DEMOGRAPHY Volume 7, Number 2 May 1970

TWO MEASURES OF GEOGRAPHIC LOCATION AND THEIR RELATION TO INCOME

James N. Morgan
Department of Economics, and Survey Research Center of the Institute for Social Research, The University of Michigan, Ann Arbor 48106

Ismail Sirageldin
School of Public Health and Department of Economics, Johns Hopkins University, Baltimore, Maryland 21205

Abstract—Two different ways of categorizing people's geographical location are compared as to their relation to earned incomes of family heads, after adjustment for education, age, sex and race. One, the traditional code, uses the size of the city in which the family lives. The other, originally suggested by Bernard Lazerwitz, focuses on the distance to, and the size of, the central city of the nearest standard metropolitan statistical area. Both appear useful, and their joint effects are examined.

Some years ago Dr. Bernard Lazerwitz (1960) suggested that city size was not an adequate description of where a man lived and that distance from the center of the nearest metropolitan area was a useful addition to our arsenal of variables.

Since that time both measures have been prepared and used in studies done by the Survey Research Center. In studies of income it has appeared that even taking account of other variables, an association remained between where a man lived and how much he earned. It is not the purpose of this note to pursue that analysis or attempt to determine whether the location is a cause or an effect of difference in income. We propose to present some data on the separate and joint relations between these two main measures of geographic location and income.

Since there are other background factors which clearly are not affected *by* geographic location or income, but do affect income, the analysis removes their effects from income and asks whether location variables are associated with the residual income differences among individuals. In this way we can test whether income differences associated with location are more than a simplistic way of accounting for differences in the clearly exogenous demographic characteristics of those who live there.

Of course this will still not tell us whether people in some areas earn more than one would expect (given their backgrounds) because the relatively well off moved there or because the location provides better job opportunities or other advantages.

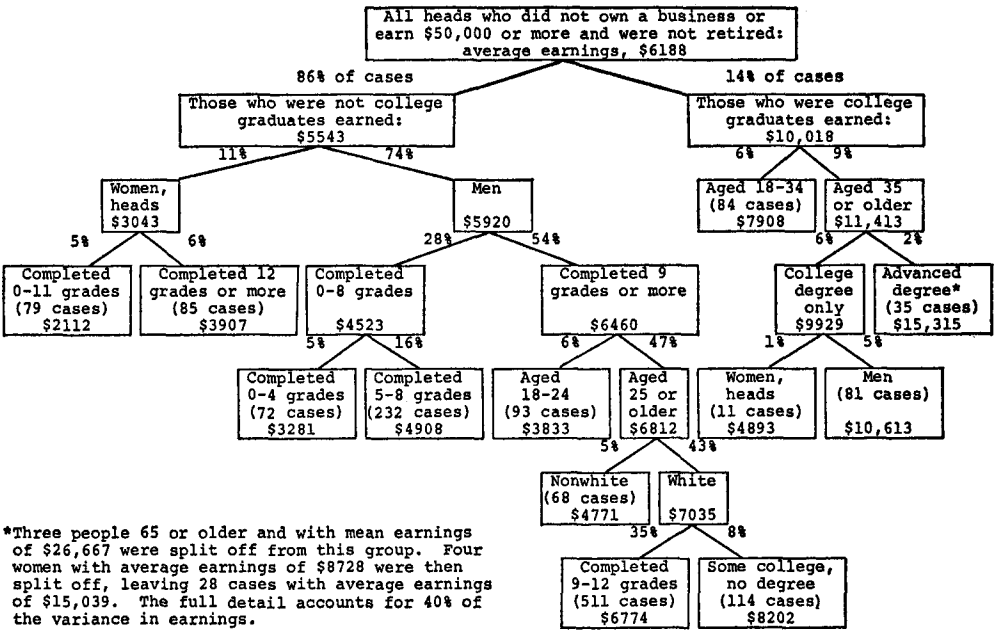
In order to keep the analysis focused on earnings, not also on differences in labor force participation of wives, we use the annual earnings of *heads* of families. Our data come from a representative national probability sample of some 2200 families interviewed in the spring of 1965. (For details of the design and findings of the original study see: Morgan, Sirageldin and Baerwaldt, 1966. The original study was financed by a grant from the Carnegie Corporation.) Since the original purpose of the study

was to investigate how people earned a living and accepted change, a great deal of attention was paid to work and earnings in the interview.

Since the factors which affect earnings do not operate additively, we used a flexible procedure to estimate an expected income for each individual family head on the basis of age, education, race and sex and then took as the variable to be associated with geographic location the residual difference between the individual's actual and his expected income. The omission of occupation from the list of background factors is not an economist's bias but reflects our desire to remove the effects only of things which could affect income and/or location but could not be affected by them. Had we removed the effects of occupation we could have been accused of eliminating an effect of location that worked through occupation—highly specialized people being forced to be in densely populated areas. Indeed this effect has been studied extensively by sociologists.

