

American Economic Association

Guessing and the Error Structure of Learning Models

Author(s): Michael K. Salemi and George E. Tauchen

Source: *The American Economic Review*, Vol. 70, No. 2, Papers and Proceedings of the Ninety-Second Annual Meeting of the American Economic Association (May, 1980), pp. 41-46

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/1815437>

Accessed: 01/08/2009 20:57

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=aea>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Economic Review*.

<http://www.jstor.org>

Guessing and the Error Structure of Learning Models

By MICHAEL K. SALEMI AND GEORGE E. TAUCHEN*

This paper is broadly concerned with problems associated with the use of test score data to infer the relative strength of inputs to the production of learning. It should be of interest to those employing a typical strategy of economic education research: pretesting students, applying some special educational treatment to a subset of students, and posttesting the students. By this strategy the researcher enquires whether the treated group learned more or learned more efficiently. This paper specifically addresses several concerns raised by Thomas Johnson by laying out a set of structural equations and thereby modelling the probability of answering a test item correctly and by dealing in a novel way with the hypothesis that some test questions are more difficult than others.

The key structural equation of the model is the learning production function which explains student learning in terms of student aptitude and study time. Because student aptitude cannot be observed, a standard procedure has been to use the pretreatment test score as an aptitude proxy. William Becker and Salemi have pointed out that such a procedure can seriously bias estimates of the production function parameters. The approach of this paper is to model explicitly the probability distribution of the test score of a student conditional on his ability level. This approach is similar to that used to estimate models with qualitative dependent variables. But the test score model suggests a convenient approximation and a

correction to least squares which together deliver consistent estimates of the structural parameters.

I. The Structural Equations

A. The Test Score Model

It is assumed that a student's ability to master economic concepts can be represented by a one-dimensional index, A . Although A cannot be observed directly, precourse test data permit the researcher to control for A . Consider a multiple choice test containing M questions divided on a priori grounds into K difficulty categories with M_k questions in the k th category.¹ Following Frederick Lord and Melvin Novick, define the indicator variable I and α_k so that

(1a)

$$I \begin{cases} > \alpha_k & \text{if a student knows the answer to} \\ & \text{a question of difficulty level } k \\ < \alpha_k & \text{otherwise} \end{cases}$$

Further define the dichotomous variable Q_{jk} so that

(1b)

$$Q_{jk} \begin{cases} = 1 & \text{if the student answers the } j\text{th} \\ & \text{question in the } k\text{th difficulty} \\ & \text{category correctly} \\ = 0 & \text{if not} \end{cases}$$

*Assistant professors, University of North Carolina-Chapel Hill and Duke University, respectively. This paper reports the highlights of work discussed in more detail in a paper available from us on request. We thank but absolve from responsibility Bill Becker, Lee Hansen, Tom Johnson, John Siegfried, and the members of the North Carolina State University econometrics seminar. We thank Augusto Lopez-Claros for competent help in carrying out the estimation.

¹We have the Test of Understanding of College Economics (TUCE) in mind. The TUCE designers identify three difficulty levels of questions: those requiring students to recognize and understand economic concepts; those requiring students to apply a single concept; and those requiring students to apply concepts in a complex way.

The distinction between I and Q_{jk} is important because a student will generally guess the answer when he does not know it. The link between aptitude and the indicator function is given by

$$(2) \quad I = A + U$$

where U is a random variable with symmetric cumulative distribution function F .² Equation (2) says that whether or not a student knows the answer to a test question depends both on the student's aptitude and a variety of omitted factors the total effects of which are represented by U .

Define $\pi_k(A)$ to be the probability that a student with aptitude A correctly answers a question of difficulty level k . Let λ be the probability that he guesses correctly if he does not know the answer. It is assumed that λ does not vary with aptitude or with the difficulty level of the question. Using the symmetry of F one may deduce

$$(3)$$

$$\pi_k(A) \equiv \text{Prob}(Q_{jk} = 1) = \lambda + (1 - \lambda)F(A - \alpha_k)$$

which says that the probability that a student answers correctly a question with difficulty level k is positively related to A and λ , and inversely related to α_k .

Define S_k to be a student's score on the questions comprising the k th difficulty category. We assume that S_k is the sum of M_k independent Bernoulli trials each with probability of success equal to $\pi_k(A)$. Then, conditional on A , S_k is approximately normal with mean of $M_k\pi_k(A)$ and variance of $M_k\pi_k(A)(1 - \pi_k(A))$. As seen in (3), S_k is a non-linear function of aptitude. To put the model in linear form, Taylor approximate F about zero, the mean value for U .

$$(4) \quad F(A - \alpha_k) \doteq F(0) + F'(0)(A - \alpha_k)$$

²Specifying properties of the cumulative distribution function F amounts to choosing units in which to measure A . Below we impose $F(0) = .50$ and $F'(0) = (\sqrt{2\pi})^{-1}$ which imply that the difference between A and α_k is measured in units of standard deviations of U .

