1977 sales is used as the dependent variable. For estimating 1982 sales, $\hat{SN}$ is not practical. We present the performance of $\hat{SN}$ for two reasons. One, it provides an idea of how $\hat{S}$ would perform in the year immediately following a census and two, the model change over successive five year periods can be observed.

The results in Table 2 reveal that $\hat{S}$ and $\hat{SN}$ are superior to the other methods for almost all of the measures of performance. SYN uses the estimated 1982 state retail sales obtained from 5). Its overall performance is inferior to the other methods. SYN assumes that the 1977 retail sales of each county in the state has changed in the same ratio as that of the state's sales. Because $\hat{R}$ does not require a 1982 retail sales estimate for states it would have been preferred if it exhibited the best measures of performance. However, in our limited work, that was not the case.

There is no particular reason why the ratio-correlation method (r.e. $\hat{S}$, $\hat{SN}$, $\hat{R}$) should perform better than the synthetic estimator, e.g. SYN1. Depending on the measure of performance, e.g. MARE, SYN1 does perform better than $\hat{S}$ or $\hat{R}$. We observe also that the α and β measures reveal that the regression methods almost always do better than the synthetic methods for the large (in sales) counties. In Table 2, the measure SUM x 10^{-6} represents the total retail sales for the state obtained by summing individual county estimates. The corresponding census counts are located in Table 1.

The regression equations $\hat{S}$ versus $\hat{SN}$ have changed in varying degrees. While the explanatory variables have remained the same (except for California where number of establishments has been added) the R^2 are lower for $\hat{SN}$. How well $\hat{SN}$ will perform when evaluated against census data for 1987 is of considerable interest.

VI.c. All counties by state

Based on the results concerning SYN in VI.b above, SYN was not produced for the remaining states. Instead, measures of performance were computed for $\hat{S}$, $\hat{R}$, SYN1 and the hybrid (H). This was done for nearly all states. The exceptions were Alaska, Hawaii and Maine. As already mentioned, difficulties in matching data due to geographic re-coding forced elimination of Alaska counties. Hawaii, with four counties, was not combined for regression purposes (see below) with a "nearby" state and so SYN1 was used as a default. For Maine, SYN1 was also used as the regression fit yielded a rather low $R^2 = .20$.

States with less than 14 county data points were not modelled separately but combined with geographically adjacent states. Hence, Delaware with three counties was combined with Maryland counties and the resulting regression used for Delaware counties. A separate regression was computed for Maryland. Similarly, Rhode Island, Connecticut and Massachusetts were grouped together; also Vermont and New Hampshire. We also arbitrarily declined to produce a hybrid estimator for states with less than 14 county data points. The number of county data points used in regression ranged from 14 in Massachusetts and Arizona to 243 in Texas. The R^2 ranged from .40 to .93. All regression models used payroll as an explanatory variable and nine models contained an additional variable (either population or the number of establishments).

Models for counties by state for each of the 47 states in which regression models were developed can be found in the Appendix. Because notational changes have been made, we briefly discuss the summary for the state of Alabama. On the first line (N=67) refers to 67 county data points. The matrix that immediately follows is a correlation matrix where

C16 ≡ number of establishment ratios

C17 ≡ sales ratios

C18 ≡ payroll ratios

C19 ≡ number of employees ratios

C20 ≡ population ratios,

all for 1977/1972. The regressions $\hat{S}$ and $\hat{R}$ follow with t-statistics in parentheses below the estimated regression coefficients. Q3 at the top of the measures table is the 75 percentile point for 1977 county sales and Synth1 is also termed Syn1 in this section.

IV.d Summary of measures over states

In the following discussion of the performance of the different methods of county sales estimation under various measures across states, each state is treated equally. That is, no allowance is made for the number of counties in the state nor its size of sales. The unit of analysis is the state. In Table 3 below, we present comparisons between selected pairs of methods. The entries in the table are the number of states where the measure of the second designated method of the given pair is superior.

Table 3. Comparisons of Performance of Pairs of Methods When Estimating for Counties by States for Four Measures

Measure	Strategy Pairs							
	$(\hat{R},N)$	$(\hat{S},N)$	(T,N)	(H,N)	$(\hat{R},\hat{S})$	$(\hat{R},H)$	$(\hat{S},H)$	(T,H)
MARE	N=31	N=32	N=24	N=22	S=29	H=30	H=29	H=20.5
WARE	N=27	N=20.5	N=20	N=20	S=34	H=30	H=28	H=19.5
α	N=27.5	N=23	N=18.5	N=18.5	S=34	H=33	H=32	H=20.5
β	N=17.5	N=16	N=15	N=11	S=32	H=29.5	H=26	H=18.5
	47	47	47	42	47	42	42	42

The last line in Table 3 denotes the total number of states over which the performances are being compared. For example, in comparing methods $\hat{S}$ and N under the WARE measure, the WARE for synthetic estimator N was less than or equal to that for $\hat{S}$ for 20.5 states (the .5 was the result of ties in the measure). The number 47 at the bottom of the column represents the number of states over which the comparisons were made. Based on Table 2, it appears that T and H are best overall with T slightly better than H. The synthetic estimator N is best when using the MARE measure.

Initial modelling and evaluation of a few states revealed that $\hat{S}$ consistently provided smaller α and β measures. This motivated the hybrid strategy. Completing all the states we find that β for $\hat{S}$ is indeed smaller than that for N in a large number of states (31 versus 16) but only marginally so for the α measure (24 versus 23). For both measures, H is better than either $\hat{S}$ or N.

Consider the performance of the combined strategy T. Recall that for 17 states, method $\hat{S}$ was used. When comparing the performance of $\hat{S}$ and N over these 17 states we observed that MARE(10), WARE(13), α(12) and β(11) where the figures in parenthesis represent the number of states where $\hat{S}$ had the better measure. Of the 27 states where H was used and comparing with $\hat{S}$ we have MARE(23), WARE(22), α(25) and β(21) where the figures are for H. Clearly H is a better choice than $\hat{S}$. When comparing N versus H for these same 27 states, the measures are MARE(16), WARE(14), α(12.5) and β(8) where the figures refer to N. The synthetic is better than H for MARE and WARE. This suggests that if MARE is the important criterion that a strategy T' that uses N in place of H would be preferred. Should T' be compared with N, the measures would be MARE(21.5), WARE(19.5), α(19.5) and β(20.5). From Table 2, the comparable measures for T versus N are MARE(24), WARE(20), α(18.5) and β(15). Both of these latter comparisons are over 47 states.

A brief perusal of the measures indicate that the MARE for counties in a state range from a little over 2% for Connecticut to a little over 13% in Nevada. While the MARE appears reasonable, we observe that for some counties extremely large absolute errors (AREs) are possible. For example, the three largest AREs for N over all counties was experienced in Georgia, Nebraska and South Dakota with AREs of 1.101, .837 and .849, respectively. Fortunately, the 1982 Census sales for these counties were 444, 1377 and 1272, respectively while the median sales of counties in their respective states were 38,380 24,607 and 15,181. Hence, the counties possessed very low sales volume and in particular they would not be affected by use of such estimates in allocation measures based on sales. Note that it is not possible to pre-identify such counties with high AREs within the present intended application.

VII. **Discussion**

We considered issues such as other forms that could be used for the dependent variable, limitations of the results presented and other applications.

VII.a. Other forms of dependent variables

One could take a cross-sectional approach and model 1977 county sales, 1977 proportion of county to state sales (P) or some transformation of the proportion versus other 1977 explanatory variables. We attempted all of the above and none performed as well as the regression. While we obtained good fits of a regression using arc sin(P) and log (P/(1-P)), their resulting measures of performance were not as good as that of $\hat{S}$. Using 1977 county sales as a dependent variable in a regression was inferior for estimating counties with low sales values. Because the dependent variables in our regression models are positive, we checked all regression models for positive predictions when the estimated intercept term was negative. Fortunately, in our work all predictions were positive. A way to overcome a deficiency of this sort is to use "logistic" regression.

VII.b. Limitations of inference

It was hypothesized that economic census retail trade county by state data could provide a means of obtaining county estimates in postcensal years by also using current data (County Business Patterns, Census Bureau's population and current survey sales estimates). In this research, we used 1972 and 1977 census data for modelling purposes and 1982 census data served as the target values. Modelling of 1977 and 1982 data served as a "best case" example when assessed against 1982 data. In this case, the results provide a glimpse of how well the model could perform in the year following the census assuming that the model would not deteriorate in one year.

It would be desirable to have county sales data in some intercensal years to be able to assess how well the model performs (for example, years 1978 to 1981). The reader will note that we are modelling five year changes (as measured by ratios) and assuming that, for application purposes, less than five year changes are reflected in the model as well. It does not follow that because five year changes are modelled reasonably well that one to four year changes will also be covered by the model. This is an assumption that is yet to be verified. The issue is stability of the regression coefficients. A topic that could possibly be explored when using the state as the unit of analysis. We discuss this in the next section.

