DOCUMENT RESUME

ED 097 403 UD 014 592

AUTHOR Kadane, Joseph B.; And Others
TITLE An Econometric Model for Estimating IQ Scores and Environmental Influences on the Pattern of IQ Scores Over Time.
INSTITUTION American Statistical Association, Washington, D.C.
SPONS AGENCY National Science Foundation, Washington, D.C.; Office of Education (DHEW), Washington, D.C.
PUB DATE 73
NOTE 10p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Elementary School Students; *Environmental Influences; *Intelligence Level; Kindergarten Children; *Longitudinal Studies; *Models; Multiple Regression Analysis; Predictive Validity; *School Role; Secondary School Students; Statistical Analysis; Urban Schools
IDENTIFIERS Pennsylvania; Pittsburgh

ABSTRACT

This paper offers a preliminary analysis of the effects of a semi-segregated school system on the IQ's of its students. The basic data consist of IQ scores for fourth, sixth, and eighth grades and associated environmental data obtained from their school records. A statistical model is developed to analyze longitudinal data when both process error and measurement error must be accounted for. IQ tests are used in this paper as convenient measures of a certain kind of performance thought to be important for success in schools and certain kinds of jobs. Most of the environmental variables included in the model can be construed to measure the nature or degree of contact with mainstream culture. The data were collected in the summer of 1971 from the cumulative school records of all students who had just finished the ninth grade in the Pittsburgh public school system. The time period examined is nine years between 1962 and 1970, during which time a proportion of the group passed from kindergarten to eighth grade in the Pittsburgh system. IQ tests were administered during this period to children in kindergarten, fourth, sixth, and eighth grades. The tests administered were the Detroit (kindergarten), Kuhlmann-Anderson (fourth grade), Otis Beta FM (sixth grade), and Otis Lennon (eighth grade). (Author/JM)

BEST COPY AVAILABLE

AN ECONOMETRIC MODEL FOR ESTIMATING IQ SCORES AND ENVIRONMENTAL INFLUENCES ON THE PATTERN OF IQ SCORES OVER TIME*

Joseph B. Kadane, Carnegie-Mellon University
Timothy W. McGuire, Carnegie-Mellon University
Peggy R. Sanday, University of Pennsylvania
Richard Staelin, Carnegie-Mellon University

1. INTRODUCTION

We offer in this paper a preliminary analysis of the effects of a semi-segregated school system on the IQ's of its students. We offer it with educational policy objectives in mind. Our basic data consist of IQ scores for a panel of children at kindergarten, fourth, sixth, and eighth grades and associated environmental data obtained from their school records. We developed a statistical model to analyze longitudinal data when both process error and measurement error must be accounted for. Our statistical model can be used on longitudinal data with other measures than IQ.

We are aware of confusion about just what IQ is, or, put another way, whether IQ is anything but what an IQ test measures. While we use IQ tests in this paper, we use them as convenient measures of a certain kind of performance thought to be important for success in schools and certain kinds of jobs. Sanday (1972 a,b,c; 1973) gives a critique of IQ tests. She says that "the content of test items is often related to experience and learning which only middle and upper class children would be likely to be exposed to" (Sanday 1972 a: 420). This suggests that the nature and degree of contact with mainstream culture would have an impact on IQ scores. We interpret our results with this theory in mind. Most of the environmental variables included in our model can be construed to measure the nature or degree of contact with mainstream culture.

2. STATISTICAL MODEL

In structuring our model, we quickly found that we had to distinguish two different phenomena, measurement error (different measures of IQ of the same person on successive days) and process error (individual variability from our notion of how IQ's develop and change over time).

We begin with measurement error. Let Z_i^j be the i-th student's test IQ score at grade j ($j=1$ for kindergarten, 2 for 4th grade, 3 for 6th grade, 4 for 8th grade), and X_i^j be the i-th student's true (but unobservable) IQ score at grade j. Then

$$Z_i^j = X_i^j + u_i^j , \quad (1)$$

where u_i^j is the measurement error. We made the standard assumptions for u_i^j: i.e., $E(u_i^j)=0$, $E(u_i^j)^2=\tau^j$, $E(u_i^j u_i^{j'}) = E(u_i^j u_{i'}^j) = 0$, u_i^j distributed multivariate normal. In other words the measurement error for each student has a variance τ^j, which causes the test scores to differ from the true score but is uncorrelated with any other test score and is independent of the student's true score. (This concept is usually referred to as the standard error of measurement, and $\tau^j = (15)^2(1-r) \approx 25$, where r is the reliability.) Equation (1) is rather firmly rooted in our notion of what measurement error is. Notice that stopping here gives a model with more parameters than data.