The method of estimating expected income was to use a computer algorithm which operates sequentially, using reduction in unexplained variance as its criterion, and dividing the sample into smaller and smaller subgroups by a series of dichotomous splits. (For a detailed specification of the computer program which carried out the search process and examples of its use, see Sonquist and Morgan, 1964.) No restrictive assumptions about additivity, linearity, or symmetry are imposed so that one group may be split on a different predictor than some parallel group, or using different classes of the same predictor. The subclasses of each predictor are either left in their natural order or arranged according to the mean of the dependent variable, and the computer then tries each of the K-1 possible splits: first group against the rest, first two against the rest, and so forth. Comparing the best split with each predictor, over all the predictors, it selects the best of the best, actually divides the sample, and proceeds to repeat the process with each subgroup thus generated until no further split will reduce the unexplained variance by as much as some fraction, usually .006, of the original total sum of squares. The alternative would have been some standardization procedure, the most elaborate versions of which generally amount to predicting an expected value using a least squares regression on the explanatory variable. Even if the explanatory variables were converted into a series of dichotomies (dummy variables) so that their effects did not have to be assumed to be linear, such a procedure still assumes that they are additive and symmetrical (any factor which matters, matters for the whole population equally), a highly dubious assumption. The results of the search algorithm appear in Figure 1, the end-group averages being the "expected values" from which residuals are to be analyzed. Thinking of the end-group averages as the subgroup means in a one-way analysis of variance, the between-group variance is some 40 percent of the total, so that those groups (plus two other very small subgroups which were split off) account for 40 percent of the variance in head's earned income. If we had, instead, used a regression, with dummy variables for each of the classes of age, education, race and sex, we would not have accounted for any more of the variance.

Least squares procedures are distorted by extreme cases so we excluded from this analysis those who earned $50,000 or more. And because of conceptual and measurement difficulties we also excluded those not in the labor force and those who owned a business. Our purpose, however, is not tests of significance but comparisons of the explanatory power (in reducing unexplained variance) of two measures of geographic location, and variance comparisons are much more robust under departures from nor-



*Three people 65 or older and with mean earnings of $26,667 were split off from this group. Four women with average earnings of $8728 were then split off, leaving 28 cases with average earnings of $15,039. The full detail accounts for 40% of the variance in earnings.

FIGURE 1.—Earned Income of Family Heads by Education, Age, Sex and Race, for 1465 Heads Who Earned $1–49,999 in 1964, Did Not Own a Business, and Were Still in the Labor Force in Early 1965

mality than significance tests. Any differences which account for visible amounts of variance would be significant, in any case, with a sufficiently large sample.

If we now compute for each family head the difference between his actual earnings and the mean for the subgroup of Figure 1 in which he falls, we can re-pool these residuals or "unexplained differences in earnings" and analyze them for different locations.

The most common classification of place is according to the size of the city in which the individual lives. The first column of Table 1 shows for such a classification the actual average annual earnings of the sample of family heads living in each size city, and the average deviation from what would have been expected given the information in Figure 1. The average difference could be thought of as the difference between the actual average and the average of what would have been predicted for those individuals using Figure 1. For those who may be tempted to think of the differences as explained by past mobility, it should be remembered that only *some* of those now in urban areas grew up on a farm, and they arrived in a city to compete (at some disadvantage) with those who had grown up in the city. Those who are best off grew up in the city in the first place (Lansing and Morgan, 1967).

The six locational classes account for some 5 percent of the variance of actual earnings and 6 percent of the (smaller) variance of the residuals. The relevant F-ratios would be 16 and 21, clearly significant even allowing for design losses in the sampling not taken into account in significance test procedures. For those interested in the sampling errors of the individual average deviations it must be remembered that they depend on the extent of design losses in a complex sample, and that these design losses are greater the *larger* the number of cases on which the mean is based. (The full effect of the clustering is felt with the full

TABLE 1.—Annual Earnings of Family Heads, and Average Differences from Expected Annual Earnings, by Size of Place and Residential Belt

[For family heads who were not retired in early 1965, earned between $1 and $49,999 inclusive in 1964 and did not own a business]