VII.c. Other applications

The methods presented can also be used for estimating total wholesale sales and service industries receipts. The CBP also provides the same explanatory variables for these groups at the state and county level as was provided for retail sales. In addition, at the state level, it is possible to model detailed kinds of retail sales, wholesale sales and service industries receipts. The choice of estimator will depend on model fit.

In the case of detailed kinds of retail sales, some state monthly survey estimates are available. As explained earlier and illustrated in Table 1 for total retail sales, we did not use the information because the regression estimates were more accurate. Had they been used, other estimation techniques could have been applied. For example, the stability of the regression coefficients in the state regression model can be assessed by using those state survey estimates provided by the current survey. Since these are provided monthly, ratios of annual intercensal estimates can be modelled.

This information could then be used to update the regression coefficients of $\hat{S}$, the state regression. (See Ericksen (1974) and Swanson (1980)). Or, as was done in a demographic context, the survey and regression estimates could be combined on the basis of their variances, Ericksen and Kadane (1985), Fuller and Harter (1987) and Isaki et. al. (1987)

VII.d. Timing

Census results for 1987 are available around mid 1989 at which time 1982 to 1987 data can be modelled. This 1.5 year delay period is also experienced by the CBP. The population and current survey estimates are obtainable sooner. Hence, it is the CBP data that affects the timeliness of small area estimation. The proposed small area methods can produce county retail sales at the same time that the CBP data is available.

VIII. Summary and Future Work

It was our aim to propose and evaluate methodologies leading to small area estimates of economic data in postcensal years. In the absence of directly estimated county sales totals, the hybrid estimator and in particular the strategy T appears to be a useful method for providing such estimates when evaluated over a five year lag. The main advantage of the regression and synthetic methods considered is that all of the required data input are available to the public.

This report covered estimation of total retail sales of employers although the methods provided can be used for other economic statistics. We have taken the position of a national planner in the use of explanatory variables available for use in modelling. It is likely that individual states have, at their disposal, other explanatory variables peculiar to their data capture systems. Sales tax has been mentioned as a possible item. An individual state modeller can easily include such explanatory variables in constructing estimators of the sort mentioned in the report. Rather than allowing for special treatment for individual states, we attempted to develop an overall strategy here.

Since the 1987 Economic Census results will soon become available it is possible to repeat the numerical work as reported. This will add much needed information on the utility of the approach presented. It is hoped that the current report provides an illustration of the performance of the small area estimates. Ultimately, it is the user who must decide whether the accuracy of the estimates is adequate for his/her purposes. We have also mentioned methods that could be developed for providing state level estimates of detailed characteristics. All of these areas could be investigated in the future.

IX. References

1. Ericksen, E.P. and Kadane, J.B. (1985), "Estimating the Population in a Census Year - 1980 and Beyond," Journal of the American Statistical Association, 80, pg. 867-875.

2. Ericksen, E.P. (1974), "A Regression Method for Estimating Population Changes of Local Areas," Journal of the American Statistical Association vol. 69, pg. 867-875.

3. Fay, R.E. III and Herriot, R.A. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," Journal of the American Statistical Association, vol. 74, pg. 269-277.

4. Fuller, W.A. and Harter, R.M. (1987), "The Multivariate Components of Variance Model for Small Area Estimation," in Small Area Statistics: An International Symposium edited by R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh, John Wiley and Sons, New York.

5. Hawkes, William, Jr. (1985), "Census Data Quality -- A User's View," Proceedings of the Bureau of the Census' First Annual Research Conference, pg. 177-192, Reston, VA.

6. Isaki, C.T., Schultz, L.K., Smith, P.J. and Diffendal, G.J. (1987), "Small Area Estimation Research for Census Undercount," in Small Area Statistics: An International Symposium edited by R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh, John Wiley and Sons, New York.

7. Levy, P.S. (1971), "The Use of Mortality Data in Evaluating Synthetic Estimates," Proceedings of the American Statistical Association, Social Statistics Section, Washington, D.C., pg. 328-331.

8. Mandell, M. and Tayman, J. (1982), "Measuring Temporal Stability in Regression Models of Population Estimation," Demography, vol. 19 no. 2, pg. 135-146.

9. Martin, J.H. and Serow, W.J. (1978), "Estimating Demographic Characteristics Using the Ratio-Correlation Method," Demography, vol. 15 no. 2, pg. 223-233.

10. Namboodiri, N.K. and Lalu, N.M. (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, vol. 36 no. 2, pg. 187-194.

11. Namboodiri, N.K. (1972), "On the Ratio - Correlation and Related Methods of Subnational Population Estimation," Demography, vol. 9, pg. 443-453.

12. O'Hare, W. (1976), "Report on a Multiple Regression Method for Making Population Estimates," Demography, vol. 13, pg. 369-379.

13. Purcell, N.J. and Kish, L. (1980), "Postcensal Estimates for Local Areas (or Domains), International Statistical Review, vol. 48, pg. 3-18.

14. Rosenberg, H. (1968), "Improving Current Population Estimates Through Stratification," Land Economics, vol. 44, pg. 331-338.

15. Schmitt, R.C. and Crosetti, A.H. (1964), "Accuracy of the Ratio-correlation Method for Estimating Postcensal Population," Land Economics, vol. 30, pg. 279-281.

16. Snow, E.C. (1911), "The Application of the Method of Multiple Correlation to the Estimation of Post-Census Populations," Journal of the Royal Statistical Society 74, pg. 575-620.

17. Swanson, D.A. (1980), "Improving Accuracy in Multiple Regression Estimates of Population Using Principles from Causal Modelling," Demography, vol. 17 pg. 413-427.

18. U.S. Bureau of the Census (1987), County Business Patterns, 1985. U.S. Government Printing Office, Washington, D.C.

19. U.S. Bureau of the Census (1984), 1982 Census of Retail Trade - Geographic Area Series, U.S. Government Printing Office, Washington, D.C.

20. U.S. Bureau of the Census (1985), Local Population Estimates Series P-26, No. 83-22-C, U.S. Government Printing Office. Washington, D.C.

X. Appendix

X.a. Variance of a predictor (eg., eq. (5))

Rather than using the cumbersome notation of section III which possessed the redeeming feature of being self-explanatory, we introduce notation that provides for concise exposition of the variance of the predictor. For the i^{th} small area, its predictor of total sales is $\hat{Y}_{2i}$

where $\hat{Y}_{2i} = X_{21i} \hat{\boldsymbol{\beta}}_0 P_{1i} Y_s$ and (a.1)

where

$\mathbf{X}_{21i}$ is the i^{th} row of the N x (n+1) matrix $\mathbf{X}_{21}$ that contains the explanatory variables

$\hat{\boldsymbol{\beta}}_0$ is an (n+1) x 1 vector of estimated regression coefficient based on the census data

P_{1i} is a scalar value - it is the proportion of the most recent census sales of the i^{th} small area and

Y_s is the estimate of total for the larger area.

Because the regression is used to predict future results, the appropriate measure of precision of $\hat{Y}_{2i}$ is the variance of $\hat{Y}_{2i} - Y_{2i}$. It can be shown that if $\hat{\boldsymbol{\beta}}_0$ and Y_s are independent that

$$MSE(\hat{Y}_{2i}) = \left[(\mathbf{X}_{2i}(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{2i} + 1)\sigma^2\right]P_{1i}^2\left[Var(Y_s) + E\left[Y_s\right]^2\right] \qquad (a.2)$$

where

$\mathbf{X}_{11}$ is the N x (n+1) matrix of explanatory variables used to estimate $\hat{\boldsymbol{\beta}}_0$ and

$V(Y_s)$ is the variance of the predictor (or estimator) of retail sales. When Y_s is a predictor of sales, $V(Y_s)$ can be produced along the lines of (a.2).

It has also been suggested that in (a.2) for model usage K years apart, k = 1,..,5 that $\sigma^2 K/5$ could be used in place of σ^2 to reflect increasing dispersal of data over time.

Estimation of (a.2) can be accomplished by using $\hat{\sigma}^2$, the usual estimator of σ^2 under the linear model; the survey estimate or, if a predictor is used for Y_s, the proposed estimator using the Anuual Retail Trade variance estimate for the U.S.; replacing $E[Y_s]$ with Y_s.

