An Inductive Approach to Production-function Modeling: A Comparison of Group Method of Data Handling (GMDH) and Other Neural Network Methods

Bruce Baker
Department of Teaching and Leadership
12-1A Bailey Hall
The University of Kansas
Lawrence, KS 66045
(785) 864-9844
bdbaker@ukans.edu
www.soe.ukans.edu/baker

DRAFT
DO NOT CITE OR QUOTE WITHOUT PERMISSION

ANNUAL MEETING OF THE AMERICAN EDUCATION FINANCE ASSOCIATION
MOBILE, ALABAMA
MARCH 13, 1998

An Inductive Approach to Production-function Modeling: A Comparison of Group Method of Data Handling (GMDH) and Other Neural Network Methods

Bruce Baker
The University of Kansas

Abstract

This study addresses the susceptibility of the traditional theoretical framework of the educational production-function to political influence and personal bias. The politics of the production-function are becoming increasingly obvious in educational research as the battle over whether money matters in improving educational productivity continues (Hanushek, 1989-1997; Hedges et. al., 1994,1996). One common thread of the studies on both sides of the money issue is the predominantly deductive approach that is used. Through the purely deductive approach, researchers are driven to test the statistical significance of preconceived notions, political or personal in origin, about how the educational system works. As an alternative approach, inductive, non-linear pattern recognition algorithms are applied to school productivity data from the state of Vermont. These algorithms, known as Neural Networks, yield optimal solutions that differ both in functional form and parameter significance from many of our preconceived notions.

Introduction

DOES MONEY MATTER? or is this really even the question? And if it's not, then why do we keep asking it, over and over (Hanushek, 1989-'97; Hedges, Laine and Greenwald, 1994, 1996)? For his most recent meta-analysis of this question, Hanushek (1997) combines the results of <u>90</u> studies regarding the relationship between schooling inputs and student outcomes. Betts (1996) uses approximately <u>60</u> related studies similarly in his assessment of the relationship between school spending, spending-related factors and student earnings. Yet, while these meta-analyses provide us with an insightful overview of existing educational productivity research, and a relative balance of political perspectives, we seem to be covering little new ground with regard to understanding the system of schooling as a whole.

Structurally, most studies of educational productivity consist of some statistical test of the effects of a selected set of schooling inputs (S) and student inputs (X) on student outcomes (Q). These tests are most often performed within a standard, deductive, hypothesis-testing framework. In addition to typical production-function analysis, there are a number of other methods for which the end result is quite the same. Among these methods are rate of return analyses (Betts, 1996) and cost-function analyses (Duncombe and Miner, 1996). At the most basic level, each of the methods rely on the determination of some mathematical relationship between spending and learning or spending and earning.

When it comes to whether money matters, it is quite easy to see how there can be both political and personal incentives for making arguments toward one side or the other of the issue. While it is expected that the cumulative outcomes of our biases will result in a balanced pool of useful information, eventually revealing a middle-ground, the tendency has been for the pool of results to become more polarized over time. How do we go about breaking the gridlock on this question? And, more importantly, should our efforts continue to be so narrowly focussed on the many permutations of this question? Perhaps our problems lie more deeply in the processes of educational research itself. Perhaps the methods by which we generate and subsequently test the “important questions of today", themselves, are too restrictive.

Potential Pitfalls of Exclusively Deductive Processes

As researchers in the social sciences, we find ourselves tightly bound to testing the rationality of existing and proposed theories. A vast amount of our research consists of statistical hypothesis tests based on some theoretical framework. Research lacking such a framework is unlikely to be granted credibility in the educational research community. Ultimately, in quantitative analyses, these theoretical frameworks are operationalized into formal mathematical models[1], equation types[2], and even functional forms[3]. But, where do these theories and models come from?

The research process is designed such that new theories are intended to evolve from novelties in previous findings, that these new theories will be validated in subsequent studies that may themselves reveal new novelties, and so on. Two distinct problems may inhibit this process. First, ongoing research on a

---

[1] Median Voter Model, Political Support Maximization Model etc.

[2] Production-Function etc.

given theoretical framework may at some point become stagnant, revealing little novelty in its findings, yet persisting nonetheless, "spinning the wheels of the research process." Second, the selection and combination of previous findings toward the generation of novel ideas must occur solely within the mind of the researcher(s).

Statistical Methods: Technical Limitations and Potential Political Influence

Beyond the formation of the research question is a limited set of statistical methods. Linear correlation and multiple linear regression analysis are among the methods that lend themselves to traditional statistical hypothesis testing. The basic objective, in either case, is to determine whether the prescribed inputs and outcomes share a statistical relationship (generally linear) other than "0."[4] Yet, for statistical hypothesis testing to occur, two key a priori decisions must be made: 1) the parameters to be tested, and 2) the shape of the frame (functional form) to be applied to the relationship. While these issues are seemingly technical, both parameter selection and determination of functional form are susceptible to the influence of personal and/or political biases.

The question of specific parameters to be tested is an applied extension of the research question. Among the important considerations are: 1) how to

---

[3] Linear, Logarithmic, Exponential, Multiplicative

[4] with the usual acceptable probability of being wrong set at 5%.

measure the constructs in question and 2) how to organize the parameters into an equation structure that will provide us with the appropriate test statistics to identify their significance. Verstegen (1997) points out that even "order of entry" can play a significant role in discriminating between the effects of school and non-school inputs.

