What is R2 all about?

Dalson Britto Figueiredo Filho[1]

José Alexandre Silva

Enivaldo Rocha

Abstract

Which substantive meaning we can draw from the R2 statistic? The coefficient of determination is defined as the sum of squares due to the regression divided by the sum of total squares. Usually, R2 is interpreted as representing the percentage of variation in the dependent variable explained by variation in the independent variables. This definition is found by both econometrics and statistics handbooks and is widely accepted among quantitative scholars. However, this view is challenged by one of the most distinguished political science methodologists. Gary King argues that the R2 is highly misused as a measure of the influence of X on Y. This paper analyzes the R2 statistic using a non-technical approach. Our purpose is to provide an intuitive understanding of the coefficient of determination and its major shortcomings.

There is a joke that states that if someone asks you about your R2 statistic this is a proxy of little knowledge in both econometrics and statistics[2]. This is because there are some criticisms regarding the use of R2 as an indicator of the influence of X on Y (Ascombe, 1973; Achen, 1977; King, 1986). In econometrics, Kennedy (2008) argues that "R2 is measured either as the ratio of the 'explained' variation to the 'total' variation […] and represents the

[1] About writers: Dalson Britto is a PhD political science candidate at Federal University of Pernambuco (UFPE), José Alexandre Silva is Professor at Federal University of Goiás (UFG) and a PhD political science candidate at UFPE. Enivaldo Rocha is Professor at UFPE. We are grateful to Natalia Leitao, Brian Robinson and Manoel Galdino for their useful suggestions in our draft version. Finally, we are thankful to referees of Leviathan Journal by their useful comments. This research was supported by both CAPES and CNPQ public agencies.

[2] Given the pedagogical purpose of this paper, we relied on different quotations from specialized literature. In addition, we minimized algebraic applications of the concepts since our aim is to provide an intuitive basis to understand them. Readers that want to check more technical approach should follow the references.

percentage of variation in the dependent variable 'explained' by variation in the independent variables" (Kennedy, 2008 p. 14). Wooldridge (2009) states that "R2 is the ratio of the explained variation compared to the total variation; thus, it is interpreted as the fraction of the sample variation in y that is explained by x" (Wooldridge, 2009 p. 44). In statistics, Moore and McCabe (2009) points out that "the square of the correlation, R2, is the fraction of the variation in one variable that is explained by least-squares regression on the other variable" (Moore and McCabe, 2009 p. 137). Similarly, Schroeder *et al*. (1986) argue that "R2, the coefficient of multiple determination, measures the percentage of the variation in the dependent variable which is explained by variations in the independent variables taken together" (Schroeder *et al*., 1996 p. 33). This same meaning is found in handbooks of applied statistics and in the help sections of different statistical software.

However, this view is challenged by one of the most prominent political science methodologists. Gary King (1986) argues that R2 is highly misused as a measure of the influence of X on Y. He states that the more accurate interpretation is that "R2 is a measure of the spread of points around a regression line, and it is a poor measure of even that" (King, 1986 p. 675). If scholars aim to evaluate their models they should check others statistics aside from R2. According to King (1986), "most of the useful information in R2 is already available in other commonly reported statistics. These other statistics are more accurate measures: They can directly answer theoretical questions. R2 cannot" (King, 1986 p. 678). However, this is not what usually happens. A large share of quantitative literature devotes little attention to other statistics, giving to much attention to the "proportion of dependent variable explained by the model". What is R2 all about? The principal aim of this paper is to analyze the R2 statistic using a non-technical approach. Our purpose is to provide an intuitive understanding regarding the coefficient of determination and its major shortcomings.

The remainder of the paper consists in three sections. First, we briefly outline the main criticisms regarding the coefficient of determination. This is followed by a replication of Anscombe`s (1973) data where we explain our understanding about the proper utilization of R2. We close with some concluding remarks on explanation and the use of statistical inference in political science.

