

ON QUANTITATIVE HISTORY: THE POVERTY INDEX FOR MEXICO*

James W. Wilkie
University of California, Los Angeles

This is a response to the recent critique of my book *The Mexican Revolution: Federal Expenditure and Social Change Since 1910*, 2nd ed. (Berkeley, 1970) by Felix G. Boni and Mitchell A. Seligson.¹ Their factor analysis of my data on poverty suggests, first, that their methodology is more “sensitive” than mine; second, that they have confirmed the thrust of an earlier analysis by Thomas E. Skidmore and Peter H. Smith;² and third, that the task of quantitative history is to reduce data of “unwieldy proportions” so that it is subject to greater understanding.

It is not my intention in this response to fault the critique for its attempt to test or find new meaning in my data (indeed, it is revealing). Rather, by showing how factor analysis has been misused and misunderstood I wish to make a number of points that need to be understood by scholars who tend to place undue confidence in such methodology. In arguing here that the critique has not achieved any of its goals, I aim to show the positive values of factor analysis, going beyond the analytic problems inherent in the critique. As William P. McGreevey has noted (1974), the generation of new data from original statistics given in *The Mexican Revolution* involves the “process of scholarly interaction which marks a significant and positive change in the field of Latin American history.” In this light, the data given in the critique have merit.

FACTOR ANALYSIS OF POVERTY IN MEXICO

Factor analysis of data on poverty in my book demonstrates how seven indicators (for Mexico’s six decennial-census years in each of 32 federal states and territories) are related to one another (see Table 1 for methodology). In the critique two “factors” or patterns emerge as being apparently correlated with certain indicators to the extent shown by 14 “factor

*In the course of developing this paper, I am grateful for critiques from three colleagues at UCLA: Bruce H. Herrick (Economics), J. Ward Keesling (Education), and Raymond J. Jessen (Business Statistics and Public Health).

clustering of factor loadings given in Table 1. If we can be persuaded that two factors do emerge and have been appropriately given conceptual meaning, then it could be argued that criticism of my book is warranted. In that case, I might have used one factor (containing five clustered indicators) instead of seven separate indicators, thus showing poverty as having an underlying structure rather than as being an unpatterned phenomenon.

Data in Table 1 are interpreted in the critique to argue that indicators 1 to 5 cluster together to measure poverty, the factor loadings ranging from .659 to .857 (a difference of .198). It is argued that because indicators 6 and 7 (percentage of persons barefoot and speaking only an Indian language) show extremely low factor loadings (.134 and .109, respectively), these two indicators lie in a pattern different from that of the other five indicators.

Moreover, it is hypothesized that items 6 and 7 identify, if crudely, the extent of "Indianness" and not the extent of poverty in Mexico. Here the barefoot population variable and language variable cluster together to measure Indianness, with factor loadings ranging only from .843 to .854 (a difference of .011).

Having found that my multidimensional indicators of poverty can be reduced to Factor I, Boni and Seligson offer two conclusions. First, they see Skidmore and Smith (1970: 78) as having correctly suggested that I am dealing with two dimensions, one of which differs radically from the other in that it restricts an individual's "life chances" in a much more fundamental way—this dimension involves "poverty." Second, in confirming my view that federal capital investment (1959-63) under President López Mateos did not go to states with the most poverty, they claim that this is true only in terms as measured by Factor I—Factor II (Indianness) being statistically non-significant.

REINTERPRETING TABLE 1

Let us reexamine the factor loadings with regard to the clusters that may be found for each of the two factors identified in the critique. In analyzing Factor I, the critique includes factor loadings for indicators 1 to 5 which are separated from indicators 6 and 7 by .525 points between the low and high numbers in the respective clusters. Yet in analyzing Factor II, the critique ignores the same kind of dramatic gap (.436), this time between the low in the cluster of indicators 3 to 7 and the high in the cluster of indicators 1 and 2. In short, if the same logic were followed in analyzing Factor II (Indianness) as in analyzing Factor I (Poverty), the Indianness range would have been expanded beyond items 6 and 7 to include indicators 3, 4, and 5,

which show relationships (.531, .612, and .520) more in consonance with indicators 6 and 7 than with items 1 and 2, where they are grouped in the critique.

