ESTIMATION OF CHILDREN IN POVERTY IN NEW JERSEY

Thomas Vietorisz, New School for Social Research
and
Robert Mier, University of Illinois at Chicago Circle

I. INTRODUCTION

Public Law 93-389, The Elementary and Secondary Education Act, mandates the annual distribution of close to $2 billion of federal funds to local school districts.[1] The primary basis for the distribution of these funds is the number of children in poverty (CIP), aged 5-17 inclusive, in each school district at the time of the last decennial census (1970). This basis is slightly modified by the inclusion of children in families receiving Aid to Families with Dependent Children (AFDC) if the family income exceeds the Government's most recent official poverty line after receipt of such welfare funds. The principal objective of this study, one of several commissioned by the Department of Health, Education and Welfare (HEW) to test the feasibility of updating CIP counts, has been the testing and determination of an improved method for estimating the number of children in poverty at the sub-state level based on relationships between labor market attachment, poverty, and specific quality of life indicators uncovered by earlier work of the Research Center for Economic Planning. This earlier study established strong statistical associations between labor market attachment, poverty, and a series of indicators measuring deterioration of the quality of life and the incidence of such social pathologies as family break-up, ill physical and mental health, crime, and poor housing.[2] The final report inferred a causal link between family incomes earned in the labor market and the prevalence of poverty, even though such causality, for well-known fundamental reasons, can never be conclusively established on the basis of statistical observations alone.[3]

II. METHODS OF PROCEDURE

The specific objectives of the study reported here included:

* Establishing regression specifications that had a high degree of explanatory power in describing the county-by-county distribution of children in poverty, as counted by the 1970 census.
* Estimating the regression parameters for the State of New Jersey.
* Applying the estimated parameters for prediction of the county-by-county distribution of children in poverty in New Jersey in 1974 and 1976.
* Comparing the predictions with the Khan-Miller-BEA estimates for New Jersey counties for 1974 and 1976.[4]

The original intent in undertaking this study was to utilize, among the findings of the earlier work, the relationship between labor market attachment and poverty, while using information on quality of life and social pathologies for background purposes only. Constraints on data availability, coupled with the tight time schedule of the project, forced a modification of this strategy, shifting the main emphasis to quality of life and social pathology indicators as predictors of poverty.

The critical data set in regard to this decision was the Continuous Work History Sample of the United States Social Security Administration. The methodology proposed for quantification of the labor market attachment of principal wage earners in households turned largely on the use of the 1-percent sample annual Employee-Employer ("Ee-Er") file. Immediately following initiation of the project, it was learned that the Social Security Administration had temporarily embargoed the release of all information in this file to new users. This embargo applied both directly, in the form of suspension of tape file sales, and indirectly, by denying permission for existing users to share their information with new users. The embargo had its origin in administrative relationships between the Internal Revenue Service and the Social Security Administration. In mid-March, at the time of initiation of the study, this embargo was expected to remain in force at least until May, with no assurance that far longer delays were excluded.

Since the Social Security tapes could not become available in time for this project, labor market attachment could be quantified on a current basis only by means of data from the Unemployment Insurance and Employment Service systems of the state. As it turned out, these data were also severely limited in availability. Even in states where such data are more readily available, such as New York, unemployment and related data are poor predictors of long-term poverty. In the earlier study of the Research Center for Economic Planning, referred to above, the fraction of poverty directly explained by unemployment variation turned out to be of the order of one percent, and was totally without statistical significance.[5]

An indicator related to unemployment that does, in fact, predict poverty far more effectively is "subemployment," a concept that broadly reflects both a lack of work opportunities and a prevalence of jobs that offer less than family support level wages. The subemployment concept combines officially defined unemployment, the long-term discouraged jobless who have stopped looking for work, involuntary part-time workers who are unable to find full-time jobs, and workers earning substandard wages (by one of several possible definitions).[6] If subemployment data could be obtained on a current basis, they would be considerably more useful than unemployment for predicting poverty. Unfortunately, the necessary components are available only once a year, and then only on a nationwide basis, from the Current Population Survey.