X. b. Summary Selections - CBP County Total Retail by State

U.S. State model $\equiv \hat{S} = .124 + .870\ C18 \quad R^2 = .952$ (a.3)

$(31.28) \quad \hat{\sigma} = .0256$

ARTS '82 $\equiv 1047454656 \times 10^3$ (employers)

$V('82) = 53760906756096 \times 10^6$

CENSUS '82 $\equiv 1039028736 \times 10^3$

I. 1982 State Retail Estimates (Employers) Using $\hat{S}$ (Sales in 10^6)

State		1982 Census (C) $x10^6$	1982 $\hat{S}$ $x10^6$	$(\hat{S} - C)/C$ (error)
1.	AL	13927	14424	.0357
2.	AK	3152	2864	-.0913
3.	AZ	13585	13835	.0183
4.	AR	8693	8996	.0348
5.	CA	120755	122028	.0105
6.	CO	16209	16329	.0074
7.	CT	15472	15119	.0228
8.	DE	3076	2997	-.0254
9.	DC	2614	2664	.0189
10.	FL	54539	53172	-.0250
11.	GA	23755	23422	-.0140
12.	HI	5101	5008	-.0181
13.	ID	3927	4117	.0483
14.	IL	49671	49853	.0036
15.	IN	23170	24249	.0465
16.	IA	12319	12877	.0453

State		1982 Census (C) $\times 10^6$	1982 $\hat{S}$ $\times 10^6$	$(\hat{S}-C)/C$ error
17.	KS	10540	10897	.0338
18.	KY	13922	14485	.0404
19	LA	19442	19825	.0197
20.	ME	5168	5172	.0007
21.	MD	20657	20323	.0161
22.	MA	28222	26754	-.0520
23.	MI	38454	39384	.0241
24.	MN	19129	19008	-.0063
25.	MS	8655	9091	.0502
26.	MO	21048	21056	.0004
27.	MT	3825	3980	.0404
28.	NE	6774	7045	.0398
29.	NV	5253	5446	.0368
30.	NH	5239	5135	-.0197
31.	NJ	35503	33807	-.0477
32.	NM	6161	6159	-.0003
33.	NY	70458	67371	-.0438
34.	NC	24082	24298	.0089
35.	ND	3276	3227	-.0151
36.	OH	45461	46714	.0275
37.	OK	15526	15744	.0139
38.	OR	12282	12937	.0532
39.	PA	49223	48821	-.0081

State	1982 Census (C) $\times 10^6$	1982 $\hat{S}$ 10^6	$(\hat{S}-C)/C$ error
40. RI	4061	4002	-.0143
41. SC	12072	12238	.0137
42. SD	2879	3001	.0426
43. TN	18826	19329	.0267
44. TX	80324	80629	.0038
45. UT	6179	6194	.0024
46. VT	2528	2561	.0127
47. VA	24217	24189	-.0011
48. WA	19599	20619	.0520
49. WV	7276	7752	.0653
50. WI	20028	20187	.0079
51. WY	2747	2898	.0547

Note: $\hat{S}$ above is predicting the ratio of 1982 state share of retail to 1977 state share of retail. To obtain state retail, $\hat{S}$ is multiplied by 1977 state share of sales and then by ARTS '82. The column headed 1982 $\hat{S}$ is the 1982 estimated state retail sales for employers based on the model in (a.3).

Besides the model in (a.3) we also considered

$\hat{R} = .198 + .892\ C18$ $\hat{\sigma} = .0406$,

the hybrid estimators and Synthl.

The correlations among the variables (see the text for definition of C16-C20) are presented below.

Corr (N=51)

	C16	C17	C18	C19
C17	.644			
C18	.618	.976		
C19	.680	.949	.968	
C20	.791	.803	.784	.823

The measures of performance of the estimators are as follows -

Measures Q3 = 16,061,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid ($\hat{S}$)	Hybrid ($\hat{R}$)
MARE	.0294	.0257	.0268	.0252	.0252
median	.0249	.0228	.0229	.0242	.0236
min	.0006	.0012	.0004	.0008	.0008
max	.1113	.0955	.0913	.1113	.1113
α	35366	15806	16037	16453	17084
$\beta \times 10^3$	.0183	.0152	.0152	.0146	.0144
WARE	.0315	.0210	.0215		

The value Q3 is the 75 percentile 1977 sales value for states. On the basis of the measures in the table, Synthl was clearly not as good as the others. We chose $\hat{S}$ even though its measures were not as good as $\hat{R}$ or the hybrids in an effort to be consistent with our strategy T. It is likely that using $\hat{R}$ instead of $\hat{S}$ would not appreciably affect the results.

X.c. Individual County '82 Retail Estimator Summary by State

In this section the results of modelling county ratios of proportion and level using $\hat{S}$ as the base are presented. For example, in using the ratios of proportion for Alabama (AL) $\hat{S} = 14424 \times 10^6$ is used to convert '82 county proportion sales to level. (See page a3).

The estimators $\hat{S}$, $\hat{R}$ and Synthl are computed for each county and the measures of performance are applied. $\hat{S}$ is the regression using ratios of proportions (1977 to 1972). $\hat{R}$ is the regression using ratios of level (1977 to 1972). Synthl is a synthetic estimator that applies (1982 to 1977) ratios of payroll level to 1977 sales. It is somewhat similar to $\hat{R}$ without an intercept (in general form).

Within each state the counties are modelled and selected models are applied to produce '82 county retail estimates. States with a small number of counties are usually combined with other states for model construction. Selection of independent variables are based on data plots including partial residual plots. Counties with disclosure problems are automatically excluded. Other data points represented by a small number of establishments or large residuals are also excluded.

For states with adequate number of counties, a hybrid estimator is also constructed. The Hybrid estimator is a mixture of Synthl and $\hat{S}$ (or $\hat{R}$) and depends on the R^2 as well. Roughly speaking, we have observed that for counties within a state the following strategy performs well - (This strategy is denoted T in what follows)

(1) Use $\hat{S}$ if R^2 .80 and n > 20

(2) Use Hybrid ($\hat{S}$) if $R^2 < .80$ and n > 20

(3) Use $\hat{R}$ if $R^2 > .40$ and $14 < n < 20$

(4) Use Synth1 otherwise

This is not to imply that the other estimators will not perform better than the strategy proposed. In fact, the other estimators do sometimes perform better. However, overall, the strategy performs well especially when compared with individual estimators. The summary and rankings of the performance of all estimators are provided in III. The measures of performance are used to rank the estimators by state.

In this section, the county estimators are summarized by state, alphabetically. The above strategy (1) - (4) is rather loose as a) R^2 near .80 was designated somewhat arbitrarily as was b) n at 20 and 14. Motivation for Hybrid ($\hat{S}$) can be found in III. The Hybrid ($\hat{S}$) uses Q3 ≡ 75 percentile on 1977 sales. That is, the Hybrid ($\hat{S}$) is $\hat{S}$ for all counties with 1977 sales exceeding Q3 and is equal to Synth1 otherwise.

The measure "state total" is the summation of the county estimates over all counties in the State.

1. Alabama (N=67)

Corr

	C16	C17	C18	C19
C17	.701			
C18	.665	.825		
C19	.672	.826	.852	
C20	.414	.513	.372	.423

$$\hat{S} = -.158 + .624\ C18 + .522\ C20 \qquad R^2 = .731$$

$$(10.54) \qquad (3.52) \qquad \hat{\sigma} = .0683$$

$$\hat{R} = -.259 + .607\ C18 + .800\ C20 \qquad \hat{\sigma} = .1115$$

Measures

Q3 = 143,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0785	.0836	.0824	.0787
median	.0558	.0551	.0540	.0558
min	.0019	.0006	.0006	.0013
max	.4117	.4433	.4419	.4117
α	919	886	860	822
$\beta \times 10^3$	.0660	.0484	.0484	.0480
WARE	.0503	.0438	.0420	.0416
State total $\times 10^{-6}$	13928	14328	14301	14263
error	.0001	.0288	.0268	.0241
'82 Census	13927			

3. Arizona (N=14)

Corr

	C16	C17	C18	C19
C17	.486			
C18	.455	.918		
C19	.436	.535	.676	
C20	.681	.647	.530	.241

$\hat{S} = -.1611 + 1.1728\ C18$ (8.01) $\quad R^2 = .842 \quad \hat{\sigma} = .0463$

$\hat{R} = -.2775 + 1.1664\ C18 \quad \hat{\sigma} = .0797$

Measures

Q3 = 240,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0463	.0445	.0478	.0486
median	.0416	.0423	.0429	.0390
min	.0018	.0054	.0187	.0018
max	.1897	.0877	.1035	.1897
α	783	770	1525	1506
$\beta \times 10^3$	.0558	.0502	.0502	.0604
WARE	.0179	.0195	.0358	.0344
State total $\times 10^{-6}$	13645	13705	13976	13935
error	.0044	.0088	.0287	.0257
'82 Census $\times 10^{-6}$	13585			

4. Arkansas (N=75)

Corr

	C16	C17	C18	C19
C17	.671			
C18	.647	.849		
C19	.645	.779	.851	
C20	.608	.506	.376	.534

$\hat{S} = -.087 + .675\ C18 + .405\ C20$ $\quad R^2 = .762$

$\quad\quad (12.37) \quad (3.50)$ $\quad \hat{\sigma} = .0798$

$\hat{R} = -.1455 + .650\ C18 + .617\ C20$ $\quad \hat{\sigma} = .133$

Measures

Q3 = 76,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0786	.0728	.0854	.0790
median	.0575	.0513	.0564	.0560
min	.0008	.0010	.0020	.0010
max	.4821	.3584	.4099	.4822
α	552	456	585	536
$\beta \times 10^3$	.0627	.0526	.0522	.0527
WARE	.0515	.0434	.0478	.0451
State total $\times 10^{-6}$	8733	8700	8953	8887
error	.0046	.0008	.0299	.0223
'82 Census $\times 10^{-6}$	8693			