1. A brief review on R2 statistic[3]

According to Anderson-Sprecher (1994), “the coefficient of multiple determination, R2, is a measure many statisticians love to hate. This animosity exists primarily because the widespread use of R2 inevitably leads to at least occasional misuse” (Anderson-Sprecher, 1994 p. 113). While the controversy over R2 has its origin in the statistics literature (Kavalseth, 1985; Helland, 1987; Willett and Singer, 1988; Lavergne, 1996; Korn and Simon, 1991; Scott and Wild, 1991; McGuirk and Driscoll, 1995), the R2 debate is important to all fields of knowledge that employ linear regression models. McGregor (1993) argues that “there is little wonder that the regression model has achieved its preferred status in the social sciences” (McGregor, 1993 p. 801/802)[4]. In addition, he states that “the regression model dominates empirical work in political science. Rough evidence of this can be found in a review the American Political Science Review: almost every article that displayed findings in tabular form used some form of regression analysis” (McGregor, 1993 p. 801)[5]. In fact, the attractiveness of the regression model can be partially explained by its capacity to summarize the relationship among different variables in a systematic and parsimonious approach. Therefore, since the use of regression models have been increasing in the social sciences in general and political science in particular, it is important to understand the controversial role of R2 and the substantive meaning scholars can draw from it.

According to King (1986), both the Pearson correlation coefficient and determination coefficient have serious shortcomings. He argues that “in most practical political science situations, it makes little sense to use these statistics.

[3] Luskin (1984; 1991a; 1991b) and King (1986; 1991) provide an excellent introduction regarding the role of R^2 statistic in political science.
[4] Despite the apparent positive quotation, McGregor (1993) presents a strong criticism to the use of regression models in political science.
[5] This statement cannot be extended to Brazilian political science literature since the utilization of quantitative methods is very limited. For example, Soares (2005) reported that less than 3% of all cases analyzed display some quantitative technique.

They do not measure what they appear to; they can be highly misleading" (King, 1986 p. 669). Similarly, Achen (1977) states that one of the main limitations of the correlation coefficient is its inability to be compared among samples. He argues that "correlations cannot be compared across samples: two correlations can differ because the variances in the samples differ, not because the underlying relationship has changed" (Achen, 1977 p. 807). Since R2 is the square of the correlation coefficient, King (1986) argues that "all of the criticisms of the correlation and standardized regression coefficients apply equally to the R2 statistic" (King, 1986 p. 675). The principal problem is that the variance in the population that the sample was drawn can strongly influence R2 magnitude. Therefore, there is no guarantee that a high value of the coefficient of determination is synonymous with a good fit since the difference can be explained by sample variance. According to Achen, (1977) "when we compare different samples, then, a small R2 gives no guarantee that a relationship is weak, nor is a large R2 a guarantee that it is strong. The same structure can yield any R2 depending on the variances of the independent variables" (Achen, 1977 p. 807). Kennedy (2008) also argues that "R2 is sensitive to the range of variation of the dependent variable, so that comparisons of R2s must be undertaken with care […] A consequence of this is that it makes no sense to compare R2 across different samples" (Kennedy, 2008 p. 27). Therefore, we cannot employ R2 neither as ultimate model quality indicator nor to compare different samples.

Kavalseth (1985) suggests that R2 statistic is an efficient estimator since it is not employed in models without intercept and/or when the linearity assumption has been violated. Scott and Wild (1991) argue that R2 can be used when all assumptions of least squares model have been met and warn that "The use of R2 is particularly inappropriate if the models are obtained by different transformations of the response scale" (Scott and Wild, 1991 p. 121). For Anderson-Sprecher (1994), "The R2 measures should not be used to evaluate models that are based on different measures of variation" (Anderson-Sprecher, 1994 p. 116). McGuirk and Driscoll (1995) argue that R2 is important but "The size of R2 and adjusted R2 are poor specification indicators since correctly specified models can have 'low' R2 values and misspecified models often have 'high' R2 values (McGuirk and Driscoll, 1995 p. 319). In addition, if Y mean is not stationary "R2, as typically calculated, is a biased and inconsistent measure of goodness of fit" (McGuirk and Driscoll, 1995 p. 319) [6].