Failure to develop a parallel analysis of the two clusters in Factor I and the two clusters in Factor II is not, however, the main problem in the critique. A closer look at the factor loadings in Table 1 suggests that each factor is composed of three clusters and not two as claimed. Here is where scholarly interpretation remains paramount: The computer can give us the objective correlations, but it is the intellect that must decide what constitutes a cluster.

In my view, the indicators cluster as shown in Table 1 by spaces setting off three groups: (a) items 1 and 2; (b) items 3 to 5; (c) items 6 and 7. Reading Factor I (Poverty) downward from the high through the middle to low cluster—and Factor II (Indianness) upward the same way—we find high levels in the .800s, middle levels in the .500s to .700s, and low levels in the .100s or below. This three-cluster view shows no distinction between Factors I and II. Indeed, Boni and Seligson note that indicators 3, 4, and 5 fall in the middle levels for both factors; but in order to develop a case, they ignore their own admission that in addition to being indicators of poverty, the variables 3, 4, and 5 are to some extent also related to the notion of Indianness. Thus, in choosing to exclude this middle level as related to Factor II, it is denied that Indianness is related to the Poverty Index (in spite of their recognition that “most Indians are poor”). Justification that the middle level cluster can be considered *only* as belonging to Factor I (Poverty) is made on the grounds that it is more highly correlated with poverty than with Factor II, Indianness. Since the difference in correlation is not great (and ignores the cluster value in factor analysis), the critique can only be considered an attempt to force a distinction between the conceptual categories of Factors I and II where, in fact, no such distinction exists. And had Boni and Seligson read Woodrow Borah (1954), it is doubtful that they would have chosen to conceptually identify Factor II as involving “Indianness.”

We may summarize by suggesting that since conceptual clarification cannot be developed by distinguishing between the Poverty Factor and the Indianness Factor, the challenge of my use of seven poverty indicators is not successful. Are we to conclude that the critique is wasted? H. M. Blalock states that

where it is difficult to identify the final factors even after rotation has been accomplished, factor analysis may not be worth the effort involved. Like other statistical techniques, it should be used as a tool which may possibly contribute to the clarification of theory, but it cannot be expected to serve as a substitute for sound theoretical thinking (1960: 388-89).

Taking the opposite view, I could say that in this case the critique is useful if only in showing that their factor analysis is not applicable to my data.

If the Boni and Seligson factor analysis fails, then also their argument related to Mexican federal expenditure³ fails, the latter argument being built upon the former. Thus, their following conclusion is invalidated: "The size of a state's indigenous population has little relevance to the distribution of federal expenditures, while the extent of its poverty does." But one can find positive value in their computation by using it, for example, to challenge their conclusion that Skidmore and Smith's criticism has been substantiated.

SKIDMORE AND SMITH REEXAMINED

Although in analyzing my Poverty Index it is true that Skidmore and Smith suggested the existence of two dimensions, they are not the same two found by Boni and Seligson. As the former authors noted (1970: 78) with regard to my seven indicators:

One might ask whether . . . three [indicators] (illiteracy, knowledge of only an Indian language, residence in towns under 2,500) do not differ radically from the other four in that they restrict an individual's mobility and therefore his "life-chances" in a much more fundamental way. Curiously enough Wilkie gives *equal* weight to each of the seven elements. It is worth noting that this index would measure as exactly equal the "social change" induced by one tortilla-eater taking to bread and one illiterate learning to read and write. Meanwhile there is no social change if a barefoot man dons sandals, but there is social change if a sandal-wearer puts on shoes!

As Boni and Seligson themselves note, the factor loadings or correlations given here in Table 1 are immediately relevant to answer the Skidmore and Smith question; therefore, it is puzzling why the question is not directly treated. (The matter is all the more puzzling not only because Smith was an acknowledged consultant in this case but also because he must have seen the contradictions in their interpretation while at the same time becoming aware of contradictions in the earlier analysis he and Skidmore prepared.) In directly pointing out that the dimensions postulated by Skidmore and Smith are contradicted by Boni and Seligson, it is pertinent to interject the comments of Woodrow Borah and Sherburne F. Cook, whose statistical analysis of all of my items except 2 and 5 contradicts both sets of critics. Borah and Cook (1966: 975-78) write:

We use only two indices of poverty in our analysis since speaking an Indian language and wearing sandals or going barefoot rather than wearing shoes give adequate evidence of the proportion and location of the population living in poverty. . . .