The next step in our model specification is to state how we think the "true IQ's", X_i^j, change over time in response to the environment and changes in it. This is done in the following equations:

$$(2) \quad \begin{aligned} X_i^j &\sim N(\underline{W}_i^j \beta^j, \sigma^j), & j=1, \\ X_i^j &\sim N(X_i^{j-1} + \underline{W}_i^j \beta^j, \sigma^j), & j=2,3,4, \end{aligned}$$

where $\underline{W}_i^j$ is a vector of demographic and environmental variables such as race, sex, SES of peers, etc. (discussed in section 4) and β^j is a vector of weights. In other words, the student's true score centers around the previous true IQ (except at kindergarten) modified by the effects of demographic and environmental factors.

What we mean by the above is that at kindergarten (before the test) we have no hard information about the child's true IQ score. Thus we express our beliefs in the form of a distribution (normal) with a mean (based on the demographics) and a variance. For the other years we also have opinions on the child's true IQ scores. These center around his previous true unobserved score plus the effect of contacts with the environment since our last estimate.

Reprinted from the 1973 Social Statistics Section Proceedings of the American Statistical Association©

Equations (2) are our priors about the true but unobservable IQ scores for each student. These priors involve parameters (technically called hyperparameters) which can be estimated (see section 3) from the observed data, namely the four IQ scores and the vector of demographic and environmental variables.

We are not completely diffuse about our knowledge of some of the parameters of the prior. In particular, we believe that the variance σ^1 is quite large (i.e., about 200) since we have little relevant information about the child before he or she enters the school system. The variances σ^2, σ^3, and σ^4, however, should be much smaller (i.e., somewhere near 20-30), since we have a) at least one observed test score and b) measures on a number of variables which might influence changes from the previous true score. The model as stated acknowledges our uncertainty about the unobservable true IQ score by defining X_i^j as a random variable which is completely described (in probability density terms) only after the $\underline{\beta}$'s and $\underline{\sigma}$'s are known.

The parameter space for the model specified by (1) and (2) can be divided into two parts, the "true" IQ's, which will be written $\underline{X}$, and the structural parameters $\underline{\beta}$, $\underline{\sigma}$, and $\underline{\tau}$, and it will be denoted by

$$\underline{\theta} = (\underline{\beta}, \underline{\sigma}, \underline{\tau}).$$

Suppose that our prior on $\underline{\theta}$ is $f(\underline{\theta})$. Some has been said about this prior. However, for the argument below, f will be left unspecified. The joint density of all the observations and parameters is

$$(3) \quad f(\underline{X},\underline{\theta},\underline{Z}) = f(\underline{Z}|\underline{X},\underline{\theta})\, f(\underline{X}|\underline{\theta}) f(\underline{\theta}).$$

Therefore, using Bayes theorem, the posterior distribution of the parameters given the data is

$$f(\underline{X},\underline{\theta}|\underline{Z}) = \frac{f(\underline{Z}|\underline{X},\underline{\theta}) f(\underline{X}|\underline{\theta}) f(\underline{\theta})}{\int f(\underline{Z}|\underline{X},\underline{\theta}) f(\underline{X}|\underline{\theta}) f(\underline{\theta})\, d\underline{\theta}\, d\underline{X}}$$

$$\propto f(\underline{Z}|\underline{X},\underline{\theta}) f(\underline{X}|\underline{\theta}) f(\underline{\theta}).$$

In one sense, the posterior distribution (4) gives our new opinion, after taking the data into account, about all the phenomena under study. However, for our data this distribution is almost impossibly multidimensional, since $\underline{X}$ has almost 7000 components and $\underline{\theta}$ has over 50 components. Therefore, we chose to consider the marginal posterior distribution of $\underline{\theta}$ given $\underline{Z}$:

$$f(\underline{\theta}|\underline{Z}) = \int f(\underline{X},\underline{\theta}|\underline{Z})\, d\underline{X}$$

$$\propto \int f(\underline{Z}|\underline{X},\underline{\theta}) f(\underline{X}|\underline{\theta}) f(\underline{\theta})\, d\underline{X}$$

$$(5) \quad = f(\underline{\theta}) \int f(\underline{Z}|\underline{X},\underline{\theta}) f(\underline{X}|\underline{\theta})\, d\underline{X}$$

$$= f(\underline{\theta})\, f(\underline{Z}|\underline{\theta}).$$

Since $f(\underline{Z}|\underline{X},\underline{\theta})$ is assumed to be normal and linear in the mean in $\underline{X}$, and since $f(\underline{X}|\underline{\theta})$ is again normal, the above integral is normal and can be computed by inspection as follows.

