

Bureau of Business Research  
Graduate School of Business Administration  
University of Michigan

February, 1971

**STEPWISE REGRESSION ANALYSIS APPLIED TO  
REGIONAL ECONOMIC RESEARCH**

WORKING PAPER NO. 27

by

Dick A. Leabo  
Professor of Statistics  
University of Michigan

**FOR DISCUSSION PURPOSES ONLY**

None of this material is to be  
quoted or reproduced without  
the express permission of the  
Bureau of Business Research

#### BACKGROUND OF THIS PAPER

An earlier version of this paper was given before the 130th Annual Meeting of the American Statistical Association, Business and Economics Section, December 28, 1970. It will be published in the Proceedings of the Business and Economics Section.

The U.S. Department of Commerce produces an annual economic time series for each state which is called personal income.<sup>1/</sup> The Regional Economics Division of the Office of Business Economics also prepares estimates on a county basis. Clearly, such data can be useful to industry in making capital budgeting decisions. For example, a firm might be interested in evaluating the sales potential of a particular area prior to considering potential expansion into that area. At the local, state, or national government level the data have economic ramifications for public policy decision makers.

#### Objectives and Methodology

The basic purpose of this paper is to measure the stability and sensitivity of a region's income, that is, which source of income has the greatest effect on the final level of personal income in a state <sup>2/</sup> and how sensitive each source is to fluctuations of the nation's economy.

In order to accomplish these aims, a multiple linear stepwise regression was run to determine the stability of the ten major

---

<sup>1/</sup> See "State Personal Income, 1969," Survey of Current Business, Apr., 1970, p. 14 ff.

<sup>2/</sup> Michigan data for the years 1948-69 are used in this paper. Source: U.S., Department of Commerce, Office of Business Economics.

sources of Michigan's personal income.<sup>3/</sup> The sensitivity of the state's income to national economic conditions was measured by relating each of these major sources to various national economic indicators.<sup>4/</sup> The aggregate series used were: gross national product (GNP), total U.S. unemployment, industrial production, manufacturing and trade sales, retail sales, total new construction, manufacturers' shipments, inventories and new orders, and durable goods sales.

#### Limitations

Certain problems are unique to the task of correlating time series data, especially if one wishes to apply any error formulas or tests

---

<sup>3/</sup>Several excellent applied textbooks dealing with regression and correlation include: Norman Draper and Harry Smith, Applied Regression Analysis (New York: John Wiley & Sons, Inc., 1966); Karl A. Fox, Intermediate Economic Statistics (New York: John Wiley & Sons, Inc., 1968); and David S. Huang, Regression and Econometric Methods (New York: John Wiley & Sons, Inc., 1970).

<sup>4/</sup>The computer program utilized was the UCLA BMD2R adapted by the Statistical Research Laboratory at the University of Michigan. The data used, as explained later, were relative first differences. For a discussion of the validity of a stepwise procedure (e.g., as opposed to the forward-selection procedure) see Draper and Smith, Applied Regression Analysis, chap. vi. The stepwise technique begins with a simple correlation matrix and enters into regression the independent variable most highly correlated with the dependent variable. Using the partial coefficients generated with respect to the other variables, the program then selects the next variable to enter the model. It is worth noting that in the stepwise method, a variable which may have been the best single one at an earlier step might, at a later step, be rejected (on the basis of an F test) because of the relationship between it and other variables now in the regression. Any variable judged not significant in terms of improving the regression equation is rejected. This stepwise procedure is continued until the program runs out of independent variables or until there are no more to be included or excluded.

of significance.<sup>5/</sup> The two major problems are, of course, the nonindependence of successive observations and the effect of the relationship of long-term trend upon the correlation coefficient.

The formulas for reliability, i.e., the standard error of estimate and the general appraisal of the correlation coefficient by F, are based upon the theory of random sampling and a normally distributed population. Briefly, the theory assumes that each observation in a sample is selected purely at random from all the items in the universe (i.e., each possible item has an opportunity of being chosen). It further assumes that successive samples are chosen in such a manner that values found in one sample have no relation or connection with the values found in the next sample.