Equations (3) and (4) and the normality of S_k imply that the relationship between aptitude and test score is

$$(5) \quad S_k = C_{1k} + C_{2k}A + e_k$$

where $C_{1k} \equiv M_k(\lambda + (1 - \lambda)F(0) - (1 - \lambda)F'(0)\alpha_k)$, $C_{2k} \equiv M_k(1 - \lambda)F'(0)$, and where $E(e_k|A) = 0$. It is important to our approach that (5) holds for both pretest and posttest scores.

B. The Learning Production Function

It is assumed that the learning process may be thought of as a production process in which the output is increased economics aptitude. The inputs to the learning process are student time, student ability at the beginning of the course, and a vector of characteristics describing the learning environment.

We suppose that the production function has the form

$$(6) \quad L \equiv A' - A = \theta_1 A + \theta_2 T + \theta_3 EN \\ + \theta_4 A \cdot T + \theta_5 A \cdot EN + \theta_6 T \cdot EN$$

where learning L is the increment to pre-course aptitude A ; T is student study time, EN is a vector of variables describing the learning environment, and the θ_i are the parameters of the learning production function (θ_3 , θ_5 , and θ_6 are vectors conformable with EN). A priori one would expect to find θ_1 , θ_2 , and θ_4 greater than zero. Of particular interest are the marginal learning product of aptitude equal to $\theta_1 + \theta_4 T + \theta_5 \cdot EN$ and the marginal learning product of student study time equal to $\theta_2 + \theta_4 A + \theta_6 EN$. Each of these terms has components representing the direct marginal contribution of the factor, the marginal contribution due to the interaction of aptitude and time, and terms which permit the marginal product to be different in various learning environments.

Note that (6) assumes an exact production function so that all the error in the model is due to U in equation (2). This assumption simplifies estimation of the model since it implies that the error in the

test score equation, (5), has the same distribution in the postscore and prescore cases.

II. The Corrected Least Squares Procedure

The basic idea of this section is that the use of student precourse test scores as a proxy measure for aptitude requires a correction to least squares in order to achieve consistent estimation of the production function parameters. The correction is based on certain population regressions implied by the model and amounts to a pretreatment of the data. In the psychometric literature, population regressions are often used to address the question of correcting test scores to obtain more accurate rankings of students by ability. In contrast, we use these population regression corrections not to rank students but rather to obtain consistent estimates of the parameters of interest. A further point developed in this section is that a priori information which permits test questions to be grouped into difficulty categories also permits a test of the model's validity because it introduces restrictions across the parameters of the regression equations.

A. The Prescore as a Proxy for Aptitude

Equation (5) may be inverted to yield aptitude as a function of a student's pretest score.

$$(7) \quad A = (S_k - C_{1k} - e_k) / C_{2k}$$

There are two points to be made by way of interpreting (7). First, S_k is an error-ridden proxy for aptitude as the presence of e_k in (7) makes clear. Second, explicitly taking account of guessing via the parameter λ alters the implication for aptitude of a change in S_k . For example, the model predicts that a one point increase in the pretest score *ceteris paribus* should be interpreted as an aptitude increase of $(M_k \cdot F'(0)(1 - \lambda))^{-1}$.

To derive the relationship between the posttreatment test score and the pretreatment test score, substitute (7) and (6) into the version of (5) valid for the posttreatment

score and obtain

$$(8) \quad S'_k = -C_{1k}\theta_1 + (1 + \theta_1)S_k \\ + (C_{2k}\theta_2 - C_{1k}\theta_4)T + (C_{2k}\theta_3 - C_{1k}\theta_5)EN \\ + \theta_4 T \cdot S_k + \theta_5 EN \cdot S_k + C_{2k}\theta_6 T \cdot EN \\ + e'_k - (1 + \theta_1 + \theta_4 T + \theta_5 EN)e_k$$

for $k = 1, 2, \dots, K$. Equation (8) is the model's prediction concerning the interrelationship among the observable variables but it is not a proper regression equation since by (5) e_k and S_k are correlated. In (5), $E(e_k|A) = 0$, but in (8) the conditional expectation of e_k given S_k , T , and EN is not, and it is this fact which implies that estimation of (8) in its uncorrected form will not yield consistent estimates of the parameters of interest.