While it is quite easy to see how parameter selection can be used to promote political biases regarding spending, the political undertones of functional forms are somewhat more elusive. Cohn & Geske (1990) note: “A linear relationship between X inputs and the Q outputs would be empirically valid to the extent that the curvature of the total output function is only mildly violated by employing a linear approximation.”[5] A common economic assertion with regard to shape of the “actual curve” is that “the education production, like all well-behaved production-functions[6], is subject to diminishing returns”(Betts, 1996).[7] This behavior is generally well captured (e.g. displays a significant, positive coefficient less than 1.0) by applying a log-log specification of wages relative to per pupil spending (Betts, 1996). Others have replaced the log of spending with a

[5] p. 166

[6] It is somewhat ironic that we consider a production to be "well behaved" based on the extent to which it conforms to our theoretical expectations and functional forms as this assessment assumes that we truly know how a production-function is supposed to behave.

[7] p. 163

quadratic function, achieving a similar interpretation (Johnson & Stafford, 1973).[8]

While a significant parameter under either of these circumstances signifies a statistically adequate fit[9], it does not necessarily signify the best, or even most appropriate fit. Nor are we likely to find the best fit by these methods. In particular, where patterns of diminishing returns are expected, placing a rigid frame (predefined functional form) over a data set is unlikely to reveal local optima, or important decision points. In rate of return studies, where determinations are made along a log-log production-function curve, results at different levels of inputs can be nearly as misleading as the linear specification.[10] The differential rates of response of outcomes to changes in inputs in a priori non-linear specifications can have serious political implications. For example: At what point along the production-function curve does spending more yield negligible performance returns? The answer to this question varies depending on the shape of the curve which will vary depending on the functional form selected.

New Approaches to Modeling

To synthesize the issues presented in the previous section, the key problem

[8] It is important to note that differential treatment of inputs and outcomes in the same model may compromise the actual relationship between the variables.

[9] At its most basic level, the significant parameter simply signifies that the fit is better than "nothing."

[10] Mathematically, the only difference that exists in applying the log-log specification is that we are forcing a slightly bent rigid frame onto the data rather than an entirely

with using purely deductive processes for studying systems as complex as our education system is that we must assume that we actually understand how the system works, by way of existing theory, before we even begin our analyses. In light of the lack of novelty of our recent findings, and continued indecision regarding the role of fiscal resources, perhaps we should humbly begin to recognize our ignorance, take a step back, and ask the more basic question: "What makes schools tick?"[11]

This, however, is a dauntingly inductive question and one for which we have not necessarily developed or applied methodologies in the field of education. We may, however, gain some insights from theory and methods of other fields. The possibilities now range from simple pattern recognition algorithms (Farlow, 1984), or Neural Networks, to exploratory modeling (Bankes, 1993), to complex adaptive systems modeling (Minar, Burkhart, Langton and Askenazi, 1997). Each of these methods employs a balance of inductive experimentation and deductive testing in order to gain a deeper understanding of a system as a whole.

The example of inductive analysis presented in this study barely scratches the surface of what is currently possible. In Bankes' typology of data-driven, model-driven, and question-driven exploratory modeling, the example that

---

straight one.

[11] Allow me to operationalize this phrase to mean: How do the multiplicity of potential factors related to student performance collectively and interactively come together to yield successful students as measured by indicators of academic performance?

follows is primarily data- driven, with the simple goal of revealing patterns and relationships in the complex system of schooling.[12] In other circles these methods might be referred to as "data mining" or "knowledge extraction from data" (Lemke, 1997).

Such "data mining" methods have been considered theoretically unacceptable in many areas of economic and educational research and are often informally referred to as "going fishing." Yet how or why, then, have they gained acceptance in fields such as financial analysis (Lemke, 1997), medicine (Buchman et. al., 1994) and real estate valuation (Worzala, Lenk and Silva, 1995)? One could contend that the difference in the competitive nature of the fields has something to do with it. In financial analysis in particular, to achieve the competitive edge, it is extremely important for the analyst to find emerging patterns and understand the complexity of the financial system in ways that other analysts do not. In educational policy research, there is little or no competitive edge to be gained by greater understanding of the system because the system itself is generally non-competitive. Thus, while we recognize the importance of developing a greater understanding of the system, we are not similarly driven to seek out competitive advantage in analyzing the system.

[12] It should be noted that because these analyses are all ultimately deterministic and lacking measures of uncertainty, in Bankes' typology they would more likely be classified as Consolidative Modeling than actual exploratory modeling. For a more detailed discussion see Bankes (1993) p. 435.

Neural Networks are a class of self-organizing algorithms typically used for knowledge extraction. The algorithms are referred to as Neural Networks because, to some extent, they mimic the pattern recognition processes of the brain. There are, however, many types of neural networks, the most common being backpropagation algorithms which identify non-linearities in data through an iterative estimation procedure similar to maximum likelihood estimation.[13] Other Neural Network methods include the Generalized Regression Neural Network (GRNN) (Specht, 1991), a single pass non-linear estimation algorithm and Group Method of Data Handling (GMDH) (Farlow, 1984), a self-organizing polynomial optimizing algorithm. Lemke (1997) and Liao (1992) note the particular value of the GMDH algorithm for inductive data mining. Each of these methods is applied in the following analysis to determine its relative usefulness, in combination with currently accepted methods, for understanding educational productivity.

