1988

Non Linear Pathmodels

Ab Mooijaart*
Frits Meijerink*
Jan de Leeuw**

*Department of Psychology, University of Leiden, Hooigracht 15, 2312 KM, Leiden, the Netherlands

**Department of Psychology and Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles CA 90024, USA.

In this paper we discuss non-linear path models for continuous or smooth data. Instead of formulating a non-linear model for the relationships between the variables, i.e. a non-linear structural-equation-model, the observed variables are transformed non-linearly. For the transformed variables a linear path model is assumed. The transformations are defined by B-splines. An example is given from physics.

1. Introduction

In this paper we discuss recursive path models for continuous data, i.e. we are dealing with structural equations models without latent variables. See for discussion of structural equations models with latent variables: LISREL developed by Jöreskog (1982), PLS developed by Wold (1982) and some ALS method given by De Leeuw (1984). In econometrics these models are called errors-in-variables models, see Aigner, Hsiao, Kapteyn and Wansbeek (1983).

Path models can be conceived as more complicated linear (multiple) regression models. For these regression models there exist generalizations to non-linear regression models, see Jennrich (1969) and Wu (1981). Another type of non-linear regression models is the Box-Cox model, Box and Cox (1964, 1982). In the Box-Cox model the dependent variable is transformed non-linearly according to a one-parameter family of functions. It is along this line that we propose a generalization of the linear path models. To stress the differences of our model with the Box-Cox model we mention: 1. in our model all variables, i.e. the exogene and endogene variables, may be transformed and 2. our model is not restricted to one-parameter functions.

This approach of transforming the variables is very much alike the non-linear multivariate analysis as discussed extensively by Gifi (1981, 1984). In these analyses there is a combination of transforming the variables optimally (in a well defined sense) and the standard linear multivariate techniques, like Regression Analysis, Principal Component Analysis, Canonical Correlation Analysis and other techniques. One of the main interests of these non-linear techniques is that they can deal with variables of different measurement levels. So, for instance, ordinal variables are transformed in such

a way that the ordinal information is not destroyed. For literature about non-linear multivariate analysis and our model, see De Leeuw(1986 a,b &c). In the model we present here we are dealing with continuous or smooth data which may be transformed too. A model which can be compared with our model is PATHALS , see Coolen and De Leeuw (1987). In PATHALS , however, the variables are discrete and a least squares loss-function is used, whereas we use a maximum likelihood function.

The transformations we use are polynomial or piece-wise polynomial transformations. These transformations will be defined by so-called B-splines. For a general discussion of splines, see De Boor (1978) and for some applications in the field of regression see Winsberg and Ramsey (1980) and Stone (1985a, 1985b). Spline functions are interesting because they can approximate non-linear functions very efficiently. A further nice feature of splines is that splines form a linear space, by which it is possible, although non-linear functions are approximated, to use the well-known linear algebra and corresponding computations.

The basic idea of our model is: let there be several unobserved random variables, $\mathbf{y}_j$, for which a path model is specified to describe the relationships between these variables. (In our notation random variables will be denoted by a bold face letter.) Instead of these $\mathbf{y}_j$ variables, variables called $\mathbf{x}_j$ are observed. The latter variables are strictly, monotonously related with the unobserved random variables $\mathbf{y}_j$, say $\mathbf{x}_j = f_j(\mathbf{y}_j)$. We further assume that the $\mathbf{y}$ variables are multivariate normally distributed. Because of this assumption and the assumption of monotonicity it is easy to write down the likelihood function for the observed $\mathbf{x}$ variables. This likelihood function is some function of the structural parameters (path coefficients and error variances) and the spline coefficients. An algorithm will be given to estimate the parameters.

This paper ends with an illustration of some empirical data.