This is why the present study intended to rely on the Social Security file tapes for defining labor market attachment. The subemployment concept embodies the effects of "secondary" labor markets in generating poverty, but is not avail-

able on a current basis by small areas. The Social Security sample tapes are available on a reasonably current basis and include information --industry codes, occupation codes, and wage levels--that allow construction of a proxy for the relative prevalence of primary and secondary jobs in an area. Unemployment data, even in combination with some related wage or benefit exhaustion information, cannot yield such a proxy.

Given this severe data problem, it has become necessary to shift emphasis to another aspect of the nexus between labor market attachment, poverty, and the quality of life uncovered by the earlier study of the Research Center for Economic Planning. Fortunately, this earlier study disclosed that just as good a profile of poverty can be assembled from quality of life indicators as from labor market attachment indicators, and--as it has turned out--the components of quality of life indicators are available on a far more timely basis for sub-state areas than are components of labor market attachment indicators. Guided by the findings compiled and extensive experience with poverty data gained in this earlier study, a major effort was thus made to collect the best possible indicators for the conditions that are known to be fundamentally involved in the existence and reproduction of poverty.

Among the many conditions known to be involved in both being caused by and causing poverty, the principal ones are physical and mental health; educational levels; drug and crime problems; and housing conditions. The data collection effort was therefore centered on obtaining a good representation of such variables in the data set on which the regressions and projections were based, and are presented in Table 1.

TABLE 1

Explanatory Variables: Data Availability at the County Level in New Jersey

Series No.	Item	Years
1	Population (census and estimates)	1970-75
2	Deaths under 1 year	1965-75
3	Deaths under 28 days	1965-75
4	Maternal deaths	1965-75
5	Stillbirths	1965-75
6	Illegitimate births	1965-75
7	Homicides	1970-75
8	Suicides	1964-75
9	Syphilis cases, all stages	1965-75
10	Syphilis cases, Pri-sec	1965-75
11	Syphilis, early latent	1965-75
12	Gonorrhea cases	1965-75
13	No. of AFDC children	1965-76
14	AFDC Assistance payments	1965-76
15	Pub. Schl enrollment, K-12	1965-76
16	P.S. enrollment K-12, black	1962-76
17	P.S. enrollments K-12, hispanic	1962-76
18	Paroch. Schl. enrollment 1-8	1972-75
19	Unemployment insurance, all weeks claimed	1968-76
20	Unemployment rate, seas. adjusted	1970-76

III. MODEL SPECIFICATION AND TESTING

The objective in the statistical correlation and regression specifications was to identify variables that met the following requirements:

* They were highly correlated with the 1970 CIP count.
* They also showed a high correlation even when lagged by several years. This is a very desirable characteristic, since current data often become available only with a long time lag; therefore, if an explanatory variable is to serve for project purposes, it should predict well for some years ahead.
* It is plausible to assume that they were causally interrelated with poverty. To the extent that a variable satisfies this requirement, the statistical relationship can be expected to hold broadly across the United States as a whole, rather than being tied to the characteristics of the state of New Jersey, for which this case study was being undertaken.
* They were available within the data set for a sufficient number of years prior to the 1970 Census to allow the confirmation of lagged correlations referred to above.

Of the variables included in the set, the following correlate approximately at the .95 level, with time lags of at least two years, with the 1970 CIP count:

Variable Number	Description
2	Deaths under 1 year
3	Deaths under 28 days
4	Maternal deaths
5	Stillbirths
6	Illegitimate births
9	Syphilis, all stages
10	Syphilis, primary and secondary
11	Syphilis, early latent
12	Gonorrhea
13	Number of AFDC children
14	AFDC assistance payments
16	Public school enrollment, black

All of these correlations, including the last one, were statistically significant at the .05 level.

By contrast, the unemployment rate correlated with the CIP count only with a coefficient of .5 and was statistically not significant.