A Primer on Neural Networks

Model Structure

There are two basic differences between simple linear regression models and Neural Network models. The first is that the linear regression model, by

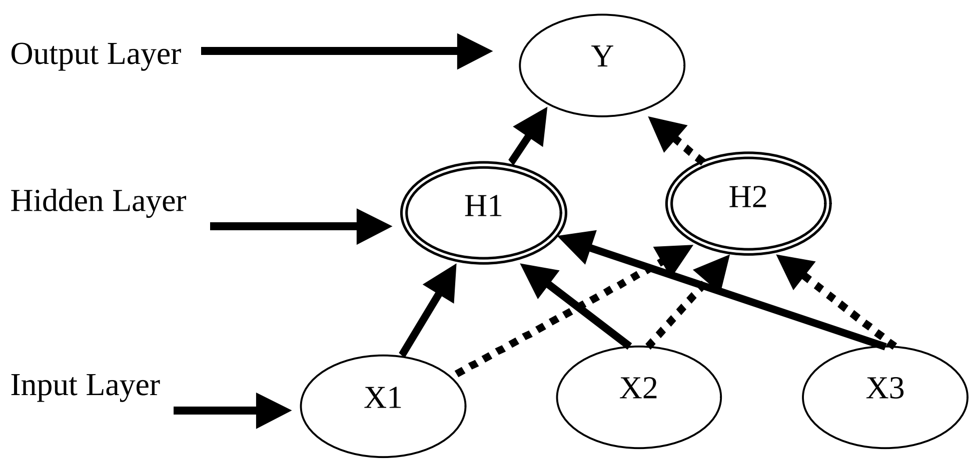
[13] The primary difference being the number and type of coefficients estimated and the existence of the "Hidden Layer" where non-linear connections are established between inputs.

definition, is linear in its parameters and the second is that the linear regression model contains no additional functions such as those which are found in the additional layers of a neural network (McMenamin, 1997). In Neural Network terms, the simple linear model:

$$\mathbf{Y = XB + u}$$

can be viewed as a "single output, feed forward system with no hidden layer and with a linear activation function" (McMenamin, 1997). To dissect this description of the linear model, let us compare the simple linear form with that of a relatively simple three layer feed forward Neural Network:

Figure 1: Network Diagram with 3 Inputs and Two Hidden Neurons



Adapted from McMenamin (1997)

The general form for the network presented in Figure 1 is:

$$\mathbf{Y = F[H_1(X), H_2(X), \ldots, H_N(X)] + u}$$

where the dependent variable, Y is a function, F, of explanatory variables, X, which have been re-scaled through a series of neural network functions, H. In general, these functions consist of a logistic, or "S" shaped activation function, sometimes referred to as a "hidden layer transfer function" (McMenamin, 1997).[14] These functions may also be referred to as *Squashing* functions (Rao and Rao, 1993).

Note in Figure 1 that all inputs feed forward into each hidden layer neuron. This is why some refer to Neural Network models as *connectionist* models (Buchman et. al., 1994). Duplication of inputs in the middle layer neurons appears to create irresolvable multicollinearities in the model. Duplication of inputs presents greater concern when the goal is to interpret the parameters and weightings of the middle layer. Most frequently, however, neural networks are applied for predictive purposes, rather than inference. The advantage gained by including two or more different mathematical treatments of the same inputs is that some of the inputs may be emphasized in one neuron, while others are emphasized in another neuron. Likewise, the degree of non-linearities and interactions between inputs in neurons may vary.

---

[14] A hyperbolic tangent function can also be used (Rao and Rao, 1993)

Estimation

While the model structure for Neural Networks is a departure from traditional regression modeling, methods for estimating network coefficients can be quite similar. In neural networks, the coefficients are referred to as connection strengths, or weights, constant terms are called biases and, at times, slopes are called tilts (McMenamin, 1997). In general, estimation procedures consist of some type of iterative algorithm that converges on a solution identified by a preset criterion such as mean square error. For example, a typical backpropagation network begins by randomly assigning weights. The residuals of the output are then assessed and the weights adjusted in the appropriate direction[15], until the network converges on the best predicting deterministic solution.

Another estimation procedure that is gaining popularity uses genetic algorithms to select optimum equation structures from a pool of randomly generated equations. *Neuroshell 2*, the software used in this study, includes this algorithm for selecting smoothing parameters in type of hybrid neural network known as a Generalized Regression Neural Network (GRNN).[16] While consistent predictive accuracy is generally attainable by such methods, equation structures and coefficient values may vary widely from one model to the next, making the

[15] Momentum terms are used to keep the weights changing in the established direction. One benefit of these terms is that they often keep the network from getting stuck in local minima (Rao and Rao, 1993)

[16] See *Neuroshell 2* User's manual p. 138.

inferential value of the models questionable.

Model Testing

With most Neural Networks, data are segregated into two classes: *Training Data* and *Test Data*. The training set data are the data to which the weights or coefficients are initially applied. In regression, training data are equivalent to the sample data used for estimation. For neural networks to converge on a generalizable solution through iteration there must be a test set, or extracted non-sample data, against which prediction accuracy is compared. The test set may be randomly extracted from the larger sample or, in the case of time-series, may consist of the most recent few events. *Neuroshell 2* includes an option to extract an additional subset of data called the *production set* (WSG, 1995 p. 101). The production set may include predictors for which the outcome measure (Y) is still unknown. The trained network is applied to the production set predictors to determine the new predicted outcomes.

A commonly expressed concern over flexible non-linear estimation methods is the tendency to overfit sample data (Murphy, Fogler and Koehler, 1994). In regression, as the number of predictors approaches the number of cases, we can achieve a near perfect fit to the outcome measure, but sacrifice the significance of the individual parameters, inferences that can be drawn from parameters, and the ability to generalize. It is assumed that due to the relatively high number of weights in Neural Networks that overfit would be equally likely and yield similar complications. Murphy, Fogler and Koehler (1994) note that as tolerance, nodes or layers are increased in backpropagation networks, while *training set* errors are asymptotic, *test set* errors fail to improve beyond an identifiable optimum. These findings provide bases for using *test set* error to optimize the model structure during training.

