A MULTI-DISCIPLINARY INVESTIGATION OF LINEARIZATION DEVIATIONS IN DIFFERENT REGRESSION MODELS

ABSTRACT

Aim: This study's objective is to examine linearization deviations in various regression models using a multidisciplinary approach.

Methods: Curve estimate models' accuracy for each type of data was tested using social, financial, and medical data sets.

Results: Although the power of a given model reduces with non-normal distributions, linearization in the social sciences is more efficient and has less variation from regression points in parametric equations. However, single-center distributions in the social sciences typically lead to nonparametric distributions. When compared to social sciences and health sciences, the effectiveness of linearization in financial sciences is higher. The original essence of financial techniques and models is frequently present, and they test their presumptions. Financial models and assumptions are more linear than those seen in real life since they correspond to artificial systems that individuals have developed, making them better suited for predetermined formulas. The rise in linearization issues in the field of finance, which has been increasingly common in recent years, is a symptom of this together with behavioral finance. Studies in the realm of health and well-known models were discovered to have the highest linearization deviations. Exponential or growth functions exhibit the highest linearization deviations in processes like growth, proliferation, and the spread of disease or pandemics. The data display considerable departures from normality and linearization, particularly in animal trials with very small statistical units or research conducted on a particular population.

Conclusion: Despite research on R2's explanatory capacity in regression, there aren't enough studies in various fields that concentrate on R2's departures from linearization. Additionally, no study was located in which the subject's mathematical foundation was examined and cross-compared across various data sets. As a result of this feature, the research represents a field first. The research's ability to pragmatically assess the distinctions between disciplines, made possible by its multi-disciplinary nature, is another unique aspect of the work.

Keywords: Linearization, Regression, Deviation, Multi-disciplinary.

1. Introduction

In its most general definition, linearization is the closest expression of an equation to the linear at a specific point. In other words, linearization is a linear approximation to a given equation. Although mathematical functions have different distributions specific to their structures, linearization is required to analyze and generalize these formulas (Asghari et al, 2022; Negrello vd, 2021; Mangasarian and Meyer, 2006; Negrut and Ortiz, 2006; Horowitz, 2002).

It is strongly advised to refrain from referring to linearization using terms from the 19th century, such as differentials or other mathematical concepts. A function that has been linearized at a certain point has the same number of variables as a linear function. The concept of differential forms was developed in 20th century mathematics, and while it is a valuable mathematical idea, it is one that is only applicable to subsequent courses that build on multivariable calculus, such as Riemannian geometry. The idea of differentials dates back to a time when some aspects of calculus were still unclear. It has endured and can be found in several calculus textbooks (Sridhar et al., 2013; Knill, 2011; Sturtevant, 2009; El-Kalaawy and Ibrahim, 2008; Andrisani, 2003).

The term differential, which was used instead of linearization in previous studies, brings with it continuous or unstable saddle results in some function types. On the other hand, linearization is simply finding the linear form of an equation. In this research, it was aimed to investigate linearization deviations in different regression models from multi-disciplinary approach.

2. Methods

Social, financial and medical data sets were used to test curve estimation models and their accuracy for each data sets. The World Bank Country Reports repository was used for research parameters in Türkiye data sheet. Research parameters and the World Bank Codes with year ranges were shown in the Table 1.

Table 1. Research parameters and the World Bank Codes with year ranges

Research Code	Type	The World Bank Parameter Name	Years
Social Field (SF)			
SF_SEP	Independent	School enrollment, primary (% gross)	1990-2019
SF_UF	Dependent	Unemployment, female (% of female labor force) (national estimate)	1990-2020
Economy Field (EF)			
EF_GDP	Independent	GDP per capita, PPP (current international $)	1990-2021
EF_CPI	Dependent	Consumer price index (2010 = 100)	1990-2021
Health Field (HF)			
HF_MR	Independent	Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%)	2000-2019
HF_LEB	Dependent	Life expectancy at birth, total (years)	1990-2020

Baseline characteristics of parameter series and their distributions were evaluated with Kolmogorov Smirnov test for normality and Augmented Dickey Fuller (ADF)-Phillip Perron (PP) unite root tests. Models were defined as in the following (Turanlı and Yılmaz, 2023):

$$y=\beta_0 + \beta_1 x \quad \text{(Linear)}$$
$$y=\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \text{ (Cubic)}$$
$$y=\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 \text{ (Quadratic)}$$
$$y=a^{\beta 0 + \beta 1 x} \text{ (Exponential)}$$
$$y= a(1+r)^t \text{ (Growh function)}$$

Curve estimation models were used to test relationship between dependent and independent parameters. SPSS 25.0 for Windows and Stata MP 14.0 for windows were used for analysis.

