Economics of Education Review, Vol. 10, No. 1, pp. 7–17, 1991.
Printed in Great Britain.

0272–7757/91 $3.00 + 0.00

Characteristics of Education Production Functions: An Application of Canonical Regression Analysis

KWABENA GYIMAH-BREMPONG* and ANTHONY O. GYAPONG†

*Economics Department, Wright State University, Dayton, OH 45435-0001, U.S.A.; and †Economics Department, Wayne State University, Detroit, MI 48202, U.S.A.

Abstract — This paper uses data from the State of Michigan and canonical regression analysis to investigate the effects of socioeconomic characteristics (SEC) of communities in the production of high school education. We find that SEC have positive and significant impacts — impacts that are independent of school resources — on the output of education. However, these independent effects are very hard to ascertain because of the high degree of correlation between SEC and school resources. We also find that education of parents is the only variable that can be used as a proxy for all SEC without misspecifying the education production function.

I. INTRODUCTION

THIS PAPER uses data from school districts in the state of Michigan and canonical regression (Vinod, 1968) to investigate (i) whether socioeconomic characteristics of communities (SEC) contribute significantly to the production of educational outcomes in ways that are independent of their effects on school resources and (ii) whether there is one specific variable that can be used to proxy SEC.[1] Instead of treating SEC as scale variables, as most studies have, we treat them as inputs into the education production process. We specify and estimate two joint education production functions — with and without SEC — and test to see if the two sets of equations are different. We then investigate which variable(s) could be used to proxy SEC in education production functions without misspecification.

Empirical research on education production functions have found student background characteristics and SEC to be positively related to educational outcomes.[2] However, SEC variables are likely to be correlated with school resources. Communities with high incomes and high educational attainments are likely to provide more and better quality school resources than poor communities. Given this correlation, it is not clear whether SEC exert any influence on educational outcome independent of school resources. SEC may have a dual role in the production of education: they affect the quantities and quality of school resources and also directly affect the output of schools. It is the latter effect that is of major concern to us in this paper. Very little attention has been devoted to testing the importance of SEC's direct contribution to education production in the economics of education literature.

Another issue we investigate is which variable(s) can be used to proxy SEC if these variables are found to have a significant effect on education production. There are several variables that can be used to proxy SEC; none of which is likely to be a perfect measure. Researchers have used combinations of such variables to capture the essential characteristics of communities or families.[3] However, these variables are likely to be highly correlated, causing collinearity problems when a number of them are included in a regression equation. We investigate whether reducing the number of SEC variables in an education production function results in misspecification.

The same set of school resources and SEC are used to produce many educational outputs. Moreover, it is possible to have complex non-causal

[Manuscript received 5 April 1989; revision accepted for publication 5 April 1990.]

relationships among the educational outputs. For example, mathematics professors have noted that students' performance in mathematics is correlated with their performance in English. This correlation may or may not be causal. Education production should therefore be treated as a joint production process. Most researchers have specified and estimated single equation non-joint education production functions. A few researchers, Boardman *et al.* (1977) and Murnane *et al.* (1981), have estimated simultaneous equation models of education production. However, simultaneous equation models are not appropriate for estimating joint production of education since they assume that the relationship among the educational outcomes is causal.

Though Chizmar and Zak (1984) have specified and used canonical regression to estimate a joint production function for education, they did not investigate the importance of SEC in the production process. Previous researchers have based their conclusions of the importance of SEC in education production on simple *t*-tests. However, it is possible that SEC variable(s) may indicate statistical significance even though they (it) may not add any explanatory power to the equation. To our knowledge, this is the only study of education production that explicitly tests for the independent effects of SEC on the production of education. Some researchers (e.g. Cohn *et al.*, 1989) have used flexible functional forms to investigate the technology of joint production of education based on cost functions. This study offers an alternative approach to the analysis of joint production when cost data is not available to the researcher. The results of investigating the importance of SEC in education production function has important policy implications for improving educational outcomes in the United States and elsewhere.

The rest of the paper is organized as follows. Section II introduces the education production function to be estimated and discusses the estimation procedure. Section III discusses the data while Section IV presents and discusses the econometric results and policy implications. Section V concludes the paper.

II. THE MODEL

The theoretical foundations of this paper lie in the works of Hanushek (1971, 1979), Summers and Wolfe (1977), Margo (1986), Oates (1981) and Hamilton (1983). Hanushek (1971, 1979) and Summers and Wolfe (1977) argue that SEC should enter the education production function as inputs because they influence the outcome of the educational process by complementing purchased inputs. Summers and Wolfe (1977) argue that educated parents are likely to complement the efforts of the school as well as provide additional reading materials, while parents who cannot read or write are unable to help their children even if they desired to do so. At the general level of the production of local public goods, Oates (1981) and Hamilton (1983) have argued that SEC should be considered as non-purchased inputs in the production of local public goods since they affect the productivity of purchased inputs. SEC should therefore be treated as fixed inputs in the production of local public goods. In this paper, we follow this argument and include SEC as non-purchased inputs in our production function.