Neural Network Architectures

Until recently, backpropagation neural networks made up approximately 80% of all neural network applications (Caudill, 1995a). Use of backpropagation has declined due to the relatively long required training times for the iterative algorithm and the development of new, quicker estimation procedures (McMenamin, 1997). Figure 1 displays a structure common to backpropagation neural networks. Permutations of this structure include: (1) number of layers of the network (2) numbers of neurons in each layer and (3) numbers and locations of the connections. Backpropagation has been proven an effective tool for both time-series prediction (Hansen and Nelson, 1997; Lachtermacher and Fuller, 1995) and cross-sectional prediction (Buchman et. al., 1994; Odom and Sharda, 1994; Worzala, Lenk and Silva, 1995).

In addition to backpropagation, this study applies two hybrid methods that rely on neural network estimation methods to identify optimal non-linear regression models. Specht (1991) developed a flexible form of non-linear regression referred to as the Generalized Regression Neural Network (GRNN). GRNN removes the necessity to specify a functional form by making use of the probability density function of the observed data. The GRNN model interpolates the relationship between each input, and between the input and outcome measures, applying a smoothing parameter, α, to each relationship to moderate the degree of non-linearity. Optimized models generally include different smoothing parameters for each input (Specht, 1991). The difficulty is in estimating these parameters.

Two methods have generally been employed for estimating smoothing parameters: (1) the holdout method (Specht, 1991) and (2) genetic algorithms (WSG, 1995. pp. 198-205). The holdout method involves using randomly removed samples as a test set for the prediction accuracy of the model. Genetic Algorithms involve the random creation of sets of equations, followed by fitness testing and selective breeding; that is, equations with poor predictive power cease to exist, while smoothing parameters in the “fit” equations are randomly recombined to create a new pool to begin the next cycle. GRNNs have not been used as widely as backpropagation but are recommended for use with sparse data and are less sensitive to the scale of the data (Caudill, 1995c; Specht, 1991). GRNN has been proven effective as a prediction tool (Buchman and Kubos, 1994).

A.G. Ivakhnenko, in 1966, proposed an algorithm for fitting polynomial regression equations.[17] The algorithm, known as Group Method of Data Handling, estimates the best fit polynomial via the Kolmogorov-Gabor polynomial:

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x_i x_j x_k$$

Where $X(x_1,x_2, \ldots x_m)$ is the vector of inputs and $A(a_1,a_2, \ldots a_m)$ is the vector of coefficients or weights (Liao, 1992). While it might seem that the infinitely complex polynomial would produce the most accurate predictive model, Farlow (1984) indicates that the relationship between model complexity (on the X axis) and prediction accuracy (on the Y axis) is actually “V” shaped. Unlike backpropagation neural networks, GMDH neural networks generally apply linear scaling [-1,+1] to all input data.[18]

Methodologies

Data from the Vermont State Reports (1996-1997) were used in this study. Data on socioeconomic characteristics of the student population, structural characteristics of the schools and student achievement were generally available

[17] See Farlow, 1984

[18] X’ = 2*(X – Min)/((Max – Min) – 1)

with the school as the basic unit (See Appendices A and B). One major shortcoming of the data was that only cross-sectional analysis could be performed at this time, negating the possibility of truly understanding the dynamics of the system, which of course, must occur over time. Yet, because this is primarily a test of methodologies and because many other production-function analyses have been similarly performed on cross-sectional data, this shortcoming in some ways actually increases the comparability of the findings.

Four methods were applied to the data: 1) multiple linear regression, 2) backpropagation, 3) GRNN and 4) GMDH. Multiple linear regression methods were employed to determine the extent to which the Vermont data conform to our usual production-function expectations. In addition, data were entered both in their original scale and then log transformed (ln).

All of the models were assessed according to two criteria: 1) prediction accuracy and 2) inferential value. The prediction accuracy of the models was used as an initial screening measure to determine which models deserved additional attention with regard to inferential value. The prediction accuracy of the models was determined by randomly withholding sets of schools from the model specification process. Then, the specified models were used to predict the performance of students in the withheld schools. This process was repeated ten times for each of the elementary schools (10 withheld per trial) and the high schools (5 withheld per trial). Models were checked for consistency of structure, and the Mean Absolute Percent Error of prediction was used to assess prediction accuracy. This concluded the inductive phase of the analyses.

Because induction alone is of similar value to deduction alone, the issues raised or patterns identified by the neural network algorithms were then individually analyzed via traditional deductive methods. Given the results of the preliminary analyses, parameters, interactions and non-linearities identified in particular via the GMDH method were tested for their statistical significance using MLR.

Results and Discussion

Is performance predictable?

Table 1 displays the results of the prediction accuracy assessment. With the given data, the GMDH algorithm consistently outperformed all three other methods. At the same time, the multiple linear regression model produced the least accurate predictions. Significant differences also exist between the predictability of the high school data using the SAT as the outcome measure, and the elementary school data using an aggregate of the math portfolio assessments. The results in Table 1 indicate that each pattern recognition algorithm has extracted more information from the data set than the linear regression model. The relative predictive power of the nonlinear models suggests the possibility of non-linear relationships within the data that are actually stronger than many of the presumed linear (or log-linear) relationships.

Table 1: Comparative Predictive Capabilities of Regression and Neural Network Models

	High School		Elementary	
	Untreated	Transformed (ln)	Untreated	Transformed (ln)
Linear Regression	6.5	12.7	28.9	28.5
Backpropagation	6.2	6.1	17.7	15.5
GRNN	4.9	4.2	27.6	26.4
GMDH	**2.1**	**1.9**	**14.4**	**8.5**

MAPE (Mean Absolute Percent Error)

Differences in Model Structures?