2. The model

Let x_j be an observed random variable, where j = 1,...,p. The linear path model, however, is not defined for x_j, but for y_j, where $x_j = f_j(y_j)$. It is supposed that the function f_j is a strictly monotously increasing function. So there is an inverse function $g_j(.) = f_j^{-1}(.)$. Consequently we can write $y_j = g_j(x_j)$. The random variables y_j are collected in a p-dimensional vector y. The following assumptions with respect to y are made

$$y \sim N(0, \Sigma(\theta)) , \tag{1}$$

where the covariance matrix of y is formulated as a function of the unknown parameters collected in vector θ. The latter function, $\Sigma(.)$, defines in this paper the path model for the transformed y variables. The dendity function for these variables can be written as

$$D_y (y;\theta) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp[-y'\Sigma^{-1}y/2] . \tag{2}$$

However, because the x variables are observed instead of the y variables, we write the density function of x as

$$D_x(x;\theta) = (2\pi)^{-p/2} |\delta y/\delta x'| \, |\Sigma|^{-1/2} \exp[-g'(x)\Sigma^{-1}g(x)/2] , \tag{3}$$

where $\delta y/\delta x'$ is the Jacobian. Obviously, because it holds $y_j = g_j(x_j)$, the Jacobian is diagonal, so $|\delta y/\delta x'| = \prod_j \delta y_j/\delta x_j$.

Let matrix Y of order (n x p), with the scores on the y variables ,for n sample units, and let Y_j be column j of matrix Y. By using B-spline functions Y_j can be written as

$$Y_j = Q_j\delta_j , \tag{4}$$

where Q_j consists of the basic splines, and δ_j are the unknown spline coefficients. See for an extensive discussion of splines de Boor (1978). The basic idea of using splines is that the relationship between two variables, **x** and **y**, is described by several polynomials. Spline functions are also called piecewise polynomal functions. For instance, variable **x** is divided in several intervals and within each interval a best, in some sense, fitting polynomial for describing **y** is looked for. Instead of using one polynomial of a high degree, several polynomials of low degree are utized to approximate the relationship between variables. Spline functions are some efficient compromise between approximation of stepfunctions and polynomial approximations. See for a discussion Gifi (1981, pp. 289-292). The user must specify the order of the polynomial, say k, and the intervals. The order of a polynomial is defined as the degree of the polynomial +1. Although there are methods also to find optimal intervals, we will fix the intervals in our approach. Intervals are seperated by so-called knots and the number of intervals for a variable will be denoted as I, i.e. the number of knots minus 1. The set of splines forms a linear space of dimensionality $r = I + k - 1$. So in (4) matrix Q_j has the order $(n \times r_j)$, where r_j is the dimensionality of the splines for variable y_j. It holds $r_j = I_j + k_j - 1$, where I_j and k_j are analogously defined for variable y_j. Note that this matrix Q_j is a function of the scores on the variable x_j.

A sufficient condition for a strictly increasing function g_j is

$$\delta_{j1} < \delta_{j2} < \ldots < \delta_{jr_j}, \tag{5}$$

which is a necessary condition for $k_j \leq 3$ (see de Boor (1978)). In this paper we will be dealing with the class of spline functions in which (5) holds. Because (5) is not a necessary condition for monotonic increasing functions, this means that we are dealing with a subset of all monotonic increasing spline functions. By imposing the order restrictions of the δ's (4) can be rewritten as

$$Y_j = Q_j T_j \gamma_j = G_j \gamma_j , \tag{6}$$

where T_j is of order (r_j x r_j) and $[T_j]_{kl} = 1$, for $k \geq l$, and $[T_j]_{kl} = 0$, otherwise, and $\gamma_j > 0$, for $j = 2,...,r_j$.

By the choice of the splines in this paper a property of Q_j is that the sums of the rows are one, i.e. $Q_j 1_{r_j} = 1_n$. So, because the first column of T_j is 1_{r_j}, it follows

$$Y_j = 1_n \gamma_{j1} + G_j^* \gamma_j^*, \tag{7}$$

where γ_{j1} is the first element of γ_j, γ_j^* is γ_j minus the first element, and G_j^* is matrix G_j minus the first column. More generally, we write $Y = G\Gamma$, with $G = (G_1;...;G_p)$ and Γ a matrix with columns the vectors γ_j, augmented with zeros. Alternatively, we can write $Y = 1_n\tau' + G^*\Gamma^*$, where $t' = (\gamma_{11}, ..., \gamma_{p1})$, $G^* = (G^*_1;...;G^*_p)$ and Γ^* is Γ minus the rows with elements γ_{j1}.