Using the methods of Daniel and Wood, the following three linear regression estimates were found appropriate in terms of functional form and model efficiency.[7] Standard deviations are shown beneath the respective coefficient in parentheses; single, double, and triple stars denote coefficients that are statistically significant at the .05, .01, and .001 levels respectively.

$$CIP = -1165.5 + \underset{(5.0)***}{70.2\ DEATH < 28}; \quad R^2 = .91;$$

$$F = 195.43.$$

Explanatory variable lagged three years; coefficient Statistically significant at the .001 level.[8]

$$CIP = 584.2 + \underset{(10.0)**}{36.1\ STILLB} + \underset{(3.24)***}{19.9\ SYPH};$$

$$R^2 = .96; \text{ overall } F = 205.95$$

All explanatory variables lagged three years; coefficients significant at the .01 level.

$$CIP = -3842.1 - 222.7\ DEATH{<}28 - 16.3\ STILLB$$
$$(66.1)^{**} \qquad (20.4)$$
$$+ 212.6\ DEATH{<}1 + .9\ UNEMP - 12.5\ ILLEG$$
$$(60.3)^{**} \qquad (.52) \qquad (5.75)^{*}$$
$$+ 1581.1\ MATDEATH + 39.0\ SYPH$$
$$(403.0)^{***} \qquad (12.9)^{**}$$
$$R^2 = .98;\ \text{overall } F = 115.0$$

All variables except unemployment lagged three years; unemployment lagged two years. Not all coefficients statistically significant.

As the first regression result shows, a single explanatory variable, that of deaths under 28 days, explains 91 percent pf the variation in CIP counts; two variables, stillbirths and total syphilis cases, esplain 96 percent; and seven variables explain 98 percent. In the regression with seven variables, the coefficients of some of the variables are seen to be negative. This is at first sight counter-intuitive, since deaths under 28 days, stillbirths, and illegitimate births are individually each positively correlated with CIP. The phenomenon becomes readily understandable, however, when it is noted that all the variables that individually are highly correlated with CIP are in turn highly intercorrelated among themselves. This is illustrated in Appendix A showing the results of principle components analysis of the data set.

The principle components analysis yielded two primary factors. Factor 1 can be interpreted as a quality-of-life factor. It has very high loadings of the eleven health and welfare variables: four load above .98 and all but one above .90, with the remaining one--maternal deaths--still showing a high loading at .83. The two unemployment variables, conversely, show low loadings on this factor. The second factor is identified as an unemployment factor. Both unemployment variables load at .73 on the second factor, while the health and welfare variables show very low loadings, rarely reaching .20 at the most.

The interpretation of these results is entirely in line with the anticipated causal relationship between quality of life and social pathology indicators and poverty. It is particularly noteworthy that data taken from birth and death certificates offer such a powerful explanatory device. This leads directly to an important consideration flowing from the current study. Since data from birth and death certificates are available on a current basis, and since they are such excellent explanatory variables for CIP, it might be useful to undertake an effort to obtain uniform country-wide tabulations of birth and death certificates by area of residence on a timely basis. With a moderate effort, data from such certificates could be tabulated directly on a county-by-county or even a school district-by-school district basis, and could be used as CIP correlates at a local area level, hopefully down to the local school district.

IV. ESTIMATES OF CHILDREN IN POVERTY

Estimates of children in poverty (CIP) for the year 1974 and 1976 have been prepared by projecting the cross-sectional regression results obtained for the year 1970. In each case, the 1970 CIP count by county in New Jersey was used as the base, and positive or negative changes were added as calculated from the regression equation, substituting into this equation the corresponding changes in the explanatory variables.

Two different regression specifications were used for projection purposes. Both used the same three explanatory variables--deaths under 28 days, weeks of unemployment insurance claimed, and the count of AFDC children. In specification A, the variables were averaged over several years, while in specification B, they referred to single years. Details will be found in Appendix

Appendix B shows the count-by-count projection results. The most notable feature of the projections is the contrast between the very high R^2's of the cross-sectional regression equations and the changeability of the coefficients over time. For reference purposes Appendix B includes the 1974 and 1976 Khan-Miller-BEA estimates.

A) State Total Counts

As can be seen in the row of state totals, these totals are highly sensitive to the time period to which the explanatory variables refer. Specification A, whose explanatory variables are averaged over a period lagged from 1 to 5 years (in the case unemployment, 1 to 2 years), leads to a projection of a state total less than half of that projected by specification B. The latter has explanatory variables that are lagged only 2 or 3 years.