5. California (N=58)

Corr

	C16	C17	C18	C19
C17	.812			
C18	.789	.930		
C19	.597	.613	.698	
C20	.488	.417	.391	.606

$\hat{S} = .069 + .924\ C18$ $\quad R^2 = .866$

(18.98) $\quad \hat{\sigma} = .0532$

$\hat{R} = .1144 + .957\ C18$ $\quad \hat{\sigma} = .0883$

Measures

Q3 = 1,129,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0423	.0570	.0479	.0403
median	.0253	.0472	.0347	.0228
min	.0000	.0025	.0004	.0000
max	.2204	.2184	.2058	.2205
α	1571	1575	1004	723
$\beta \times 10^3$	.0093	.0078	.0078	.0059
WARE	.0226	.0174	.0139	.0126
State total $\times 10^{-6}$	118926	122818	121350	120971
error	-.0151	.0171	.0049	.0018
'82 Census $\times 10^{-6}$	120755			

6. Colorado (N=63)

Corr

	C16	C17	C18	C19
C17	.785			
C18	.732	.792		
C19	.727	.682	.884	
C20	.755	.677	.542	.527

$\hat{S} = -.142 + .539\ C18 + .580\ C20 \qquad R^2 = .714$

$\qquad\qquad (7.32) \qquad (4.26) \qquad \hat{\sigma} = .1645$

$\hat{R} = -.239 + .525\ C18 + .871\ C20 \qquad \hat{\sigma} = .2770$

Measures

Q3 = 63,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0868	.1007	.0980	.0857
median	.0569	.0566	.0513	.0569
min	.0033	.0004	.0010	.0047
max	.5086	.8079	.8368	.5086
α	345	695	507	369
$\beta \times 10^3$	.0212	.0313	.0304	.0219
WARE	.0246	.0383	.0305	.0285
State total $\times 10^{-6}$	16157	15738	16054	16062
error	-.0032	-.0290	-.0100	-.0091
'82 Census $\times 10^{-6}$	16209			

7. Connecticut (N=8)

Corr (N=27)

	C16	C17	C18	C19
C17	.626			
C18	.771	.936		
C19	.534	.768	.791	
C20	.789	.463	.624	.402

$\hat{S} = .303 + .695\ C18$ $\qquad R^2 = .875$ *

(13.25) $\qquad \hat{\sigma} = .0314$

$\hat{R} = .428 + .743\ C18$ $\qquad \hat{\sigma} = .0442$

Measures (N=8)

	Synth1	$\hat{R}$	$\hat{S}$
MARE	.0373	.0226	.0232
median	.0442	.0252	.0241
min	.0017	.0052	.0028
max	.0621	.0292	.0322
α	5974	1071	1634
$\beta \times 10^3$	.0346	.0297	.0300
WARE	.0533	.0231	.0283
State total $\times 10^{-6}$	14647	15189	15088
error	-.0533	-.0182	-.0248
'82 Census $\times 10^{-6}$	15472		

* $\hat{S}$ and $\hat{R}$ used N=27 counties of CT, RI and MA combined in constructing the model.

8. Delaware (N=3)

Corr (N=27)

	C16	C17	C18	C19
C17	.687			
C18	.754	.925		
C19	.664	.815	.886	
C20	.295	.543	.517	.312

$\hat{S} = .015 + .976\ C18 \qquad R^2 = .855$ *

$\qquad\qquad (12.16) \qquad \hat{\sigma} = .0584$

$\hat{R} = .0224 + 1.006\ C18 \qquad \hat{\sigma} = .0584$

Measures (N=3)

	Synth1	$\hat{R}$	$\hat{S}$
MARE	.0502	.0295	.0266
median	.0577	.0368	.0295
min	.0210	.0010	.0068
max	.0717	.0507	.0434
α	4221	2017	1465
$\beta \times 10^3$	.1222	.1122	.1127
WARE	.0616	.0407	.0354
State total $\times 10^{-6}$	2886	2950	2973
error	-.0616	-.0407	-.0335
'82 Census $\times 10^{-6}$	3076		

* $\hat{S}$ and $\hat{R}$ used 27 counties of DE and MD combined in constructing the model

10. Florida (N=67)

Corr (N=67)

	C16	C17	C18	C19
C17	.602			
C18	.589	.951		
C19	.421	.872	.818	
C20	.446	.532	.442	.525

$\hat{S} = .0511 + .708\ C18 + .237\ C20$ $\qquad R^2 = .919$

$\qquad\qquad (22.47) \qquad (3.51)$ $\qquad \hat{\sigma} = .0627$

$\hat{R} = .082 + .708\ C18 + .323\ C20$ $\qquad \hat{\sigma} = .101$

Measures (N=67) Q3 = 465,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0668	.0721	.0588	.0645
median	.0425	.0567	.0374	.0425
min	.0023	.0001	.0019	.0023
max	.3111	.2039	.2070	.3110
α	1702	3341	1245	1233
$\beta \times 10^3$	.0138	.0121	.0123	.0127
WARE	.0397	.0597	.0315	.0318
State total $\times 10^{-6}$	52628	51366	53044	53086
error	-.0530	-.0582	-.0274	-.0267
'82 Census $\times 10^{-6}$	54539			

11. Georgia (N=155)

Corr

	C16	C17	C18	C19
C17	.608			
C18	.665	.856		
C19	.663	.847	.931	
C20	.624	.459	.525	.552

$\hat{S} = .273 + .707\ C18$ $\qquad R^2 = .733$

(20.49) $\qquad \hat{\sigma} = .1305$

$\hat{R} = .418 + .692\ C18$ $\qquad \hat{\sigma} = .2000$

Measures (N=155) Q3 = 67,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.1008	.0985	.1026	.1004
median	.0652	.0653	.0705	.0651
min	.0000	.0001	.0002	.0000
max	1.101	1.2663	1.3439	1.1014
α	896	1283	830	833
$\beta \times 10^3$	.0320	.0320	.0316	.0315
WARE	.0622	.0779	.0560	.0554
State total $\times 10^{-6}$	22913	22237	23079	23055
error	-.0349	-.0634	-.0280	-.0290
'82 Census $\times 10^{-6}$	23743			

12. Hawaii (N=4)

	Synth1
MARE	.0302
median	.0277
min	.0098
max	.0557
α	3271
$\beta \times 10^3$	.2770
WARE	.0484
State total $\times 10^{-6}$	4901
error	-.0393
'82 Census $\times 10^{-6}$	5101

13. Idaho (N=41)

Corr

	C16	C17	C18	C19
C17	.626			
C18	.698	.867		
C19	.528	.736	.819	
C20	.557	.541	.507	.390

$\hat{S}$ = .147 + .823 C18 R^2 = .751

(10.84) $\hat{\sigma}$ = .0818

$\hat{R}$ = .2696 + .828 C18 $\hat{\sigma}$ = .150

Measures

Q3 = 63,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0919	.0934	.0883	.0925
median	.0669	.0773	.0643	.0655
min	.0002	.0001	.0033	.0062
max	.4527	.4622	.4645	.4528
α	447	454	389	435
$\beta \times 10^3$	.1114	.0842	.0875	.1006
WARE	.0453	.0534	.0460	.0479
State total $\times 10^{-6}$	3945	4048	3993	3986
error	.0066	.0327	.0186	.0169
'82 Census $\times 10^{-6}$	3920			

14. Illinois (N=102)

Corr

	C16	C17	C18	C19
C17	.443			
C18	.409	.873		
C19	.504	.709	.801	
C20	.117	.195	.054	-.011

$\hat{S} = -.1228 + .705\ C18 + .40\ C20 \qquad R^2 = .785$

$\qquad\qquad\qquad (18.53) \qquad (3.17) \qquad \hat{\sigma} = .058$

$\hat{R} = -.1870 + .721\ C18 + .602\ C20 \qquad \hat{\sigma} = .088$

Measures

Q3 = 170,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0611	.0873	.0725	.0610
median	.0426	.0635	.0470	.0413
min	.0000	.0008	.0000	.0025
max	.4146	.3721	.3385	.3043
α	1982	1149	1082	1033
$\beta \times 10^3$	.0217	.0214	.0211	.0192
WARE	.0566	.0274	.0367	.0358
State total $\times 10^{-6}$	47422	50236	49120	48929
error	-.0452	.0114	-.0111	-.0149
'82 Census $\times 10^{-6}$	49671			