With respect to other indications of poverty, [our correlations] show that the proportion of illegitimacy has a significant positive relation with proportion of illiteracy and proportion of people who continue to speak or who know an Indian tongue. Similarly there is a strong negative correlation between proportion of illegitimacy and proportion of people who wear shoes. These correspondences combine to indicate the prevalence of *unión libre* and illegitimacy among the poorer, the rural, and the Indian segments of the Mexican population. Persons with all of these characteristics are concentrated in the south of the country and along the West Coast. It is almost impossible to separate the Indian from the rural poor since in general Indians tend (in terms of adoption of new ways of European industrial and urbanized civilization) to be the more backward, as well as the more poverty-stricken, of the rural population.

Difference concerning the salient aspects of Mexican poverty as seen by these three sets of analysts is shown (along with the bases for their views) in Table 2.

TABLE 2 *Comparative Views on the Essential Indicators of Mexican Poverty*

Observers	Indicators from Table 1	Basis for Analysis
Borah & Cook (1966)	Indian-speakers* Barefoot persons & sandal-wearers	} Correlations against illegitimacy
Skidmore & Smith (1970)	Only Indian-speakers Isolated persons Illiterates	
Boni & Seligson (1973)	Sandal-wearers Tortilla-eaters Isolated persons Illiterates Persons with no sewage disposal	} Factor analysis

*Includes persons speaking Indian and Spanish as well as only Indian.

Of the two items from my list of seven indicators not included by Boni and Seligson, the item "only Indian speakers" appears on the Skidmore and Smith list and (with modification) on the Borah and Cook list; the item "barefoot persons" appears on the Borah and Cook list. Thus, in one way or another, the recurrence of each of the seven indicates that it is reasonable to include all the items in my index.

If Skidmore and Smith seem surprised that in my index there is

social change only when a sandal-wearing person changes to shoes (and not when a barefoot person dons sandals), perhaps it is because they see a natural progression in social conditions: The poor move up in status by the adoption of sandals, a necessary stage before they move up to the next highest level by putting on shoes. Although this view might sometimes be the case, it does not take into account geographical conditions. My view (expressed in terms of climate in *The Mexican Revolution*) is that geographical factors tend to account for the footwear for the poor: Although the Skidmore and Smith assumption tends to be valid for tropical areas, it is invalid, for example, in non-tropical Querétaro and Chihuahua. And since the condition of going barefoot or of wearing sandals does not involve a culture shift from traditional to modern clothes and/or outlook on personal health, I note also that the important shift is from not wearing shoes to wearing them. Moreover, persons wearing sandals are not necessarily any more protected from soil-borne roundworm infection than barefoot persons (Holvey, 1972).

In contrast with Borah and Cook, Boni and Seligson would exclude the barefoot indicator from the idea of poverty, including it as an indicator of Indianness. The validity of that correlation is best refuted by analysis of the 1940 population census in which the Mexican government asked (for the first and only time) about the cultural characteristics of persons speaking an Indian language. It is notable that 3.6 times more Mexicans regularly went barefoot than only spoke an Indian language (Wilkie, 1970a: 212, 223). One-fifth of the populace speaking only Spanish went barefoot (about the same ratio as those speaking Spanish and Indian); and 21.3 per cent of only Indian speakers wore sandals (México, 1940: 35). Boni and Seligson clearly have not understood the meaning of the Mexican data. Nevertheless, they have helped us to answer the original question posed by Skidmore and Smith: Two radically distinct dimensions of poverty do *not* appear in Table 1.

With regard to my seven indicators, I believe that the Skidmore and Smith reference to tortilla-eaters and sandal-wearers (quoted above in their question) is misleading. In *The Mexican Revolution* I do not say that the eating of tortillas per se indicates poverty (as they imply), but that those who regularly eat tortillas instead of wheat bread tend to eat a traditional low-animal-protein diet of beans, rice, *atole*, and chili instead of a modern diet of meat, fish, milk, and eggs. Indeed, analysis of the 1960 population census (in which for the first time a question was included to find out how many persons do not regularly eat meat, fish, milk, and eggs) allows us to develop in Table 3 a correlation matrix linking those regularly eating tortillas to those eating a low-protein complex of foods.

