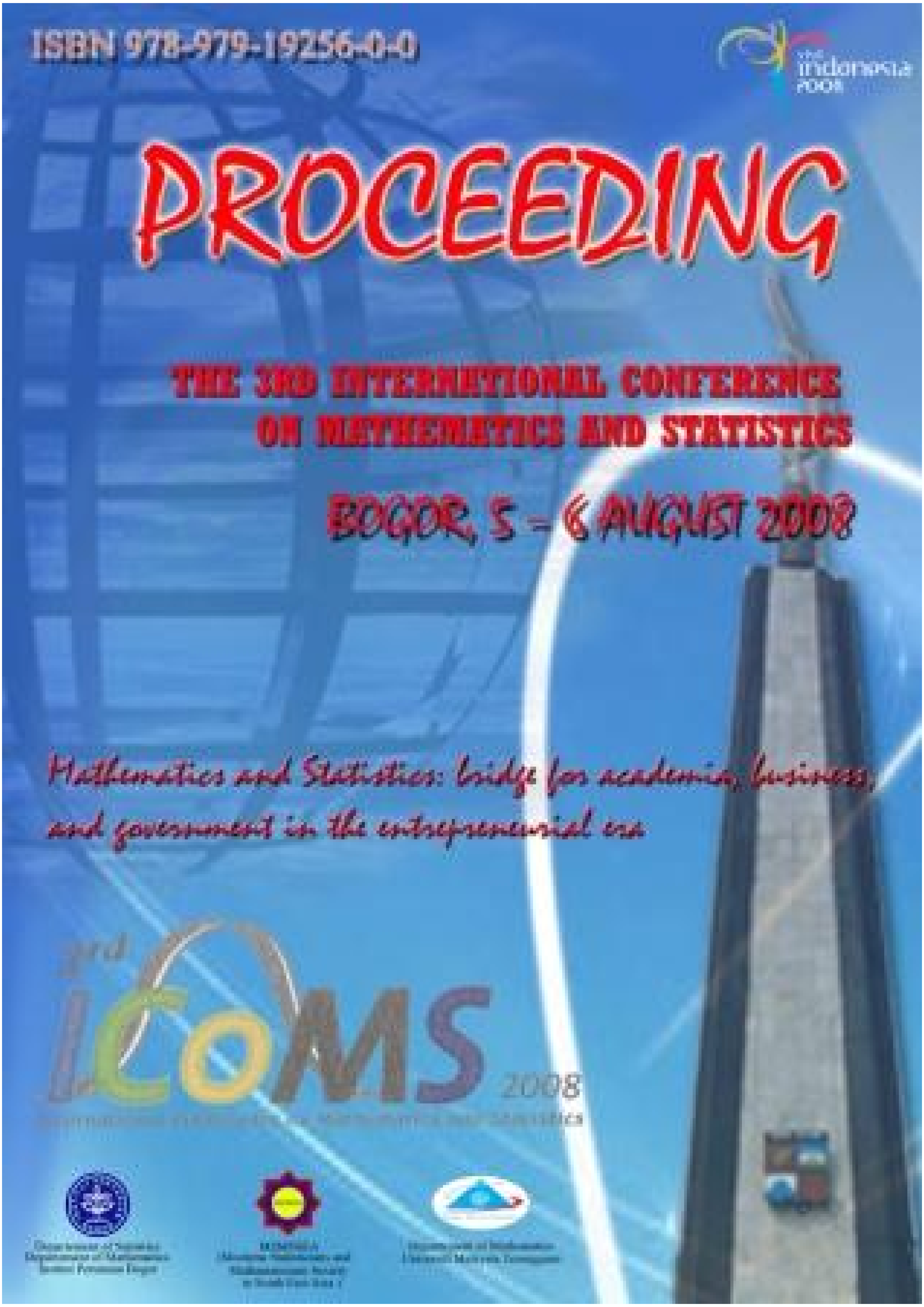

ISBN 978-979-19256-0-0

PROCEEDING

THE 3RD INTERNATIONAL CONFERENCE ON MATHEMATICS AND STATISTICS

BOGOR, 5 – 6 AUGUST 2008

Mathematics and Statistics: bridge for academia, business, and government in the entrepreneurial era