Let

$$\epsilon_i^1 = X_i^1 - \underline{W}_i^1 \underline{\beta}^1,$$

$$\epsilon_i^j = X_i^j - X_i^{j-1} - \underline{W}_i^j \underline{\beta}^j, \qquad j=2,3,4.$$

Then the ϵ's are normal and independent with zero mean and variances $(\sigma^1, \sigma^2, \sigma^3, \sigma^4)$. Substituting into (1) and transforming,

$$(6) \quad \begin{aligned} Z_i^1 &= \underline{W}_i^1 \underline{\beta}^1 + u_i^1 + \epsilon_i^1, \\ \Delta Z_i^j &= Z_i^j - Z_i^{j-1} = \underline{W}_i^j \underline{\beta}^j + u_i^j - u_i^{j-1} + \epsilon_i^j, \qquad j=2,3,4; \end{aligned}$$

i.e., changes in observed IQ scores are functions of environmental factors.

Let

$$\underline{Z}_i^\Delta = (Z_i^1,\ Z_i^2 - Z_i^1,\ Z_i^3 - Z_i^2,\ Z_i^4 - Z_i^3)$$

and

$$\underline{m}^i = (\underline{W}_i^1 \underline{\beta}^1,\ \underline{W}_i^2 \underline{\beta}^2,\ \underline{W}_i^3 \underline{\beta}^3,\ \underline{W}_i^4 \underline{\beta}^4).$$

Then

$$(7) \qquad \underline{Z}_i^\Delta | \underline{\theta} \sim \mathcal{N}(\underline{m}^i, \underline{V}),$$

where

$$\underline{V} = \begin{bmatrix} \tau^1 + \sigma^1 & -\tau^1 & 0 & 0 \\ -\tau^1 & \tau^2 + \tau^1 + \sigma^2 & -\tau^2 & 0 \\ 0 & -\tau^2 & \tau^3 + \tau^2 + \sigma^3 & -\tau^3 \\ 0 & 0 & -\tau^3 & \tau^4 + \tau^3 + \sigma^4 \end{bmatrix}.$$

Notice that (7) performs the integration in (5) painlessly as the convolution of two normal distributions.

The combination of measurement error and process error in our model makes it a special case of models involving unobservable variables (see Goldberger (1973), Griliches (1973), and Joreskog (1970), and the references cited there). One distinction between their approach and ours is that we can examine the posterior distribution of the unobserved variables. One way of doing that in this case is to calculate

$$f(\underline{X}|\underline{Z}) = \int f(\underline{X},\underline{\theta}|\underline{Z})d\underline{\theta}$$

and to note that the posterior on the students' intelligences $\underline{X}$ will be approximately independent over students. This possibility, although interesting, is not pursued further here.

3. ESTIMATION

Estimation of the parameter space $\underline{\theta} = (\underline{\beta},\sigma,\tau)$ is based on the fact that the system of equations (6) (or (7)) is in the form of four seemingly unrelated regression equations. Were the covariance matrix $\underline{V}$ completely general, it would be exactly in the form studied by Zellner (1962). However the zeros in the upper-right and lower-left corners of $\underline{V}$ pose a problem not explicitly considered there.

Zellner proposes that each equation be estimated separately using ordinary least squares, yielding consistent but asymptotically inefficient estimates of the β's. The residuals from these regressions can then be used to obtain consistent estimates of the covariance matrix. Finally, use of the estimate of the covariance matrix thus obtained in a generalized least-squares framework yields consistent and asymptotically efficient estimates of the β's.

Use of this method on the system (6) will also yield consistent and asymptotically efficient estimates of the β's because the estimate of $\underline{V}$ will be consistent under the model (6). Alternatively, using the first round residuals to estimate the diagonal elements and elements just off the diagonal of $\underline{V}$, and zero to estimate the other elements of $\underline{V}$, is also consistent; hence the resultant β's from the application of generalized least squares also are consistent and asymptotically efficient. This second alternative seems to us more in keeping with the model, so we estimated it that way.