Correlations are extremely high for Mexico's thirty-two political units, including urban places of more than 2,500 persons (C and D), as well as rural places (E and F). These correlations are all above .950, only .50 away from a perfect 1.000 positive relationship.

TABLE 3 *Correlation Matrix for Persons*
(1) Regularly Eating Tortillas and
(2) Not Regularly Eating Meat, Fish, Milk, and Eggs-1960

Persons Regularly Eating Tortillas†	Persons Not Regularly Eating Meat, Fish, Milk, and Eggs†		
	Total (A)	Urban (C)	Rural (E)
Total (B)	.952	.637	.947
Urban (D)	.826	.953	.651
Rural (F)	.902	.482	.957

Source: Data from Mexico (1960: 280-281), computed with UCLA Health Sciences Computing Facility Program for Correlation.

†Computed from aggregate data for Mexico's 32 political entities.

ON APPLYING QUANTITATIVE TECHNIQUES TO QUANTITATIVE HISTORY

Although we have refuted Skidmore and Smith's query concerning two distinct dimensions, Boni and Seligson did not fully understand the question, otherwise they might have tried to develop statistical dimensions for each of the seven indicators instead of factoring out one dimension made up of two indicators. Thus, they could have tested Skidmore and Smith's hypothesis that my items should have been weighted to attach greatest importance to, for example, the isolated-population indicator.

Significantly, economist Clark W. Reynolds (1970: 46) has seen the isolated-population indicator to be the least important. With regard to my Poverty Index for Mexico as a whole, he notes that

since the D. F. [Federal District] primarily consists of Mexico City and its suburbs, the index is biased downward in this case by the inclusion of the percentage of population in communities of less than 2,500 as one of the seven characteristics. This is, perhaps, the weakest of the indicators in any case.

If we test Reynold's concern for biased data by eliminating the Federal District from total calculation for isolated living conditions, we find that in 1960 the national Poverty Index was 34.1, not 33.1 percent. This small downward bias (if it is a bias) is more than compensated for by the value of

including the capital city in the index—as Oscar Lewis (1952) has shown, poor people live no worse (and may live better) in urban slums than in rural slums. (For discussion of Mexico City's appeal to migrants, see Cornelius, 1975; and Wilkie, 1970c.) From another view, the Poverty Index for the Federal District was 8.8 in 1960 or 10.3 if the "downward bias" is omitted by reducing the number of indicators from seven to six. In a regional analysis of 32 states and territories (regardless of population size), the use of the 10.3 figure as a substitute for 8.8 in the Federal District does not change the national average, which is 34.6 regardless of which of the two figures is used.

Although full data are supplied in *The Mexican Revolution*, neither Skidmore and Smith nor Boni and Seligson attempted to recalculate the index by assigning their own values to each indicator, despite the fact that both sets of scholars have propagated their brand of quantitative analysis as *the* solution to historical analysis. Moreover, the latter team of researchers has apparently introduced a serious methodological problem into my data by eliminating the time-series element. Rather than collapsing data for all six census years into an "average" across all census years, they might have compared factor loadings for each census year or developed longitudinal factor analysis in order to help understand the dynamics of the historical process (Glass et al., 1972).

If Boni and Seligson have left the historical element out of their application of quantitative techniques, Skidmore and Smith have not recognized that weighting of indicators depends upon the intended use of the index to be developed. In certain circumstances it would be appropriate to use only one indicator by itself (as each is developed in my book); in some circumstances the Poverty Index could be the most useful (greatest weight being given to each indicator which has the largest variance, also as in my book); or in other cases the Index could be related to other criteria in order to determine what weighting is appropriate.

It is important to note, in any case, that the Poverty Index is defined only in terms of its component items; and if any items be added or subtracted, not only the results of the index but its definition would change. And, to quote from *The Mexican Revolution* (pp. xxix-xxx):

The definition of poverty . . . does not necessarily deal with individual poverty. Persons included in the index may exhibit several characteristics of poverty and yet have a relatively high income. Nevertheless, collectively speaking, the integration of the Mexican nation is greatly impeded by the persistence of a high level in characteristics of poverty. Social modernization, along with economic development, is required in order to raise general standards of living. The Poverty Index seeks to measure decrease in the collective level of social deprivation in Mexico at different historical times.