3. Estimation of the parameters

Define x_i en y_i as row i of X and Y, respectively. Then the likelihood can be written as

$$L = L(X;\Theta,\Gamma) = (2\pi)^{-np/2} |\Sigma|^{-n/2} \{\prod_i \prod_j \delta y_{ij}/\delta x_{ij}\} \exp[-\sum_i g'(x_i)\Sigma^{-1}g(x_i)/2], \tag{8}$$

and so -2 log L is

$$-2 \log L = np(\log 2\pi) + n\log |\Sigma| + \sum_i g'(x_i)\Sigma^{-1}g(x_i) - 2\sum_i\sum_j \log(\delta y_{ij}/\delta x_{ij}). \tag{9}$$

Let Z be a matrix with element $z_{ij} = \log(\delta y_{ij}/\delta x_{ij})$ and $\delta y_{ij}/\delta x_{ij} = g'_{1ij}\gamma_j$, where g'_{1ij} is row i of the matrix of derivatives of the splines called G_{1j}. Notice that the first column of G_{1j} consists of zero elements only. The function to be minimized is now

$$f_{\Theta,\Gamma} = f(X;\Theta,\Gamma) = n\log |\Sigma| + tr\{G\Gamma\Sigma^{-1}\Gamma'G'\} - 2(1_n' Z 1_p), \tag{10}$$

with the unknown parameter matrices Θ and Γ. The algorithm we use to minimize this function is some alternating method, i.e. first the function is minimized w.r.t. Θ, for fixed Γ, then the function is minimized w.r.t. Γ, for fixed Θ. This process is repeated till convergence is reached. The two steps will be discussed seperately.

Step 1: for given Γ let S be defined as $\Gamma'G'G\Gamma/n$, then the function to be minimized is

$$f_\Theta = f(S;\Theta) = n\log|\Sigma| + n\mathrm{tr}\{S\Sigma^{-1}\}. \tag{11}$$

Minimizing this function w.r.t. Θ is a standard problem. This problem arizes also in maximum likelihood methods for normally distributed variables. Because we are here dealing with recursive path models only, the unknown model parameters can be solved easily by a series of least squares problems, see Wold (1954).

Step 2: for given Θ, and so for given Σ, the function to be minimized is

$$f_\Gamma = f(X;\Gamma) = \mathrm{tr}\{G\Gamma\Sigma^{-1}\Gamma'G'\} - 2(1_n'Z1_p). \tag{12}$$

This function can be written as

$$f_{\tau,\Gamma^*} = \mathrm{tr}\{(1_n\tau' + G^*\Gamma^*)\Sigma^{-1}(1_n\tau' + G^*\Gamma^*)'\} - 2(1_n'(\log G_1{}^*\Gamma^*)1_p).$$

Instead of minimizing f_{τ,Γ^*}, we minimize f_{Γ^*} which is defined as $f_{\Gamma^*} = \min_\tau f_{\tau,\Gamma^*}$. It is simply to prove, by taking derivatives of f_{τ,Γ^*} w.r.t. τ and by setting this vector of derivatives equal to zero, that an optimal estimate of τ is

$$\tau = -\Gamma^{*\prime}G^{*\prime}1_n/n. \tag{13}$$

Define $E = I_n - 1_n1_n'/n$, and $\bar{G} = EG^*$ (i.e. in $\bar{G}$ are the columns in deviations of the means) then the function to be minimized is

$$f_{\Gamma^*} = tr\{\bar{G}\Gamma^*\Sigma^{-1}\Gamma^{*\prime}\bar{G}'\} - 2(1_n'(\log G_1{}^*\Gamma^*)1_p). \quad (14a)$$

A more convenient way of writing (14a) is

$$f_{\Gamma^*} = \sum_j\sum_k \sigma^{jk}\gamma_j^{*\prime}A_{jk}\gamma_k^* - 2\sum_i\sum_j \log(g'_{1ij}\gamma_j^*), \quad (14b)$$

where σ^{jk} is element (i,j) of Σ^{-1}, and $A_{jk} = \bar{G}_j'\bar{G}_k$.

The estimation method we have chosen for estimating the vectors γ_j^*, is some modified Newton-Raphson method. This method is designed to achieve quadratic convergence when the Hessian matrix is positive definite, see, e.g. Luenberger (1973). The reason for a modified method is that there do hold some order restrictions for the elements of γ_j^*. In this method we write for an estimate of γ_j^*

$$\gamma_j^{*(k+1)} = \gamma_j^{*(k)} - \alpha\ \{H(\gamma_j^{*(k)})\}^{-1}\nabla f_{\Gamma^*}(\gamma_j^{*(k)}), \quad (15)$$

