MIGRATION TO AND FROM STATE ECONOMIC AREAS IN THE INTERIOR SOUTHEAST: AN EXPERIMENT USING POOLED CROSS-SECTION DATA*

Richard J. Olsen

ORNL-NSF Environmental Program
and
University of Tennessee (on leave)

The paucity of meaningful time-series data for small geographic regions has frequently confined analysis to cross-section data at one point in time; for example, an interest in the growth of SMSAs has led to the cross-section analysis of the differences between SMSAs of various sizes. Yet we all know that variation over space is not the equivalent of variation over time. The Knoxville SMSA of 1980 may not, even under the most favorable circumstances, look like the Atlanta SMSA of 1970. The purpose of this paper is to recommend the use of pooled cross-section analysis and to illustrate its application to one dynamic component of regional growth--migration. In addition, the problem of heteroscedastic disturbances is treated through the selective application of weighted regression analysis.

THE BASIC MODEL

As a starting point, a simple single-equation model of migration has been postulated. Symbolically, the model is

(Model I)
$$\text{INMIG} = f(\text{Char. Place}, \text{Char. Persons}, u)$$
$$\text{OUTMIG} = g(\text{Char. Place}, \text{Char. Persons}, v)$$

where INMIG represents the annual rate of in-migration to a particular State Economic Area (SEA) and OUTMIG represents the annual rate of out-migration from an SEA. "Char. Place" represents measures of the characteristics of the place, such as the unemployment rate, while "Char. Persons" represents measures of the characteristics of the persons, such as median years of schooling. u and v are random variables that represent the cumulated effects of omitted explanatory variables, measurement errors in INMIG and OUTMIG, and the basic random element in human behavior.

While one might choose to group the explanatory variables in some other way, the distinction between "Char. Place" and "Char. Persons" is useful. The economic theory and past empirical studies of migration[2] suggest that a particular Char. Place measure should be positively correlated with INMIG while negatively correlated with OUTMIG, or vice versa. For example, a high unemployment rate should both raise out-migration and lower in-migration. On the other hand, a particular Char. Persons measure should be positively (or negatively) correlated with both OUTMIG and INMIG. For example, a high level of education should raise both out-migration and in-migration.

POOLED CROSS-SECTION ANALYSIS

While it appears appropriate to fit Model I separately to 1950 and 1960 census data for State Economic Areas in the Southeast, the frequent warning[3]

*Research sponsored by the National Science Foundation RANN Program under Union Carbide Corporation's contract with the U.S. Atomic Energy Commission.

that cross-sectionally estimated parameters are apt to over- or underestimate the sensitivity of a dependent variable (in this case, migration rates) to change in socioeconomic conditions over time has suggested the modification of Model I to use pooled cross-section data. The revised model is

(Model II) INMIG = f (Char. Place, Char. Persons, Pooled, u)
OUTMIG = g (Char. Place, Char. Persons, Pooled v)

where the Pooled variable is a shift factor which takes on the value 0 for observations during 1950 and the value 1 for observations during 1960.

The basic rationale of this procedure is shown in Figure 1.

Let the solid portions of lines A and B, respectively, represent the responsiveness of out-migration rates to changes in an explanatory variable X as estimated from separate sets of 1950 and 1960 cross-section data. It would then be very tempting to forecast out-migration rates over time along either of these lines (A or B) and perhaps into the dashed portion of each. This procedure, however, is incorrect since comparable time-series data would yield a fitted curve such as C. It is line C, which shows less or more response in OUTMIG to a change in X, that should be used as a guide to out-migration dynamics over time.

Since suitable time-series data are not available for small geographic areas, our MODEL II, which utilizes pooled cross-section data[4], may be thought of as a pseudo time-series analysis. It is an attempt to minimize the over-or under-responsiveness of parameters estimated from individual cross sections while effectively utilizing the abundant supply of explanatory variables present in each cross section to explain signficant shifts over time.