However, it is apparent that the level, for example, of GNP, or of a state's personal income in any given year, is not completely independent of that of the previous year. Rather, it is probably approximately the level established in the preceding year but adjusted by new or changing factors. However, because of the nature of the data being studied, this autocorrelation does not necessarily mean that measures of correlation and of range of error in such problems are invalidated. Forces of nature--such as temperature and rainfall, which affect crop production--the threat of wars or prospect of peace, and perhaps many other variables remove some of this built-in relationship. That is, the level of the time series being analyzed is not necessarily geared to previous years and therefore may be regarded as representing a reasonably "random" sample.

---

<sup>5/</sup> See Mordecai Ezekiel and Karl A. Fox, Methods of Correlation and Regression Analysis (3rd ed.; New York: John Wiley & Sons, Inc., 1959), chap. xx.

And, if the observations for a given year have no particular relation with items which might have been "selected" other years on the basis of different influences by forces of nature, we may then be reasonably confident that this method is a useful technique in spite of its theoretical weaknesses.

In addition, most of the trend influences were removed by the use of relative first differences. In order to measure the impact of any trend remaining in the data, a stepwise regression was run; time was used as one of the independent variables, but it was judged not significant and never entered the regression at the .005 level.<sup>6/</sup> For practical purposes this means that the relative first difference data tend to measure the cyclical impact. Certainly, some irregular probably remains, but these random shocks can usually be identified. If trend has been removed from economic data, the analysis does not show the relation among the variables as such but rather the relation of the series with respect to their tendency to deviate from their respective secular growth.

Some authors in the 1920s <sup>7/</sup>suggested rather strongly that no attempt should be made to measure the reliability of the coefficient of multiple determination, and asserted that error formulas do not apply to time series. This position no longer is accepted. In this study a general appraisal of the correlation was provided through the F test of significance. In a direct correlation of time series data, no attempt should be

---

<sup>6/</sup> Also, when the relative first differences were correlated with time, the coefficient of determination was .01, indicating that almost all of the trend was removed.

<sup>7/</sup> See Ezekiel and Fox, Methods of Correlation and Regression Analysis, pp. 325-30, and G.U. Yule, "Why Do We Sometimes Get Nonsense-Correlations between Time Series? A Study in Sampling and the Nature of Time Series," Journal of the Royal Statistical Society, LXXXIX, (1926), 1-64.

made to measure the reliability. However, the reliability of two or more cycles may be so measured whenever deviations from trend approximate normal distributions. Then the cycles are considered direct measures of certain variable forces and the correlation indicates the degree to which one series covaries with another, above what might have been expected by chance.

The use of a rectilinear relationship is justified or considered appropriate because any curvilinear effect that might be present probably has been eliminated with the trend. If the above techniques are used, the original problems of correlating time series are minimized and the impact of the cyclical swings upon Michigan's personal income may more readily be determined.

#### Analysis of Data and Conclusions

Our basic objectives were to measure the stability and sensitivity of a region's income. To analyze the stability of Michigan's income, it is necessary to determine which component of total income has the greatest effect on the final level.

#### Measuring the stability of a region's income

A multiple linear stepwise regression was run in order to obtain a realistic picture of the importance of the various industrial sources of Michigan's personal income. The basic data used (relative first differences) were U.S. Department of Commerce estimates of wages and salaries for nine major components. The independent variables included were:

$X_1$  -- construction

$X_2$  -- manufacturing

- X<sub>3</sub>--wholesale and retail trade
- X<sub>4</sub>--finance, insurance, and real estate
- X<sub>5</sub>--transportation and public utilities
- X<sub>6</sub>--services
- X<sub>7</sub>--federal government (military)
- X<sub>8</sub>--federal government (civilian)
- X<sub>9</sub>--state and local government

