

CAHIER 8513

Functional Forms and Educational  
Production Functions\*

by

Claude Montmarquette  
and  
Sophie Mahseredjian\*\*  
Université de Montréal

March 1985

\*We are indebted to Richard Laferrière and Rachel Houle for useful suggestions. Remaining errors are the sole responsibility of the authors. Research financed by grants from CRSH (Ottawa) and FCAR (Quebec).

\*\*Centre de recherche et développement en économie, Université de Montréal, C.P. 6128, succursale A, Montréal, Québec, H3C 3J7.

---

Ce cahier est publié conjointement par le Département de science économique et par le Centre de recherche et développement en économie de l'Université de Montréal.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.C. pour l'aide et le soutien à la recherche.

## ABSTRACT

Different functional forms are proposed and applied in the context of educational production functions. Three different specifications - the linear, logit and inverse power transformation (IPT) - are used to explain first grade students' results to a mathematics achievement test. With IPT identified as the best functional form to explain the data, the assumption of differential impact of explanatory variables on achievement following the status of the student as a low or high achiever is retained. Policy implications of such result in terms of school interventions are discussed in the paper.

## RESUME

Dans cet article, diverses formes fonctionnelles sont proposées et appliquées dans le contexte des fonctions de production en éducation. Trois spécifications différentes - linéaire, logit et "inverse power transformation" (IPT) - sont appliquées à un modèle expliquant le rendement en mathématiques des élèves de première année au primaire. L'hypothèse d'effets différenciés des variables explicatives sur le rendement des élèves selon leur niveau de réussite est mise en évidence par le modèle IPT. Les implications de ce résultat vis-à-vis des politiques d'intervention au niveau scolaire sont discutées dans le texte.

## S-Curves and the Learning Process at School

Functional forms have for some time become a focus of interest in applied economics in such diversified areas as the demand for money, transport and local public goods, studies on technological changes and diffusion process, manufacturing production functions, etc. It is somewhat surprising that in the study of production functions in education no similar interest has been developed, and particularly, that in the literature very little has been proposed on the learning process at school other than through a linear specification.<sup>1</sup>

In this paper, we propose the reference to S-curve functional forms. In the context of a production function in education, S-curves almost called for themselves since the output variable is limited in values ranging from 0% to 100% with the inconvenience these limitations force on the error terms of the specification.

---

<sup>1</sup> In the education literature one exception is the paper by Sørensen and Hallinan (1977). See also, the critical comment of Hauser (1978) on this last paper. In their book, McKenzie and Staaf (1974) referred very briefly to an "S"-shaped learning curve. Becker (1983) in his three part series on research methodology in economic education refers to a probit specification in the analysis of course grades. Hanushek (1979, p. 372), however, does not seem to consider this question as an important empirical issue mainly because of the limited ranges of variation observed for the education variables, "distinguishing among alternative functional forms is often impossible".

Furthermore, generally speaking, limited success has been obtained with the usual linear regression analysis of the determinants of academic achievement and particularly in establishing the importance of specific school intervention programs.<sup>2</sup> Also numerous authors have been concerned with the influence of socio-economic variables, school effects and school intervention variables on high versus low achievers.<sup>3</sup> For example, it is generally believed that a teacher who has more experience and schooling might produce different results on the student's academic achievement according to the student's success at school.

In regression analysis, one way to deal with this problem is to introduce interaction variables between the test results of the student, usually the output variable of the model, and the concerned input (explanatory) variables. This procedure leads to a nonlinear specification with as many parameters to be estimated as there are interaction variables specified. Another possibility is to split the sample between high and low achievers and run separate regressions on each group. The problem here is to determine where to split the sample and how to compare the coefficient estimates of the two regressions.

---

2 See Hanushek (1981) for a very comprehensive review.

3 See Summers and Wolfe (1977), Murnane (1975).

The reference to S-curve functional forms is a clear alternative.