However, including SEC in the education production function introduces problems of interpretation of coefficients since it makes it difficult to identify the independent effects of school resources on education output. To illustrate this point, suppose education output (Y) is a function of school resources (X) and SEC (Z); i.e. $Y = Y(X,Z)$. Suppose further that X is a function of Z and other variables (Q); i.e. $X = X(Z,Q)$. This implies that $Y = Y(X(Z,Q),Z)$. From the last equation, $\mathrm{d}Y/\mathrm{d}X = (\partial Y/\partial X)(\partial X/\partial Z) + \partial Y/\partial X$ which is a combination of the marginal product of school resources and the effects of Z on school resources. This makes it difficult to identify the independent effects of school resources on educational outputs. To circumvent this problem, we specify and estimate two education production functions — with and without SEC and test to see if the two equations are different.

We consider education as a production process in which inputs — school resources (**X**), student characteristics (**W**), and SEC (**Z**) are used to produce a vector of education outputs (**Y**). We assume that a strictly quasi-concave education production function exists. The education production function in its implicit form is given as:

$$\mathbf{G}(\mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{Z}) = 0 \; g_x, g_w \geq 0 \qquad (1a)$$

where all variables are as defined above. We expect the coefficients of **X** and **W** to be at least nonnegative in order to satisfy the weak essentiality condition. The vector **Z** can take forms that enhance the productivity of purchased inputs (e.g. education of parents) or it could take forms that decrease the productivity of purchased inputs (e.g. crime rate). If **Z** assumes forms that enhance (decrease) the productivity of purchased inputs, then we expect it to have a positive (negative) marginal product. If SEC have no effect on educational outputs, then the education production function is:

$$\mathbf{H}(\mathbf{Y}, \mathbf{X}, \mathbf{W}) = 0,\ h_x,\ h_w \geq 0 \quad (1b)$$

where all variables are as defined above. Equations (1a) and (1b) are the two equations we estimate and compare in investigating the importance of SEC in the production of education.

To estimate Equation (1) econometrically, we have to provide a specific functional form. We specify a joint education production function of the Cobb Douglas functional form as:[4]

$$\mathbf{Y}^{\alpha} - \lambda\mathbf{X}^{\beta}\mathbf{W}^{\gamma}\mathbf{Z}^{\theta}\mu = 0 \quad (2a)$$

where μ is a stochastic error term, α, λ, β, γ, and θ are coefficients to be estimated, and all other variables are defined above. Setting θ to zero gives the corresponding equation for (1b). We label the resulting Equation (2b). If **Z** is important in the production of education, Equations (2a) and (2b) will be different; otherwise there will be no difference between the two equations. The marginal elasticity of output i with respect to input j in (2) is:

$$ME(Y_i,X_j) = \partial \ln Y_i/\partial \ln X_j = \beta_j/\alpha_i. \quad (3)$$

The corresponding marginal product of input j in the production of output i is given as:

$$MP(Y_i,X_j) = (Y_i/X_j)ME(Y_i,X_j) = (Y_i/X_j)(\beta_j/\alpha_i). \quad (4)$$

The marginal rate of transformation between the two outputs is:

$$\partial Y_i/\partial Y_j = -[(\alpha_j/Y_j)/(\alpha_i/Y_i)]. \quad (5)$$

We use two measures of output — ACT scores in mathematics (*ACTM*) and English (*ACTE*) to proxy educational output. This test is taken in the 12th grade and therefore measures the cumulative output of the school system. To account for prior knowledge (achievements) of students, we include student scores on an achievement test given in the 10th and 11th grades to measure their proficiency in English (*READ*) and mathematics (*MATH*). The vector **W** should theoretically include innate student characteristics such as IQ and motivation. Data on these characteristics are not available to us. We assume, however, that these characteristics will be captured by past student achievements as measured by *MATH* and *READ*. We have four school inputs — per student expenditure on instruction (*INST*), support services (*SUPT*), capital (*CAP*), and pupil–teacher ratio (*PT*). We expect *MATH*, *READ*, *INST*, *SUPT*, and *CAP* to be positively related to student test scores. We expect *PT* to be negatively related to test scores since larger class sizes imply that teachers have less time to interact with each student.

There are several variables that could be used to proxy SEC. Of these, previous research has revealed that income (*INC*), educational attainment of the adult population (*EDHS*), poverty (*POV*), and crime rates (*CRIME*) are significant determinants of student performance. Though none of these variables is likely to be a perfect measure of SEC, a combination of these variables will provide a good measure of a large aspect of the characteristics of the student's learning environment. We initially include all these variables in our education production function.