All the parameters of the system (6) are identified except for σ^4 and τ^4. However the sum $\sigma^4 + \tau^4$ is identified. (See Kadane(1972) for an explanation of identified functions on the parameter space.)

4. IMPLEMENTATION OF THE MODEL

The data were collected in the summer of 1971 from the cumulative school records of all students who had just finished the ninth grade in the Pittsburgh public school system. The time period examined is nine years between 1962 and 1970, during which time a proportion of the group passed from kindergarten to eighth grade in the Pittsburgh system. IQ tests were administered during this period to children in kindergarten, fourth, sixth, and eighth grades.

The tests administered were the Detroit (kindergarten), Kuhlmann-Anderson (fourth grade), Otis Beta FM (sixth grade) and Otis Lennon (eighth grade).

3762 children took at least one IQ test, and 2,067 children took all four tests. This latter group excludes children assigned to special education classes for the slow learner, since such children are not given these IQ tests after they are assigned to such classes. It also excludes children who moved into or out of the school system. These students may have been exposed to different cultural influences than those who were enrolled in the school system for the full nine years. The applicability of our results to children who have moved and slow learners is a topic for future research. We used only the records of the 1713 children which are complete on all the independent variables.

Table 1 (see top of next page) lists the variables used in W_1^1, with their means and standard deviations.

The Sex variable is scored 2 for female, 1 for male. SES is measured by the Hollingshead (1957) Two Factor Index of Social Position, which assigns each individual an index value according to occupation and education (with occupation weighted more heavily). Hollingshead (1957:10) suggests that social class position be determined on the basis of index score as follows.

Table 2

Relation of Social Class to SES Index as Suggested by Hollingshead

Social Class		Range of SES Scores
Upper	I	11-17
	II	18-27
	III	28-43
	IV	44-60
Lower	V	61-77

Notice that the higher the Social Class, the lower the SES index. SES of peers is the average SES for all kindergarteners in the school of the child. Because Pittsburgh in 1962 had neighborhood kindergartens, we take this variable to represent the SES of the neighborhood the child was raised in. The head of household variable is scored -1 if both parents are in the house, +1 otherwise. Race of student is scored 0 for white and 1 for non-white. Non-whites in Pittsburgh are almost entirely

Table 1. Variables Used in Kindergarten Equation

	Variable Name	Mean	Standard Deviation
1.	Constant	1.0	0.0
2.	Sex	1.51	0.50
3.	Number of Siblings	2.74	2.06
4.	SES of parents	53.02	14.32
5.	SES of peers	53.24	9.23
6.	Head of Household	-.75	.65
7.	Race of Student	.380	.486
8.	% Black in School	35.4	39.0
9.	Race ·% Black	29.3	40.8
10.	(% Black in School)^2	2775.	3786.
11.	Race ·(% Black)^2	2521.	3823.

black. % black is the school average of the race variable, multiplied by 100. Thus the proportion of non-whites in our sample (38%) approximates the sample average proportion of non-whites in the school (35.4%). If each school had the same proportion of blacks, the standard deviation of percent black in school would be zero. In a completely segregated system which has a school average of 35.4% blacks, the standard deviation would be $\sqrt{(35.4)\cdot(64.6)} = 47.8$. Thus the actual standard deviation of 39.0 is evidence of a high degree of segregation.

Table 3 gives a cross-tabulation of SES with race for the entire group of 2,067 students.

Table 3

Cross-Tabulation of
Students by SES and Race

Index Score of Social Position

Race	11-37	38-57	58-77	TOTAL
Blacks Number	30	182	663	875
% of Blacks	3	21	76	100
Whites Number	268	502	422	1192
% of Whites	23	42	35	100

Table 3 shows that there is a relationship between race and SES, with blacks having higher SES, and hence lower class, than whites. In the group of 1713 children chosen for intensive analysis, the correlation between race and SES of parents is .41.

The remaining three variables are higher order terms and interactions of the previous ones.

Table 4 lists the variables used in W_1^2, with their means and standard deviations (see top of next page for Table 4).

Variables 1, 2, 3, 4, 6, and 7 are the same as in kindergarten. However because school-mates need not be the same as in kindergarten, variables 5, 8, 9, 10 and 11 are not the same. Variable 3 is actually the number of siblings when the child entered kindergarten, and for that reason does not change in 4th grade. Variable 12 is the student-faculty ratio of the school of the child, averaged over the five years from kindergarten to 4th grade. Variable 14 is the change in the percent of blacks in the school from kindergarten to fourth grade, and variable 15 is variable 14 times variable 7.