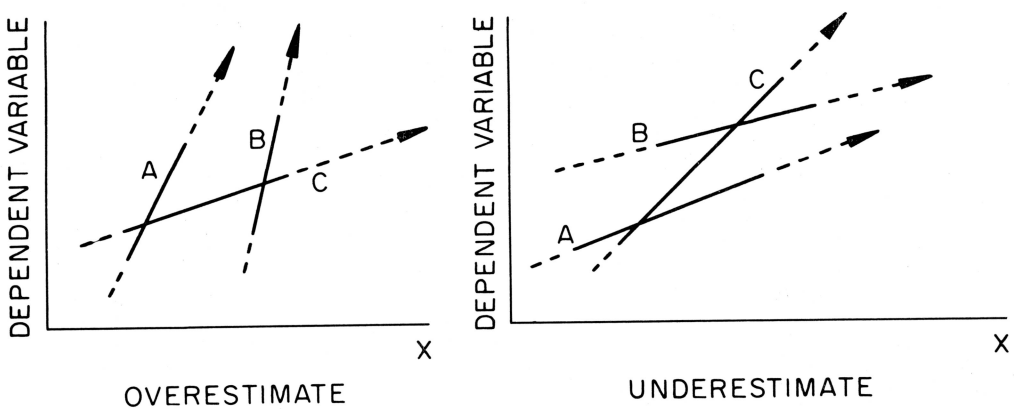
WEIGHTED REGRESSION ANALYSIS.

Early in our attempts to fit and test both Models I and II it became apparent the assumption of the Ordinary Least Squares (OLS) Regression Model that u has a constant variance was not being met. The residual variation about the fitted regression line varied systematically with State Economic Area size measured in terms of population. A moment's thought will confirm that this is just what one should expect when using sample data which varies in sample size between observations. When the sample size varies from a low of 2, 300 to a high of 64, 000, as it did in our study area, [5] one can reasonably expect the smaller samples to contain an intrinsically larger variability simply because of sampling error, which is known to be reduced in size as the sample size increases.

In these circumstances, the use of the Weighted Least Squares (WLS) Model is more appropriate. As a consequence of using the more appropriate WLS instead of the inappropriate OLS, (1) parameter estimates from the use of our sample of data are different, (2) estimates of the standard errors of the papameters are unbiased, and (3) as a result of (2), statistical tests of parameter significance are valid. All of the empirical results reported below were derived by using the WLS procedure where each observation of only those variables subject to sampling error (i.e., resulting from the 20 or 25 percent sample census enumeration procedure) has been weighted by the $\sqrt{P_i / \sum_i P_i}$ where P_i is the population of the i th SEA. [6]

DATA FOR STATE ECONOMIC AREAS IN THE INTERIOR SOUTHEAST

Our analysis has been confined to cross-section data from the 1950 and 1960 Census of Population for each of approximately 67 State Economic Areas in the interior Southeastern United States (Figure 2). The Study area was


DEPENDENT VARIABLE
A
B
C
X
OVERESTIMATE
DEPENDENT VARIABLE
C
B
A
X
UNDERESTIMATE

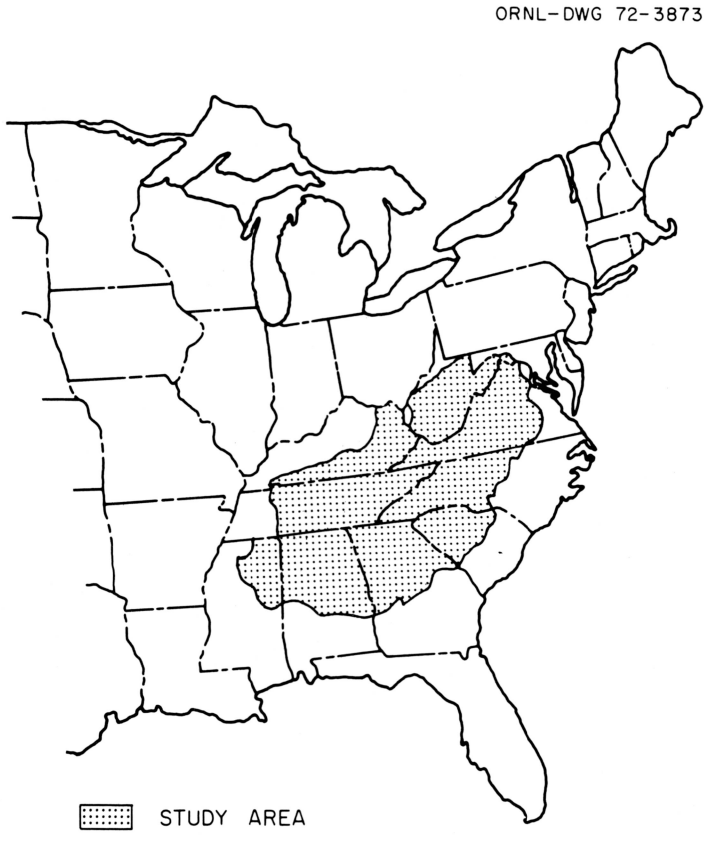


FIGURE 2

The Interior Southeast as Defined for this Study

geographically restricted in this manner to control by exclusion for what may be rather unique migratory conditions present in the South Atlantic and Gulf Coastal areas, as well as areas adjacent to major river systems (the Mississippi and Ohio Rivers in particular). It was our hope that holding some factors constant in this fashion would permit the use of a less complicated set of a relationships to explain geographic mobility while maintaining a satisfactory amount of explanatory power.