Our estimation procedure follows Vinod's adaptation (Vinod, 1968) of Hotelling's canonical correlation analysis to the estimation of joint production functions.[5] We use canonical regression because it is able to handle more than one dependent variable in a single equation as well as take into account possible non-causal inter-relationships among the dependent variables — tasks that classical regression cannot perform. Estimating separate equations for each output neglects relationships among the outputs, while estimating a simultaneous equation model assumes that the relationship among the dependent variables is causal — an assumption we are not able to make given the current state of our knowledge of the education process. Moreover, both separate regressions and simultaneous equation models are likely to neglect aspects of joint production technology, such as scope economies, while canonical regression takes these into consideration.

If $k_i = a_i y_i$ and $p_j = b_j x_j$ are linear combinations of y_i and x_j respectively, the simple correlation between k_i and p_j is the ith canonical correlation between k_i and p_j (r_i) and a_i and b_j are the canonical coefficients of the Y and X vectors respectively. The coefficients of Equation (2) are related to the canonical coefficients in the following way:

$$\hat{\alpha}_i = \alpha_i, \ \beta_j = \beta_j r_i \qquad (6)$$

where r_i is the first canonical correlation. Vinod (1968) demonstrates that this approach yields consistent estimates for joint production while OLS estimates are biased.

III. DATA

This study employs cross-sectional data for 175 school districts with a population of 1000 or more in the state of Michigan. The data are school district averages for the 1986/1987 academic year. This sample was chosen in part because of the availability of very rich data for that year. Besides average test scores for 1986/1987, it also provides scores on a proficiency test administered to 10th grade students two years earlier. This allows us to model education production in value added terms.

School output is measured by average ACT scores in mathematics (*ACTM*) and English (*ACTE*) of graduating high school students in a district during the 1986/1987 academic year. Though we recognize that standardized test scores may not reflect all the outputs of complex organizations such as the school system, we think that these test scores reflect those aspects of school outputs that are readily quantifiable. Second, as Hanushek (1979) has argued, most evaluations of school systems are based on these test scores. Most important, both students and parents judge the quality of schools based on these scores. Data for *ACTM* and *ACTE* were obtained from the files of American College Test Program, *ACT: High School Profiles, 1987* (ACT Program, Iowa City, Iowa, 1987).

Data for *INST*, *CAP*, *SUPT*, and *PT* were obtained from Michigan State Board of Education, *Michigan K-12 School Districts Ranked by Selected Financial Data 1985/86* (Lansing, Michigan, 1987). *MATH* and *READ* are the average score on proficiency tests administered to all 10th grade students in the fall of 1985 as part of the Michigan Educational Assessment Program (MEAP). Data for *MATH* and *READ* were obtained from Michigan Department of Education, *Intermediate School District Report, 1985/86* (Lansing, Michigan, 1986).

The four variables we use to represent socioeconomic characteristics of the school district are *INC*, *POV*, *CRIME*, and *EDHS*. We measured educational attainment as the percentage of the adult population (25 years and older) that has completed 12 years or more of education, while *INC* is the median family income in a school district. *POV* is the percentage of families in a school district that has incomes below the poverty line in 1980, and *CRIME* is the Federal Bureau of Investigation's (FBI) index crime rate for a school district. The data for *INC* and *EDHS* were obtained from Michigan Department of Commerce, *Business Information, 1985* (Lansing, Michigan, 1986). *CRIME* data were obtained from Michigan Department of State Police, *Crime in Michigan: Uniform Crime Report*, 1985 (Lansing, Michigan, 1986). *POV* data were obtained from Bureau of Census, *1980 Census of Population*, Vol. 1, *Characteristics of the Population: Michigan* (Washington D.C., Bureau of the Census: September 1981).

There were 175 school districts in our sample. However, we could not get data on *POV* and *CRIME* for 23 of the districts. We therefore have 152 usable observations for the study. Table 1 shows summary statistics of the data used in our estimation. The data show considerable variation in school resources and socioeconomic characteristics across school districts in Michigan. All variables were normalized around their means for estimation purposes.

IV. ECONOMETRIC RESULTS

Parameter estimates of the canonical regression for the full model are presented in Table 2. Both the χ^2 and Rao's F statistic indicate a rejection of the null hypothesis that all coefficients are equal to zero at any reasonable significance level. Canonical redundancy analysis also indicates that the full model explains about 57% of the variance of test scores.