The variables used in the sixth and eighth grade equations were the same as in the fourth grades, and are given below in Table 5.

Table 4. Variables Used in 4th Grade IQ Equation

	Variable Name	Mean	Standard Deviation
1.	Constant	1.0	0.0
2.	Sex	1.51	0.50
3.	Number of Siblings	2.74	2.06
4.	SES of parents	53.02	14.32
5.	SES of peers	53.33	9.37
6.	Head of Household Missing	-.75	.65
7.	Race of Student	.380	.486
8.	% Black in School	35.5	38.4
9.	Race ·% Black	29.5	40.5
10.	(% Black)^2	2735.	3711.
11.	Race ·(% Black)^2	2514.	3759.
12.	Student-faculty ratio, K to 4	32.5	2.96
13.	# changes of school, K to 4	1.89	1.03
14.	Δ % Black, K to 4	.26	18.2
15.	Race ·(Δ% Black, K to 4)	.21	14.4

Table 5. Variables used in 6th and 8th Grade Equations

	Variable Name	6th Mean	6th Standard Deviation	8th Mean	8th Standard Deviation
1.	Constant	1.0	0.0	1.0	0.0
2.	Sex	1.51	.50	1.51	.50
3.	Number of Siblings	2.74	2.06	2.74	2.06
4.	SES of parents	53.02	14.32	53.02	14.32
5.	SES of peers	53.21	9.30	53.30	9.09
6.	Head of Household Missing	-.75	.65	-.75	.65
7.	Race of Student	.380	.486	.380	.486
8.	% Black in School	35.9	39.8	34.9	35.7
9.	Race ·% Black	29.8	41.6	26.7	38.9
10.	(% Black)^2	2878.	3913.	2490.	3617.
11.	Race ·(% Black)^2	2619.	3942.	2220.	3688.
12.	Student-faculty ratio	29.1	3.69	23.4	4.41
13.	# changes of school	.13	.37	.81	.58
14.	Δ% Black	.45	13.4	-.69	17.64
15.	Race ·Δ% Black	-.49	11.2	-1.97	13.87

5. RESULTS

$\hat{V}$, the estimate of the covariance matrix V, is a consistent estimate of V under this model. We obtained

$$\hat{V} = \begin{bmatrix} 236.2 & -145.7 & 0 & 0 \\ -145.7 & 168.4 & -27.3 & 0 \\ 0 & -27.3 & 65.6 & -21.8 \\ 0 & 0 & -21.8 & 61.2 \end{bmatrix}.$$

From $\hat{V}$, the following consistent estimates can be derived:

$$\hat{\tau}^1 = 145.7 \qquad \hat{\sigma}^1 = 70.5$$
$$\hat{\tau}^2 = 27.3 \qquad \hat{\sigma}^2 = 0$$
$$\hat{\tau}^3 = 21.8 \qquad \hat{\sigma}^3 = 16.5$$
$$\sigma^4 + \tau^4 = 39.4$$

The first thing that strikes one about these estimates is that $\hat{\sigma}^2$ is surely too low, that $\hat{\sigma}^1$ is probably too low, and that both are consequences of $\hat{\tau}^1$ being too high. Were $\hat{\tau}^1$ close to the anticipated value of 25 or so, $\hat{\sigma}^1$ would be close to 200, and $\hat{\sigma}^2$ would be about 120, which is high but not unreasonable. These results reveal a weakness in our model. Quite possibly there is non-independence between ε^1 and ε^2, u^1 and u^2, or between the ε's and u's. We leave these possibilities as topics for future research. Any fuller parametrization of V involving zeros where we have put them will lead to $\hat{V}$ being a consistent estimate for V, and hence our estimates of the regression coefficients would still be consistent and asymptotically efficient. As a result, despite this weakness in the model we think the regression coefficients given in Table 6 may be of some interest.

Caution should be exercised in the interpretation of the race and percent-blacks-in-school variables because of the presence of higher-order terms in different ways below.

A few things stand out from Table 6. First, the results on the sex variable indicate that women have an advantage through 4th grade which is lost by the time 8th grade is completed. This is in accord with literature that women mature physically more rapidly than men, although the faster pace of loss between 6th and 8th, compared to 4th to 6th, indicates the possibility of negative relative conditioning of women around intellectual matters.