Negligible significance was found among the predictors in the full linear regression models. Only parent level of education (ln transformed model) displayed significance ($p<.10$) in the high school model and only teacher salary displayed significance in the elementary school model ($p<.10$), providing an intriguing split decision on the traditional argument over the relative roles of socio-economic factors and fiscal input factors.

Deductive analysis of non-linear and interaction parameters revealed by the GMDH algorithm, however, paint a somewhat different and more complex picture of schooling in Vermont (See Table 2). While in many cases the parameters selected by GMDH were similar to those identified by MLR, the natures of the relationships were dramatically different. For example, with the un-transformed data, a strong third order, diminishing return relationship for school size, was revealed ($p<.05$). Interestingly, significance for all three orders of this parameter were retained when the natural log transformation was applied ($p<.10$). Somewhat more elusive, yet no less interesting, were the interaction terms, such as the three-way interaction between parent level of education, school size, and dropout rate with the un-transformed data.

The results of the elementary school analyses were significantly more complicated (See Table 3). While only one parameter displayed any level of significance in the linear model, multiple questions were raised regarding non-linear and interaction parameters with both the untreated and log transformed data. While the third order length of school year relationship was moderated by transformation, numerous non-linearities related to spending issues (TEASAL, AVGCS), wealth (PVPS) and school level inputs (ENROLL) emerged. In addition, two- and three-way interactions for many of these variables displayed significance. Again, the question of the interactive nature of school size and student background characteristics (such as parent level of education) was raised.

Table 2: Comparison of MLR Results with Additive Polynomial Produced by GMDH. (High School Models)

Parameter	**Linear Regression**	**Linear Regression (ln)**	**GMDH (Polynomial and Nonlinear terms)**	**Coefficient**	**GMDH (ln) (Polynomial and Nonlinear terms)**	**Coefficient**
EDUC	3.062	0.0628*	EDUC	3.023***	EDUC	0.053
ENROLL	.296	0.026	ENROLL	-1.824	ENROLL	-1.980*
			$ENROLL^2$	4.600**	$ENROLL^2$	0.475*
			$ENROLL^3$	-1.45**	$ENROLL^3$	-0.037*
DROP	-5.683	-02024	DROP	5.461***	DROP	0.106
			$DROP^2$	-3.853***	$DROP^2$	-0.006
			$DROP^3$	0.967***		-
LSY	4.688	0.803	LSY	0.688	LSY	0.393
TEASAL	0.003	0.172*	TEASAL	-0.575*		
LSD	-31.02	0.027	LSD	0.097	LSD	5.641
					LSD^2	-1.491
INSPP	-0.013	-0.043		-		-
SPECED	1.910	0.003		-		-
ATTEND	2.895	0.786	ATTEND	0.043		-
FRLUN	0.504	-0.005		-	FRLUN	0.003
PUPTEA	-2.562	-0.106	PUPTEA	-0.611**		-
TECHED	-0.173	-0.010		-	TECHED	0.125**
					$TECHED^2$	-0.058*
					$TECHED^3$	0.011**
STUCMP	0.366	0.014		-		-
CONST	1095	-2.322	CONST	-1.644	CONST	1.943
			INTERACTION TERMS			
			ENROLL DROP	-1.583***	DROP FRLUN	-0.038
			ENROLL EDUC	-2.712***	ENROLL DROP	0
			DROP EDUC	-2.465***	ENROLL FRLUN	0
			ENROLL DROP EDUC	2.506***	ENROLL DROP FRLUN	0.007
			LSY LSD	-0.095	TECH DROP	-0.041**
			TEASAL PUPTEA	0.689**	LSY LSD	0

*p<.10, **P<.05, ***P<.01

1. GMDH models without log transformation of data still involve data scaling according to the formula: X' = 2*(X – Min)/((Max – Min) – 1) where X' is the scaled form and max and min values are determined ± 2 standard deviations

2. Coefficients in the Ln Transformed models may appear quite large because the model is constructed using the product of the ln terms rather than the ln of the products.

Table 3: Comparison of MLR Results with Additive Polynomial Produced by GMDH. (Elementary School Models)

Parameter	Linear Regression	Linear Regression (ln)	GMDH (Polynomial and Nonlinear terms)	Coefficient	GMDH (ln) (Polynomial and Nonlinear terms)	Coefficient
TEASAL	-0.001*	-0.132	TEASAL	0.755	TEASAL	-54.12***
			$TEASAL^2$	-0.023	$TEASAL^2$	1.513***
			$TEASAL^3$	0.033	$TEASAL^3$	0
FRLUN	0.189	0.090	FRLUN	2.873	FRLUN	0
			$FRLUN^2$	-2.141		
			$FRLUN^3$	0.837**		
AGIIND	21.92	0.222	AGIIND	0.163	AGIIND	-4170**
					$AGIIND^2$	-1.580
STUCMP	-0.354	-0.046	STUCMP	-0.723	STUCMP	1.042
					$STUCMP^2$	-0.495
					$STUCMP^3$	0.067
INSPP	0.001	0.088	INSPP	-0.168	INSPP	0.191*
SPECED	0.288	0.053	SPECED	0.040	SPECED	-0.765
PVPS	-0.000	-0.022	PVPS	-0.468	PVPS	-39.67***
			$PVPS^2$	0.401	$PVPS^2$	5.369***
			$PVPS^3$	-0.076	$PVPS^3$	-0.218***
PPOV	-0.188	-0.061	PPOV	0.020	PPOV	0.431
			$PPOV^2$	-0.262	$PPOV^2$	-0.316
EDUC	0.143	0.070	EDUC	0.075	EDUC	-100.1***
					$EDUC^2$	0.428
					$EDUC^3$	-0.045
LSY	0.506	0.468	LSY	-7.574***	LSY	-2.148
			LSY^2	4.768***	LSY^2	0
			LSY^3	-0.906***	LSY^3	-
AVGCS	-0.329	0.057	AVGCS	0.326	AVGCS	259.1***
			$AVGCS^2$	-0.433	$AVGCS^2$	-80.78***
			$AVGCS^3$	0.171	$AVGCS^3$	9.304***
LSD	-0.871	-0.112	LSD	-0.053	LSD	18.10**
ENROLL	0.015	0.014	ENROLL	2.170	ENROLL	-83.03***
			$ENROLL^2$	-1.560	$ENROLL^2$	0.100**
			$ENROLL^3$	0.301		
RETENT	-0.152	-0.009	RETENT	1.971**	RETENT	-0.034
			$RETENT^2$	-0.278*		
ATTEND	-0.171	0.743	ATTEND	-0.412	ATTEND	15.20
PUPTEA	-0.062	-0.202	PUPTEA	-0.218		
CONSTANT	-28.99	-1.071	CONSTANT	3.246	CONSTANT	215.5