The coefficients of *MATH* and *READ*, their marginal elasticities, as well as their marginal products are positive and significantly different from zero as expected. Secondly, the magnitude of these coefficients, marginal elasticities, and marginal pro-

Table 1. Summary statistics of education resource data in Michigan

Variable	Mean	S.D.
ACTE	18.07	1.36
ACTM	17.41	2.10
MATH	70.28	11.05
READ	83.17	7.73
INC	21,870.79	6,285.92
CRIME	296.35	1,496.81
INST	1,860.35	646.19
CAP	189.79	138.98
SUP	1,211.36	350.66
PT	21.74	2.49
EDHS	69.30	9.56
POV	7.44	4.89
$N = 174$		

Variable definitions: *ACTE, ACTM*, ACT scores in English, mathematics; *MATH, READ*, prior achievement (10th and 11th grade) in mathematics and English; *INC*, median family income; *CRIME*, FBI index crime rate; *EDHS*, percent of adult population with 12 or more years of schooling; *POV*, percentage of population with income below poverty line; *INST*, per pupil instructional expenditure; *CAP*, per pupil capital expenditure; *SUP*, per pupil expenditures on instructional support; *PT*, pupil–teacher ratio.

ducts are similar to those obtained by other researchers (Hanushek, 1971). This shows that past student achievements are important determinants of students' performance. In effect, the educational system "builds on" students' prior knowledge. The marginal rate of transformation between the two outputs, calculated at the means of the outputs is −0.8181.

Of the school resources, only *CAP* and *PT* have the expected coefficients while *INST* and *SUPT* have unexpected negative coefficients. With the exception of *CAP*, all these coefficients are significantly different from zero and are similar to those obtained by Boardman *et al.* (1977) and Dolan and Schmidt (1987), among others. The negative coefficients of *INST* and *SUPT* seem to suggest that increasing school resources will have a negative impact on student performance. This is completely at variance with school officials' plea to increase school resources in order to improve the performance of the school system. This tentative conclusion is also different from the results of researchers who find that schools do make a difference in student achievement.[6]

Of the SEC variables, *INC* and *EDHS* have positive coefficients while *POV* has negative coefficient as expected. The coefficient of *EDHS* is significant while those of *INC* and *POV* are insignificant. *CRIME* has an unexpected positive and significant coefficient, giving the impression that higher crime rates are associated with higher student performance, all things equal. If this were the case, the crime-ridden inner city schools should outperform their suburban counterparts in test scores — a conclusion not borne out by test scores.

One thing that stands out from Table 2 is the implausible coefficients of *INST*, *SUPT*, and *CRIME*. The negative coefficients of these inputs seem to violate the weak essentiality condition for variable input use in production. What are the possible sources of these sign reversals? Collinearity among the variables easily suggests itself. It can reasonably be expected that high income school districts can afford to spend more per pupil for instruction, instructional support, and supervision than poor school districts. The Pearson correlation coefficients between *INC* on the one hand, and *INST*, *POV*, and *SUPT*, on the other, in the sample are 0.59, −0.60, and 0.50 respectively. The correlation coefficient between *INST* and *SUPT* is 0.81. These high degrees of correlation among the explanatory variables could cause multicollinearity problems.

To investigate whether collinearity is the cause of sign reversals in the estimated coefficients in Table 2, we reestimated Equation (2a) without *INC* and *SUPT*. Coefficient estimates of this modified equation are presented in Table 3. The magnitude of the coefficients as well as the marginal products and marginal elasticities in Table 3 are similar to those in Table 2. We used canonical redundancy analysis to test for equality of the equations in Tables 2 and 3.[7] The calculated F statistic is 1.07. With 4 and 278 degrees of freedom, we are unable to reject the null hypothesis of equality of the two regression equations. We conclude that *INC* and *SUPT* do not add additional explanatory power to the education production function.

Comparing the signs of the coefficients in Table 3 to those in Table 2, we note some differences. In Table 3, we find that all the coefficients have the expected signs. In particular, *INST* now has the expected positive coefficient while *CRIME* has a negative coefficient. All the other coefficients retain their expected signs. Moreover, with the exception

Table 2. Canonical coefficient estimates of education production function: full model

Variable	Parameter estimates	ME* (*ACTM*)	ME (*ACTE*)	MP‡ (*ACTM*)	MP (*ACTE*)
ACTM	0.4607				
ACTE	0.5845				
MATH	0.1692 (1.927)†	0.3673	0.2895	0.0910	0.0744
READ	0.2950 (3.277)	0.6403	0.5047	0.1340	0.1097
INST	−0.1129 (1.897)	−0.2451	−0.1932	−0.0023	−0.0019
CAP	0.0319 (0.574)	0.0692	0.0546	0.0063	0.0052
SUP	−0.1530 (2.284)	−0.3321	−0.2618	−0.0038	−0.0039
PT	−0.1318 (1.419)	−0.2861	−0.2255	−0.2291	−0.1874
INC	0.0748 (0.390)	0.1624	0.1280	0.0001	0.0001
EDHS	0.3615 (4.537)	0.7847	0.6185	0.1971	0.1613
POV	−0.0816 (0.927)	−0.1771	−0.1396	−0.4144	−0.3391
CRIME	0.1776 (1.718)	−0.3855	−0.3038	−0.0226	−0.0185
$N = 151$	$r_1 = 0.7560$		$X^2 = 145.207$		Rao's $F = 9.073$

*ME, marginal elasticity of output; with respect to input *j*. MP, marginal product of input *j*.