15. Indiana (N=92)

Corr

	C16	C17	C18	C19
C17	.367			
C18	.515	.747		
C19	.468	.610	.794	
C20	.315	.189	.318	.381

$\hat{S} = .2645 + .723\ C18$ $\quad R^2 = .558$

(10.66) $\quad \hat{\sigma} = .0681$

$\hat{R} = .422 + .756\ C18$ $\quad \hat{\sigma} = .109$

Measures

Q3 = 148,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0544	.1064	.0701	.0563
median	.0467	.1019	.0613	.0477
min	.0006	.0025	.0035	.0006
max	.2029	.2750	.2232	.2029
α	585	2194	814	639
$\beta \times 10^3$	.0252	.0212	.0209	.0191
WARE	.0338	.0821	.0433	.0382
State total $\times 10^{-6}$	23179	25029	23963	23778
error	.0004	.0802	.0342	.0262
'82 Census $\times 10^{-6}$	23170			

16. Iowa (N=99)

Corr

	C16	C17	C18	C19
C17	.540			
C18	.556	.816		
C19	.544	.662	.791	
C20	.383	.399	.415	.341

$\hat{S} = .238 + .744\ C18$ $\qquad R^2 = .666$

(13.90) $\qquad \hat{\sigma} = .0674$

$\hat{R} = .400 + .754\ C18$ $\qquad \hat{\sigma} = .113$

Measures

Q3 = 75,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0695	.1046	.0800	.0729
median	.0480	.0817	.0519	.0512
min	.0012	.0001	.0007	.0008
max	.3516	.3947	.3563	.3516
α	527	1052	634	563
$\beta \times 10^3$	.0430	.0433	.0426	.0415
WARE	.0449	.0716	.0499	.0473
State total $\times 10^{-6}$	12273	13062	12635	12523
error	-.0038	.0603	.0257	.0166
'82 Census $\times 10^{-6}$	12319			

17. Kansas (N=102)

Corr

	C16	C17	C18	C19
C17	.382			
C18	.249	.734		
C19	.502	.632	.611	
C20	.209	.186	.092	.266

$\hat{S} = .177 + .797\ C18$ $\qquad R^2 = .539$

(10.66) $\qquad \hat{\sigma} = .1018$

$\hat{R} = .292 + .790\ C18$ $\qquad \hat{\sigma} = .1677$

Measures

Q3 = 52,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0870	.0831	.0862	.0876
median	.0663	.0656	.0577	.0700
min	.0002	.0009	.0015	.0002
max	.3948	.4297	.4496	.3949
α	284	313	314	294
$\beta \times 10^3$	.0271	.0296	.0286	.0276
WARE	.0304	.0386	.0335	.0329
State total $\times 10^{-6}$	10450	10411	10591	10546
error	-.0046	-.0083	.0089	.0045
'82 Census $\times 10^{-6}$	10540			

18. Kentucky (N=120)

Corr

	C16	C17	C18	C19
C17	.432			
C18	.409	.867		
C19	.532	.710	.784	
C20	.376	.441	.367	.366

$\hat{S} = .217 + .780\ C18$ $\qquad R^2 = .752$

(18.91) $\qquad \hat{\sigma} = .0949$

$\hat{R} = .371 + .791\ C18$ $\qquad \hat{\sigma} = .1620$

Measures

Q3 = 70,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0974	.1241	.1098	.1010
median	.0676	.0794	.0698	.0692
min	.0009	.0002	.0000	.0009
max	.4382	.6288	.5811	.4382
α	686	1124	824	750
$\beta \times 10^3$	.0470	.0393	.0393	.0384
WARE	.0506	.0738	.0557	.0539
State total $\times 10^{-6}$	14087	14838	14537	14458
error	.0119	.0658	.0442	.0386
'82 Census $\times 10^{-6}$	13922			

19. Louisiana (N=64)

Corr

	C16	C17	C18	C19
C17	.536			
C18	.480	.869		
C19	.609	.796	.813	
C20	.369	.427	.371	.383

$\hat{S} = .299 + .697\ C18$ $\qquad R^2 = .755$

(13.83) $\qquad \hat{\sigma} = .0622$

$\hat{R} = .526 + .692\ C18$ $\qquad \hat{\sigma} = .1096$

Measures

Q3 = 160,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0662	.0740	.0730	.0638
median	.0441	.0694	.0680	.0456
min	.0005	.0002	.0005	.0005
max	.4004	.2747	.2818	.4004
α	1113	1149	1111	1090
$\beta \times 10^3$	.0563	.0515	.0500	.0514
WARE	.0433	.0455	.0445	.0436
State total $\times 10^{-6}$	19539	19811	19801	19718
error	.0050	.0190	.0185	.0442
'82 Census $\times 10^{-6}$	19442			

20. Maine (N=16)

Corr

	C16	C17	C18	C19
C17	.342			
C18	.247	.445		
C19	.069	.331	.533	
C20	.227	.220	.180	-.159

$\hat{S} = .504 + .473\ C18$ $\qquad R^2 = .198$ *

$\qquad (1.86)$ $\qquad \hat{\sigma} = .0471$

$\hat{R} = .819 + .503\ C18$ $\qquad \hat{\sigma} = .0766$

Measures

Q3 = 64,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0523	.0805	.0682	.0584
median	.0343	.0689	.0585	.0382
min	.0144	.0069	.0064	.0064
max	.2004	.2503	.1758	.2004
α	853	1776	1336	1274
$\beta \times 10^3$	.1012	.2430	.2355	.2203
WARE	.0421	.0529	.0539	.0516
State total $\times 10^{-6}$	4995	5356	5048	5043
error	-.0336	.0364	-.0232	-.0243
'82 Census $\times 10^{-6}$	5168			

* No explanatory variable was reasonably correlated with C17, but we forced C18 into the model. Clearly, Synth1 is the default.

21. Maryland (N=24)

Corr

	C16	C17	C18	C19
C17	.670			
C18	.739	.923		
C19	.647	.803	.878	
C20	.266	.524	.494	.280

$\hat{S} = -.019 + 1.008\ C18$ $\qquad R^2 = .851$

(11.23) $\qquad \hat{\sigma} = .0606$

$\hat{R} = -.0297 + 1.034\ C18\ \hat{S} = .094$

Measures Q3 = 400,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0561	.0505	.0441	.0498
median	.0440	.0354	.0257	.0347
min	.0044	.0096	.0059	.0044
max	.2304	.2228	.2115	.2304
α	2833	1933	1202	1517
$\beta \times 10^3$	.0338	.0340	.0340	.0367
WARE	.0520	.0411	.0286	.0320
State total $\times 10^{-6}$	19611	19859	20142	20026
error	-.0506	-.0386	-.0249	-.0306
'82 Census $\times 10^{-6}$	20657			

22. Massachusetts (N=14)

Corr

	C16	C17	C18	C19
C17	.892			
C18	.980	.920		
C19	.689	.691	.733	
C20	.853	.705	.837	.505

$\hat{S} = .366 + .632\ C18$ $\qquad R^2 = .847$

(8.16) $\qquad \hat{\sigma} = .0397$

$\hat{R} = .510 + .678\ C18$ $\qquad \hat{\sigma} = .055$

Measures $\qquad$ Q3 = 2,000,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0716	.0493	.0505	.0625
median	.0712	.0580	.0613	.0712
min	.0044	.0044	.0047	.0044
max	.1627	.0988	.1010	.0950
α	15584	7465	7451	9298
$\beta \times 10^3$	.1401	.0879	.0871	.0679
WARE	.0783	.0518	.0520	.0618
State total $\times 10^{-6}$	26025	26785	26784	26488
error	-.0778	-.0509	-.0510	-.0615
'82 Census $\times 10^{-6}$	28222			

23. Michigan (N=83)

Corr

	C16	C17	C18	C19
C17	.476			
C18	.468	.875		
C19	.454	.755	.775	
C20	.305	.677	.543	.516

$\hat{S}$ -.0825 + .694 C18 + .367 C20 R^2 = .823

(12.83) (5.11) $\hat{\sigma}$ = .0543

$\hat{R}$ = -.128 + .691 C18 + .562 C20 $\hat{\sigma}$ = .084

Measures Q3 = 218,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0611	.0678	.0621	.0573
median	.0529	.0548	.0545	.0474
min	.0007	.0030	.0000	.0007
max	.2413	.3087	.2989	.2413
α	1437	1083	909	.896
$\beta \times 10^3$	.0262	.0234	.0233	.0234
WARE	.0453	.0352	.0323	.0322
State total $\times 10^{-6}$	37197	39143	38588	38380
error	-.0327	.0179	.0035	-.0019
'82 Census $\times 10^{-6}$	38454			

