Chapter 9
The Acceptable R-Square in Empirical Modelling for Social Science Research

Peterson K. Ozili
https://orcid.org/0000-0001-6292-1161
Central Bank of Nigeria, Nigeria

ABSTRACT

This chapter examines the acceptable R-square in social science empirical modelling with particular focus on why a low R-square model is acceptable in empirical social science research. The paper shows that a low R-square model is not necessarily bad. This is because the goal of most social science research modelling is not to predict human behaviour. Rather, the goal is often to assess whether specific predictors or explanatory variables have a significant effect on the dependent variable. Therefore, a low R-square of at least 0.1 (or 10 percent) is acceptable on the condition that some or most of the predictors or explanatory variables are statistically significant. If this condition is not met, the low R-square model cannot be accepted. A high R-square model is also acceptable provided that there is no spurious causation in the model and there is no multi-collinearity among the explanatory variables.

1. INTRODUCTION

This paper examines the acceptable R-square in empirical social science research. It focuses on why a low R-square model is acceptable in empirical social science research.

DOI: 10.4018/978-1-6684-6859-3.ch009

As a general principle, an econometric model is considered to have a high predictive power if the model has a high R-square or adjusted R-square (Gujarati, Porter, & Gunasekar, 2012). This general principle often gives the scientist some confidence that the explanatory variables in the model are good predictors of the dependent variable (Hill, Griffiths, & Lim, 2018).

Many social scientists, who follow this principle, are often excited when their models report a high R-square and they get worried when their models report a very low R-square. Their worry is further amplified when they learn that statisticians and scientists in the pure sciences will dismiss a model as "weak", "unreliable" and "lacking a predictive power" if the reported R-square of the model is below 0.6 (or 60 percent when expressed in percentage). In this paper, I address this issue and show that empirical modelling in social science has a different purpose compared to empirical modelling in the pure science.

The rest of the paper is structured as follows. Section 2 discuss the imperfect nature of social science. Section 3 highlights the different range of R-square. Section 4 presents the conclusion.

2. LITERATURE REVIEW

There is adequate literature about the R-squared. The literature about R-squared shows some of its applications. Miles (2005) showed that the R-squared and the adjusted R-squared statistics are derived from analyses based on the general linear model (e.g., regression, ANOVA), and they represent the proportion of variance in the outcome variable which is explained by the predictor variables in the sample (R-squared) and an estimate in the population (adjusted R-squared). Hagle and Mitchell (1992) suggest a refinement to the R-squared called the pseudo R-squared. They suggest that the corrected Aldrich-Nelson pseudo R-squared is a good estimate of the R-squared of a regression model because of its smaller standard deviations and range of errors, and its smaller error of regression. They also point out that the Aldrich-Nelson correction to the R-squared is more robust when the assumption of normality is violated. However, they cautioned that the pseudo R-squared should be used with caution because even a good summary measure can be misinterpreted; therefore, it was suggested that the pseudo R-squared should be used in conjunction with other measures of model performance. Chicco et al (2021) suggested that the use of the R-squared statistic as a standard metric to evaluate regression analyses is popular in any scientific domain. This is because the coefficient of determination (or R-squared) is more informative and truthful than other goodness of fit measures. Cameron and Windmeijer (1997) show that R-squared type goodness-of-fit summary statistics have been constructed for linear models using a variety of methods. They

Related Content