†Absolute value of *t* statistics in parentheses. Standard errors used in calculating the *t* statistics are based on a multivariate regression of a linear confination of the outcome on the independent variables. The "delta" method is then used to calculate the standard errors. For more on the "delta" method, see Billingstey, P. (1979) *Probability and Measure*, New York: John Wiley & Sons. We thank Mark Kennet for helping us with the "delta" method.

‡MP calculated at the means of variables.

of *CAP*, all coefficients are significantly different from zero. This contrast with the results in Table 2 where some of the coefficients had the wrong signs. We conclude that the "wrong" signs of some of the coefficients on Table 2 is mainly due to collinearity problems. We note that some researchers (Boardman *et al.*; 1977, Dolan and Schmidt, 1987) have obtained similarly negative coefficients for school resources when they included SEC in the education production function. The "wrong" signs on the coefficients of school resources should be interpreted with caution and unless further investigation is conducted, should not be the basis for policy recommendations.

There is an alternative interpretation of the negative coefficients of the school resource variables when the education production function includes SEC. It is possible that measured variations in quantifiable school inputs are not highly correlated with variations in forces (but non-measurable) that affect school output. In effect, there is a large element of measurement error in the school resource variables resulting in a downward bias of their coefficients. The downward bias resulting from this error in variables is large enough to reverse the signs of the coefficients. This interpretation is consistent with the point made by Hanushek (1981).[8] We note however that when we estimated the education production function without the school resource variables, the resulting equation was significantly different from the full model in the statistical sense.[9]

Although initial indications from the estimates in Tables 2 and 3 are that SEC are positively related to student test scores, it is possible that they are just preempting other variables rather than providing significant independent contributions to the explanation of the variance in student performance. To further explore the contributions SEC make to the explanation of the variance in educational

Table 3. Canonical coefficient estimates of education production function: model without *INC* and *SUP*

Variable	Parameter estimates	ME* (*ACTM*)	ME (*ACTE*)	MP‡ (*ACTM*)	MP (*ACTE*)
ACTM	0.4544				
ACTE	0.5906				
MATH	0.1705 (1.962)†	0.3752	0.2887	0.0929	0.0742
READ	0.2936 (3.347)	0.6461	0.4971	0.1352	0.1080
INST	0.1908 (2.212)	0.4199	0.3231	0.0039	0.0031
CAP	0.0312 (0.544)	0.0686	0.0528	0.0063	0.0050
PT	−0.1095 (1.512)	−0.2410	−0.1854	−0.1930	−0.1541
EDHS	0.3796 (4.970)	0.8354	0.6427	0.1615	0.1676
POV	−0.1051 (1.440)	−0.2313	−0.1780	−0.4165	−0.4323
CRIME	−0.1178 (1.712)	−0.2592	−0.1995	−0.0152	−0.0122
$N = 151$	$r_1 = 0.7516$		$X^2 = 142.143$		Rao's $F = 11.10$

*ME, marginal elasticity of output *i* with respect to input *j*.
†Absolute value of *t* statistics in parentheses.
‡MP, marginal product of input *j* in the production of output *j*. Marginal products are calculated at the means of variables.

Table 4. Canonical coefficient estimates of education production function: restricted model

Variable	Parameter estimates	ME* (*ACTM*)	ME (*ACTE*)	MP‡ (*ACTM*)	MP (*ACTE*)
ACTM	0.3257				
ACTE	0.7134				
MATH	0.2641 (2.859)†	0.8109	0.3402	0.2009	0.0952
READ	0.4374 (4.715)	1.3429	0.6131	0.2811	0.1332
INST	0.0636 (1.529)	0.1953	0.0892	0.0008	0.0009
CAP	0.0532 (0.853)	0.1633	0.0746	0.0149	0.0071
SUP	0.0817 (0.729)	0.2508	0.1145	0.0036	0.0017
PT	−0.0708 (0.934)	0.2174	0.0992	0.1741	0.0825
$N = 151$	$r_1 = 0.6449$		$X^2 = 107.845$		Rao's $F = 10.604$

*ME, marginal elasticity of output *i* with respect to input *j*, *i* = *ACTM*, *ACTE*. MP, marginal product of input *j*.
†Absolute value of *t* statistics in parentheses.
‡Calculated at means of variables.

output, we estimate Equation (2b) and compare it to the results obtained for Equation (2a).