Measurements on some of the relevant variables were available directly from the 1950 and 1960 Census of Population; other variables had to be constructed from the census data directly available. Both sets are summarized in Table 1. One should note that: (1) observations on all of these variables were available for each of the State Economic Areas in the Southeast for each of the census years, 1950 and 1960 (except where noted)[7]; (2) much of this information was originally from a 20 percent or 25 percent sample within the census enumeration procedures and, therefore, is subject to sampling error; (3) the data reflect census definitions and may be different from the theoretically best definitions; and (4) employment data and, therefore, our male skills index data were as of place of residence, not place of work.

From the point of view of applying pooled cross-section analysis to migration; the need for comparably defined data from each cross-section limited our analysis to eight age groups; beginning with age 20, and to total migrants rather than males and females separately. We could have studied whites and nonwhites separately but chose not to do so at this stage of our analysis. Furthermore, since annual migration rates were available from the 1950 Census of Population while five-year rates were available for 1960 [8], we were forced to crudely annualize the 1960 five-year rates. While we chose to annualize by dividing by five, a smaller number (e.g., 3) might have been a more appropriate divisor since persons leaving an SEA and then returning within the five-year period are really migrants but do not get counted as such. While the annualization procedure used does have some effect on the regression results, the specific annualization process that should be used is still to be determined. [9]

POOLED CROSS-SECTION RESULTS BY AGE

Best results to date include as explanatory variables: (1) the difference between local and national male unemployment rates, (2) population density, (3) Armed Forces personnel as a percent of the population, (4) college students as a percent of the population, (5) an index of male skills, (6) nonwhites as a percent of the population, (7) adjusted local minus national labor participation rates, and (8) the pooled variable. The reversal of signs between in-migration and out-migration analyses implies that the unemployment and population density variables are measures of "characteristics of place." All remaining variables appear to be measures of "characteristics of persons."

The coefficients of the unemployment variable are negative for in-migration and positive for out-migration as one would expect a priori. The size and significance of its coefficient indicate that the unemployment variable is much more important to out-migration than to in-migration. In both cases the size of the coefficient declines with age, implying that younger persons respond much more strongly to a given difference between local and national employment conditions.

The other "characteristics of place" measure, population density, is more significant to in-migration than to out-migration. The signs of the coefficients indicate that a higher population density raises in-migration for all age groups while reducing out-migration of the highly mobile 20 to 24 year old group. What is probably being captured here is the rural to urban movement of the past decades.

Table I

Data Used In Interior Southeast Migration Analysis

(1) Total Population
(2) Males 14 Years Old and Over
(3) Male Inmates of Institutions
* (4) Males in Civilian Labor Force
* (5) Unemployed Males
* (6) Armed Forces
* (7) Median Years of School Completed--Persons 25 Years Old and Over
* (8) Median Family Income (for 1949 and 1958, respectively)
* (9) Percent of State's Population Born in State
**(10) Percent of Persons 25 Years Old and Over With Some College Training
(11) Land Area
(12) Percent of Population Classified as Nonwhite
(13) Percent of Population Residing in Urban Areas
*(14) College Students--for 1950, Number Enrolled Aged 18-29
--for 1960, Number Enrolled Aged 18-34
(15) Mean January Temperature
*(16) Male Employment by Occupation
**(17) Male Skills Index--See Text
*(18) 1950 Annual In-migration Rate--By Age
*(19) 1950 Annual Out-migration Rate-By Age
*(20) 1960 Five Year In-migration Rate--By Age
*(21) 1960 Five Year Out-migration Rate--By Age

**(22) 1960 Annual In-migration Rate = $\frac{(20)}{5}$

**(23) 1960 Annual Out-migration Rate = $\frac{(21)}{5}$

**(24) Male Unemployment Rate = $\frac{(5)}{(4)} \times 100$

**(25) Male Labor Participation Rate = $\frac{(4)}{(2)-(3)} \times 100$

(26) Population Density = $\frac{(1)}{(11)}$

---

* Based on a 20% (1950) or 25% (1960) sample within the census enumeration procedure.