3. Results

Curve fit estimation model analysis results for social field sample data of the World Bank were given in the Table 2.

Table 2. Curve fit estimation model analysis results for social field sample data of the World Bank

	Model Summary					Parameter Estimates			
Equation	**R Square**	**F**	**df1**	**df2**	**p value**	**Constant**	**b1**	**b2**	**b3**
Linear	0.150	4.945	1	28	0.034	39.866	-.294		
Logarithmic	0.155	5.132	1	28	0.031	149.116	-30.114		
Inverse	0.160	5.320	1	28	0.029	-20.329	3075.705		
Quadratic	0.228	3.995	2	27	0.030	525.088	-9.924	0.048	
Cubic	0.228	3.995	2	27	0.030	525.088	-9.924	0.048	0.000
Compound	0.106	3.303	1	28	0.080	117.328	.976		
Power	0.109	3.440	1	28	0.074	1115766.860	-2.523		
S	0.113	3.578	1	28	0.069	-.279	258.334		
Growth	0.106	3.303	1	28	0.080	4.765	-.025		
Exponential	0.106	3.303	1	28	0.080	117.328	-.025		

R^2 ranges from 0.228 to 0.106 indicating deviation around 12.2%. Cubic and quadratic are the best fitted models, whereas compound, growth and exponential are the less fitted models.

Curve fit estimation model analysis results for economy field sample data of the World Bank were given in the Table 3.

Table 3. Curve fit estimation model analysis results for economy field sample data of the World Bank

	Model Summary					Parameter Estimates			
Equation	**R Square**	**F**	**df1**	**df2**	**p value**	**Constant**	**b1**	**b2**	**b3**
Linear	0.881	221.774	1	30	0.000	9151.094	82.324		
Logarithmic	0.498	29.750	1	30	0.000	9460.494	2122.673		
Inverse	0.125	4.304	1	30	0.047	17087.253	-896.144		
Quadratic	0.901	132.139	2	29	0.000	8200.025	115.146	-0.130	
Cubic	0.938	140.396	3	28	0.000	9184.095	25.948	0.794	-0.002
Compound	0.842	159.639	1	30	0.000	9704.352	1.005		
Power	0.550	36.650	1	30	0.000	9595.569	0.134		
S	0.162	5.816	1	30	0.022	9.655	-0.061		
Growth	0.842	159.639	1	30	0.000	9.180	0.005		
Exponential	0.842	159.639	1	30	0.000	9704.352	0.005		

R^2 ranges from 0.938 to 0.125 indicating deviation around 81.3%. Cubic is the best fitted model, whereas inverse is the less fitted model. However, inverse and S models are out of acceptable tresholds as over 50% difference from other models. Curve fit estimation model analysis results for health field sample data of the World Bank were given in the Table 4.

Table 4. Curve fit estimation model analysis results for health field sample data of the World Bank

	Model Summary					Parameter Estimates			
Equation	**R Square**	**F**	**df1**	**df2**	**p value**	**Constant**	**b1**	**b2**	**b3**
Linear	0.935	258,659	1	18	0.000	63,842	-,622		
Logarithmic	0.939	276,650	1	18	0.000	216,296	-46,127		

Inverse	0.943	295,577	1	18	0.000	-28,415	3415,391		
Quadratic	0.958	192,793	2	17	0.000	321,902	-7,603	0.047	
Cubic	0.958	192,793	2	17	0.000	321,902	-7,603	0.047	0.000
Compound	0.947	320,654	1	18	0.000	231,329	,966		
Power	0.950	339,777	1	18	0.000	1133935,557	-2,571		
S	0.952	358,810	1	18	0.000	,301	190,281		
Growth	0.947	320,654	1	18	0.000	5,444	-,035		
Exponential	0.947	320,654	1	18	0.000	231,329	-,035		

R^2 ranges from 0.958 to 0.939 indicating deviation around 0.19%. Cubic and quadratic are the best fitted models, whereas linear is the less fitted model.