3rd ICoMS 2008

International Conference on Mathematics and Statistics

Department of Statistics
Department of Mathematics
Institut Pertanian Bogor

[illegible]
(Association Statisticians and
Mathematicians Society
in South East Asia)

Department of Mathematics
Universiti Malaysia Terengganu

POSTER

The 3"* International Conference on Mathematics and Statistics (ICoMS-3)
Institut Pertanian Bogor, Indonesia, 5-6 August 2008

The ACE Algorithm for Optimal Transformations in Multiple Regression

Kusman Sadik

Department of Statistics, Institut Pertanian Bogor Jl. Meranti, Wing 22 Level 4,
Kampus IPB Darmaga, Bogor 16680 - Indonesia
e-mail : kusmans@ipb.ac.id

Key Words: Alternating conditional expectations, non-restrictive transformations, parametric transformations, multivariate analysis, generalized additive models.

1. Introduction

We employed the Alternating Conditional Expectations (ACE) technique (Breiman & Friedman, 1985) to relax the assumption of model linearity. By generating non-restrictive transformations for both the dependent and independent variables, ACE develops regression models that can provide much better model fits compared to models produced by standard linear techniques such as Ordinary Least Squares. The ACE transformations can reveal new information and insights on the relationship between the independent and dependent variables.

The objective of fiilly exploring and explaining the effect of covariates on a response variable in regression analysis is facilitated by properly transforming the independent variables. A number of parametric transformations for continuous variables in regression analysis have been suggested (Box and Tidwell, 1962;Kruskal, 1965; Mosteller and Tukey, 1977; Cook and Weisberg, 1982; Carroll and Ruppert, 1988; Royston, 2000).

In this paper, we introduce the ACE algorithm for estimating optimal transformations for both response and independent variables in regression and correlation analysis, and illustrate through two examples that usefulness of ACE guided transformation in multivariate analysis. The power of the ACE approach lies in its ability to recover the functional forms of variables and to uncover complicated relationships.

2. The Alternating Conditional Expectation Methods

Non-parametric regression techniques are based on successive refinements by attempting to define the regression surface in an iterative fashion v^ile remaining 'data-driven' as opposed to 'model-driven'. These non-parametric regression methods can be broadly classified into those v\4iich do not transform the response variable (such as Generalized Additive Models) and those which do ACE.

The ACE module produces an output of graphical transformations for the dependent and independent variables. ACE will also indicate the adjusted and imputed *p-value* of the model based on these graphical transformations (all *p-values* are really only the computational counterparts of the *p-values* in a standard regression model, but they are useful for between-model comparisons as well as for variables selection). An ACE regression model has the general form:

$$\theta(Y) = \alpha + \sum^{p} \phi_i(X_i) + \varepsilon$$

where ^ is a function of the response variable, *Y,* and *(jh* are fiinctions of the predictors JC, *i* = 1, ...,/>.. Foragivendataset consisting of a response variable 7 and predictor variables Jfi,... , *Xp,* the ACE algorithm starts out by defining arbitrary measurable mean-zero transformations *0{Y),* (^i(Xi),... , 0^(X^).The error variance *{i)* that is not explained by a regression of the transformed dependent variable on the sum of

$$\varepsilon^2(\theta, \phi_1, \ldots, \phi_p) = E\left\{\left[\theta(Y) - \sum_{i=1}^{p} \phi_i(X_i)\right]\right\}^2$$

The ACE algorithm approaches this problem by minimizing the ^. If only there is a predictor variable *X,* so it needs minimizing ^{^(7) - (>*{X)}Y*• For fixed *0,* the minimizing is 0(X) = *E{eiY) \ X]*.
For fixed 0 minimizing is *Q{Y)* = E{0(X) | 7}. This is the key idea in the ACE algorithm, it begins with some starting fiinctions and alternates between these two steps until convergence (Hastie dan TibshIrani, 1990).

transformed independent variables is (under the constraint, (r)J = i)