** Indirectly based on a 20% or 25% sample.

Table II

Estimated In-Migration Functions for the Interior Southeast
Using Pooled 1950 and 1960 Cross-Section Data

Dependent Variable (Age Group)	Constant	Local Minus National Male Unemployment Rate	Population Density (00's/sq.mi.)	Percent Armed Forces Are of Population	Percent College Students Are of Population	Index of Male Skills in 1000's	Percent of Population That is Non-White	Adjusted Local Minus National Male Labor Part. Rate	Pooled	Overall F Value (Degrees of Freedom)
IN-MIGRATION RATE (20-24)	.340	-.172 (2.978)***	.047 (5.664)***	.780 (12.47)***	.929 (6.77)***	.3 (2.533)**	.00058 (.659)	-.0414 (.835)	-.377 (15.246)***	94.3 (8,119)
INMIG (25-29)	.333	-.075 (1.346)	.031 (3.927)***	.399 (6.598)***	.503 (3.799)***	.45 (3.773)***	.00013 (.150)	-.0450 (.941)	-.321 (13.443)***	50.6 (8,119)
INMIG (30-34)	.229	-.044 (.966)	.020 (3.033)***	.273 (5.570)***	.231 (2.152)**	.40 (4.136)***	.00093 (1.349)	-.0463 (1.192)	-.229 (11.796)***	36.1 (8,119)
INMIG (35-39)	.181	-.077 (2.249)**	.015 (3.039)***	.199 (5.366)***	.115 (1.407)	.34 (4.588)***	.00046 (.878)	-.0374 (1.271)	-.172 (11.706)***	36.1 (8,119)
INMIG (40-44)	.149	-.064 (2.293)**	.010 (2.427)**	.147 (4.892)***		.33 (6.798)***	.00048 (1.179)	-.0267 (1.267)	-.153 (12.970)***	41.5 (7,120)
INMIG (45-54)	.115	-.054 (2.513)**	.007 (2.155)**	.080 (3.418)***		.26 (6.789)***	.00053 (1.679)*	-.0318 (1.944)*	-.123 (13.448)***	43.1 (7,120)
INMIG (55-64)	.088	-.025 (1.275)	.008 (2.671)***	.076 (3.509)***		.21 (6.067)***	.00046 (1.559)	-.0448 (2.961)***	-.107 (12.618)***	38.0 (7,120)
INMIG (65 + over)	.062	-.018 (.890)	.012 (4.309)***	.045 (2.127)**		.24 (6.803)***	.00055 (1.888)*	-.0286 (1.917)*	-.097 (11.553)***	37.2 (7,120)

NOTE: Student t values are in parentheses below coefficients.

* Significantly different from zero with probability .90.
** Significantly different from zero with probability .95.
*** Significantly different from zero with probability .99.

Table III

Estimated Out-Migration Functions for the Interior Southeast
Using Pooled 1950 and 1960 Cross-Section Data

OUT-MIGRATION Age Group	Constant	Local Minus National Male Unemployment Rate	Population Density (00's/sq.mi.)	Percent Armed Forces Are of Population	Percent College Students Are of Population	Index of Male Skills in 1000's	Percent of Population That is Non-White	Adjusted Local Minus National Male Labor Part. Rate	Pooled	Overall F Value (Degrees of Freedom)
20-24	.394	.706 (8.804)***	-.022 (1.897)*	.056 (.638)	-.164 (.861)	1.26 (7.316)***	.00215 (1.767)*	-.376 (5.458)***	-.316 (9.173)***	29.6 (8,119)
25-29	.252	.262 (4.478)***	.0002 (.025)	.673 (10.595)***	.533 (3.829)***	.60 (4.757)***	.00120 (1.361)	-.244 (4.849)***	-.230 (9.144)***	39.6 (8,119)
30-34	.194	.154 (4.315)***	.005 (.963)	.291 (7.531)***	.232 (2.739)***	.55 (7.266)***	.00062 (1.148)	-.120 (3.913)***	-.195 (12.723)***	44.8 (8,119)
35-39	.164	.089 (3.191)***	.006 (1.614)	.175 (5.791)***	.078 (1.174)	.44 (7.409)***	.00017 (.393)	-.067 (2.792)***	-.143 (12.032)***	35.1 (8,119)
40-44	.144	.093 (3.822)***	.004 (1.049)	.116 (4.398)***		.38 (8.917)***	.00017 (.468)	-.055 (2.984)***	-.132 (12.756)***	35.1 (7,120)
45-54	.116	.073 (3.644)***	.0025 (.891)	.072 (3.290)***		.31 (8.780)***	.00014 (.480)	-.028 (2.481)**	-.115 (13.508)***	35.8 (7,120)
55-64	.101	.064 (3.770)***	.002 (.762)	.044 (2.368)**		.27 (9.068)***	.00008 (.333)	-.018 (1.382)	-.108 (15.007)***	40.5 (7,120)
65 + over	.083	.039 (2.536)**	.003 (1.426)	.019 (1.105)		.23 (8.509)***	.00028 (1.247)	-.018 (1.555)	-.089 (13.539)***	35.4 (7,120)