Parameter estimates for (2b) are presented in Table 4. The coefficients in Table 4 are all of the expected signs. However, the coefficients are larger in absolute magnitude than their counterparts in Tables 2 and 3. For example, the coefficients of *MATH* and *READ* are 54.9 and 49% respectively higher in Table 4 than in Table 2. Possibly, these higher coefficients indicate that these variables are picking up the effects of the excluded SEC variables in addition to their own contribution to educational output. While the exclusion of SEC from the education production function may not lead to sign reversal of the coefficients of the remaining variables, it may result in the inflation of these coefficients. Also, with the exception of *INST*, the coefficients of all school resources are statistically insignificant.

An F test to test the hypothesis of equality between the restricted and the full education production function against the alternative that they are different has a calculated F statistic of 9.41 if the equation in Table 2 is the maintained hypothesis and 17.73 if the modified equation in Table 3 is the maintained hypothesis. With 8 (4) and 278 (282) degrees of freedom, we reject the null hypothesis at $\alpha = 0.01$ or better. Excluding SEC from the education production function will therefore result in misspecification and hence biased parameter estimates.

We have concluded that SEC are significant determinants of education output and that these SEC variables are highly correlated. Is it possible to use one of these variables to proxy SEC without causing misspecification; and if so, which of the SEC variables should be chosen for that purpose? There is no theory to guide us in investigating this question so we resort to statistical experimentation. We reestimate (2a) with only one SEC variable at a time and compare the resulting estimates with the full model. Parameter estimates for these models are presented in Table 5. Model 1 refers to the equation with *INC* as the SEC variable; Model 2 has *EDHS*, Model 3 has *POV*, while Model 4 has *CRIME* as the respective SEC variables.[10]

With the exception of *INST* in Model 1, all the coefficients in Table 5 have the expected signs. An analysis of the canonical structure also indicates that with the exception of *INST* in Models 1 and 3 and *SUP* in Models 2–4, all the explanatory variables contribute significantly to the explanation of education output in all the models. Comparing the coefficients in Table 5 to those in Tables 2–4, one

Table 5. Canonical coefficient estimates of education production function: modified equations

	Parameter estimates			
Variable	Model 1	Model 2	Model 3	Model 4
ACTM	0.2869	0.4381	0.3352	0.3510
ACTE	0.7493	0.6065	0.7045	0.6840
MATH	0.1929	0.1542	0.2296	0.2791
	(2.114)*	(1.974)	(2.540)	(3.024)
READ	0.4164	0.2964	0.3618	0.4378
	(4.643)	(3.98)	(3.885)	(4.748)
INST	−0.0787	0.0150	0.0018	0.0400
	(0.643)	(0.675)	(1.016)	(1.328)
CAP	0.0268	0.0143	0.0500	0.0641
	(0.443)	(1.250)	(0.826)	(1.028)
SUP	0.1816	0.0882	0.1338	0.0920
	(1.618)	(0.867)	(1.218)	(0.830)
PT	−0.1648	−0.0704	−0.1010	−0.0836
	(2.090)	(1.872)	(2.336)	(1.865)
INC	0.2285	—	—	—
	(3.389)			
EDHS	—	0.4061	—	—
		(5.637)		
POV	—	—	−0.2330	—
			(3.140)	
CRIME	—	—	—	−0.0693
				(0.923)
N	151	151	151	151
r_1	0.7041	0.7439	0.7001	0.6672
X^2	149.3687	134.1150	147.6269	140.5767
Rao's F	10.202	12.190	10.021	9.301

*Absolute value of t statistics in parentheses.

observes that while the coefficients are similar in absolute magnitudes to those in Tables 2 and 3, they differ from those in Table 4. *F* tests reject the null hypothesis that none of the SEC variables adds to the explanatory power of the educational production function at $\alpha = 0.01$ even though the coefficient of *CRIME* is insignificant in Model 4.[11] We can conclude that each SEC variable contributes significantly to educational output.

Can we conclude from the above results that each of Models 1–4 is the statistical equivalent of the full model? We investigate this question by maintaining the full model as the null hypothesis and each of Models 1–4 as an alternate hypothesis and conduct *F* tests. This test is both a specification and model selection test since it is able to discriminate among different models and also check for correct specification. With calculated *F* statistics of 20.53, 2.93, 21.11, and 29.33 for Models 1, 2, 3, and 4 respectively and 6 and 278 degrees of freedom, we reject the null hypothesis of equality between the full model and Models 1, 3, and 4, but not for Model 2. While *INC*, *POV*, and *CRIME* are important determinants of student performance, none of them by itself is sufficient to completely characterize the socioeconomic environment of the student in education production.

Nonrejection of the null hypothesis for Model 2 implies that the education of parents captures all the essential characteristics of the socioeconomic environment in which the student learns. For the specification and estimation of education production functions one can safely use the education of parents or of the adult population to proxy SEC without misspecification.