The dependent variable (X<sub>0</sub>) was total Michigan personal income. Undoubtedly, these data more or less guaranteed a certain degree of correlation; however, this is not really the significant part of the problem. What is interesting here is the weight each major component has in the determination of the final level of Michigan's personal income. In other words, the measure of correlation is not the important statistical term in this case, but rather the order (as determined by the stepwise regression) in which each industrial source enters the model.<sup>8/</sup>

Table 1 ranks the nine major sources of personal income according to the order they entered the stepwise regression model. That variations in the cyclical swing of manufacturing wages and salaries have the greatest impact upon the cyclical fluctuations of Michigan's income will not be a shock to anyone. In fact, manufacturing wages and salaries alone account for over 85 per cent of the variations. All other sources combined account

---

<sup>8/</sup> Incidentally, the coefficient of multiple determination (R<sup>2</sup>) was found to be .96 and R = .98. The standard error was 1.24 per cent. Another measure, the beta net regression coefficients, also is useful in assessing the importance of each independent variable. The betas are more meaningful than the regular regression coefficients (b) because the dissimilarity of the units (no problem here) is eliminated. The betas also can be tested for statistical reliability.



TABLE 1

Importance of Nine Major Sources of  
Michigan's Personal Income, Ranked by Order of Entry  
into Stepwise Multiple Regression\*

Rank	Independent Variable (Wages and Salaries)	Regression Coefficient	Standard Error
1	Manufacturing	0.42790	0.100
2	Finance, insurance, and real estate	0.34446	0.176
3	Wholesale and retail trade	0.32702	0.206
4	Federal, military	0.06355	0.072
5	Construction	-0.03529	0.064
6	Transportation and public utilities	-0.08896	0.196
7	State and local government	0.04584	0.111
8	Federal, civilian	0.00884	0.044
9	Services	-0.03342	0.193

\*The constant in the multiple regression equation is -0.61743;  
 $R^2 = .96$ ; standard error of estimate = 1.24297 per cent.

for only 11 per cent of the fluctuations (4 per cent is unexplained by these nine variables).

What might be equally interesting and important is the sequence in which the remaining eight variables entered the regression. The finance, insurance, and real estate category entered after manufacturing, followed by the wholesale and retail trade income. Surprisingly, the federal military component of Michigan's personal income came fourth, ahead of both construction and transportation and public utilities. The wages and salaries of the service industry were the last variable to be included and explain very little of the fluctuations in the dependent variable. The negative regression coefficients of the sources for services, construction, and transportation and public utilities suggest a contracyclical effect upon the state's personal income and, in that sense, contribute to its stability. However, the volatile nature of the construction industry probably offsets, in some years, any stabilizing influence of this source.

Ten years ago a similar analysis showed the services' wages and salaries to be the second most important source with respect to cyclical impacts. Manufacturing was in first place then as now. The construction industry was in the middle in the 1960 analysis as it is now. The finance, insurance, and real estate component and also wholesale and retail trade are more influential than ten years ago in affecting total Michigan personal income. In fact, the wholesale and retail trade category has virtually traded places with the service industry in terms of its significance to the regression model.

Measuring the sensitivity of Michigan's personal income

The sensitivity of Michigan's personal income to national economic conditions was studied through the eight linear multiple step-wise regressions which used relative first difference data again. Each major source, as well as the total personal income, was correlated with selected national economic indicators to determine the affects of cyclical swings in the nation's economy upon various industrial sectors of the Michigan economy. Will a cyclical rise or decline in general business conditions (from long-term trend) have a large or small effect upon the personal income originating in various industries? In a period of recession would the manufacturing sector in Michigan suffer relatively more than construction or trade? The following analysis attempts to suggest some answers to such questions. We will discuss each industry in order of its relative importance to the total as indicated by Table 1. (Also see Tables 2, 3, and 4 below.)