24. Minnesota (N=87)

Corr

	C16	C17	C18	C19
C17	.350			
C18	.315	.667		
C19	.460	.649	.761	
C20	.272	.218	.085	.098

$\hat{S} = .357 + .636\ C18$ $\quad R^2 = .445$

(8.25) $\quad \hat{S} = .0789$

$\hat{R} = .585 + .654\ C18$ $\quad \hat{S} = .1291$

Measures

Q3 = 100,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0568	.1091	.0762	.0553
median	.0447	.1129	.0712	.0424
min	.0027	.0008	.0007	.0007
max	.3695	.3701	.3163	.3695
α	1186	1148	873	745
$\beta \times 10^3$	.0471	.0465	.0452	.0354
WARE	.0625	.0483	.0536	.0497
State total x 16^{-6}	18359	19712	18907	18730
error	-.0403	.0305	-.0116	-.0209
'82 Census x 10^{-6}	19129			

25. Mississippi (N=80)

Corr

	C16	C17	C18	C19
C17	.414			
C18	.527	.909		
C19	.595	.785	.893	
C20	.388	.492	459	.469

$\hat{S} = .235 + .763\ C18$ $\qquad R^2 = .826$

(19.26) $\qquad \hat{\sigma} = .0672$

$\hat{R} = .3687 + .737\ C18$ $\qquad \hat{\sigma} = .106$

Measures

Q3 = 70,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0992	.0891	.1117	.1013
median	.0732	.0665	.0818	.0760
min	.0007	.0012	.0007	.0032
max	.4266	.4191	.4850	.4266
α	759	600	926	841
$\beta \times 10^3$	.0764	.0690	.0690	.0693
WARE	.0543	.0523	.0675	.0642
State total $\times 10^{-6}$	8862	8671	9081	9013
error	.0245	.0024	.0498	.0491
'82 Census $\times 10^{-6}$	8650			

26. Missouri (N=115)

Corr

	C16	C17	C18	C19
C17	.495			
C18	.558	.809		
C19	.633	.697	.784	
C20	.422	.463	.484	.498

$\hat{S} = .182 + .814\ C18$ $\quad R^2 = .655$

(14.65) $\quad \hat{\sigma} = .0947$

$\hat{R} = .283 + .834\ C18$ $\quad \hat{\sigma} = .148$

Measures $\quad$ Q3 = 75,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0909	.1184	.1129	.0920
median	.0635	.0851	.0795	.0689
min	.0003	.0013	.0024	.0018
max	.4512	.5110	.4908	.4513
α	1363	1037	1019	848
$\beta \times 10^3$	.0591	.0467	.0476	.0403
WARE	.0703	.0520	.0531	.0488
State total $\times 10^{-6}$	20350	21313	21160	21036
error	-.0331	.0126	.0053	-.0006
'82 Census $\times 10^{-6}$	21048			

27. Montana (N=53)

Corr

	C16	C17	C18	C19
C17	.487			
C18	.527	.635		
C19	.365	.466	.725	
C20	.310	.371	.300	.348

$\hat{S} = .335 + .619\ C18$ $\quad R^2 = .403$

(5.87) $\quad \hat{\sigma} = .0992$

$\hat{R} = .594 + .630\ C18$ $\quad \hat{\sigma} = .1759$

Measures

Q3 = 40,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0926	.1405	.1146	.0954
median	.0631	.1256	.0776	.0706
min	.0037	.0061	.0001	.0053
max	.3613	.4784	.3800	.3614
α	227	695	354	267
$\beta \times 10^3$	.0582	.1056	.0929	.0700
WARE	.0360	.0750	.0471	.0426
State total $\times 10^{-6}$	3845	4046	3802	3803
error	.0050	.0580	-.0057	-.0055
'82 Census $\times 10^{-6}$	3824			

28. Nebraska (N=85)

Corr

	C16	C17	C18	C19
C17	.123			
C18	.255	.722		
C19	.231	.568	.716	
C20	.351	.111	.161	.174

$\hat{S}$ = .307 + .668 C18 R^2 = .522

(9.52) $\hat{\sigma}$ = .120

$\hat{R}$ = .497 + .677 C18 $\hat{\sigma}$ = .193

Measures

Q3 = 38,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0962	.1231	.1067	.0960
median	.0583	.0977	.0789	.0685
min	.0008	.0001	.0000	.0000
max	.8374	.5516	.5489	.8374
α	293	446	314	299
$\beta \times 10^3$	.0435	.0445	.0431	.0430
WARE	.0402	.0550	.0375	.0444
State total $\times 10^{-6}$	6744	7030	6857	6811
error	-.0032	.0390	.0134	.0067
'82 Census $\times 10^{-6}$	6766			

29. Nevada (N=14)

Corr

	C16	C17	C18	C19
C17	.712			
C18	.776	.761		
C19	.799	.855	.754	
C20	.793	.771	.844	.744

$\hat{S} = .380 + .601\ C18$ $\quad R^2 = .579$

(4.06) $\quad \hat{\sigma} = .1257$

$\hat{R} = .732 + .632\ C18$ $\quad \hat{\sigma} = .2425$

Measures

Q3 = 78,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0817	.1303	.1168	.0844
median	.0383	.0701	.0511	.0413
min	.0058	.0083	.0115	.0077
max	.2953	.5349	.4800	.2953
α	810	2820	1407	913
$\beta \times 10^3$	.0978	.1970	.1950	.1132
WARE	.0356	.0722	.0384	.0365
State total $\times 10^{-6}$	5323	5529	5337	5327
error	.0272	.0670	.0301	.0281
'82 Census $\times 10^{-6}$	5182			

30. New Hampshire (N=10)

Corr (N=23)

	C16	C17	C18	C19
C17	.433			
C18	.471	.895		
C19	.329	.684	.709	
C20	.186	.351	.363	.430

$\hat{S} = .091 + .906\ C18$ $\qquad R^2 = .80$ *

(9.22) $\qquad \hat{\sigma} = .0368$

$\hat{R} = .147 + .930\ C18$ $\qquad \hat{\sigma} = .0594$

Measures (N=10)

	Synth1	$\hat{R}$	$\hat{S}$
MARE	.0434	.0250	.0267
median	.0501	.0286	.0316
min	.0027	.0011	.0040
max	.0710	.0455	.0485
α	1308	488	553
$\beta \times 10^3$	.0494	.0554	.0552
WARE	.0454	.0246	.0268
State total $\times 10^{-6}$	5002	5134	5118
error	-.0452	-.0200	-.0231
'82 Census $\times 10^{-6}$	5239		

* The regression models are based on NH and VT counties combined.

31. New Jersey (N=21)

Corr

	C16	C17	C18	C19
C17	.794			
C18	.808	.964		
C19	.805	.899	.947	
C20	.626	.735	.796	.662

$\hat{S} = .0451 + .948\ C18$ $\qquad R^2 = .929$

$\qquad (15.73)$ $\qquad \hat{\sigma} = .036$

$\hat{R} = .0657 + .988\ C18$ $\qquad \hat{\sigma} = .0526$

Measures

Q3 = 1,592,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0818	.0512	.0544	.0747
median	.0827	.0473	.0508	.0710
min	.0202	.0126	.0143	.0202
max	.1467	.1094	.1127	.1270
α	12415	5165	5757	8932
$\beta \times 10^3$	.0425	.0362	.0362	.0397
WARE	.0812	.0489	.0523	.0675
State total $\times 10^{-6}$	32620	33775	33650	33105
error	-.0812	-.0487	-.0522	-.0675
'82 Census $\times 10^{-6}$	35503			

32. New Mexico (N=32)

Corr

	C16	C17	C18	C19
C17	.574			
C18	.670	.879		
C19	.580	.609	.738	
C20	.572	.427	.548	.503

$\hat{S} = .2078 + .771\ C18$ $\qquad R^2 = .773$

(10.11) $\qquad \hat{\sigma} = .070$

$\hat{R} = .368 + .753\ C18$ $\qquad \hat{\sigma} = .124$

Measures

Q3 = 138,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0846	.0897	.0872	.0864
median	.0552	.0637	.0573	.0552
min	.0010	.0264	.0150	.0010
max	.5728	.5277	.5278	.5727
α	536	693	567	536
$\beta \times 10^3$	.0731	.0847	.0817	.0766
WARE	.0412	.0492	.0412	.0396
State total $\times 10^{-6}$	6017	5961	6033	6033
error	-.0235	-.0326	-.0209	-.0209
'82 Census $\times 10^{-6}$	6161			

33. New York (N=62)

Corr

	C16	C17	C18	C19
C17	.585			
C18	.484	.820		
C19	.506	.764	.897	
C20	.537	.511	.473	.517

$\hat{S} = .078 + .346\ C16 + .600\ C18 \qquad R^2 = .719$

$\qquad\qquad (3.12) \qquad (8.89) \qquad \hat{\sigma} = .0596$

$\hat{R} = .101 + .474\ C16 + .623\ C18 \qquad \hat{\sigma} = .0778$

Measures

Q3 = 72,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0659	.0459	.0468	.0558
median	.0514	.0468	.0353	.0388
min	.0026	.0011	.0000	.0026
max	.5119	.1467	.1722	.5119
α	11067	3940	2623	3963
$\beta \times 10^3$	.0517	.0336	.0336	.0432
WARE	.0868	.0505	.0405	.0453
State total $\times 10^{-6}$	64542	67686	69239	68262
error	-.0840	-.0393	-.0173	-.0312
'82 Census $\times 10^{-6}$	70458			