One of the surprising results in this study is that whenever we include *INC* in the production function, the coefficient of *INST* reversed sign while excluding *INC* changed the coefficient of *INST* to the expected positive sign. We have argued above that this may be due to collinearity between *INST* and *INC*. An examination of the canonical structure reveals that when *INC* is included in the education production function, *INST* does not contribute anything whatsoever to the explanation of education output. The reverse is true when *INC* is deleted from the production function. This is further evidence of the problem caused by the collinearity between *INC* and *INST*. It appears that *INC* does not have any effect on education output independent of its effects through the provision of school resources.[12]

Does our result depend on the method of estimation? We investigate this question by estimating separate equations for *ACTM* and *ACTE* using OLS techniques. The results are presented in Table 6. The coefficients of *MATH*, *READ* and the SEC variables are similar to those in Table 2 even though they differ in absolute values. However, it is interesting to note that the school resource variables have insignificant coefficients in the *ACTM* equation while they are negative and significantly different from zero in the *ACTE* equation — results that differ a little bit from our observations in Tables 2 and 3. It is possible that the separate equation approach does not capture some characteristics of joint production such as economies of scope.

The finding that socioeconomic characteristics of school districts (parents) are important inputs in the production of high school education is consistent with the result of previous research that finds that SEC are significant factors in the production of education (Hanushek, 1971; Dolan and Schmidt,

Table 6. OLS estimates of education production

Variable	Parameter estimates *ACTM*	*ACTE*
CONSTANT	−0.0302	0.0338
	(0.513)*	(0.586)
MATH	0.2611	0.0981
	(2.709)	(1.037)
READ	0.1899	0.3673
	(1.966)	(3.878)
INST	0.0132	−0.1970
	(0.108)	(1.640)
CAP	0.0743	−0.0024
	(1.195)	(0.040)
SUP	−0.1697	−0.1295
	(1.504)	(1.170)
PT	−0.0289	−0.1986
	(0.360)	(2.518)
INC	0.0132	0.1176
	(0.109)	(0.985)
EDHS	0.4247	0.3456
	(4.634)	(3.844)
POV	−0.1017	−0.0614
	(0.991)	(0.610)
CRIME	0.1718	0.1025
	(2.258)	(1.373)
F	15.523	16.641
R^2	0.4919	0.5105

*Absolute value of *t* statistics in parentheses.

1987; Murnane *et al.*, 1981; Tuckman, 1971; among others). Failure to include these variables as inputs in the production of education results in misspecification of the education production function. The policy implication of our result is that the search for educational excellence should focus on schools as well as on improving the larger socioeconomic environment of the student.

Our result that education of the adult population can be used in place of all SEC variables implies that researchers can stick to parsimony without misspecifying their models. Second, it allows researchers to reduce collinearity problems in estimating education production functions. Third, because of collinearity problems, previous research results should be viewed with caution. We agree with Brown and Sacks (1975) that school resources do matter, but their importance may not be revealed unless the researcher is careful in modelling the education production function.

V. CONCLUSION

This paper used data from the state of Michigan and canonical regression to investigate the importance of socioeconomic characteristics of communities in the production of high school education. We found socioeconomic characteristics to be important inputs in the production of education. This means that researchers who exclude SEC from their education production functions may have misspecified their models. Conclusions based on such coefficients are, at best, questionable. Furthermore, of the four SEC variables employed, only education of the adult population can be used to represent all the essential characteristics of communities in so far as the production of education is concerned. This makes it easier for researchers to reduce collinearity without misspecifying the education production function.

We found that school resources positively influence student performance. However, because of the collinearity between school resources and socioeconomic characteristic of communities, it is very difficult (and sometimes impossible) to disentangle the independent effects of school resources from those of SEC. Extreme caution should therefore be exercised in interpreting the coefficients of school resources without further investigation. The policy implication that flows from this study is that the search for educational excellence should be fought on two fronts: the school with more and better quality resources and the student's environment outside the school.

Acknowledgements — We thank the staff of the Michigan State Board of Education, Lansing, Michigan, for providing some of the data used in this study and an anonymous referee for helpful comments. We thank Mark Kennet for helping us on the "delta" method and Greg Willis for providing excellent research assistance. We alone, however, are responsible for any remaining errors.