Explicitly, let's consider a logit function linking the test result  $Y$  of student  $i$  to a set of  $k$  explanatory variables  $X$ :

$$Y_i = \frac{1}{1 + \exp(-\sum_{\lambda=1}^k \beta_{\lambda} X_{i\lambda})} \quad 1)$$

or

$$\ln\left(\frac{Y_i}{1-Y_i}\right) = \sum_{\lambda=1}^k \beta_{\lambda} X_{i\lambda}.$$

The equation (1) implies that  $0\% < Y_i < 100\%$ . For any  $X_i$ , say  $X_{ij}$ , the effect of  $X_{ij}$  on  $Y_i$  depends on the level of  $Y_i$  as:

$$\frac{\partial\left(\frac{\partial Y_i}{\partial X_{ij}}\right)}{\partial Y_i} = \beta_j \left[ \frac{-2 Y_i^2 + Y_i}{1 - Y_i} \right] \exp \left[ -\ln\left(\frac{Y_i}{1 - Y_i}\right) \right]. \quad (2)$$

With the inflexion point being at  $Y_i = 50\%$ , for  $Y_i < 50\%$ , the expression (2) is positive for a positive  $\beta_j$ . In other words, the effect of  $X_{ij}$  on  $Y_i$  increases with  $Y_i$  up to the level of 50%. This effect of  $X_{ij}$  on  $Y_i$  decreases for  $Y_i$  greater than 50%. So, with this functional form there is an a priori assumption that it is easier for any effective variable to improve the grades of a low achiever than those of a high achiever.

This point is crucial: since the linear case assumes no distinctive effect of a given variable between low and high achievers, the point estimate of the effect of this variable tends to average out over all students. If empirically the differential impact assumption is supported by the data, then one might erroneously conclude of no effect from a linear specification whereas the variable might in fact significantly affect the low achievers.

One appealing feature of the logit specification is its linearization when coping with estimation problems. One of its drawbacks is that the inflexion point is necessarily at 50%.

However, at the cost of a nonlinear estimation problem, we can go around this last restriction with the inverse power transformation specification (referred as IPT) proposed by Gaudry (1981):

$$Y_i = \frac{[\lambda \exp(\sum_l \beta_l X_{il}) + 1]^{1/\lambda + \mu}}{[\lambda \exp(\sum_l \beta_l X_{il}) + 1]^{1/\lambda + 1 + \mu}}, \lambda \neq 0$$

(3)

$$= \frac{\exp[\exp(\sum_l \beta_l X_{il})] + \mu}{\exp[\exp(\sum_l \beta_l X_{il})] + 1 + \mu}, \lambda \rightarrow 0$$

with the constraints

$$\lambda \exp \left( \sum_{\ell} \beta_{\ell} X_{i\ell} \right) + 1 > 0$$

and  $-\left[ \lambda \exp \left( \sum_{\ell} \beta_{\ell} X_{i\ell} \right) + 1 \right]^{1/\lambda} < \mu$  if  $\lambda \neq 0$ ,

$$-\left\{ \exp \left[ \exp \left( \sum_{\ell} \beta_{\ell} X_{i\ell} \right) \right] \right\} < \mu$$
 if  $\lambda \rightarrow 0$ ,

where  $\lambda$  and  $\mu$  are respectively the symmetry and level parameters of this S-curve. These constraints imply  $0\% < Y_i < 100\%$ . With this functional form, the inflexion point is determined by the data. Furthermore, with  $\lambda = 1$  and  $\mu = -1$ , the logit model becomes a special case of the IPT model.

### An Empirical Illustration

Using a sample of 872 first graders attending the Montreal francophone school system in 1979-80 we applied the linear, logit and IPT specifications to a model explaining students' results to a mathematics achievement test. In Table 1, we present the variables of the model with their sample means and standard deviations. The variables retained are traditional in education production function studies. They are related to the students' personal characteristics, socio-economic and school factors. The class size variable is an interesting intervention variable which has produced controversial results in the literature.<sup>4</sup>

<sup>4</sup> See Hanushek (1981).