In summary, our literature review suggests that: as the coefficient of determination (R2) depends on the correlation coefficient (r) it can be influenced by the difference variance across samples (Achen, 1977); For this reason, R2 statistics cannot be used to compare different samples (King, 2001); R2 statistic should not be employed to analyze models without intercept (Kavalseth, 1985); R2 should not be used when ordinary least regression basic assumptions are violated (Scott and Wild, 1991); R2 should not be used when the mean of Y is not stationary (McGuirk and Driscoll, 1995); R2 does not guarantee god fit model. Similarly, a small R2 is not a material proof of misspecified model (Achen, 1977; McGuirk and Driscoll, 1995). After considering these claims we have some doubts: (1) how can scholars offer an overall estimative of the effect of their independent variables on Y without R2 statistic? They should use the unstandardized coefficients (Ascombe, 1973; Achen, 1977; King, 1986). (2) How can scholars evaluate the quality of their estimated coefficients? They should evaluate the standard error. The smaller the error is, the more precise the estimate is (King, 1986). (3) And to evaluate the full model? Scholars should look to the F statistic. They can evaluate the errors and estimate the confidence interval. High spread among regression line is a proxy for a low R2 (King, 1986). Finally, if the scholar really wants to use R2? King (1986) answers: "if your goal is to get a big R2, then your goal is not the same as that for which regression analysis was designed" (King, 1986 p. 677) [7]. He concludes, "The purpose of regression analysis and of all parametric statistical analyses is to

[6] Regarding time series regression, Gary King warned us that taking two models,

$$Y_t = \alpha + bY_{t-1} + cX_t + \varepsilon_t \quad (1)$$

$$Y_t - b^*Y_{t-1} = \alpha^* + c^*X_t + \varepsilon^*_t \quad (2)$$

and trying experiments we should observe that c is very similar to c*, as is all relevant information about the two equations, including the fit to the data. However, we also should find that the R^2 is completely different in the two equations. Usually the coefficient of determination is much larger in the equation 1 than in the equation 2. To illustrate this argument we employed data from Jones (1914) about evolution of both the mink and muskrat populations (1850-1911). In model 1 we reached a coefficient *c* of .0159 and c* of .0174. However, regarding R^2 statistic we found .4476 for model 1 and .1264 for model 2. In addition, in model 1 we reached a Durbin-Watson statistic of 1.838 while in the model 2 we reached 2.0112. On substantive grounds, this means that R^2 cannot inform us regarding serial autocorrelation in the model 1.

[7] Kennedy (2008) argues that "It is worth reiterating that searching for a high R^2 or a high R¯2 runs the real danger of finding, though perseverance, an equation that fits the data well but is incorrect because it captures accidental features of the particular data set at had (called "capitalizing on chance") rather the true underlying relationship" (Kennedy 2008: 89).

estimate interesting population parameters. The best regression model usually has an R2 that is lower than could be obtained otherwise” (King, 1986 p. 677).

2. Empirical example: replicating Anscombe`s (1973) data

Another precaution scholars should take is to avoid interpreting R2 statistic without prior graphically analyzing their data. Anscombe's canonical example shows not only the same correlation coefficient, but also the same value for any other summary statistic (F, standard error[8], b, beta, etc). We replicate data to emphasize our argument[9].