*p<.10, **P<.05, ***P<.01

1. GMDH models without log transformation of data still involve data scaling according to the formula: X' = 2*(X – Min)/((Max – Min) – 1) where X' is the scaled form and max and min values are determined ± 2 standard deviations

2. Coefficients in the Ln Transformed models may appear quite large because the model is constructed using the product of the ln terms rather than the ln of the products.

Table 3 (Continued)

INTERACTION TERMS (ELEMENTARY MODEL)

Parameter	**Linear Regression**	**Linear Regression (ln)**	**GMDH (Polynomial and Nonlinear terms)**	**Coefficient**	**GMDH (ln) (Polynomial and Nonlinear terms)**	**Coefficient**
			RETENT FRLUN	-1.174*	ENROLL TEASAL	9.919***
			RETENT TEASAL	-1.383*	ENROLL EDUC	31.44***
			TEASAL FRLUN	-1.162	TEASAL EDUC	9.191***
			RETENT FRLUN TEASAL	1.369*	ENROLL TEASAL EDUC	-2.992***
			AVGCS FRLUN	-0.601	LSY AGIIND	805.9**
			AVGCS AGIIND	-0.168	AVGCS AGIIND	1475**
			FRLUN AGIIND	-0.485	LSY AVGCS AGIIND	-285.4**
			AVGCS FRLUN AGIIND	0.559	STUCMP AGIIND	-0.025
			ENROLL PPOV	0.382	LSD PVPS	-2.362**
			LSY PVPS	-0.105	ENROLL ATTEND	-4.478
			ATTEND STUCMP	0.495	PPOV EDUC	-0.115
			TEASAL INSPP	0.177	ENROLL RETENT	0.016
			PUPTEA SPECED	0.268	SPECED PVPS	0.092
			ENROLL SPECED	-0.241	AVGCS EDUC	0.883*
					INSPP AGIIND	0.587
					AVGCS TEASAL	-2.759**

*p<.10, **P<.05, ***P<.01

1. GMDH models without log transformation of data still involve data scaling according to the formula: X' = 2*(X – Min)/((Max – Min) – 1) where X' is the scaled form and max and min values are determined ± 2 standard deviations

2. Coefficients in the Ln Transformed models may appear quite large because the model is constructed using the product of the ln terms rather than the ln of the products.

Parsimonious Models

In order to reduce the collinearities in both linear and non-linear models, and gain a better understanding of the relative significance of the key predictors identified, parsimonious models were constructed and tested. Table 3 and Table 4 display options for the linear parsimonious models. As expected, given the relatively small "n" for each sample, both parameter significance and the *adjusted R-square* values for the models were improved. In the high school model, however, only two significant parameters were revealed in each of the parsimonious forms. The school size factor (ENROLL) remained significant in all models ($p<.05$). Either dropout rate or parent level of education, which display collinearity in the full model, could be included as the second key feature of the model. In the elementary parsimonious linear model, only the adjusted gross income index factor for the community remained significant.

Table 4: Parsimonious High School Models (OLS)

Model	**Adjusted R^2**	**Predictors**	**Coefficient (P)**
Model 1: Untreated data	.508	DROP	-10.614***
		ENROLL	0.600***
		INTERCEPT	979.2***
Model 2: Untreated data	.481	EDUC67	7.062***
		ENROLL	0.319**
		INTERCEPT	876.49***
Model 3: Ln Transformed data	.525	ENROLL	0.035***
		EDUC67	0.073***
		INTERCEPT	6.521***

*P<.05, **P<.01, ***P<.001

Table 5: Parsimonious Elementary School Models (OLS)

Model	**Adjusted R^2**	**Predictors**	**Coefficient (P)**
Model 1: Untreated data	.044	AGIIND	13.92*
		INTERCEPT	5.562***
Model 2: Ln Transformed data	.056	EDUC67	0.115**
		INTERCEPT	3.406***
Model 3: Ln Transformed data	.051	AGIIND	0.338**
		INTERCEPT	3.785***

*P<.05, **P<.01, ***P<.001

Table 6 displays the parsimonious non-linear models based on the parameters revealed by the GMDH algorithm. In the full models (Tables 2 and 3), parameter significance was difficult to assess due to the relatively large number of terms given the relatively small data set. In the parsimonious high school non-linear model, as might be expected, the *adjusted R-square* continued to display a stronger fit than the linear model. While the fit of the model was driven by the same factors (school size and dropout rates) the form of each relationship was different. In addition, there was recognition of a three-way interaction (P< .01) between school size, dropout rates and parent level of education that may yield amplified effects on student outcomes.