3. Data Simulated

In this study, we apply the ACE technique to a synthetic example - case for which we know the correct answers - to demonstrate how the ACE algorithm can be used to identify the functional relationship between dependent and independent variables. This synthetic example is a multivariate case with four predictor and 250 observations generated from the following model

$$Y \ = 5 + \exp(X_1) + |X_2| + X_3^2 + X_4^3 + \varepsilon$$

where X_1 ~ uniform(-5,5), X_2 ~ uniform(-20,20), X_3 ~ uniform(-5,5), and X_4 ~ uniform(-4,4) and ε is independently drawn from a standard normal distribution N(0, 1). Note that the plots in Figure 1 do not reveal any obvious functional forms for either the dependent variable or predictors, even though X_1, X_2, X_3, and X_4 are statistically independent. Under such circumstances, direct application of linear regression is not appropriate.

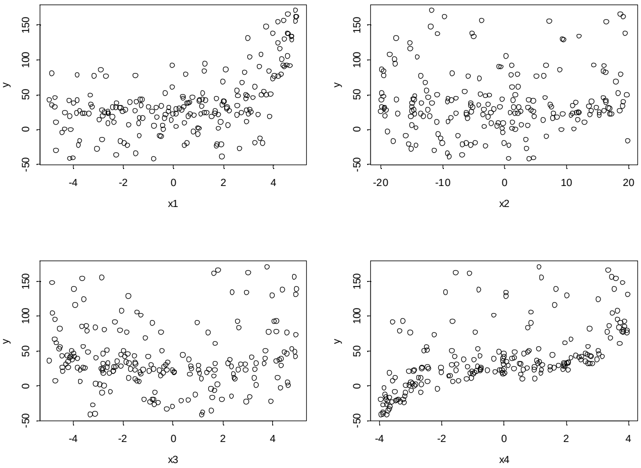

Figure 1. Scatterplots of simulated dataset (each predictor v.s. dependent variable)

To check if the ACE algorithm can recover these functions, we applied the algorithm to this simulated data set and the results are plotted in Figure 2. Clearly, ACE is able to recover the corresponding functions. A regression of the transformed dependent variable on all the transformed covariates results in all parameter coefficients of the independent variables being positive and close to:

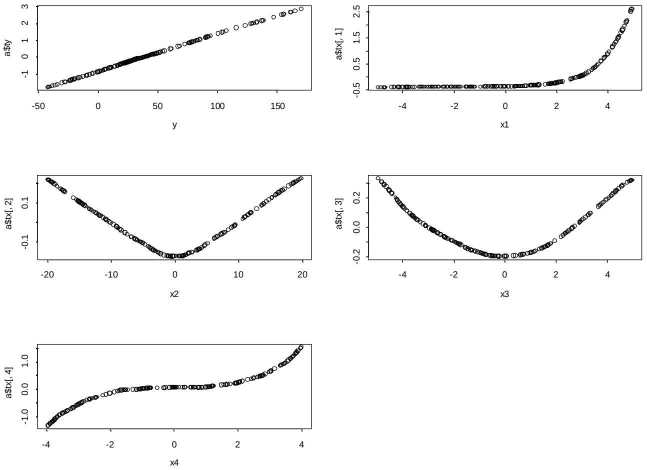

Figure 2. Scatterplots of ACE algorithm results (optimal transformation)

This results is indicating that the optimal parametric transformations have achieved. The ACE transformed variables has an adjusted R^2 of 0.995, considerably better than the value of 0.651 obtained using OLS. Note that in theory ACE cannot produce a worse fit than ordinary regression, because if no transformations are found to be necessary (i.e., the ordinary regression model is appropriate), then ACE would simply suggest nearly linear transformations for all the variables.

4. Conclusions

The ACE algorithm is a non-parametric automatic transformation method that produces the maximum multiple correlation of a response and a set of predictor variables. The approach solves the general problem of establishing the linearity assumption required in regression analysis, so that the relationship between response and independent variables can be best described and existence of non-linear relationship can be explored and uncovered. An examination of these results can give the data analyst insight into the relationships between these variables, and suggest if transformations are required.

The ACE plot is very useful for understanding complicated relationships and it is an indispensable tool for effective use of the ACE results. It provides a straightforward method for identifying functional relationships between dependent and independent variables. There will often be a number of potential candidates for transformation of a variable suggested by the ACE plot that fit the data well according to R^2.