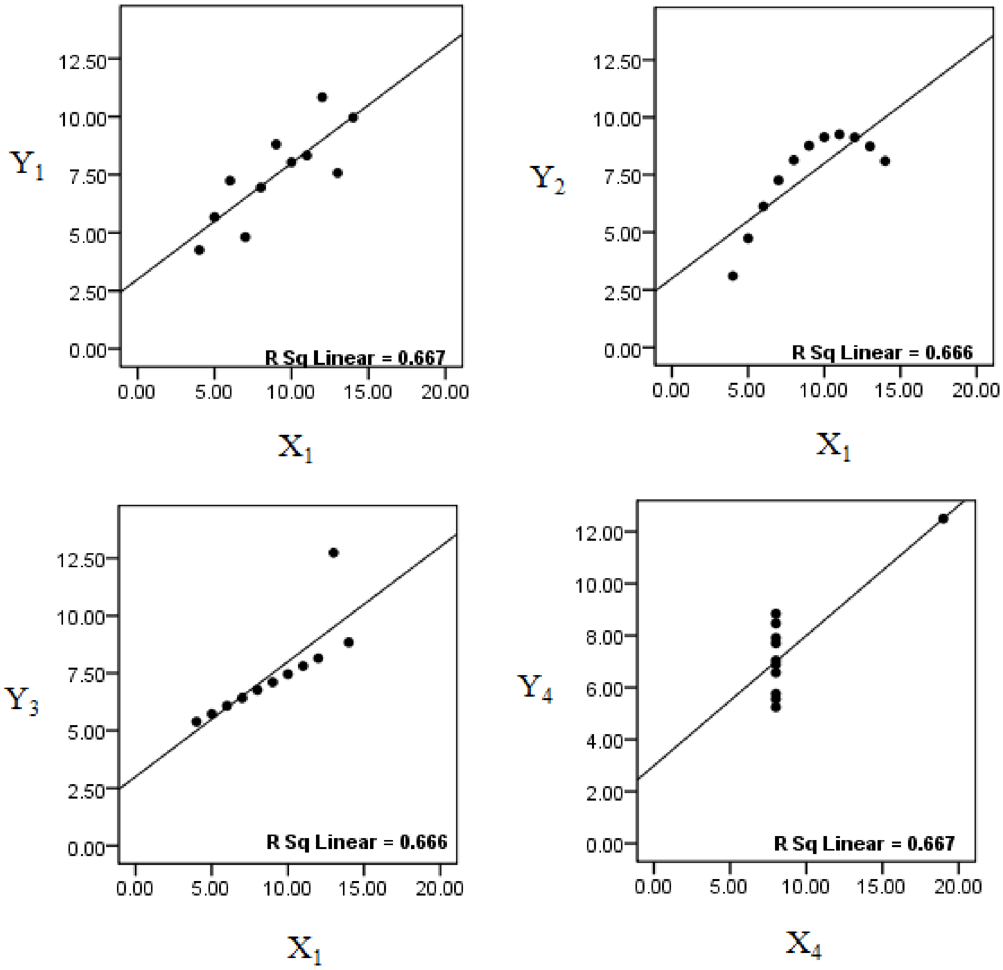

Figure 1 - Different relationships for the same R2

[8] Hair et al (2009) argue “that standard error is the expected distribution of an estimated regression coefficient. The standard error is similar to the standard deviation of any set of data values, but instead denotes the expected range of the coefficient across multiple samples of the data” (Hair et al, 2009: 157).
[9] Source: Authors` elaboration based on Anscombe (1973).

All four graphs display the same R2: .667. However, the nature of the relationship between the variables is quite different. In graphs 1, 2 and 3 the independent variable (X1) is the same and the dependent variable varies (Y1, Y2 and Y3). To make our case, we created a new variable based on Anscombe's (1973) data. X5 is negatively correlated with both Y1 (-.816) and X1 (-.648). We run a bivariate model using Y1 as dependent variable and X1 as independent variable (Model 01). Then, we run another model including the new independent variable (X5) (Model 02). The table below summarizes this information.

Table 01 – Comparing models using Anscombe (1973) data[10]

Statistics	Model 01	Model 02
R	0.816	0.889
R^2	0.667	0.809
Adjusted R^2	0.629	0.761
Std. Error of the Estimate	1.237	0.993
F	17.990	16.916
Sig	0.002	0.001
df	10	8

Model 02 shows both a higher correlation coefficient (.889) and a higher coefficient of determination (.809) when compared with model 01 (.816 and .667, respectively). The standard error of the estimate of the model 02 is smaller (.993) than the error of model 01 (1.237). Finally, model 02 reached a significance level (.001) more reliable than model 01 (.002). Based on this evidence, should we consider model 02 better than model 01? Following King`s (1986) tips, we can simply declare that the new model shows a better fit when

[10] Adjusted R^2 is a measure similar to regular R^2 but it controls for both number of cases and variables included in the model. Regular R^2 will always increase by adding new variables regardless of their contribution to model proper specification. According to Hair et al (2009), "modified measure of the coefficient of determination that takes into account the number of independent variables and sample size. Although the addition of independent variables will always cause the coefficient of determination to rise, the adjusted coefficient of determination may fall if the added independent variables have little explanatory power of if the degrees of freedom become too small. This statistic is quite useful for comparison between equations with different numbers of independent variables, differing sample sizes, or both" (Hair et al, 2009: 170). The Std. Error of the Estimate is a measure A measure of how much the value of a test statistic varies from sample to sample. It is the standard deviation of the sampling distribution for a statistic. For example, the standard error of the mean is the standard deviation of the sample means. According to Hair et al (2009), "the standard error of the estimate (SE_E) is a measure of the variation in the predicted values that can be used to develop confidence intervals around any predicted value. It is similar to the standard deviation of a variable around its mean, but instead is the expected distribution of predicted values that would occupy multiple samples of the data were taken" (Hair et al, 2009: 157).

compared to model 01. In other words, in the new model the spread of points around the regression line is smaller when we compare it with model 01.