NOTES

1. Chizmar and Zak (1984) use a similar approach in estimating a joint production function for education.
2. Hanushek (1979) provides an excellent review of the literature on the modeling of education production.
3. See Dolan and Schmidt (1987), Hanushek (1971), and Perl (1973) for examples of such formulation.
4. We tried a translog specification but extreme collinearity problems made it impossible to estimate a translog production function.
5. Details of this procedure are contained in Vinod (1968, 1976). Chizmar and Zak (1984) have also applied this procedure to the estimation of education production function. For brevity, we only present the outlines of the procedure here. The reader is referred to the above sources for details.
6. We describe the result here as "tentative" because we return to investigate it later in the paper drawing different conclusions.
7. Dillon and Goldstein (1984) suggest that the appropriate statistics for hypothesis testing in canonical correlation analysis are those based on canonical redundancy analysis. All tests in this paper are therefore based on canonical redundancy statistics.
8. We thank an anonymous referee of *The Review* for drawing our attention to this alternative interpretation.
9. Coefficient estimates for this equation are not reported because of space consideration. They are however available from the authors upon request.

10. Because of space considerations, we only present the estimated coefficients without the accompanying marginal elasticities and products.
11. The calculated *F* statistics are 8.70, 26.53, 8.16, and 7.45 for Models 1, 2, 3, and 4 respectively, with 1 and 278 degrees of freedom each.
12. It is possible that *INC* does not really enter the production function; it may indicate demand for education. However, we are unable to separate this demand for education from the production function. We will follow this line of inquiry in another study in the near future.

REFERENCES

BOARDMAN, A.E., DAVIS, O.A. and SANDAY, P.R. (1977) A simultaneous equation model of the educational process. *J. Public Econ.* **7**, 23–49.

BOARDMAN, E.A. and MURNANE, R.J. (1979) Using panel data to improve estimates of the determinants of educational achievement. *Sociology of Education* **52**, 113–121.

BROWN, B.W. and SACKS, D.H. (1975) The production and distribution of cognitive skills within schools. *J. polit. Econ.* **83**, 571–593.

BROWN, B.W. and SACKS, D.H. (1980) Production technologies and resource allocations within classrooms and schools: theory and measurement. In *The Analysis of Educational Productivity*, Vol. 1: *Issues and Microanalysis* (Edited by THOMAS, J.A. and DREEBEN, R.). Chicago: University of Chicago Press.

CHIZMAR, J.F. and ZAK, T.A. (1984) Canonical estimation of joint education production functions. *Econ. Educ. Rev.* **3**, 37–43.

COHN, E., RHINE, S.W.R. and SANTOS, M.C. (1989) Institutions of higher education as multiproduct firms: economies of scale and scope. *Rev. Econ. Statist.* **71**, 284–290.

DILLON, W.R. and GOLDSTEIN, M. (1984) *Multivariate Analysis: Methods and Applications*. New York: John Wiley & Sons.

DOLAN, R.C. and SCHMIDT, R.M. (1987) Assessing the impact of expenditure on achievement: some methodological and policy considerations. *Econ. Educ. Rev.* **6**, 285–299.

HAMILTON, W.B. (1983) The flypaper effect and other anomalies. *J. Public Econ.* **22**, 347–361.

HANUSHEK, E.A. (1971) Teacher characteristics and gains in student achievement: estimation using micro data. *Am. Econ. Rev.* **61**, 280–288.

HANUSHEK, E.A. (1979) Conceptual and empirical issues in the estimation of education production functions. *J. Hum. Resour.* **14**, 351–388.

HANUSHEK, E.A. (1981) Throwing money at schools. *J. Policy Analysis and Management* **1**, 19–41.

HANUSHEK, E.A. (1986) The economics of schooling. *J. Econ. Lit.* **24**, 1141–1177.

KUYLEN, A.A.A. and VERHALLEN, T.M.M. (1981) The uses of canonical analysis. *J. econ. Psychol.* **1**, 217–237.

MARGO, R.A.(1986) Educational achievement in segregated school systems: the effects of 'separate-but-equal'. *Am. Econ. Rev.* **76**, 794–801.

MURNANE, R.J., MAYNARD, R.A. and OHLS, J.C. (1981) Home resources and children's achievement. *Rev. Econ. Statist.* **63**, 369–377.

OATES, W.E. (1981) On local finance and Tiebout Model. *Am. Econ. Rev.* **71**, 93–98.

PERL, L.J. (1973) Family background, secondary school education expenditure, and student ability. *J. Hum. Resour.* **8**, 156–180.

SEBOLD, F.D. and DATO, W. (1981) School funding and student achievement: an empirical analysis. *Public Finance Quarterly* **9**, 91–105.

SUMMERS, A. and WOLFE, B. (1977) Do schools make a difference. *Am. Econ. Rev.* **67**, 639–652.

TUCKMAN, H.P. (1971) High school inputs and their contribution to school performance. *J. Hum. Resour.* **6**, 490–509.

VINOD, T. (1968) Econometrics of joint production. *Econometrica* **36**, 322–336.

VINOD, T. (1976) Canonical ridge and econometrics of joint production. *J. Econometrics* **4**, 147–166.

WINKLER, D. (1975) Educational achievement and school peer group composition. *J. Hum. Resour.* **10**, 189–204.