4. Discussion and Conclusion

Although linearization in social sciences is more effective and has less deviation from regression points in parametric equations, the power of given model decreases with non-normal distributions. However, distributions in social sciences in single centers generally results in nonparametric distributions (Chavan and Kulkarni, 2017; Dalgaard, 2008).

In financial sciences, linearization is more effective than health sciences, and less effective than social sciences. Financial approaches and models often had their original nature, and they test their assumptions. Since finance refers to the systems that people have created artificially, financial assumptions and models are more linear than real life, and they are more suitable for determined formulas. Along with behavioral finance, which has become widespread today, the increase in linearization problems in the field of finance is a sign of this (Kothari, 2010).

Studies in the field of health and established models were found to be the studies with the highest deviations in terms of linearization. In processes such as growth, proliferation, and spread of disease or pandemics, linearization deviations are greatest in exponential or growth functions. Especially in animal experiments with very small statistical units or studies performed on a specific group, the data show high deviations from normality and linearization (Fathollahi-Fard vd, 2021).

Although there are studies on regression explanatory power R^2, there are not enough studies focused on deviations from regression linearization in different disciplines. In addition, no study was found in which the mathematical infrastructure of the subject was tested and cross-compared in different data sets. Therefore, the research is an important study in the field due to this feature. Another originality of the research is that it allows the pragmatic evaluation of the differences between disciplines, thanks to its multi-disciplinary feature.

References

Andrisani D. (2003). *Linearization of Nonlinear Equations.* Purdue University, Collage of Engineering, West Lafayette, IN.

Asghari, M., FathollahiFard, A.M., Mirzapour Al-e-hashem, S.M.J. and Dulebenets, M.A. (2022). Transformation and Linearization Techniques in Optimization: A State-of-the-Art Survey. *Mathematics,* 10(283), 1-26.

Chavan and Kulkarni, (2017). Role of Non-Parametric Test in Management & Social Science Research. *Quest International Multidisciplinary Research Journal,* 6(9), 38-52.

Dalgaard, P. (2008). *Introductory Statistics With R*. Springer.

El-Kalaawy, O. H. and Ibrahim, R. S. (2008). Linearizing Transformation and Exact Solutions Of Nonlinear Equations In Mathematical Physics. *Italian journal of pure and applied mathematics*, 23(1), 197-204.

Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M., Tavakkoli-Moghaddam, R. and Smith, N.R. (2021). Bi-level programming for home health care supply chain considering outsourcing. *J. Ind. Inf. Integr.*, 25, 100.

Horowitz, P. P. (2002). *Dynamic Systems and Feedback Class Notes for ME132*. University of California, Berkeley, CA.

Knill, O. (2011). Linearization, In Math S21a: Multivariable calculus, Harvard University Press.

Kothari, C. R. (2010). *Research Methodology*. New Age International Publishers, Mumbai.

Mangasarian, O.L. and Meyer, R.R. (2006). Absolute value equations. *Linear Algebra Appl*., 419, 359–367.

Negrello, C., Gosselet, P. and Rey, C. (2021). Nonlinearly Preconditioned FETI Solver for Substructured Formulations of Nonlinear Problems. *Mathematics*, 9, 316.

Negrut, D. and Ortiz, J.L. (2006). A Practical Approach for the Linearization of the Constrained Multibody Dynamics Equations. *J. Comput. Nonlinear Dynam*. Jul 2006, 1(3), 230-239.

Sridhar, S., Linderoth, J. and Luedtke, J. (2013). Locally ideal formulations for piecewise linear functions with indicator variables. *Oper. Res. Lett.*, 41, 627–632.

Sturtevant, T. (2009). Linearizing Equations Handout. *Wilfrid Laurier University*, <http://denethor.wlu.ca/data/linear.pdf>, Retrieved: 22.10.2022.

Turanlı, M. and Yılmaz, K. (2023). Linearization Problems in Managerial Sciences. “International Research in Social, Humanities and Administrative Sciences”, ISBN: 978-625-6971-27-1