The parsimonious non-linear elementary model yielded a variety of different significant parameters as well as differences in functional forms. While the parsimonious linear elementary models, on average, explained approximately 5% of the variance in student performance in math, the parsimonious non-linear model explained nearly 30% of the variance in math performance. This model presents some support for the importance of spending related factors such as the availability of computers (STUCMP), an inverse linear relationship, and length of school year (LSY), a third order diminishing return pattern. Teacher salaries, however, displayed an unexpected inverse relationship ($P < .01$). None of these apparently significant relationships were revealed by linear analysis.

Table 6: Parsimonious Models (GMDH)

High School Model **Adjusted R^2 = .638**		**Elementary School Model** **Adjusted R^2 = .291**	
Predictors	*Coefficient*	*Predictors*	*Coefficient*
ENROLL	-0.453	AGIIND	0.226
ENROLL^2	0.706*	ATTEND	-0.759*
ENROLL^3	-0.373**	AVGCS	0.462
LSY	1.113*	INSPP	-0.245
LSD	0.105	STUCMP	-1.125**
DROP	-1.197	TEASAL	-0.513***
DROP^2	-0.689***	FRLUN	1.340
DROP^3	0.223***	FRLUN^2	-1.550
EDUC	-0.903	FRLUN^3	0.627**
		LSY	-5.407***
		LSY^2	3.153***
		LSY^3	-0.553***
		PPOV	-0.041
		PPOV^2	-0.072
		RETENT	0.830***
		RETENT^2	-0.306***
INTERACTION TERMS			
DROP EDUC	0.194	ATTEND STUCMP	0.823*
ENROLL DROP	0.618	AVGCS AGIIND	-0.225
ENROLL EDUC	0.127	AVGCS FRLUN	-0.571
ENROLL DROP EDUC	0.323***	AGIIND FRLUN	-0.087
LSY LSD	-0.130*	AVGCS FRLUN AGIIND	0.290
		TEASAL INSPP	0.293**

*p<.10, **P<.05, ***P<.01

1. GMDH models without log transformation of data still involve data scaling according to the formula: X' = 2*(X – Min)/((Max – Min) – 1) where X' is the scaled form and max and min values are determined ± 2 standard deviations
2. Coefficients in the Ln Transformed models may appear quite large because the model is constructed using the product of the ln terms rather than the ln of the products.
3. While this analysis does not employ Ln transformation, the GMDH algorithm contains and internal scaling of the data. The differences in model "fit" and predictive accuracy between methods is negligible.

Conclusions and Recommendations

The computer-assisted inductive methods used in this experiment yielded both differences in functional form and parameter selection when compared with typical linear regression methods. The non-linearities identified could not necessarily be expected based on our current understanding of educational productivity, thus it is unlikely that we would find such patterns by traditional deductive methods. The predictive abilities of the GMDH models in particular, and the relative stability of the model structure[19] suggest that this approach in particular may deserve closer attention as a potentially useful tool for data mining in education policy analysis. GMDH provides the added benefit of generating results that can be deductively tested and interpreted by methods already commonly used.

Do these methodologies help us in any way to answer the now age-old question: "Does money matter?" While the outcomes of these models may or may not directly address the role of fiscal resources in influencing student performance, the models do, at the very least, enhance our understanding of the system as a whole. As a result, we may become more attuned to significant features of the educational system that are ultimately cost-related, such as school size in the high school model and teacher salaries in the elementary school model

[19] While each of the model outcomes are not reported here, it should be noted that for a given data set, regardless of sub-sample, the GMDH algorithm repeatedly yielded the same parameters, including non-linear and interaction terms.

(ln).[20]

Do these methodologies entirely obviate political or personal influence? While this would be a clearly impossible task, these methods do, at the very least, open the door to potential research questions that are not already embedded in the minds of policy analysts. But, like the deductive process, induction cannot exist in a vacuum. Even the example presented herein has left a number of questions unanswered regarding the interactive nature of school and student level inputs that need to be further addressed. While in the end the true nature of these interactions may evade our reasoning, taking the time to explore their meaning can only serve to enhance our understanding of educational productivity.

[20] If, in fact, these parameters hold up to more rigorous subsequent analysis.

References

Bankes, Steve (1993) *Exploratory Modeling for Policy Analysis.* Operations Research. 41(3) 435-449

Betts, Julian (1996) *Is there a link between school inputs and earnings?* In Does Money Matter? The Brookings Institution. pp. 141-191.

Buchman, Timothy; Kubos, Kenneth; Seidler, Alexander; Siegforth, Michael. (1994) *A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit.* Critical Care Medicine. 22 (3)

Caudill, Maureen (1995a-February) *Part 1: The View from Now.* In Using Neural Networks. AI Expert. 5-12

Caudill, Maureen. (1995b-February) *Part 3: Putting Time in a Bottle.* In Using Neural Networks. AI Expert. 19-24

Caudill, Maureen (1995c-February) *Part 7: GRNN and Bear It.* In Using Neural Networks. AI Expert. 47-52

Cohn, Elchanan and Geske, Terry. (1990) *Production and Cost Functions in Education* in The Economics of Education: Third Edition. Pergamon Press, NY. 159-210

Duncombe, William; Miner, Jerry (1996) *Productive Efficiency and Cost-Effectiveness: Different Perspectives on Measuring School Performance.* Symposium on Cost-Effectiveness in Education. New York State Board of Regents. Albany, NY.

Farlow, S.J. (1984) Self-organizing Method of Modeling: GMDH Type Algorithms. Marcel Dekker. New York.