Table 2 summarizes the order of the importance of the independent variables when correlated with total Michigan personal income and each of the major sources of income except federal military and civilian expenditures. (The latter two sources were not deemed important enough to develop separate regression models.) In the models studied, Michigan personal income seems most sensitive to fluctuations in GNP; industrial production (INDPRD); manufacturing and trade sales (MFGTRS); and manufacturer's shipments, inventories, and new orders (MMSGIO). Given the fact that Michigan basically has an industrial economy oriented towards manufacturing production, this is not an unexpected finding. However, it should be of interest to governmental and business policy makers because

TABLE 2  
Step Number at Which Independent Variable Entered Regression

Independent Variable	Dependent Variable								Frequencies of Ranks (out of 8)						
	TOMPI	MFGXO	FINIRE	WRTRAD	CONSOXO	TRNSPU	SLGOVT	SERVCS	1,	2,	or 3	4,	5,	6,	or 7
GNP	2	6	7	6	2	3	3	2	5	3					
TOTUEM	6	DNER†	3	5	3	5	DNER	4	2	4‡					
INDPRD	4	1	2	2	7	4	1	3	5	3					
MFGTRS	3	5	5	4	1	2	2	1	5	3					
RETSLS	5	2	4	3	4	6	4	DNER	2	5‡					
MFGSIO	1	3	1	1	5	1	NIM	5	5	2					
TNCONS	NIM*	NIM	NIM	NIM	6	NIM	NIM	NIM	0	1§					
TCCOUN	NIM	NIM	8	NIM	NIM	NIM	NIM	NIM	0	1§					
MDONFU	NIM	NIM	6	NIM	NIM	NIM	NIM	NIM	0	1§					
DGSALS	NIM	4	NIM	NIM	NIM	NIM	NIM	NIM	0	1§					

\*NIM = not in model.  
†DNER = did not enter regression at .005 level.  
§ Last four independent variables were never used in more than one model each.  
# One or more independent variable did not enter regression.

TABLE 3

Summary of Sensitivity Regressions  
Data: Relative First Differences

Dependent Variable*	Standard Deviation	Mean of Dependent Variable	Coefficient of Multiple Determinant	Standard Error
Total Michigan personal income (TOMPI)†	4.95	6.35	.87	2.14
Manufacturing (MFGXO)‡	8.04	6.76	.52	6.66
Finance, insurance, and real estate (FINIRE)‡	3.33	7.93	.55	2.90
Wholesale and retail trade (WRTRAD)†	4.65	5.94	.74	2.86
Construction (CONSXO)§	10.69	8.03	.69	7.44
Transportation and public utilities (TRNSPU)†	4.41	5.39	.79	2.39
State and local government (SLGOVT)	4.72	9.33	.43	3.99
Services (SERVCS)†	4.15	7.69	.73	2.51

\*Independent variables are: (1) GNP, (2) TOTEM, (3) INDPRD, (4) MFGTRS, (5) RETSLS, (6) MFSIO, (7) MDONFU, (8) TCCOUN, (9) TNCONS, and (10) DGSALS.

†Independent variables in model: (1) thru (6).

‡Independent variables in model: (1) thru (8).

§Independent variables in model: (1) thru (6) plus (9).

||Independent variables in model: (1) thru (5).

#Independent variables in model: (1) thru (6) plus (10).

TABLE 4

Fluctuations in Michigan Personal  
Income Associated with U.S. Economic Indicators\*

Income Source	R <sup>2</sup>	If Economic Indicators Rise 5 Per Cent (Percentage)	If Economic Indicators Fall 5 Per Cent (Percentage)
Total Michigan personal income	.87	+6.3	+1.3
Manufacturing	.52	+7.2	-3.7
Finance, insurance, and real estate	.55	+8.0	+3.8
Wholesale and retail trade	.74	+5.6	-0.7
Construction	.69	+5.2	-28.9
Transportation and public utilities	.79	+5.3	+2.6
State and local government	.43	+7.6	+9.9
Services	.73	+6.8	+3.0

\*See Table 3 for specific independent economic variables related with each source. Because of lower coefficients of determination, the reliability of some regression equations is open to question. This is especially true for state and local government wages and salaries where less than one-half of the variation is explained.

it usually means that Michigan's income, either total or by component, rises faster on the cyclical upswing than the nation's economy. The construction industry, which is harder hit by a recession than any other in Michigan, drops faster and further on the downswing (see Table 4) than the national average although in a growing economy it moves upward at about the same rate. Retail sales (RETSLS) and total U.S. unemployment (TOTUEM) were not as significant in influencing cyclical swings in the states' income. In the case of TOTUEM a lag effect might be operating and, if quarterly or monthly data had been related, the influence of this independent variable probably would have been more noticeable.