Since the Poverty Index involves at least some overlap between items, we may also say that it includes a complex of “hardcore” poverty that is most difficult to eliminate precisely because governmental (or private) programs are not usually coordinated to deal with more than one aspect of poverty at the same time. Too, the wide range in averages of the seven indicators for 1960 (from high to low: 71.5, 49.3, 37.8, 31.4, 23.4, 14.3, 3.8) suggests, for example, breadth of coverage.

ON DEFINING QUANTITATIVE HISTORY

As I have attempted to show here, factor analysis developed appropriately is of value in examining questions raised by presentation of time-series data, but it is not the only *type* of quantitative analysis possible, as Boni and Seligson would have us believe. Perhaps it is convenient to think of my analysis as utilizing classificatory (or descriptive) statistics in contrast with factor analysis which involves inductive (or predictive) statistics.⁴ On the one hand, classificatory statistics (involving especially measures of proportion and central tendency) allow us to develop multifaceted premises upon which interpretation may be based—the interpretation containing no more information than the premises taken collectively. On the other hand, inductive statistics (involving in this case factor analysis) are based upon probability: interpretation is verifiable only if all possible instances have been examined, and it contains more information than the premises or observations upon which they are based.

In noting elsewhere (Wilkie, 1973) that both kinds of statistical analysis offer important quantitative approaches to develop interpretation, I also suggest that it is a mistake to think in terms of “right” or “wrong” methodology. Rather, because there is no perfect method, we should think in terms of formulating alternative views with regard to persuading readers that the logic of one approach makes better sense than another. In any case, alternative views have merit if light is shed on the problem at hand. Seen in this view, the critique by Boni and Seligson is interesting and helpful.

It is regrettable that in their zeal to promote factor analysis, Boni and Seligson formulated their argument on the assumption that state-level data for thirty-two units and six census years are too numerous for comprehension, reducing my data to fourteen factor loadings. On the contrary, not only do we need state-level data to get at basic meaning in history, but also to carry out classificatory time-series analysis of indicators at the local level. If in analyzing microlevel data we must recognize

that the computer is enormously useful in developing statistical analysis, we must also heed L. L. Thurstone (1947: 56) who has written:

Factor analysis has its principal usefulness at the border line of science. It is naturally superseded by rational formulations in terms of the science involved. Factor analysis is useful, especially in those domains where basic and fruitful concepts are essentially lacking and where crucial experiments have been difficult to conceive. The new methods have a humble role. They enable us to make only the crudest first map of a new domain. But if we have scientific intuition and sufficient ingenuity, the rough factorial map of a new domain will enable us to proceed. . . .

In short, Thurstone reminds us that technicians who mechanically apply quantitative methods to complicated bodies of data cannot resolve problems in scholarly research. Further, it should be clear from the analyses here that those who manipulate data in the abstract may err grievously.

Although no study is ever perfect (in this case whether it be my *Mexican Revolution*, the critique by Skidmore and Smith, or the factor analysis by Boni and Seligson), one lesson emerges for scholarship: Alternative views have great value and we must recognize that they form the basis for advances in research. Thrust and parry in scholarly exchange of ideas can be enhanced if zeal is tempered by tolerance.

NOTES

1. Felix G. Boni and Mitchell A. Seligson, "Applying Quantitative Techniques to Quantitative History: Poverty and Federal Expenditure in Mexico," *LARR*, 8: 2: 105-110 (Summer 1973).
2. Thomas E. Skidmore and Peter H. Smith, "Notes on Quantitative History: Federal Expenditure and Social Change in Mexico Since 1910," *LARR*, 5: 71-85 (Spring 1970). For the response, see Wilkie (1970b).
3. For my latest view on social investment, especially as related to deceltralized expenditures, see "Recentralization: The Budgetary Dilemma in the Economic Development of Mexico, Bolivia, and Costa Rica" (1974a). This view is not discussed in the otherwise interesting methodological modification of my budgetary approach developed by Enrique A. Baloyra, "Oil Policies and Budgets in Venezuela," *LARR*, 9: 2: 28-72 (Summer 1974).
4. In making this distinction I disagree with Rummell (1970: 22-23) who sees these terms as two sides of the same coin. True, the terms are not mutually exclusive, and factor analysis in Table 1 lends itself to classification, yet the element of prediction is paramount in inductive analysis: Theoretically 75 per cent of the data for the seven poverty indicators in Table 1 can be predicted by knowing the scores of the two emergent factors. For development of classificatory statistics with regard to Latin America, see Wilkie 1974b and 1974c.