B. The Correction to Least Squares

It is possible to derive an expression for $E(e_k|S_k, T, EN)$ and thus to write a corrected version of equation (8). Assume that T and EN are strictly exogenous variables and are thus uncorrelated with e_k . Next write

$$(9) \quad E(e_k|S_k, T, EN) = b_{0k} + b_{1k}S_k \\ + b_{2k}EN + b_{3k}T$$

Equation (9) is the population regression of the error from the pretest score equation on the observed explanatory variables. Our strategy is to employ (9) to correct (8). For the general errors in variables problem this strategy will fail since $Cov(e_k, S_k)$ would be unknown; however, the structure of the test-score model implies

(10)

$$Cov(S_k, e_k) = (M_k^2 \bar{\pi}_k - M_k^2 \bar{\pi}_k^2 - \sigma_{S_k}^2) / (M_k - 1)$$

where $\bar{\pi}_k$ is a sample mean and $\sigma_{S_k}^2$ a sample variance. Equation (10) together with the sample covariances of S_k , EN , and T are sufficient information to estimate the b_{jk} .

Use (9) together with (8) to obtain

$$(11) E(S'_k|S_k, T, EN) = -C_{1k}\theta_1 \\ + (1 + \theta_1)\tilde{S}_k + (C_{2k}\theta_2 - C_{1k}\theta_4)T \\ + (C_{2k}\theta_3 - C_{1k}\theta_5)EN + \theta_4T \cdot \tilde{S}_k \\ + \theta_5EN \cdot \tilde{S}_k + \theta_6C_{2k}T \cdot EN$$

where $\tilde{S}_k = S_k - b_{0k} - b_{1k}S_k - b_{2k}EN - b_{3k}T$. A comparison of equations (11) and (8) makes it clear that a correction to the least squares procedure is required to produce consistent estimates of the structural parameters. For example, (11) shows that the probability limit of the uncorrected least squares coefficient of S_k is $(1 + \theta_1)(1 - b_{1k})$ rather than $(1 + \theta_1)$ as (8) might seem to suggest.

C. Parameter Identification and the Estimation Procedure

To simplify notation rewrite equation (11) as

$$(12) E(S'_k|S_k, T, EN) = B_{k0} \\ + B_{k1}\tilde{S}_k + B_{k2}T + B_{k3}EN + B_{k4}T \cdot \tilde{S}_k \\ + B_{k5}EN \cdot \tilde{S}_k + B_{k6}EN \cdot T$$

For $K=3$, (12) is a system of three equations and twenty-one coefficients. However, the twenty-one coefficients are each functions only of nine structural parameters $(\theta_1, \theta_2, \dots, \theta_6, \alpha_1, \alpha_2, \alpha_3)$ and the three normalizing parameters $(\lambda, F(0), F'(0))$. The restrictions that the B_{kj} obey are apparent from a comparison of (11) and (12). Estimation of (12) with and without the restrictions in force provides a natural test of the validity of the model.

The procedure we suggest to estimate the parameters is comprised of three parts. First estimate the b_{kj} of (9) and form the \tilde{S}_k . Recall that this strategy is available to us precisely because our test score model tells us how e_k and S_k covary. Second, estimate (12) by ordinary least squares using the cor-

rected prescores but not imposing the restrictions. Estimates of the residuals from these regressions can be used to test for heteroskedasticity and, if needed, to obtain consistent estimates of the variances of the error terms. The sums of squared residuals of these unconstrained regressions are also needed to form a test of the validity of the restrictions. Third, estimate the restricted model by an appropriate non-linear procedure.

III. Empirical Results

The results of applying our suggested estimation procedure to the data set described in Becker and Salemi are briefly summarized in Table 1. Because preliminary estimation suggested that the interactions between study time and environment and between study time and aptitude provided for in (6) were unimportant, they were excluded from the regressions reported. Two models were estimated. In Model I, the production function includes ability, study time, and dummy variables designating the school attended by the students. In Model II, the production function includes the Model I variables plus interactions between the school dummies and aptitude. Unconstrained estimates with both aggregated ($M=33$) test score data and disaggregated data ($M_k=11, k=1, 2, 3$) were obtained by ordinary least squares. Constrained estimates with the disaggregated data were obtained by full-information maximum likelihood (FIML) using a version of the Davidon-Fletcher-Powell algorithm described by Karl Jöreskog. The χ^2 test for heteroskedasticity suggested by T.S. Breusch and A. R. Pagan was performed using the corrected prescore data in both aggregated and disaggregated cases. These tests fail to reject the null hypothesis of no heteroskedasticity at standard significance levels.³

³Inspection of (8) will suggest immediately that heteroskedasticity may be a problem in this model. However, as we show in our more detailed paper, heteroskedasticity is likely to be less serious a problem when the errors in variables problem is more serious. For our data the errors in variables problem is quite serious and thus we are not surprised by the results of these tests.