This sensitivity has its origin in the time trands of some of the explanatory relationships. The number of deaths under 28 days has, for example, decreased substantially in recent years, and these decreases do not necessarily follow a linear pattern. Changes in the AFDC caseload, likewise, cannot be taken to change in a smooth, regular manner over time.

It is, therefore, concluded that the projections must be normalized in order to be useful. The very strong cross-sectional relationship between CIP by county and the explanatory variables can be used to predict each county's share in the state-wide CIP count, but the rapid changes of coefficients over time preclude the direct longitudinal use of cross-sectional coefficients. For the present projections, the state SIE count was used as a normalizer; in future work, normalization must be tied to the total CIP count for the U.S. as a whole, since only this total is available as a yearly basis from the CPS.

B) County-by-county Proportions

Comparison of the Khan-Miller-BEA estimates with the projections obtained shows that the projections may overstate the county-by-county changes in the CIP count since 1970. These changes have their root in shifting spatial patterns of poverty. The Khan-Miller-BEA estimates do not allow for any autonomous county-by-county

poverty shifts; they simply infer the consequences of an overall population and income increase on a spatially fixed 1970 CIP pattern.[9] Therefore, it is not possible to judge whether the differences between the Khan-Miller-BEA county-by-county proportions of CIP and the projected proportions obtained by regression methods have their origin in:

- understatement of spatial shifts by the Khan-Miller-BEA method;
- overstatement of spatial shifts by the regression method; or
- a combination of both

It is, therefore, concluded that the quality-of-life variables should be used as a group in future work, in the form of factor scores calculated for individual counties. This will merge the information contained in these explanatory variables and will greatly reduce the effect of random variations on county-by-county projections.

V. CONCLUSIONS

The conclusions flowing from the work under this contract can be summarized as follows:

* A broad data set of quality-of-life and unemployment variables has been identified that correlates strongly with CIP and is available for projections by individual counties on a reasonably current basis.
* Regressions specified on this data set in a number of different ways, using lagged or time-averaged explanatory variables, give excellent cross-sectioned explanations of the 1970 county-by-county CIP count.
* These regressions give statistically significant coefficients, at the .95 level or better, for many of the quality-of-life variables; the coefficients of the unemployment variables are, however, generally less significant.
* A factor analysis of the explanatory variables yields a quality-of-life factor with very high loadings of eleven health and welfare variables, and an unemployment factor with high loadings of the two unemployment variables.
* County-by-county projections based on the cross-sectional regression results indicate that the regression coefficients change over time. Therefore, the cross-sectional regressions should be normalized to a state or national total.
* County-by-county projections may also show instabilities owing to random variations in individual explanatory variables, especially for the smaller counties. Merging sets of variables by the device of using factor scores should improve stability in this regard.

It is, therefore, recommended that:

* The high explanatory power of the cross-sectional regressions should be confirmed, using the same data set as applied to a different state.
* The regressions should be run with variables expressed as proportions of state-wide totals, and merged into quality-of-life and unemployment factors.
* The explanatory power of the unemployment factor should be improved by including further variables, using the Social Security sample tapes if and when available.
* The regressions should be sharpened by distinguishing at least urban and rural counties in a large data set.
* If the results of future tests in one or more other states are promising, a national data set should be tested, consisting of all 3,000 counties. Within such a data set, different types of urban areas and different regions of the country should be distinguished, and the absolute CIP counts should be normalized to the U.S. total, as well as the total of all metropolitan areas.

FOOTNOTES

1. This research was sponsored under contract from the Office of the Assistant Secretary for Education (Policy Development), Department of Health, Education, and Welfare to the Research Center for Economic Planning, New York, N.Y. Assistance in statistical computation and analysis has been provided by David Less.