An important fact to keep in mind is that individually each of the six <sup>9/</sup> independent variables is highly related in a simple correlation to fluctuations in Michigan's total personal income. It is only when they are combined in a multiple relationship with a specific source of income that the relative importance of each is altered. The coefficients of partial determination also confirm this fact. One should not infer that changes in TOTUEM or RETSLS are unimportant to Michigan's economy. Rather, when their effects are combined with other independent variables we discover that these two are generally less important. It really is a matter of relative positions.

Normally, national retail sales are roughly coincident with cyclical movements of the economy. However, in the case of Michigan data this variable was only of modest value in measuring any sensitivity; it

---

<sup>9/</sup>GNP, total U.S. unemployment, industrial production, manufacturing trade and sales, retail sales, and manufacturer's shipments, inventories, and new orders.

appeared to be most closely related to manufacturing and to wholesale and retail trade wages and salaries.

Referring to Table 4 we can make the following observations:

1. While total Michigan personal income, manufacturing wages and salaries, and wholesale and retail trade wages and salaries tend to rise faster than the nation's economy on the upswing, they do not drop as far. In that sense these sources of income cushion the total.
2. Finance, insurance and real estate, transportation and public utilities, state and local government and services have a countercyclical effect on Michigan's total income. Indeed, governmental wages and salaries tend to increase faster when the nation's economy is declining than during a period of growth. However, one must keep in mind the relative importance of each source of income.
3. Finally, the tremendous fluctuations in the construction industry are obvious and well known. During periods of economic expansion this industry maintains pace with the nation. However, during periods of recession when mortgage money is usually "tight," construction wages and salaries fall almost five times as fast as the nation's overall economic activity. In an era of a serious shortage in housing, especially for low- to moderate-incomes, this industry would seem to be a vital target for governmental policy makers who could help mitigate the effects of a recession at the same time making progress in an important area of social concern.

#### Summary

The primary purpose of this paper was to demonstrate how step-wise regression and correlation analyses might be employed to measure the stability and sensitivity of a region's personal income. One must recognize that the degree of relationship in some models is relatively low, which would indicate that additional or different independent variables could reduce the unexplained variation. A second limitation, although I



do not think it serious, is that data for only twenty-one years were used. Conceivably, data for another sample of years might alter the relationships; however, the data used were for the most recent two decades. The following conclusions were discussed in the text but they might merit restatement here.

1. Manufacturing wages and salaries are the most closely related to total personal income in Michigan. Of the variation in total personal income 85 per cent is associated with fluctuations in this segment of the state's economy. The other eight industries when combined in a multiple relationship explain only 11 per cent of the variation. (See Table 1 for "rankings.")

2. The services, construction, and transportation and public utilities wages and salaries exert a countercyclical effect on the level of Michigan's total personal income. The volatile nature of the construction industry on the downswing offsets the stabilizing influence of the other two industries.

3. State and local government and federal civilian wages and salaries have only a modest impact on the stability of the state's economy.

4. Total personal income in Michigan is most sensitive to fluctuations in such aggregate indicators as GNP, industrial production, manufacturing and trade sales, and manufacturer's shipments, inventories and new orders. (See Table 2.) This reflects the fact that the state's economy is heavily oriented towards manufacturing production.

5. Total personal income, as well as the seven major components, rises at a faster rate on the upswing of a business cycle than the nation's economy generally. (See Table 4.)

6. The construction wages and salaries component of the state's income is hit the hardest during a recession; however, during the expansion phase of the cycle, it moves upward at about the pace of GNP. (See Table 4.)