To help the reader understand which coefficients are important and which are not, we calculate below in Tables 7 to 10 the predicted IQ of a white student

Table 6. Regression Estimates and Estimated Deviations

Variable	Equation 1		Equation 2		Equation 3		Equation 4	
	Est.	s.d.	Est.	s.d	Est.	s.d.	Est.	s.d.
1. Constant	128.4	2.84	-10.24	3.40	10.96	2.28	2.02	2.18
2. Sex	2.44	.745	1.28	.630	-1.621	.394	-1.80	.380
3. # siblings	-.474	.188	-.168	.159	-.126	.0996	-.0302	.0964
4. SES parents	-.208	.033	.0696	.0280	-.0282	.0178	-.0439	.0168
5. SES peers	-.148	.061	-.124	.0479	-.0832	.0324	-.0192	.0318
6. Head of household missing	-.747	.578	.229	.490	.256	.306	-.181	.295
7. Race	-7.04	3.02	9.68	2.77	-1.84	1.58	.430	1.53
8. % Black	-.00078	.0700	.0987	.0588	.0112	.0389	-.0623	.0391
9. Race · % Black	-.00391	.118	-.264	.10[illegible]	.0363	.0668	-.0807	.0653
10. (% Black)^2	-.00009	.00090	-.00108	.00080	-.00005	.0005[illegible]	.00041	.00059
11. Race · (% Black)^2	-.00003	.00116	.00239	.00102	-.00037	.[illegible]0067	.00074	.00072
12. St/fac ratio			.0970	.0752	-.00895	.0559	.0481	.0486
13. # school changes			.0593	.207	-.540	.467	.109	.334
14. Δ(% Black)			.0259	.0206	-.0235	.0256	-.0234	.0177
15. Race · [Δ(% Black)]			-.0158	.0260	.0350	.0298	.0439	.0229

with all exogenous variables at the mean for a white student and the predicted IQ for a black student at the mean for blacks. An alternative method of analysis would be to compute significance levels for the estimates. While this latter method of analysis is popular, it is also misleading (Kadane (1973)). For this reason we choose to weight most heavily the analysis of Tables 7 to 10.

Table 7. Effect of Kindergarten Regression Coefficients on Mean Black and Mean White Student

	Variable	Regression Coef.	Black Mean	White Mean	Contribution to Black Score	Contribution to White Score	Δ
1.	Constant	128.4	1	1	128.4	128.4	0
2.	Se	2.44	1.54	1.49	3.76	3.64	-.12
3.	# Siblings	-.474	3.29	2.412	-1.56	-1.14	.42
4.	SES parents	-.208	60.47	48.547	-12.58	-10.10	2.48
5.	SES peers K	-.148	60.25	49.046	-8.92	-7.26	1.66
6.	Head of Household Missing	-.747	-.663	-.813	.50	.61	.11
7.	Race	-7.04	1	0	-7.04	0	
8.	% Black	-.00078	77.17	9.806	-.06	-.01	
9.	Race ·% Black	-.00391	77.17	0.0	-.30	0	8.14
10.	(% Black)^2	-.00009	6634.06	402.77	-.59	-.04	
11.	Race ·(% Black)^2	-.00003	6634.06	0.0	-.20	0	
	Total				101.41	114.10	12.69

Table 8. Effects of Regression Coefficients of Change from Kindergarten to 4th Grade on Mean Black and Mean White Student

	Variable	Regression Coef.	Black Mean	White Mean	Contribution to Black Score	Contribution to White Score	Δ
1.	Constant	-10.238	1	1	-10.24	-10.24	0
2.	Sex	1.28	1.54	1.49	1.97	1.91	-.06
3.	#Siblings	-.168	3.29	2.412	-.55	-.41	.14
4.	SES parents	-.0696	60.47	48.547	-4.21	-3.38	.83
5.	SES peers 4	-.124	60.30	49.143	-7.48	-6.09	1.39
6.	Head of Household Missing	.229	-.663	-.813	-.15	-.19	-.04
7.	Race	9.68	1	0	9.68	0	
8.	% Black K-4	.0987	77.64	9.608	7.66	.95	
9.	Race ·% Black	-.264	77.64	0	-20.50	0	-4.18
10.	(% Black)^2	-.00108	6615.23	351.98	-7.14	.38	
11.	Race (% Black)^2	.00239	6615.23	0	15.81	0	
12.	Student/Fac. Ratio K-4	.0970	31.24	33.299	3.03	3.23	.20
13.	# Changes in Sch. K-4	.0593	2.13	1.733	.13	.10	-.03
14.	Δ% Black K-4	.0259	.545	-.93	.01	.02	.01
15.	Race ·(Δ% Black)	-.0158	.545	0.0	-.01	0	.01
	Total				-11.99	-13.72	-1.73