2. Thomas Vietorisz, "Earned Family Incomes and the Urban Crisis", a research report submitted to the Center for the Study of Metropolitan Problems, National Institute for Mental Health, January, 1976; see also Robert Mier, Thomas Vietorisz, and Jean-Ellen Giblin, "Indicators of Labor Market Functioning and Urban Social Distress", Social Economy of Cities, ed. Gary Gappert and Harold M. Rose (Beverly Hills, Sage Publications, 1975), pp. 361-394.

3. For example, David Layzer, "Is There Any Real Evidence that I.Q. Tests are Heritable?", Scientific American, July, 1975, pp. 126-128. In any regards, a "culture-of-poverty" thesis such as Banfield's would suggest a similar model for estimating children in poverty. Edward Banfield, Unheavenly City Revisited, (Boston, Little, Brown, 1974), pp. 52-76.

4. Herman P. Miller and Abdul Khan, "Methodology for Estimating the Number of Children in Poverty for States and Counties", Business Uses of Small-Area Statistics and Education's Needs and Methods for Estimating Low-Income Population (U.S. Department of Commerce, Bureau of the Census, June 1976), pp. 40-46

5. Mier, et. al.

6. Thomas Vietorisz, Robert Mier, and Jean-Ellen Giblin, "Subemployment; Exclusion and Inadequacy Indexes", Monthly Labor Review, May 1975 pp. 3-12.

7. Cuthbert Daniel and Fred S. Wood, Fitting Equations to Data (New York, Wiley-Interscience, 1971).

(Footnotes Continued)

8. The regressions equations were also specified with the CIP counts normalized as proportions to total school enrollments, and the explanatory variables normalized as proportions to population. These regressions equations yielded no improvement over those discussed in the text, either in terms of R^2s or in terms of standard deviations of the coefficients.

9. Miller and Khan

APPENDIX A

PRINCIPLE COMPONENTS RESULTS

Vbl. No.	Vbl. Name	Rotated Factor Pattern Factor 1	Factor 2
13	AFDC Children	0.98132	0.09468
14	AFDC Payments	0.98249	0.06832
20	Unemployment rate	-0.13816	0.73272
19	Unemployment insurance	0.35064	0.73508
2	Death < 1	0.93673	0.20010
3	Death < 28	0.94196	0.15960
4	Maternal death	0.83416	-0.08902
9	Syphillis All	0.98331	0.09303
10	Syphillis 1,2	0.07183	0.00095
11	Syphillis early	0.95971	0.12614
12	Gonorrhea	0.95456	-0.07979
5	Stillborn	0.90628	0.20548
6	Illegal births	0.98163	0.12196

Orthogonal Transformation Matrix

	1	2
1	0.99349	0.11392
2	-0.11392	0.99349

APPENDIX B

ESTIMATES OF CHILDREN IN POVERTY

County	Estimate for 1975 (SIE total)	1970 CIP (census)	1974 BEA	1976 BEA	1974 RCEP(A)	1976 RCEP(A)	1974 RCEP(B)	1976 RCEP(B)
		#	#	#	#	#	#	#
Atlantic	9409	5703	6268	6956	5583	5612	11512	11659
Bagen	3880	7382	7231	8461	2975	1695	10587	9391
Burlington	6111	7176	6545	7911	3567	3733	10013	10289
Camden	26869	12445	13293	15381	14105	17912	27787	32001
Conmay	2425	1477	1558	1689	1357	1530	2640	2898
Cumberland	16402	4051	4597	5249	3632	4039	7650	8157
Essex	37151	36793	39685	40691	23057	21240	61966	60427
Gloucester	4074	3796	4264	4662	2819	2124	6699	6104
Hudson	20176	19723	21208	24595	11162	12896	31750	33836
Houstendon	1649	909	1250	1266	888	1043	1543	1709
Mason	9700	7088	7241	8168	5959	5672	13279	13349
Middlesex	9700	7353	7597	7984	7633	6293	19558	14416
Monmouth	9797	9228	10551	11343	7671	6324	17377	16102
Morris	1067	3224	3536	3524	751	526	3978	3968
Ocean	8051	4153	6280	5763	4807	4800	9045	9263
Passaic	18818	11287	11653	12261	11910	10530	24279	22686
Salem	2231	2227	2343	2843	1338	1363	3273	3318
Somerset	1940	3027	2274	2276	1325	986	3163	2799
Sussex	2231	1294	1697	2001	1372	1297	2426	2393
Union	6984	7850	8299	9030	3941	4412	12004	12851
Weber	1455	1232	1300	1505	838	930	1809	1878
TOTAL	194,000	155,690	169,172	183,505	116,688	114,957	278,382	279,496