where k refers to iteration step k, $H(\gamma_j^{*(k)})$ is the matrix of second derivatives of f_{Γ^*} w.r.t. γ_j^*, evaluated in the point $\gamma_j^{*(k)}$, and $\nabla f_{\Gamma^*}(\gamma_j^{*(k)})$ is the gradient of f_{Γ^*} w.r.t. γ_j^*, evaluated in the point $\gamma_j^{*(k)}$. α is some optimal chosen step length in the direction $-\{H(\gamma_j^{*(k)})\}^{-1}\nabla f_{\Gamma^*}(\gamma_j^{*(k)})$. So the first and second derivatives are necessary. These derivatives are given by

$$\delta f_{\Gamma^*}/\delta\gamma_j^* = 2\sum_k \sigma^{jk}A_{jk}\gamma_k^* - 2\sum_i g_{1ij}(g'_{1ij}\gamma_j^*)^{-1}, \quad (16a)$$

$$\delta^2 f_{\Gamma^*}/\delta\gamma_j^*\delta\gamma_k^{*\prime} = 2\sigma^{jk}A_{jk} + 2\delta^{jk}\sum_i g_{1ij}\ g'_{1ik}(g'_{1ij}\gamma_j^*)^{-2}, \quad (16b)$$

where $\delta^{jk} = 1$ for $j = k$, and $\delta^{jk} = 0$ otherwise.

In the next section we give an algorithm for estimating γ_k^*, under the condition that $\gamma_k^* \geq 0$.

4. The Algoritm

For the sake of ease and without ambiguity, we drop all sub- and superscripts. So we are dealing with a function $f(\gamma)$, its gradient ∇f, its Hessian H and the parameter vector γ, for which it holds that $\gamma_i \geq 0$. The method we discuss here is the rapid "Manifold Suboptimization Method", as described by Zangwill (1969). However, because we are dealing with very simple restrictions ($\gamma_i \geq 0$) we will specialize this method down to our specific case.

Let, in general, the restrictions be written as

$$c_i(\gamma) \geq 0, \quad i = 1,...,m$$

and let the derivatives of the functions c_i w.r.t. γ be written as $a_i(\gamma)$. For a solution of γ some constraints are active, i.e. $c_i(\gamma) = 0$, and some are inactive, i.e. $c_i(\gamma) > 0$. Then the function $f(\gamma)$ is minimized if for the active constraints the gradient of the function can be written as

$$\nabla f(\gamma) = \sum_{i \in A} \lambda_i a_i(\gamma), \quad \text{for } \lambda_i \geq 0,$$

where A is the set of active constraints. See for a proof of this, e.g., Zangwill (1969). In our case where the restriction is $\gamma_i \geq 0$ we have $c_i(\gamma) = \gamma_i$, and so $a_i(\gamma) = 1$. From this it follows that for an optimal point

1: if restriction i is inactive, i.e. $\gamma_i > 0$, the derivative of the function w.r.t. γ_i is equal to 0, and 2: if restriction i is active, i.e. $\gamma_i = 0$, the derivative of the function w.r.t. γ_i is λ_i, which should be positive.

In the algorithm given below $B_k(\gamma)$ is defined as the set of active constraints in iteration k. So an important point in the algorithm is to check whether the gradient is positive or not for an active constraint. If not, then the corresponding constraint is made inactive and the process is continued. The whole procedure can now be summarized as follows:

Step I: choose a feasable start configuration for γ, say $\gamma^{(0)}$. Set k = 0, and $B_0(\gamma^{(0)}) = \{\emptyset\}$. See for a proper choice Remark 1 and 2.

Step II: Set k = k+1. Compute the direction $d^{(k)} = -H^{(k)-1}\nabla f^{(k)}$, and determine the optimal step-size α along this direction (see Davidon (1959)). There are two possibilities:

Step IIIa: the point $\gamma^{(k+1)} = \gamma^{(k)} + \alpha d^{(k)}$ is feasable, then there are three possibilities

Step IIIa1: $\nabla f_i(\gamma^{(k+1)}) \neq 0$ for $i \notin B_k(\gamma^{(k+1)})$, then go to Step II.

Step IIIa2: $\nabla f_i(\gamma^{(k+1)}) = 0$ for $i \notin B_k(\gamma^{(k+1)})$ and $\nabla f_i(\gamma^{(k+1)}) > 0$ for $i \in B_k(\gamma^{(k+1)})$, then an optimal γ has been reached.