Hansen, James V., Nelson, Ray D. (1997) *Neural Networks and Traditional Time Series: A Synergistic Combination in State Economic Forecasts.* IEEE Transactions on Neural Networks. 8 (4)

Hanushek, Eric (1997) *Assessing the Effects of School Resources on Student Performance: an Update.* Educational Evaluation and Policy Analysis. 19 (2)

Hanushek, Eric (1996a) *The Quest for Equalized Mediocrity: School Finance Reform Without Consideration of School Performance* in Picus, Lawrence and Wattenbarger, James. Where Does the Money Go? Yearbook of the American Education Finance Association, 1996. Corwin Press. Thousand Oaks, CA. pp. 20-43.

Hanushek, Eric (1996b) *The Productivity Collapse in Schools* in Developments in School Finance, 1996. National Center for Education Statistics. pp. 185-195.

Hanushek, Eric (1996c) *School Resources and Student Performance.* In Burtless, Gary. Does Money Matter? The Brookings Institution. pp. 44-70.

Hanushek, Eric (1996d) *Making Schools Work.* Symposium on Cost-Effectiveness in Education. New York State Board of Regents. Albany, NY.

Hanushek, Eric (1994) Making Schools Work: Improving Performance and Controlling Costs. The Brookings Institution. Washington, DC.

Hanushek, Eric (1989) *The Impact of Differential Expenditures on School Performance*. Educational Researcher. May. pp. 45-62.

Hanushek, Eric.(1986) *The Economics of Schooling: Production and Efficiency in Public Schools.* Journal of Economic Literature. Vol. XXIV

Hedges, Larry; Greenwald, Rob (1996) *Have Times Changed? The Relation between School Resources and Student Performance.* In Burtless, Gary. Does Money Matter? The Brookings Institution. pp. 74-90.

Hedges, Larry; Laine, Richard; Greenwald, Rob.(1996) *Money Does Matter: A Research Synthesis of a New Universe of Education Production-function Studies* in Picus, Lawrence and Wattenbarger, James. Where Does the Money Go? Yearbook of the American Education Finance Association, 1996. Corwin Press. Thousand Oaks, CA. pp. 44-70

Hedges, Larry; Laine, Richard; Greenwald, Rob.(1994) *Does Money Matter? A Meta-analysis of Studies of the Effects of Differential School Inputs on Student Outcomes.* Educational Researcher. 23 (3)

Johnson, George E.; Stafford, Frank P. (1973) *Social Returns to Quantity and Quality of Schooling.* Journal of Human Resources. 8 (Spring) 139-155

Lachtermacher, Gerson; Fuller, David. (1995) *Backpropagation in time-series forecasting.* Journal of Forecasting. 14 (4) 381-393

Lemke, Frank. (1997) *Knowledge Extraction from Data Using Self-Organizing Modeling Technologies.* Paper published on eSEAM'97 Conference. www.scriptsoftware.com/km/paper2.html

Liao, Tim Futing (1992). *A modified GMDH approach for social science research: exploring patterns of relationship in the data.* Quality and Quantity. Vol. 26. pp 19-38

McMenamin, J. Stuart (1997) A Primer on Neural Networks for Forecasting. Journal of Business Forecasting. Fall. 17-22

Minar, Nelson; Burkhart, Roger; Langton, Chris; Askenazi, Manor. (1997) *The Swarm Simulation System: A Toolkit for Building Multi-Agent Systems.* Santa Fe Institute. www.santafe.edu/projects/swarm/overview.

Murphy, Christopher M., Fogler, H. Russell, Koehler, Gary. (1994). *Artificial Stupidity.* Icfa Continuing Education. March 30. 44-48.

Odom, Marcus; Sharda, Ramesh (1994) *A Neural Network Model for Bankruptcy Prediction.* Unpublished Manuscript. Department of Accounting, University of Oklahoma.

Specht, Donald (1991) *A Generalized Regression Neural Network.* IEEE Transactions on Neural Networks. 2 (6) pp. 568-576

Verstegen, Deborah (1997) *Higher Level Thinking and Problem Solving Skills and Resource Inputs Into Schooling.* American Education Finance Association Annual Meeting. Jacksonville, FL.

Ward, Steve; Sherald, Marge (1995) Neuroshell 2 User's Manual. Ward Systems Group, Inc. Frederick, MD.

Worzala, Elaine; Lenk, Margarita; Silva, Ana. (1995) *An Exploration of Neural Networks and Its Application to Real Estate Valuation.* The Journal of Real Estate Research. 10 (2).

APPENDIX: VARIABLE DESCRIPTIONS
(Source: 1996&1997 Vermont School Reports. Vermont State Department of Education)

Variable Name	**Definition**	**Model**
EDUC67	Percent of parents with BA or higher	HS/ELEM
FRLUN	Percent of students receiving free or reduced lunch	HS/ELEM
AGIIND	Average adjusted gross income indexed against median district	ELEM
PPOV	Percent of students living in poverty	ELEM
PVPS	Property value per student	ELEM
SPECED	Percent of students receiving special education services	HS/ELEM
TECHED	Percent of students participating in technical education	HS
DROP	High-school drop-out rate	HS
ENROLL	Average enrollment per grade level	HS/ELEM
ATTEND	Average daily rate of attendance	HS/ELEM
RETENT	Percent of students held back from grade advancement	ELEM
LSY	Length of school year (days)	HS/ELEM
LSD	Length of school day (hours)	HS/ELEM
AVGCS	Average class size	ELEM
PUPTEA	Pupil to teacher ratio	HS/ELEM
STUCMP	Pupil to computer ratio	HS/ELEM
INSPP	Instructional expenditures per pupil	HS/ELEM
TEASAL	Teacher salary	HS/ELEM
SATT	Average combined SAT	HS
PASS_MATH	Average percent of students successful in each of four categories 1)numbers and operations, 2) geometry and measurement, 3) functions and algebra and 4) probability and statistics in portfolio assessment at the 4th grade level	ELEM