NOTES: (A) $CIP_{74,76} = CIP_{70} + 33.4 \Delta DEATH < 28_{Ave} + 0.341 \Delta UNEMP_{Ave} + 0.59 \Delta AFDC_{Ave}$

(B) $CIP_{74,76} = CIP_{70} + 32.8 \Delta DEATH < 28 + 0.541 \Delta UNEMP + 0.63 \Delta AFDC$

(C) Average of 1974 and 1976 A-type estimated percentages applied to 1975 SIE state total.

APPENDIX C

ESTIMATING METHODS FOR PROJECTIONS

Estimating Equation A

$CIP_{1970} = 435.43 + 33.42\ DEATH < 28^{*}_{Ave\ 65-69} + 0.34\ UNEMP_{Ave\ 68-69} + 0.60\ AFDC^{**}_{Ave\ 65-69}$ $R^2 = .973$

T - Statistic	(3.51)	(0.96)	(6.93)
Standard error	(9.52)	(0.35)	(0.09)

*Significant at .05 level
**Significant at .01 level

F - statistic for the entire equation is 202.49. It is significant at better than the .01 level.

$DEATH < 28_{Ave\ 65-69}$ = Average Annual Number of Infant Deaths occuring between birth and 28 days in the period 1965-1969.

$UNEMP_{Ave\ 68-69}$ = Average Monthly Insured Unemployed in the period 1968-1969.

$AFDC_{1967}$ = Average Monthly Number of AFDC Children Assisted in the period 1965-1969.

In difference, or estimatory, form this equation is:

$$CIP_{1974} = CIP_{1970} + 33.42\ (DEATH < 28_{Ave\ 69-73} - DEATH < 28_{Ave\ 65-69}) + 0.34\ (UNEMP_{Ave\ 72-73} - UNEMP_{Ave\ 68-69}) + 0.60\ (AFDC_{Ave\ 69-73} - AFDC_{Ave\ 65-69})$$

$$CIP_{1976} = CIP_{1970} + 33.42\ (DEATH < 28_{Ave\ 71-75} - DEATH < 28_{Ave\ 65-69}) + 0.34 - (UNEMP_{Ave\ 74-75} - UNEMP_{Ave\ 68-69}) + 0.60\ (AFDC_{Ave\ 71-75} - AFDC_{Ave\ 65-69})$$

Estimating Equation B

$CIP_{1970} = 423.6 + 32.85\ DEATH < 28^{*}_{1967} + 0.55\ UNEMP_{1968} + 0.63\ AFDC^{**}_{1967}$ $R^2 = .971$

T Statistic	(3.37)	(1.43)	(6.28)
Standard Error	(9.76)	(0.38)	(0.10)

*Significant at .05 level
** Significant at .01 level

F Statistic for the entire equation is 188.08. It is significant at better than the .01 level.

$DEATH < 28_{1967}$ = Infant deaths occuring between birth and 28 days in 1967.

$UNEMP_{1968}$ = Average Monthly Insured Unemployed in 1968.

$AFDC_{1967}$ = Average Monthly Number of AFDC Children Assisted in 1967.

In difference of estimating form, this equation is:

$$CIP_{1974} = CIP_{1970} + 32.85\ (DEATH < 28_{1971} - DEATH < 28_{1967}) + 0.55\ (UNEMP_{1972} - UNEMP_{1968}) + 0.63\ (AFDC_{1971} - AFDC_{1967})$$

$$CIP_{1967} = CIP_{1970} + 32.85\ DEATH < 28_{1973} - DEATH < 28_{1967}) + 0.55\ (UNEMP_{1974} - UNEMP_{1968}) + 0.63\ (AFDC_{1973} - AFDC_{1967})$$