3. Conclusion

Which substantive meaning can we draw from the R2 statistic? Almost none. It is impossible to draw substantive meaning based only on the magnitude of the coefficient of determination. In particular, R2 cannot help us to make causal claims about the relationship between the independent variables and the dependent variable. Likewise, R2 does not assist us regarding omitted variable bias. We cannot use R2 as a proxy of well specified model. R2 does not inform us if X5 is strongly correlated with X1 (colinearity problems in the data). In sum, following King's (1986) argument, R2 cannot answer theoretical questions. Thus, scholars should focus on both unstandardized coefficients and their estimated errors rather than on R2 (King, 1986).

In our view, one of the major challenges of research is to draw substantive meaning from empirical results. Statistics is undeniably a powerful tool to understand political phenomena. To extract relevant information from applied research it is necessary to fully understand the potential and limits of each methodological choice. We expect to help non-technical readers understand the role of the coefficient of determination in empirical research.

4. References

Achen, Christopher H. 1977. Measuring Representation: Perils of the Correlation Coefficient. *American Journal of Political Science* 21: 805-815.

Anderson-Sprecher, Richard. 1994. Model Comparisons and R^2. *The American Statistician* 48: 113-117.

Anscombe, Francis. 1973. Graphs in Statistical Analysis. *American Statistician* 27: 17-21.

Beck, Nathaniel. 1991. Model Selection: Are Time Series Techniques Useful in Cross-Sectional Problems? Presented at the annual meeting of the Midwest Political Science Association, Chicago.

Helland, Inge S. 1987. On the Interpretation and Use of R2 in Regression. Biometrics 43(1): 61-69.

Jones, J. Walter. 1914. *Fur-Farming in Canada*. Ottawa: Commission of Conservation.

Kennedy, Peter. 2008. *A Guide to Econometrics*. San Francisco, CA: Wiley-Blackwell.

King, Gary. 1986. How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science* 30:666-687.

King, Gary. 1991. Stochastic Variation: A Comment on Lewis-Beck and Skalaban's The R-Square. *Political Analysis* 2: 185-200.

Korn, Edward L., and **Richard Simon. 1991**. Explained Residual Variation, Explained Risk, and Goodness of Fit. *The American Statistician* 45(3): 201-206.

Kvalseth,Tarald 0. 1985. Cautionary Note About R2. *The American Statistician* 39: 279-285.

Lavergne, Pascal. 1996. The Hot Air in R2: Comment. *American Journal of Agricultural Economics* 78(3): 712-714.

Luskin, Robert C. 1984. Looking for R2: Measuring Explanation outside OLS. *Political Methodology* 10: 513-32.

Luskin, Robert C. 1991. Abusus Non Tollit Usum: Standardized Coefficients, Correlations, and R2. *American Journal of Political Science* 35(4): 1032-1046.

Luskin, Robert C. 1991. R-Square Encore. *Political Methodologist* 4: 21-23.

McGregor, James P. 1993. Procrustus and the regression model: On the misuse of the regression model. *PS: Political Science & Politics* 26: 801-804.

McGuirk, Anya, and **Paul Driscoll.** 1995. The Hot Air in R2 and Consistent Measures of Explained Variation. *American Journal of Agricultural Economics* 77: 319-328.

Moore, David S., and **George P. McCabe**. 2009. *Introduction to the Practice of Statistics*. West Lafayette, IN: W.H. Freeman Press.

Schroeder, Larry D., David L. Sjoquist, and **Paula E. Stephan.** 1986. *Understanding Regression Analysis: An Introductory Guide*. Beverly Hills, CA: Sage Publications.

Scott, Alastair., and **Chris Wild. 1991.** Transformations and R2. *The American Statistician* 45(2): 127-129.

Willett, John B., and **Judith D. Singer.** 1988. Another Cautionary Note about R2: Its Use in Weighted Least-Squares Regression Analysis. *The American Statistician* 42(3): 236-238.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. Boston, MA: South-Western College Publishing.

Submetido em 2011-01-30

Aceito em 2011-07-12