NOTE: Student t values are in parentheses below coefficients.

* Significantly different from zero with probability .90.

** Significantly different from zero with probability .95.

*** Significantly different from zero with probability .99.

The geographic mobility of at least two groups, Armed Forces personnel and college students, is not likely to be in response to economic conditions. Since the relative presence of these two groups varies between State Economic Areas and over time, we have attempted to standardize for their presence by including an Armed Forces variable and a college students variable.[10] Both variables enter both in- and out-migration functions with sizable and significant positive coefficients, implying that the presence of college students or Armed Forces personnel substantially raises both in and out-migration rates. In addition, the influence of these variables as measured by the size and significance of their coefficients is practically invariant to the addition or deletion of other variables.

Since there has been so much discussion of how education raises mobility, we have examined three alternative measures of education: (1) median years of school completed by those 25 years old and over, (2) percent of the population 25 years old and over with some college, and (3) an index of male skills. The Male Skills Index gave the best results in the sense that all of its coefficients were sizable and significantly different from zero. The index is an attempt to measure the variation in average skill level of employed males and is defined to be

$$\text{Male Skills Index}_{jt} = \frac{\sum_{i=1}^{10} W_{it} E_{ijt}}{\sum_{i=1}^{10} E_{ijt}}$$

where E_{ijt} represents male employment in occupation i in State Economic Area j at time t, and W_{it} is the national median real income of employed male civilians in occupation i at time t. Perhaps because the sample was confined to the Southeast, the Index of Male Skills seems to have its greatest impact on out-migration.

While differences in propensities to migrate are frequently found in the raw data on white and nonwhite movers, after standardization for education, age, etc., most of the differences tend to disappear. However, the percent nonwhite variable was included to see if any residual differences could be found. Generally, no racial differences in propensity to move could be found; the magnitude of all coefficients was very small and generally not significantly different from zero.

From an early stage of the analysis, the Male Labor Participation Rate (MLPR) has been included as an additional measure of labor market tightness, especially of chronically depressed conditions. But the high correlation of MLPR with the unemployment and skills variables made its impact uncertain. the belief that the MLPR variable had something important to contribute, an adjusted MLPR variable has been defined[12] in an attempt to purge the MLPR variable of the influence of unemployment and skills variables already included in the analysis. The adjusted MLPR variable has been kept in our analysis because its inclusion (1) substantially improves and clarifies the interpretation of the unemployment variable in both in-and out-migration relationships and (2) contributes directly and significantly to the explanation of out-migration. Its interpretation as a "characteristic of persons" variable, however, is not completely clear.

The Pooled variable was included to capture any time trend that re-

mained after including all previously mentioned variables. Somewhat surprisingly, all coefficients were significantly different from zero and negative. The negative sign implies that after adjusting for changes in all of the variables considered, there has been a decline over time in the propensity to migrate. The size of the decline, however, has been small.

CONCLUSIONS

While the pooled cross-section procedure illustrated in this paper is not without problems (especially those caused by changing definitions over time), there may be much truth in the forecast of L. R. Klein[4] that the spread of econometric methods to regions that have a weak data base is likely to depend on the use of time sequences of cross-sections to fill in some irreparable time-series gaps.

REFERENCES

1. This assumption may not be adequate where labor market conditions and migration rates are determined simultaneously through the dynamic interaction of labor supply and demand. A later phase of our research is using a simultaneous equations model to test the present assumption.