Step IIIa3: $\nabla f_i(\gamma^{(k+1)}) = 0$ for $i \notin B_k(\gamma^{(k+1)})$ and for some i $\nabla f_i(\gamma^{(k+1)}) < 0$ for $i \in B_k(\gamma^{(k+1)})$, then define $j \in B_k(\gamma^{(k+1)})$, for which $\nabla f_j(\gamma^{(k+1)}) = \text{min.}$, and let $B_{k+1}(\gamma^{(k+1)}) = B_k(\gamma^{(k+1)}) - \{j\}$, and go to Step II.

Step IIIb: the point $\gamma^{(k+1)} = \gamma^{(k)} + \alpha d^{(k)}$ is infeasable. Define $\mu_i = -\gamma_i^{(k)}/ d_i^{(k)}$ and $\mu^* = \min\{\mu_i \mid \gamma^{(k+1)} = \gamma^{(k)} + \mu_i d^{(k)} \geq 0\}$. Then $\gamma^{(k+1)} = \gamma^{(k)} + \mu^* d^{(k)}$ and $B_{k+1}(\gamma^{(k+1)}) = B_k(\gamma^{(k+1)}) + \{i\}$, and go to step II.

A geometrical interpretation of this method is as follows:

In step II an optimal point along the direction $d^{(k)}$ is found, without any constraints on the parameters. Then it is verified whether this point is feasible.

--If this point is feasible but it is not yet optimal, then the process continues (Step IIIa).

--If this point is feasible and the gradients for the free parameters are zero and the gradients for the fixed parameters (i.e. the parameters which are set equal to zero) are positive, then the optimal point has been found.

--If this point is feasible and the gradients for the free parameters are zero but some of the gradients for the fixed parameters are negative, then a suboptimal point has been found on a boundary. The boundary for which the gradient is most negative is looked

for and the corresponding constraint is skipped. This means the set of active boundaries is diminished with one.

-- if this point is not feasible, then a feasible point on the direction $d^{(k)}$ is looked for the one which is closest to the optimal infeasible point. This feasible point will be on a boundary and this boundary then enters as an active boundary.

Remark 1: A start configuration for γ can be found by, e.g., a least squares solution of γ_j, for j = 1,...,p, from formula (6). Obviously, by doing so transformations are found without taking into account the underlying path model. Empirical studies with these start configurations have shown that they are very appropriate.

Remark 2: If the start configuration is too far from the optimal solution, then the Hessian matrix may be not positive definite. In those situations another start configuration has to be supplied.

Remark 3: As a consequence of the used method, it may happen that a temporary solution is on a combination of boundaries for which the function is not defined. In such cases $\delta y_{ij}/\delta x_{ij} = 0$, and the logarithm is not defined. A remedy to solve this problem is starting again with another start configuration.

5. Some Statistical Remarks

6. An example

The example we discuss here was also analyzed by De Leeuw (1986c) and Van Wijk (1987). It was taken from a very interesting paper by Wilson (1926). He discusses the question if statistical methods in general, and correlational methods such as regression in particular, can help us to discover natural laws. Wilson's point of view is that correlational procedures are not of much help if they are not combined with 'antecedent rationalism', i.e. with proir knowledge about the subject matter. He illustrates his point by a physical example, taken from the work of Willard Gibbs on the equilibrium of heterogeneous substances. Gibbs derived, from theoretical considerations, a formula connecting absolute temperature x_1, pressure x_2 and density x_3 of a mixture of gases with convertible components. The formula is of the form

$$g_3(x_3) = \beta_{31} g_1(x_1) + g_2(x_2) + \beta_{30} ,$$

with $g_1(x_1) = 1/x_1$, $g_2(x_2) = \ln x_2$, and $g_3(x_3) = \ln \{A (x_3 - A)/(2A - x_3)^2\}$, where A = 2.073. β_{30} and β_{31} are two constants which must be determined empirically. Gibbs determined the constants from experiments by Cahours and Bineau, and then used the formula to predict the outcomes of 65 new experiments by Neumann. He discusses the deviations he finds in terms of the rational formula above.

Wilson uses ordinary linear least squares to predict x_3 from x_1 and x_2, and he concludes that the results of this blind approach are quite useless from the point of view of physical theory. It seems to us that this conclusion is a bit pessimistic. It turns out that with the nonlinear technique of this paper we can recover the rational transformations quite nicely. Let us illustrate this with some results.