34. North Carolina (N=100)

Corr

	C16	C17	C18	C19
C17	.538			
C18	.423	.860		
C19	.426	.701	.787	
C20	.479	.262	.266	.221

$\hat{S}$ = .140 + .531 C18 + .327 C16 $\hat{R}$ = .777

(14.57)   (4.03) $\hat{\sigma}$ = .0647

$\hat{R}$ = .222 + .542 C18 + .480 C16 $\hat{\sigma}$ = .103

Measures

Q3 = 192,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0673	.0763	.0754	.0663
median	.0410	.0511	.0498	.0410
min	.0002	.0001	.0005	.0002
max	.6492	.5849	.5845	.6491
α	809	809	794	643
$\beta \times 10^3$	.0305	.0319	.0317	.0266
WARE	.0449	.0400	.0398	.0340
State total $\times 10^{-6}$	23577	24325	24286	24115
error	-.0210	.0101	.0085	.0014
'82 Census $\times 10^{-6}$	24082			

35. North Dakota (N=51)

Corr

	C16	C17	C18	C19
C17	.684			
C18	.593	.800		
C19	.423	.662	.755	
C20	.251	.315	.239	.152

$\hat{S}$ .305 + .628 C18 R^2 = .640

(9.33) $\hat{\sigma}$ = .0901

$\hat{R}$ .542 + .610 C18 $\hat{\sigma}$ = .1602

Measures Q3 = 30,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.1030	.1071	.1074	.1106
median	.0685	.0783	.0883	.0947
min	.0011	.0013	.0056	.0011
max	.5432	.4528	.4299	.5434
α	423	616	732	728
$\beta \times 10^3$	.0960	.1370	.1170	.1240
WARE	.0636	.0832	.0956	.0928
State total $\times 10^{-6}$	3123	3084	3014	3022
error	-.0462	-.0582	-.0800	-.0770
'82 Census $\times 10^{-6}$	3275			

36. Ohio (N=88)

Corr

	C16	C17	C18	C19
C17	.454			
C18	.510	.877		
C19	.571	.717	.804	
C20	.305	.578	.529	.456

$\hat{S} = .132 + .869\ C18$ $\qquad R^2 = .769$

(16.91) $\qquad \hat{\sigma} = .04776$

$\hat{R} = .207 + .899\ C18$ $\qquad \hat{\sigma} = .07465$

Measures

Q3 = 330,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0488	.0746	.0675	.0517
median	.0387	.0676	.0565	.0431
min	.0004	.0008	.0000	.0004
max	.1782	.2154	.2046	.1782
α	1232	1839	1478	1069
$\beta \times 10^3$	.0245	.0195	.0199	.0186
WARE	.0383	.0476	.0421	.0376
State total $\times 10^{-6}$	44666	47337	46894	46337
error	-.0175	.0413	.0315	.0193
'82 Census $\times 10^{-6}$	45461			

37. Oklahoma (N=77)

Corr

	C16	C17	C18	C19
C17	.451			
C18	.596	.811		
C19	.640	.761	.784	
C20	.306	.373	.340	.369

$\hat{S} = .174 + .822\ C18$ $\qquad R^2 = .658$

(12.01) $\qquad \hat{\sigma} = .0794$

$\hat{R} = .296 + .828\ C18$ $\qquad \hat{\sigma} = .1353$

Measures

Q3 = 85,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0675	.0668	.0713	.0688
median	.0483	.0461	.0526	.0506
min	.0017	.0003	.0015	.0036
max	.2171	.2451	.2596	.2171
α	500	585	631	607
$\beta \times 10^3$	.0313	.0376	.0381	.0371
WARE	.0301	.0324	.0354	.0359
State total $\times 10^{-6}$	15629	15542	15705	15681
error	.0066	.0010	.0115	.0100
'82 Census $\times 10^{-6}$	15526			

38. Oregon (N=36)

Corr

	C16	C17	C18	C19
C17	.548			
C18	.451	.831		
C19	.376	.682	.888	
C20	.710	.608	.521	.337

$\hat{S}$ = .002 + .592 C18 + .370 C20 $\quad\hat{\sigma}$ = .733

(6.70) (2.28) $\quad\hat{\sigma}$ = .0736

$\hat{R}$ =.003 + .623 C18 + .610 C20 $\quad\hat{\sigma}$ = .134

Measures

Q3 = 228,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0631	.1709	.0655	.0628
median	.0401	.1574	.0471	.0418
min	.0037	.0419	.0002	.0037
max	.2490	.4084	.2601	.2490
α	817	6892	1059	915
$\beta \times 10^3$	.0583	.0812	.0755	.0679
WARE	.0375	.1284	.0422	.0406
State total $\times 10^{-6}$	12470	13859	12493	12444
error	.0153	.1284	.0171	.0132
'82 Census $\times 10^{-6}$	12282			

39. Pennsylvania (N=67)

Corr

	C16	C17	C18	C19
C17	.513			
C18	.561	.892		
C19	.605	.822	.916	
C20	.480	.541	.466	.520

$\hat{S} = .0311 + .664\ C18 + .300\ C20$ $R^2 = .816$

(13.49) (2.64) $\hat{\sigma} = .0515$

$\hat{R} = .0477 + .690\ C18 + .460\ C20$ $\hat{\sigma} = .0789$

Measures

Q3 = 650,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0609	.0727	.0553	.0532
median	.0457	.0667	.0442	.0332
min	.0014	.0005	.0012	.0014
max	.3758	.3303	.2951	.3758
α	4453	2468	1879	2416
$\beta \times 10^3$	.0624	.0366	.0377	.0451
WARE	.0617	.0399	.0376	.0393
State total $\times 10^{-6}$	46878	-	48800	48262
error	.0476	-	-.0086	-.0195
'82 Census $\times 10^{-6}$	49223			

40. <u>Rhode Island</u> (N=5)

<u>Corr</u> (N=27)

	C16	C17	C18	C19
C17	.626			
C18	.771	.936		
C19	.534	.768	.791	
C20	.789	.463	.624	.402

$\hat{S} = .303 + .695\ C18$ $\qquad R^2 = .875$ *

(13.25) $\qquad \hat{\sigma} = .0314$

$\hat{R} = .428 + .743\ C18$ $\qquad \hat{\sigma} = .0442$

<u>Measures</u> (N=5)

	Synth1	$\hat{R}$	$\hat{S}$
MARE	.0451	0241	.0291
median	.0409	.0205	.0296
min	.0260	.0036	.0032
max	.0663	.0537	.0641
α	1891	470	668
$\beta \times 10^3$	.0341	.1083	.1170
WARE	.0466	.0200	.0195
State total $\times 10^{-6}$	3872	4034	3996
error	-.0466	-.0066	-.0160
'82 Census $\times 10^{-6}$	4061		

* Regressions are based on combined counties of CT, RI, and MA.

41. South Carolina (N-46)

Corr

	C16	C17	C18	C19
C17	.599			
C18	.625	.855		
C19	.627	.852	.926	
C20	.476	.567	.669	.654

$\hat{S} = -.011 + .995\ C18$ $\qquad R^2 = .731$

$\qquad (10.94)$ $\qquad \hat{\sigma} = .0857$

$\hat{R} = -.018 + .982\ C18$ $\qquad \hat{\sigma} = .139$

Measures $\qquad$ Q3 = 211,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0482	.0525	.0499	.0473
median	.0350	.0482	.0349	.0346
min	.0006	.0006	.0049	.0006
max	.1705	.1952	.1646	.1705
α	436	880	382	353
$\beta \times 10^3$	.0318	.0315	.0315	.0283
WARE	.0340	.0509	.0308	.0295
State total $\times 10^{-6}$	11884	11528	12020	11976
error	-.0156	-.0451	-.0043	-.0080
'82 Census $\times 10^{-6}$	12072			

42. South Dakota (N=61)

Corr

	C16	C17	C18	C19
C17	.322			
C18	.293	.853		
C19	.436	.441	.618	
C20	.118	.135	.154	.260

$\hat{S} = .145 + .824\ C18$ $\qquad R^2 = .728$

(12.58) $\qquad \hat{\sigma} = .1043$

$\hat{R} = .234 + .800\ C18$ $\qquad \hat{\sigma} = .168$

Measures

Q3 = 25,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.1230	.1176	.1285	.1256
median	.0852	.0847	.0947	.0854
min	.0008	.0074	.0004	.0004
max	.8487	.6655	.7904	.8488
α	660	809	765	686
$\beta \times 10^{-6}$	.236	.268	.263	.241
WARE	.0627	.0722	.0710	.0688
State total $\times 10^{-6}$	2775	2684	2860	2838
error	-.0200	-.0522	.0100	.0022
'82 Census $\times 10^{-6}$	2831			