Size of place	All belts	Central cities: Of 12 largest SMA's	Central cities: Of other SMA's	Suburban areas: Of 12 largest SMA's	Suburban areas: Of other SMA's	Adjacent areas[b]	Outlying areas[c]
		annual earnings (dollars)					
All places	6188	6480	5712	7448	7297	5434	4597
Central cities of 12 largest standard met. areas (SMA's)	6480	6480	...	...	...	...	...
Other cities, 50,000 or more	6029	...	5650	6925	9386	...	...
Urban places, 10,000-49,999.	7058	...	6718	8850	7528	5660	5326
Other urban[a]	6925	...	...	8385	7236	6033	4965
Rural inside an SMA.	7593	...	...	...	6672	8705	4737
Rural outside an SMA	4478	...	...	...	...	4094	4133
		difference from expected earnings					
All places	0	628	- 125	1168	580	- 629	-1255
Central cities of 12 largest standard met. areas (SMA's)	628	628	...	...	...	...	...
Other cities, 50,000 or more	24	...	- 198	595	1714	...	...
Urban places, 10,000-49,999.	237	...	1195	1367	213	- 288	-1023
Other urban[a]	558	...	...	1496	842	52	- 936
Rural inside an SMA.	595	...	...	...	16	1270	- 673
Rural outside an SMA	-1526	...	...	...	...	-1581	-1482
		number of cases					
All places	1465	220	265	228	230	240	282
Central cities of 12 largest standard met. areas (SMA's)	220	220	...	...	...	...	...
Other cities, 50,000 or more	331	...	251	71	9	...	...
Urban places, 10,000-49,999.	262	...	14	84	57	38	69
Other urban[a]	294	...	...	73	125	40	56
Rural inside an SMA.	82	...	...	...	39	40	3
Rural outside an SMA	276	...	...	...	...	122	154

a-Urban places of 2,500-9,999; other urbanized areas.
b-Areas beyond the boundary of a suburban belt but within 50 miles of the central business district.
c-All territory more than 50 miles from the central business district of a central city, and not in the same county.

sample whereas a small group may contain only one representative of each cluster.) Using the approximation developed for the original study and the standard deviation of the residuals ($3120), we have the following estimated sampling errors. (See Morgan, Sirageldin and Baerwaldt, 1966, Appendix A, pp. 360–378a.) The sampling errors of the average *actual* earnings are 30 percent larger. Since the grouping by location preserves the clusters more, the upper limits may be closer to the truth.

There are, then, some real differences in earnings not attributable to differences in people's background characteristics but associated with the size of city in which they live. The main difference, however, is the lower income in rural areas where there is no city of 50,000 or more in the county.

The residential belt classes proposed by Bernard Lazerwitz were based on the notion that closeness to the center of a

Base (no. of cases)	Limit of sampling error: Lower	Limit of sampling error: Upper
25	$624	$735
100	312	378
400	156	202
900	104	148
1600	78	124

large metropolitan area, with its job opportunities (and higher living costs), and the size of that metropolitan area, might be more relevant than the population of the particular legal entity in which a man lived. A middle-sized city might be isolated or close to a much larger city. Even a rural area might be very close to a big city. The first rows of the upper panels of Table 1 show the actual and residual mean incomes by residential belt classifications. Thinking of the subgroup means as a one-way analysis of variance, the actual means account for 5 percent of the variance of actual incomes and the mean residuals for 7 percent of the variance of the residuals. The F-ratios are 16 and 22, clearly significant, and the same approximate sampling errors apply.

While the power of the belt code does not seem substantially greater than that of the old city size code, it does provide an important isolation of the suburban areas (whether they are cities or not) and a separation of the non-standard-metropolitan-areas according to their distance from a large city (the last two groups).

Let us now see how these two classifications fit together. One class is the same in both, the central cities of the twelve largest metropolitan areas; but the other city-size groups spread over two to four of the belt codes. Table 1 shows the actual and residual earnings data cross-classified by the two characteristics.

Once again, thinking of the 18 groups as providing a one-way analysis of variance, their means account for 13 percent of the variance in actual income, or 10 percent of the variance of the residuals. Since there are more groups, however, using up more degrees of freedom, the F-ratios are 12 and 9, smaller but still significant. Many of the groups are so small that no significance should be attributed to them separately, as one could tell by their estimated sampling errors.

What stands out most clearly is that the cities of 50,000 or more, excluding the central cities of the twelve largest metropolitan areas, vary depending on whether they are central cities or suburbs. Similarly with the smaller cities and towns—it depends heavily how close they are to the center of a large city (standard metropolitan area). On the other hand, city size seems to matter within some of the belt classifications, inversely in the case of the suburban areas of the twelve largest metropolitan areas, the residents of smaller suburbs earning more.

It would seem that both distinctions are needed. And until further analysis proves otherwise, we cannot reject the hypothesis that there may be something about location which affects a man's income—job opportunities, connections, even his image and reputation. This seems most likely in the case of the outlying regions, though even here the differences may only reflect natural selection (the more ambitious having moved to town). On the other hand, even if it seems a persuasive hypothesis that the suburbanites moved to the suburbs because they were wealthy, it may still be possible that wealthy neighbors also help one get ahead faster. Such speculations are beyond the capacity of our present data to test.

References

Lansing, John, and James Morgan. 1967. The effect of geographical mobility on income. Journal of Human Resources 2:449–460.

Lazerwitz, Bernard. 1960. Metropolitan community residential belts 1950 and 1956. American Sociological Review 25:245–252.

Morgan, James, Ismail Sirageldin, and Nancy Baerwaldt. 1966. Productive Americans. Ann Arbor: Survey Research Center, The University of Michigan.

Sonquist, John, and James Morgan. 1964. The Detection of Interaction Effects. Ann Arbor: Survey Research Center, The University of Michigan.