Table 9. Effect of Regression Coefficients of Change from 4th to 6th Grade on Mean Black and Mean White Student

	Variable	Regression Coef.	Black Mean	White Mean	Contribution to Black Score	Contribution to White Score	Δ
1.	Constant	10.96	1	1	10.96	10.96	0
2.	Sex	-1.621	1.54	1.49	-2.50	-2.42	.08
3.	# Siblings	-.126	3.29	2.412	-.41	-.30	.11
4.	SES parents	-.0282	60.47	48.547	-1.71	-1.37	.34
5.	SES peers 6	-.0832	59.82	49.243	-4.98	-4.10	.88
6.	Head of Household Missing	.256	-.663	-.813	-.17	-.21	-.04
7.	Race	-1.84	1	0.0	-1.84	0	
8.	% Black 5-6	.0112	78.49	9.822	.88	.11	
9.	Race · % Black	.0363	78.49	0.0	2.85	0	1.07
10.	(% Black)^2	-.000048	6891.77	412.65	-.33	-.02	
11.	Race · (% Black)^2	-.000369	6891.77	0.0	-2.54	0	
12.	Student/Fac. Ratio 5-6	-.00895	27.26	30.314	-.24	-.27	-.03
13.	# Changes in Sch. 4-6	-.540	.170	.109	-.09	-.06	.03
14.	Δ(% Black) 4-6	-.0235	-1.29	1.538	-.03	-.04	-.01
15.	Race•[Δ(%Black)]	.0353	-1.29	0.0	.05	0.0	-.05
	Total				-0.1	2.28	2.38

Using the Δ column especially, one can see that some variables do not matter much in their contribution to the explanation of differences between black and white IQ scores, while others matter a great deal. We have lumped all of the variables dealing with race and integration together.

We find Table 11 below to be an informative summary of Tables 7 to 10. In it, we calculate cumulative effects rather than the effects due to differences, and we lump the two SES variables together.

Table 11. Cumulative Effects of SES versus Race-Segregation on the Difference in IQ Between a Mean White and a Mean Black

	K	4th	6th	8th
SES	4.14	6.36	7.58	8.30
Race-Segregation	8.14	3.96	5.03	7.29
Net of others	.41	.64	.53	.88
Total	12.69	10.96	13.14	16.47

Thus the SES variables account for about a third of the difference at kindergarten, and for more than half the difference at 4th grade and beyond. Note that these calculations are done for fictional persons: a black whose demographic and environmental variables are at the mean for all blacks, a white whose demographic and environmental variables are at the mean for all whites.

Finally, we present a highly tentative analysis of the linear and quadratic terms of the degree of integration variables (% Black) from Table 6. Again we use the cumulative effects, which we compute separately for whites and blacks.

Table 10. Effect of Regression Coefficients of Change from 6th to 8th Grade on Mean Black and Mean White Student

	Variable	Regression Coef.	Black Mean	White Mean	Contribution to Black Score	Contribution to White Score	Δ
1.	Constant	2.02	1	1	2.02	2.02	0
2.	Sex	-1.80	1.54	1.49	-2.77	-2.68	.09
3.	# Siblings	-.0302	3.29	2.412	-.10	-.07	.03
4.	SES parents	-.0439	60.47	48.547	-2.65	-2.13	.52
5.	SES peers 8	-.0192	59.84	49.369	-1.15	-.95	.20
6.	Head of Household Missing	-.181	-.663	-.813	.12	.15	.03
7.	Race	.430	1	0	.43	0	
8.	% Black 7-8	-.0623	70.18	13.375	-4.37	-.83	
9.	Race ·% Black	-.0807	70.18	0	-5.66	0	
10.	(% Black)^2	.000406	5842.23	448.93	2.37	.18	2.26
11.	Race ·(% Black)^2	.000739	5842.23	0	4.32	0	
12.	Student/ Fac. Ratio 7-8	.0481	21.52	24.555	1.04	1.18	.14
13.	# Changes in Sch. 6-8	.109	1.654	.735	.18	.08	-.10
14.	Δ(% Black) 6-8	-.0234	-5.18	2.310	.12	.05	-.07
15.	Race ·[Δ(% Black)]	.0439	-5.18	0.0	-.23	0	.23
		Total			-6.33	-3.0	3.33