Temperature has only 9 discrete values therefor just one interval is used for this variable. The other two variables are divided into 4 intervals each. In table 1 the function values of different solutions are given. The function value is given by -2 ln likelihood. In this example we concentrate on the rational transformation of the variable density. The solutions differ in the specified order of the B-splines for this variable.

--- Insert Table 1 about here ---

It is obvious from table 1 that the fit, in terms of the ln likelihood, becomes better if the order of the splines for the variable density increases (There is one exception !!!!!, explanation: .....). In table 1 there is also a column for the explained variance, R^2, for each solution, another measure for indicating how good the prediction is. We see, roughly, that this coefficient becomes larger with the order of the splines. However, this is not necessarily so. The explanation for a possible decrease of R^2 is that we do not maximize the R^2, but the likelihood. Maximizing the explained variance can be done by specifying the least squares function, instead of the likelihood function. Another way of investigating how well our solution is, is to compare the optimal transformations of density and the rational transformation. These comparisons are given in figure 1 to **.

--- Insert figure 1 to ** about here ---

7. References

Aigner, D.J., Hsiao, C., Kapteyn, A. & Wansbeek, T.J. (1983). Latent variables in econometrics. In Griliches, Z. and Intrilligator, M.D. (Eds.), *Handbook of Econometrics*, Amsterdam, North Holland Co.

Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society, **B26**, 211-252.

Box, G.E.P. & Cox, D.R. (1982). An analysis of transformation revisited, rebutted. Journal of the Royal Statistical Society, **77**, 209-210.

Davidon, W. (1959). *Variable Metric Method for MinimizationN, Argone National Laboratory Report ANL-5990, Argone, Illinois.*

De Boor, C. (1978). *A practical guide to splines.* Berlin, Springer-Verlag.

Coolen, H. & De Leeuw, J. (1987). Least squares path analysis with optimal scaling. Report RR-87-03, Department of Data Theory, Leiden.

De Leeuw, J. (1984). *Least Squares and Maximum Likelihood for Causal Models with discrete Variables.* Report RR-84-09, Department of Data Theory, Leiden.

De Leeuw, J. (1986a). *Nonlinear Multivariate Analysis with Optimal Scaling.* Report RR-86-03, Department of Data Theory, Leiden.

De Leeuw, J. (1986b). *Nonlinear Path Analysis with optimal Scaling*, Report RR-86-04, Department of Data Theory, Leiden.

De Leeuw, J. (1986c). *Regression with optimal Scaling of the Dependent Variable.* Report RR-86-08, Department of Data Theory, Leiden.

Gifi, A. (1981, 1984). *Nonlinear Multivariate Analysis.* Department of Data Theory, University of Leiden.

Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Mathematical Statistics*, **40**, 633-643.

Jöreskog, K.G. (1982). The LISREL approach to causal model-building in the social sciences. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation. Causality, structure, prediction.* Amsterdam, North Holland Publishing Co.

Luenberger, D.G. (1973). *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Menlo Park, California.

Stone, C.J. (1985a). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689-705.

Stone, C.J. (1985b). *The dimensionality reduction principle for generalized additive models*. Technical Report 41, Department of Statistics, University of California, Berkely.

Van Wijk, M. (1987). An ML-method for Non-linear Pathmodels. Department of Psychology, University of Leiden, Doctoral Thesis.

Wilson, E.B. (1926). Empiricism and rationalism. *Science*, **64**, 47-57.

Winsberg, S. & Ramsey, J.O. (1980). Monotone transformation to additivity using splines. *Biometrika*, **67**, 669-674.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, **9**, 501-513.

Wold, H.O. (1954).Causality and econometrics. *Econometrica*, **22**, 176-177.

Wold, H.O. (1982). Soft modeling: the basic design and some extensions. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation. Causality, structure, prediction.* Amsterdam, North Holland Publishing Co.

Zangwill, W.I. (1969). *Nonlinear Programming. A unified approach.* Englewoods Cliffs, Prentice Hall.

TABLE 1

goodness of fit measures for different transformations

order of splines	function value	difference	R^2
-	381.7469	.917	
2	208.6416	.985	
3	209.0473	.985	
4	208.1766	.985	
5	207.1702	.985	
6	204.9572	.985	
7	203.5085	.985	
8	200.9457	.986	
9	198.7750	.986	