For each specification in Table 1, we present the regression coefficients and identify which variables of the model are statistically significant to explain the students' mathematical test results. Except for the father's education variable, all the signs are the same, and except for the age and absence variables, all the variables significant in one specification are also found significant in the others. These results could not discriminate among the functional forms and indeed the computation of the square product-moment correlation coefficients between fitted and observed values of the students' mathematical results could not help to identify the best specification, since it yields comparable values of 0.4046, 0.3904, 0.4053 for respectively the linear, logit and IPT specifications.<sup>5</sup>

However, the level and the symmetry parameters of the IPT specification reject the logit function, as point estimates give us  $\hat{\mu} = -0.0399$  and  $\hat{\lambda} = 2.798$ , compared to respectively -1 and 1 in the logit model.<sup>6</sup> Furthermore, a likelihood ratio test (value of 10.884) of the IPT versus logit turned out significant at the 5% level ( $\chi^2_{2;5\%} = 5.991$ ).

---

5

$$\text{Computed from the formula: } r^2 = \left[ \frac{\sum_i (Y_i - \bar{Y}) (\hat{Y}_i - \bar{\hat{Y}})}{(\sum_i (Y_i - \bar{Y})^2 \sum_i (\hat{Y}_i - \bar{\hat{Y}})^2)^{1/2}} \right]^2.$$

Note that in the linear case  $r^2$  equal to the square of multiple correlation ( $R^2$ ).

<sup>6</sup> The estimation of the IPT function was obtained by a program developed by Laferrière and Gaudry (1985).



To compare IPT with the linear alternative specification we used nonnested hypotheses tests proposed by Davidson and MacKinnon (1981).

Assuming  $H_0$ : IPT specification and  $H_1$ : linear specification, we first tried the P-test procedure which is based on regression:

$$Y_i - \hat{f}_i = \alpha(\hat{g}_i - \hat{f}_i) + \hat{F}_i b + \varepsilon_i, \quad (4)$$

where

$$\hat{f}_i = \frac{[\hat{\lambda} \exp(\sum_{\ell} \hat{\beta}_{\ell} X_{i\ell}) + 1]^{1/\hat{\lambda}} + \hat{\mu}}{[\hat{\lambda} \exp(\sum_{\ell} \hat{\beta}_{\ell} X_{i\ell}) + 1]^{1/\hat{\lambda}} + 1 + \hat{\mu}},$$

$$\hat{g}_i = \sum_{\ell} \hat{\beta}_{\ell} X_{i\ell},$$

$\hat{F}$  is a row vector containing the derivatives of  $f$  with respect to the parameters  $\beta$  for the  $i$ th observation, evaluated at  $\hat{\beta}$  and, finally,  $\varepsilon_i$  is the error term.  $H_0$  is accepted if  $\alpha$  is not significantly different from zero. The estimation of (4) gives  $\hat{\alpha}_p = 0.102$  with a  $t$  statistic of 0.288 to accept  $H_0$ : IPT specification.

Alternatively with  $H_0$ : linear and  $H_1$ : IPT, a comparable J-test based on estimating the regression  $Y_i = (1-\alpha) g_i + \alpha \hat{f}_i + \varepsilon_i$ , yields  $\hat{\alpha}_j = 0.831$  with a  $t$  statistic of 3.503 to reject  $H_0$  in favor of the IPT specification.

With IPT identified as the best functional form to explain the data, the assumption of differential impact of variables following the status of the student as a low or high achiever is therefore supported by the data. To mark the differences on policy implication of such result, we finally report in Table 1 the elasticities computed at mean values of the variables of the model. Except for the self-concept test variable, we noted substantial differences for the other statistically significant variables: the I.Q. and the mother's education variables present greater point elasticities associated with the IPT specification than with the linear function. A particularly interesting case is the class size variable which shows a much greater point elasticity (in absolute terms) with the IPT function. Since with this specific data set the IPT inflexion point is at 68%, almost exactly at the mean value of the mathematical test variable, a policy of smaller class size for all classes in our sample with average maths results lower than 68%, will be encouraged by our study but could have been ignored with a linear specification for the model.

## Conclusion

Welfare of the society and sustained economic growth are often associated by scientists and political authority with the production of a quality human capital stock. The process of learning at school is a small but important link in that production. Current state of the arts shows that the production functions of education and their implications for policy analysis is a very complex problem yet to be entirely resolved. Our approach and illustration from a functional form viewpoint should hopefully enlarge the possibilities to cope efficiently with this problem.