Table 12. Cumulative Effects of Degree of Segregation/Integration on the IQ's of Black and White Students

	Linear	Quadratic	Best	Worst	Maximum effect
Whites					
Kindergarten	-0.00078	-0.00009	0	100	.98
Fourth grade	.0979	- .00117	41.8	100	3.95
Sixth grade	.1091	- .00122	44.7	100	3.73
Eigth grade	.0468	- .00081	28.9	100	4.10
Blacks					
Kindergarten	- .00469	- .00012	0	100	1.67
Fourth grade	- .170	.00119	0	71.4	6.07
Sixth grade	- .122	.00077	0	79.2	4.83
Eighth grade	- .265	.00192	0	69.0	9.14

The magnitude of the effect at kindergarten is small and can be disregarded. But the effect of segregation grows, becoming very serious indeed for blacks, especially by eighth grade. Because we are aware of the highly contentious area these results have led us to, we emphasize that these calculations are highly tentative and speculative. One reason we are unsure of these results is that in as highly segregated a system as Pittsburgh had, we have little data for blacks in mainly white schools and vice-versa. This led to large standard deviations, especially on the quadratic terms. The optima are the ratio of the linear term to twice the quadratic, and thus the uncertainty is magnified. Perhaps new data gathered on students who have been through a more integrated school experience would help us estimate these effects better.

6. CONCLUSION

There are several kinds of conclusions to this paper. One is the specific interpretation of this data set given in section 5. A second is that the kind of model we have used can be used to ascertain the effect of any environmental change on school children s IQ. For example, some schools have experimented with open classrooms; this kind of analysis would be appropriate for finding out what effect such a change would have.

We intend to explore several kinds of further analyses on this data set. First, we plan to find out what we can do to raise our effective sample from 1713 to 2067 by doing something about missing independent variables. Second, we would like to include an analysis of achievement test scores, and data on tardiness, absence, health and behavior marks, and grades. All these variables should be endogenous, and perhaps should also enter the IQ equations. Third, we would like to look further into the variance-covariance matrix estimation. Fourth, it would be nice to have variables for the sex of teachers, and to estimate teacher quality. Also, we plan to re-estimate the parameters using the maximum likelihood method. Finally, we could investigate the estimates of true IQ s, the X's, induced by our model. Perhaps in a few years' time we might collect a similar body of data again,* now that integration is more widespread in Pittsburgh. It would be interesting to see if its effects are predicted well by our model.

References

Goldberger, Arthur S. 1974 "Unobservable Variables in Econometrics," pp. 193-213 in Frontiers of Econometrics, Paul Zarembka, editor, Academic Press.

Griliches, Zvi. 1973 "Errors in Variables and Other Unobservables," Unpublished Discussion Paper.

Hollingshead, August B. 1957 "Two Factor Index of Social Position." Published by the Author, New Haven, Connecticut.

Joreskog, K. G. 1970 "A general method for analysis of covariance structures," Biometrika, 57, 239-251.

Kadane, Joseph B. 1972 "Identification of Functions of Parameters" Unpublished Technical Report

1973 Review of "Significance Tests" and "Statistics with a View toward Applications" Journal of the American Statistical Association 68, 1025.

Sanday, Peggy R. 1972a "On the Causes of IQ Differences Between Groups with Implications for Social Policy," Human Organization, Winter, Vol. 31: 411-424.

1972b "An Alternative Interpretation of the Relationship Between Heredity, Race, Environment, and IQ," Phi Delta Kappan, December, 250-254.

1972c "A Model for the Analysis of Variation in Measured Intelligence Between Groups," Mental Tests and Cultural Adaptation, Lee Cronbach and Peter Drenth, Eds, The Hague: Mouton Press.

1973. A Diffusion Model for the Study of the Cultural Determinants of Differential Intelligence. Final Report submitted to the U.S. Department of Health, Education, and Welfare, National Institute of Education.

Zellner, Arnold. 1962 "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57, 348-368.

* We wish to express our appreciation to Russ Winer, John Snyder, and especially Dan Rosen for computational assistance. Kadane's participation was supported in part by National Science Foundation Grant GS-38609. Sanday's and Staelin's participation was supported in part by the National Science Foundation and the Office of Education.