43. Tennessee (N=93)

Corr

	C16	C17	C18	C19
C17	.603			
C18	.547	.814		
C19	.569	.675	.796	
C20	.468	.499	.450	.423

$\hat{S} = .186 + .581\ C18 + .235\ C20$ $\qquad R^2 = .70$

$\qquad\qquad (10.00) \qquad (3.26)$ $\qquad \hat{\sigma} = .0785$

$\hat{R} = .302 + .577\ C18 + .360\ C20$ $\qquad \hat{\sigma} = .127$

Measures

Q3 = 100,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0963	.1120	.1200	.0988
median	.0649	.0729	.0778	.0723
min	.0010	.0002	.0017	.0039
max	.6250	.7178	.7959	.6250
α	1074	1060	1096	843
$\beta \times 10^3$	.0574	.0385	.0455	.0369
WARE	.0503	.0505	.0499	.0465
State total $\times 10^{-6}$	18723	19521	19393	19270
error	-.0052	.0372	.0304	.0239
'82 Census $\times 10^{-6}$	18821			

44. Texas (N=243)

Corr

	C16	C17	C18	C19
C17	.607			
C18	.619	.900		
C19	.666	.755	.849	
C20	.578	.570	.575	.609

$\hat{S} = .119 + .858\ C18$ $\quad R^2 = .811$

(32.12) $\quad \hat{\sigma} = .0871$

$\hat{R} = .216 + .865\ C18$ $\quad \hat{\sigma} = .1574$

Measures

Q3 = 87,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0804	.0780	.0781	.0787
median	.0571	.0556	.0561	.0541
min	.0003	.0000	.0000	.0004
max	.3887	.3309	.3306	.3887
α	801	828	824	851
$\beta \times 10^3$	.0100	.0093	.0093	.0098
WARE	.0360	.0401	.0401	.0405
State total $\times 10^{-6}$	80037	78802	78832	78924
error	-.0029	-.0183	-.0179	-.0168
'82 Census $\times 10^{-6}$	80271			

45. Utah (N=23)

Corr

	C16	C17	C18	C19
C17	.737			
C18	.760	.939		
C19	.580	.799	.894	
C20	.713	.694	.816	.748

$\hat{S} = .196 + .794\ C18$ $\quad R^2 = .882$

(12.51) $\quad \hat{\sigma} = .0832$

$\hat{R} = .345 + .807\ C18$ $\quad \hat{\sigma} = .147$

Measures Q3 = 73,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0751	.0810	.0810	.0780
median	.0515	.0692	.067	.0515
min	.0014	.0019	.0000	.0045
max	.2937	.3633	.3656	.2937
α	626	629	637	643
$\beta \times 10^3$	.0827	.0946	.0910	.0956
WARE	.0321	.0304	.0317	.0325
State total $\times 10^{-6}$	5908	6128	6148	6134
error	-.0238	.0126	.0160	.0135
'82 Census $\times 10^{-6}$	6052			

46. Vermont (N=13)

Corr (N=23)

	C16	C17	C18	C19
C17	.433			
C18	.471	.895		
C19	.329	.684	.709	
C20	.186	.351	.363	.430

$\hat{S} = .091 + .906\ C18$ $\qquad R^2 = .80$ *

(9.22) $\qquad \hat{\sigma} = .0368$

$\hat{R} = .147 + .930\ C18$ $\qquad \hat{\sigma} = .0594$

Measures (N=13)

	Synth1	$\hat{R}$	$\hat{S}$
MARE	.0355	.0467	.0458
median	.0208	.0390	.0409
min	.0000	.0016	.0006
max	.1240	.1816	.1793
α	245	308	298
$\beta \times 10^3$	.0882	.1086	.1085
WARE	.0254	.0314	.0303
State total $\times 10^{-6}$	2493	2553	2548
error	-.0119	.0118	.0099
'82 Census $\times 10^{-6}$	2523		

* Regression models based on combined counties of VT and NH.

47. Virginia (N=127)

Corr

	C16	C17	C18	C19
C17	.704			
C18	.693	.912		
C19	.765	.880	.909	
C20	.375	.556	.506	.553

$\hat{S} = .170 + .806\ C18$ $\quad R^2 = .832$

(24.89) $\quad \hat{\sigma} = .1067$

$\hat{R} = .280 + .829\ C18$ $\quad \hat{\sigma} = .1764$

Measures

Q3 = 115,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0859	.0961	.1074	.0868
median	.0616	.0751	.0761	.0595
min	.0003	.0009	.0015	.0003
max	.6938	.5232	.5823	.6938
α	1208	1038	1161	993
$\beta \times 10^3$	.0489	.0449	.0446	.0410
WARE	.0609	.0566	.0561	.0512
State total $\times 10^{-6}$	22618	23080	23732	23490
error	-.0269	-.0070	.0210	.0106
'82 Census $\times 10^{-6}$	23243			

48. Washington (N=39)

Corr

	C16	C17	C18	C19
C17	.816			
C18	.697	.931		
C19	.716	.817	.780	
C20	.662	.617	.537	.781

$\hat{S} = .131 + .835\ C18$ $\qquad R^2 = .867$

(15.56) $\qquad \hat{\sigma} = .0802$

$\hat{R} = .238 + .863\ C18$ $\qquad \hat{\sigma} = .1457$

Measures $\qquad$ Q3 = 330,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0629	.0708	.0551	.0612
median	.0534	.0710	.0502	.0517
min	.0013	.0010	.0025	.0013
max	.1897	.1538	.1718	.1897
α	835	1482	589	632
$\beta \times 10^3$	.0256	.0227	.0226	.0243
WARE	.0274	.0457	.0249	.0257
State total $\times 10^{-6}$	20085	20473	19917	19925
error	.0248	.0446	.0162	.0166
'82 Census $\times 10^{-6}$	19599			

49. West Virginia (N=55)

Corr

	C16	C17	C18	C19
C17	.674			
C18	.575	.832		
C19	.489	.714	.803	
C20	.387	.414	.434	.393

$\hat{S} = .023 + .406\ C16 + .545\ C18$ $\qquad R^2 = .75$

$\qquad\qquad (3.45) \qquad (7.83)$ $\qquad \hat{\sigma} = .0717$

$\hat{R} = .039 + .707\ C16 + .554\ C18$ $\qquad \hat{\sigma} = .1239$

Measures

Q3 = 112,000

	Synthl	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0850	.1270	.0833	.0863
median	.0627	.1105	.0771	.0741
min	.0029	.0129	.0019	.0029
max	.2796	.4678	.3642	.2796
α	604	2277	659	741
$\beta \times 10^3$	.0620	.0555	.0525	.0539
WARE	.0528	.1211	.0591	.0612
State total $\times 10^{-6}$	7499	8119	7589	76289
error	.0306	.1158	.0429	.0484
'82 Census $\times 10^{-6}$	7276			

50. Wisconsin (N=70)

Corr

	C16	C17	C18	C19
C17	.568			
C18	.641	.751		
C19	.568	.461	.765	
C20	.445	.228	.269	.485

$\hat{S} = .347 + .652\ C18$ $\qquad R^2 = .565$

(9.39) $\qquad \hat{\sigma} = .0605$

$\hat{R} = .564 + .666\ C18$ $\qquad \hat{\sigma} = .0984$

Measures

Q3 = 196,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0584	.0958	.0666	.0575
median	.0464	.0679	.0451	.0444
min	.0003	.0005	.0001	.0003
max	.3199	3841	.3337	.3200
α	1064	1575	744	518
$\beta \times 10^3$	.0436	.0354	.0348	.0255
WARE	.0469	.0585	.0299	.0274
State total $\times 10^{-6}$	19438	21071	20234	20094
error	-.0290	.0526	.0107	.0038
'82 Census $\times 10^6$	20019			

51. Wyoming (N=23)

Corr

	C16	C17	C18	C19
C17	.588			
C18	.610	.920		
C19	.736	.735	.808	
C20	.470	.727	.793	.580

$\hat{S} = .058 + .937\ C18$ $\qquad R^2 = .846$

(10.75) $\qquad \hat{\sigma} = .0825$

$\hat{R} = .119 + .948\ C18$ $\qquad \hat{\sigma} = .168$

Measures Q3 = 85,000

	Synth1	$\hat{R}$	$\hat{S}$	Hybrid $\hat{S}$
MARE	.0541	.0601	.0542	.0544
median	.0377	.0484	.0439	.0402
min	.0012	.0015	.0009	.0012
max	.2408	.2470	.2342	.2408
α	731	951	774	770
$\beta \times 10^3$	.1473	.1503	.1495	.1460
WARE	.0606	.0751	.0636	.0624
State total $\times 10^{-6}$	2881	2922	2889	2890
error	.0489	.0636	.0517	.0522
'82 Census $\times 10^{-6}$	2747			