7. All six independent variables (GNP, TOTUEM, INDPD, MFGTRS, RETSLS, and MFGSIO) are highly related individually to fluctuations in Michigan personal income. The simple coefficients of determination are all high. It is only when these variables are combined in a multiple relationship with a specific source of income that the relative importance of each varies. Given this fact, TOTUEM and RETSLS are relatively less important in determining the final level of Michigan's personal income.

8. Total personal income, manufacturing and wholesale and retail trade wages and salaries tend to rise faster than the nation's economy. Also, they do not drop as fast on the downturn in business activity. (See Table 4.) In this sense these sources of income provide a cushioning effect for the state's economy.

9. Finance, insurance and real estate, transportation and public utilities, and state and local government wages and salaries exert a countercyclical affect on the total personal income. In fact, the latter source tends to rise faster during a recession than during a period of economic growth. While such a built-in stabilizer is good, one must remember the relative importance of each source of income. On this basis, the real impact is much less than our intuitive guess might expect.

OTHER WORKING PAPERS

Working Paper  
Number

- 1 "Reflections on Evolving Competitive Aspects in Major Industries," by Sidney C. Sufrin
- 2 "A Scoring System to Aid in Evaluating a Large Number of Proposed Public Works Projects by a Federal Agency," by M. Lynn Spruill
- 3 "The Delphi Method: A Systems Approach to the Utilization of Experts in Technological and Environmental Forecasting," by John D. Ludlow
- 4 "What Consumers of Fashion Want to Know," by Claude R. Martin, Jr.--Out of print. To be published in a forthcoming issue of the Journal of Retailing.
- 5 "Performance Issues of the Automobile Industry," by Sidney C. Sufrin, H. Paul Root, Roger L. Wright, and Fred R. Kaen--Out of print. To be published as a future Michigan Business Paper.
- 6 "Management Experience with Applications of Multidimensional Scaling Methods," by James R. Taylor
- 7 "Profitability and Industry Concentration," by Daryl Winn
- 8 "Why Differences in Buying Time? A Multivariate Approach," by Joseph W. Newman and Richard Staelin --Out of print. To be published in a forthcoming issue of the Journal of Marketing Research.
- 9 "The Contribution of the Professional Buyer to the Success or Failure of a Store," by Claude R. Martin
- 10 "An Empirical Comparison of Similarity and Preference Judgments in a Unidimensional Context," by James R. Taylor
- 11 "A Marketing Rationale for the Distribution of Automobiles," by H.O. Helmers
- 12 "Global Capital Markets," by Merwin H. Waterman

Working Paper  
Number

- 13 "The Theory of Double Jeopardy and Its Contribution to Understanding Consumer Behavior," by Claude R. Martin
- 14 "A Study of the Sources and Uses of Information in the Development of Minority Enterprise--A Proposal for Research on Entrepreneurship," by Patricia L. Braden and Dr. H. Paul Root
- 15 "Program Auditing," by Andrew M. McCosh
- 16 "Time Series Forecasting Procedures for an Economic Simulation Model," by Kenneth O. Cogger
- 17 "Models for Cash Flow Estimation in Capital Budgeting," by James T. Godfrey and W. Allen Spivey
- 18 "Optimization in Complex Management Systems," by W. Allen Spivey
- 19 "Support for Women's Lib: Management Performance," by Claude R. Martin, Jr.
- 20 "Innovations in the Economics of Project Investment," by Donald G. Simonson
- 21 "Corporate Financial Modeling: Systems Analysis in Action," by Donn C. Achtenberg and William J. Wroblewski
- 22 "Sea Grant Delphi Exercises: Techniques for Utilizing Informed Judgments of a Multidisciplinary Team of Researchers," by John D. Ludlow
- 23 "The Spanish in Nova Scotia in the XVI Century--A Hint in the Oak Island Treasure Mystery," by Ross Wilhelm
- 24 Not yet ready to be released.
- 25 "Market Power, Product Planning, and Motivation," by H. Paul Root
- 26 "Competition and Consumer Alternatives," by H. Paul Root and Horst Sylvester