2. For example, Samuel Bowles, "Migration as Investment: Empirical Tests of the Human Investment Approach to Geographic Mobility," Review of Economics and Statistics, Vol. LII, No. 4 (Nov. 1970); Michael J. Greenwood, "An Analysis of the Determinants of Geographic Labor Mobility in the United States," Review of Economics and Statistics, Vol. LI, No. 2 (May 1969) and "Logged Response in the Decision to Migrate," Journal of Regional Science, Vol. 10, No. 3 (1970); John B. Lansing and Eva Mueller, The Geographic Mobility of Labor (Ann Arbor: Institute for Social Research, 1967); Ira S. Lowry, Migration and Metropolitan Growth: Two Analytical Models (San Francisco: Chandler Publishing Co., 1966); H. S. Parnes, "Labor Force Participation and Labor Mobility, "A Review of Industrial Relations Research, Vol. I (1970); Gian S. Sahota, "An Economic Analysis of Internal Migration in Brazil," Journal of Political Economy, Vol. 76, No. 2 (March/April 1968); and Charles E. Trott, "An Analysis of Out-Migration," paper presented at the annual meetings of the American Statistical Association, Fort Collins, Colorado, August 25, 1971.

3. For example, Jacob Mincer, "Labor-Force Participation and Unemployment: A Review of Recent Evidence," in Robert A. Gordon and Margaret S. Gordon (editors), Prosperity and Unemployment, (New York: J. Wiley and Sons, (1966).

4. L. R. Klein makes the following comment in a recent attempt to chart the future development of econometrics: "Since econometric methods are now being taken up in several countries that have a weak data base, it is likely that there should be more reliance on supplementary samples of cross-section data. These can, in principle, be obtained to fill in some irreparable time-series gaps. Research on pooling will be needed in these cases, and also research on drawing inferences from time sequences of cross-section should be further developed," L. R. Klein, "Whither Econometrics?" J. Amer. Stat. Assoc. 66 (334), 415-21 (June 1971).

5. The smallest State Economic Area in our migration sample contained 9180 households in 1960, while the largest contained 255, 983. The 25 percent sample of households used for most 1960 Census of Population responses should have resulted in a sampling of approximately 2, 300 households from the smallest SEA and 64, 000 from the largest.

6. See R. J. Wonnacott and T. H. Wonnacott, Econometrics (New York, Wiley and Sons, 1970) for a more detailed discussion of the relationships between OLS and WLS.

7. U. S. Bureau of the Census, U. S. Census of Population: 1960, Subject Reports, Mobility for States and State Economic Areas, Final Report PC(2) 2-B. U. S. Census of Population: 1960, Subject Reports, State Economic Areas, Final Report PC(3)-1A. U. S. Census of Population: 1960, Subject Reports, State of Birth, Final Report PC(2)-2A. U. S. Census of Population: 1960, Volume I, Characteristics of the Population. U. S. Census of Population: 1950, Population Mobility-States and State Economic Areas, Special Report P-E #4B. U. S. Census of Population: 1950, Volume II, Characteristics of the Population.

8. For the 1950 Census persons (those in 20% of all households) were asked where they resided one year ago (1949) while for the 1960 Census persons (those in 25% of all households) were asked where they resided five years ago (1955). These responses were then (1) expanded to estimates for the population as a whole, (2) aggregated to State Economic Areas, and (3) divided by the relevant 1950 and 1960 SEA population to form annual 1950 migration rates and five-year 1960 migration rates.

9. Generally, there was a gradual change in the size and signficance of each coefficient as we moved from the results based on data annualized by 5 to those based on data annualized by 3. The absolute size of the change for a given variable was largest for the early age groups and smallest for the older age groups. There was no noticeable change in signs from those found in Tables II and III. The absolute size of the coefficients, as well as their statistical significance, increased uniformly for the first four explanatory variables; (1) local minus national male unemployment rate, (2) population density, (3) percent Armed Forces, and (4) percent college. The coefficients of the "Pooled" variable, as well as their statistical significance, declined noticeably. The overall explantory power of the relationships, as measured by the F value, remained remarkably stable.

10. The college students variable has only been included in the first four age groups.

11. This definition is similar to that proposed by Robert L. Raimon and Validmar Stoikov, "The Quality of the Labor Force," Industrial and Labor Relations Review, April 1967, pp. 391-413.

12. Adjusted MLPR is defined to be the residuals from $Y = 0.368 - \underset{(14.3)}{1.797}X_1 + 0.597X_2 + \underset{(3.0)}{0.126}X_3$, where $\underset{(2.3)}{}$ Y = Local minus national MLPR, X_1 = Local minus National Male Unemployment Rate, X_2 = Male Skills Index, and X_3 = Pooled (1950 = 0; 1960 = 1).