Table 1  
Descriptive Statistics,  
Coefficient and Elasticity Estimates

Variable	Sample Mean (Standard Deviation)	Coefficient Elasticity		
		Linear <sup>a</sup>	Logit <sup>a</sup>	IPT <sup>b</sup>
Achievement in Mathematics (scale: 0-100)	68.49 (16.40)			
Self-Concept Test (scale: 0-40)	29.96 (5.51)	0.462 <sup>c</sup> 0.2019	0.0213 <sup>c</sup> 0.1769	0.0687 <sup>c</sup> 0.2014
I.Q. in Kindergarten (scale: 0-1000)	516.96 (60.87)	0.133 <sup>c</sup> 1.0011	0.00825 <sup>c</sup> 1.1798	0.0294 <sup>c</sup> 1.4856
Absences (half day) x 10	62.62 (61.76)	-0.0139 <sup>c</sup> -0.0127	-0.000787 -0.0136	-0.00336 -0.0206
Age (months)	86.10 (4.33)	0.202 <sup>c</sup> 0.2537	0.0122 0.2901	0.0179 0.1507
Father's Education (years)	10.85 (4.47)	0.181 0.0287	-0.00596 -0.0179	-0.0367 -0.0390
Mother's Education (years)	10.43 (3.40)	0.629 <sup>c</sup> 0.0957	0.0461 <sup>c</sup> 0.1329	0.164 <sup>c</sup> 0.1678
Teacher's Experience (years)	21.25 (7.62)	-0.0747 -0.0232	-0.00216 -0.0127	-0.00748 -0.0156
Teacher's Education (years)	14.65 (1.59)	0.112 0.0239	0.0138 0.0559	0.0511 0.0734
Teacher's Absences (half day) (x 10)	298.01 (178.20)	0.00107 0.0047	0.000118 0.0092	0.000492 0.0144
School Principal's Experience (years)	11.60 (5.15)	0.0190 0.0032	0.00528 0.0169	0.0147 0.0167
Class Size	23.76 (3.62)	-0.309 <sup>c</sup> -0.1073	-0.0199 <sup>d</sup> -0.1309	-0.0687 <sup>c</sup> -0.1599

Notes:

- <sup>a</sup> The level of statistical significance was established from the usual t-statistics.
- <sup>b</sup> The level of statistical significance was established from likelihood ratio test carried out for each variable.
- <sup>c</sup> Significant at the 5% level.
- <sup>d</sup> Significant at the 10% level.

## References

- Becker, Jr. W.E. (1983), "Economic Education Research: Part III, Statistical Estimation Methods", The Journal of Economic Education 14 (3), 4-15.
- Davidson, R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", Econometrica, 49 (3), 781-793.
- Gaudry, M.J.I. (1981), "The Inverse Power Transformation Logit and Digit Mode Choice Models", Transportation Research, Part B, 15B (2), 97-103.
- Hanushek, E.A. (1979), "Conceptual and Empirical Issues in the Estimation of Educational Production Functions", The Journal of Human Resources, 14 (3), 352-387.
- Hanushek, E.A. (1981), "Throwing Money at Schools", Journal of Policy Analysis and Management, 1 (1), 19-41.
- Hauser, R.M. (1978), Comment on "Reconceptualization of School Effects", Sociology of Education, 51, 68-72.
- Laferrière, R. and M.J.I. Gaudry (1985), "IPT-BL: A Program for the Inverse Power Transformation Binomial Logit Model", Centre de recherche sur les transports, University of Montreal, Forthcoming publication.
- McKenzie, R.B. and R.I. Staaf (1974), An Economic Theory of Learning, a Public Choice Monograph, University Publications, Virginia.
- Murnane, R.J. (1975), "Impact of School Resources on the Learning of Inner City Children", Ballinger Publishing Co., Cambridge, Mass.
- Sørensen, A.B. and M.T. Hallinan (1977), "A Reconceptualization of School Effects", Sociology of Education, 50, 273-289.
- Summers, A. and B. Wolfe (1977), "Do Schools Make a Difference?", American Economic Review, 67 (4), 639-652.