BIBLIOGRAPHY

- BALOYRA, ENRIQUE A.
1974 "Oil Policies and Budgets in Venezuela." *LARR*. 9: 2: 28-72.
- BLALOCK, HUBERT M.
1960 *Social Statistics*. Chicago.
- BONI, FELIX G., AND MITCHELL A. SELIGSON
1973 "Applying Quantitative Techniques to Quantitative History: Poverty and Federal Expenditures in Mexico." *LARR*. 8: 2: 105-110.
- BORAH, WOODROW
1954 "Race and Class in Mexico." *Pacific Historical Review*. 23: 4: 331-42.
- BORAH, WOODROW, AND SHERBURNE F. COOK
1966 "Marriage and Legitimacy in Mexican Culture: Mexico and California." *California Law Review*. 54: 2: 946-1008.
- CORNELIUS, WAYNE A., JR.
1975 *Politics and the Migrant Poor in Mexico City*. Stanford. (Forthcoming).
- DOHERTY, PETER A., AND JOHN M. BALL
1971 "Central Functions of Small Mexican Towns." *Southeastern Geographer*. 11: 1: 20-28.
- GLASS, GENE V., V. L. WILSON, AND J. M. GOTTMAN
1972 *Design and Analysis of Time-Series Experiments*. Denver (University of Colorado).
- HOLVEY, DAVID N. (ed.)
1972 *The Merck Manual of Diagnosis and Therapy*. Rothway, New Jersey.
- LEWIS, OSCAR
1952 "Urbanization Without Breakdown: A Case Study." *Scientific Monthly*. 75: 1: 31-41.
- MC GREEVEY, WILLIAM PAUL
1974 "Recent Materials and Opportunities for Quantitative Research in Latin American History: Nineteenth and Twentieth Centuries." *LARR*. 9: 2: 73-82.
- MÉXICO, DIRECCIÓN GENERAL DE ESTADÍSTICA
1940 *Sexto Censo General de Población: Resumen General*. México, D. F.
1960 *Octavo Censo General de Población: Resumen General*. México, D. F.
- REYNOLDS, CLARK W.
1970 *The Mexican Economy: Twentieth-Century Structure and Growth*. New Haven.
- RUMMELL, R. J.
1970 *Applied Factor Analysis*. Evanston, Illinois.
- SKIDMORE, THOMAS E., AND PETER H. SMITH
1970 "Notes on Quantitative History: Federal Expenditures and Social Change in Mexico Since 1910." *LARR*. 5: 1: 71-85.
- THURSTONE, L. L.
1947 *Multiple Factor Analysis*. Chicago.
- WILKIE, JAMES W.
1970a *The Mexican Revolution: Federal Expenditure and Social Change Since 1910*. Berkeley.
1970b "On Methodology and the Use of Historical Statistics." *LARR*. 5: 1: 87-91.

THE POVERTY INDEX FOR MEXICO

- 1970c "La Ciudad de México como Imán de la Población Económicamente Activa, 1930-1965." In: Bernardo García et al. eds. *Historia y Sociedad en el Mundo de Habla Española: Homenaje a José Miranda*. México, D. F. (The English-language version is included in Part I of Wilkie, 1974b).
- 1973 "Alternative Views in History: (1) Historical Statistics and (2) Oral History." In: Richard E. Greenleaf and Michael C. Meyer eds. *Research in Mexican History: Topics, Methodology, Sources, and a Practical Guide to Field Research*. Lincoln.
- 1974a "Recentralization: The Budgetary Dilemma in the Economic Development of Mexico, Bolivia, and Costa Rica." In: David T. Geithman ed. *Fiscal Policy for Industrialization and Development in Latin America*. Gainesville. (Reprinted in Part I of Wilkie, 1974c).
- 1974b *Statistics and National Policy*. Los Angeles.
- 1974c *Measuring Land Reform*. Los Angeles.