TABLE 1—ESTIMATES FOR THE COEFFICIENTS AND STANDARD ERRORS OF ABILITY (θ_1) AND STUDY TIME (θ_2) IN THE LEARNING PRODUCTION FUNCTION

Model: Estimation Procedure	$\hat{\theta}_1$	$\hat{\theta}_2$
Becker and Salemi	.22(.23)	.05(.14)
A. Aggregated Test Scores Unconstrained <i>OLS</i>		
I:Uncorrected	-.307(.065)	-.003(.030)
I:Corrected	.488(1.61)	-.008(.099)
II:Uncorrected	-.307(.212)	-.009(.030)
II:Corrected	.502(.524)	-.014(.030)
B. Disaggregated Test Scores Unconstrained <i>OLS</i>		
I:Corrected ($k=1$)	12.82(2.21)	.048(.015)
I:Corrected ($k=2$)	2.63(.914)	-.034(.017)
I:Corrected ($k=3$)	.326(.271)	.006(.015)
II:Corrected ($k=1$)	9.44(5.36)	.050(.016)
II:Corrected ($k=2$)	2.72(2.46)	-.035(.017)
II:Corrected ($k=3$)	.386(.825)	.006(.015)
C. Disaggregated Test Scores Constrained <i>FIML</i>		
I:Corrected	.368(.217)	-.003(.008)
II:Corrected	.504(.300)	-.003(.008)

Notes: *NOBS*=314. The *FIML* computations impose the normalizations $\lambda=.25$, $F(0)=.50$, and $F'(0)=(\sqrt{2\pi})^{-1}$. Becker and Salemi results were computed from their Table 3, p. 84, and were obtained by them via an instrumental variables approach.

The four major findings of our econometric work are as follows: First, when uncorrected test score data are employed the *OLS* estimates of θ_1 are negative and significant. However, with the corrected data the estimates of θ_1 are positive as economic theory would predict and marginally significant. Positive estimates of θ_1 are obtained whether Model I or Model II is estimated, whether aggregated or disaggregated data are employed, and whether or not the constraints are imposed. In every case the estimated value of θ_1 is greater than that reported by Becker and Salemi. For Model I, θ_1 is the marginal product of aptitude. For Model II, however, the marginal product of aptitude is permitted to vary across schools and is equal for school j to $\theta_1 + \theta_{5j}$. Using *FIML* on the disaggregated data it was found that the marginal product estimates ranged between .680 (.376) and .206 (.163).

Second, when the disaggregated data are employed imposing the restrictions implied by the model disciplines the estimation in an important way as a comparison of parts B and C in Table 1 readily makes clear. On the one hand then, the restrictions have content important for extracting the structural parameters. On the other hand, the data do not seem to accord well with these

restrictions. A comparison of the sum of squared residuals in the constrained and unconstrained cases was performed using the standard likelihood ratio test. The appropriate statistic is estimated to be 67.53 which just exceeds the ninety-ninth percentile of the χ^2 distribution with 42 degrees of freedom.

Third, the model returns estimates of the *alphas* from equation (1a) exactly in accordance with the view that category 1 questions require the least aptitude to master and category 3 questions the most. The actual levels for the *alphas* (which depend upon the normalizations reported in Table 1) are for Model II: 2.58(1.28), 2.74(1.22), and 3.60(1.43), respectively. An examination of a joint confidence ellipse suggests that the difference between *alpha* 1 and *alpha* 2 is statistically small, while that between *alpha* 2 and *alpha* 3 is large.

Fourth, somewhat surprisingly our estimates of θ_2 , the marginal product of time in the learning production function, are insignificantly different from zero except in the unconstrained disaggregated case. This may be due in part to the fact that students reported their own study time.

In conclusion, we would stress that the prescore correction suggested here provides

a useful alternative to instrumental variables procedures for estimating learning models. We also take the view that our results recommend further studies using disaggregated test score data to estimate parsimonious structural models. We conjecture that a useful alternative to our hypothesis of a constant guessing parameter, λ , is to model λ as a function of student aptitude. Finally, we recommend that further research be designed to measure student study time more directly.

REFERENCES

- W. Becker and M. Salemi**, "The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models," *J. Econ. Educ.*, Spring 1977, 8, 77-92.
- T. S. Breusch and A. R. Pagan**, "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, Sept. 1979, 47, 1287-94.
- T. Johnson**, "Research on Economic Education: How Well is it Answering the Questions Asked?," *Amer. Econ. Rev. Proc.*, May 1979, 69, 22-27.
- K. Jöreskog**, "A Computer Program for Minimizing a Function of Several Variables," mimeo., Educational Testing Service, Princeton 1973.
- Frederic Lord and Melvin Novick**, *Statistical Theories of Mental Test Scores*, Reading, Mass. 1968.