Understanding R^2 in Regression Analysis

Herbert M Barber, Jr, PhD, PhD
Managing Partner & Chief Investment Officer
Xicon Economics, LLC

> Without data, you're just another person with an opinion. - W Edward Deming, PhD

One of the greatest challenges in basic statistics is not found in the associated mathematics but in interpreting the analyses, or moreover, the output from the analyses. Granted, the math can be complex, but once we find our way through the math, interpreting the output becomes another challenge, altogether. Greek letters, symbols, squares, cubes, superscripts, subscripts, subscripts to subscripts, coefficients, intervals, integrals, and derivatives piled on even more squiggly lines; then add the mathematics associated with engineering and econometrics to the statistical mix, and we have an analytical mix of potions and concoctions that makes the math something with which to reckon. Still, it is not the math that is challenging, though it certainly can be, rather, it is the interpretation of the output the math generates that becomes a challenge. Such is the case with countless statistical methods and their subsequent outputs. Such is the case with R^2, or goodness of fit, our topic here.

R^2 is a measure of the strength of the relationship between a regression model and its independent variable(s), recalling from your elementary statistics courses that we use independent variables to project, or forecast, dependent variables. For example, in a business setting, we may use social media advertising and subsequent sales to project gross revenue. Or, in a manufacturing setting, we may use demand to project product manufacturing. And from an investment perspective, we may use the market and a sub-sector to project equity price. We can include several variables in the mix, or only a few. However, developing the model is only part of the equation, pun intended. To appreciate how well the model projects, or how well the model is fitted to the independent variables, i.e., predictor variables, we need to understand R^2.

Definitions or synonyms for R^2 range from model fit, line of best fit, goodness of fit, coefficient of determination, to explained variance, but perhaps its most appropriate definition is that of explained variance, as R^2 notes the variance in the dependent variable explained by the independent variable(s), with the R^2 interval ranging from 0 to 1, though an R^2 finding of 1.0 may suggest that the model has been overfit, which is another issue with which to deal altogether, meaning we must now deal with noise and other potential issues. Nonetheless, the larger the value for R^2, the greater the variance in the dependent variable that is explained by the independent variable(s). For example, if $R^2 = 0.8$, the independent variable(s) explains 80 percent of the variance in the dependent variable.

While there are equations to calculate R^2, or explained variance, a quick way to determine R^2 is to square the correlation coefficient between the dependent and independent variable. For example, if the correlation coefficient between the price of MSFT and market price is 0.7, squaring the coefficient determines the explained variance. So, if r=0.7, R^2=0.49, or 49 percent. For models using more than one independent variable, however, we must adjust R^2, as using additional predictors always increases R^2, despite the possibility of the additional variable(s) having no relationship with the dependent variable. Viewed differently, haphazardly adding independent variables may overfit the model, even when increasing R^2. In such a case, we use the following correction:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right]$$

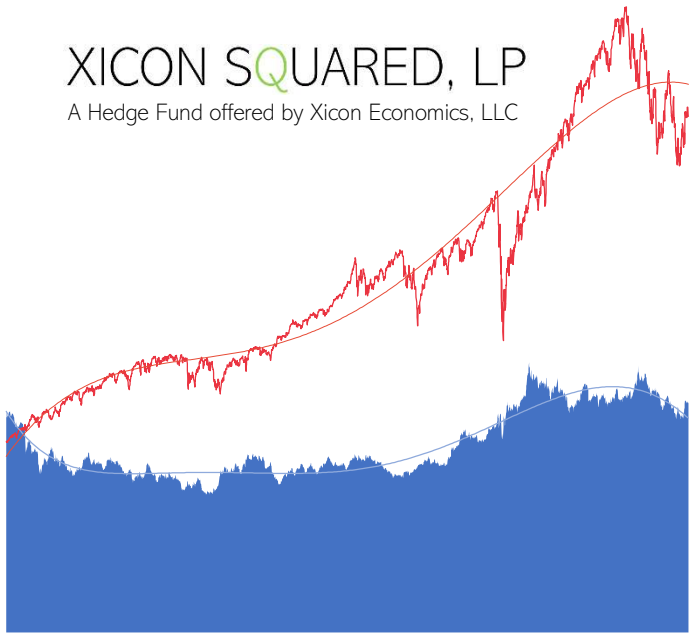
where, R^2 *= explained variance*
R^2_{adj} *= adjusted* R^2*, due to bias in* R^2
n *= number of observations/points in data set*
k *= number of independent variables in model*

While R^2 provides a strong estimation of model fit, R^2 has limitations, however. For example, R^2 does not address homoscedasticity. Homoscedasticity refers to noise or random disturbances, i.e., error, in the relationship between the dependent variables and the independent variable(s), and assumes the variation is constant, or nearly constant, across all independent variables. However, when homoscedasticity is not present, heteroscedasticity within the model's independent variables may be overweighted, controlling greater portions of variance than we prefer in regression models due to ordinary least squares' (OLS) effort of minimizing residuals (and thereby decreasing standard errors).

In investment modeling, R^2 helps researchers and analysts identify models that more strongly represent, or project, relationships within the data. In so doing, R^2 provides a means of expressing the predictive performance of a model. That said, while R^2 provides a strong indication of model fit, it remains imperative to understand R^2 prior to attempting to explain the model and its fit. Merely plowing through the software to obtain R^2, or any other output, without understanding the output allows for erroneous decisions—and in the case of financial economics, such decisions can be expensive.

Herbert M Barber, Jr, PhD, PhD serves as the Managing Partner and Chief Investment Officer of Xicon Economics. Intersecting the fields of engineering, finance, econometrics, and statistics, Dr. Barber is an expert in financial economics as it relates to the management of random walk theory and navigation of constructs surrounding efficient market hypotheses, especially within assets operating under extreme uncertainty. For over 30 years, he has provided advisory, consulting, and management of large capital investments in the private and public sectors. Dr. Barber holds 5 academic degrees, including two research doctorates.

Xicon Economics, LLC provides investment research, financial and investment advisory, and asset management for corporations and investors. We conduct scientific and applied research coupled with advanced statistical and econometric analyses and modeling to render complex financial and economic decisions to ensure investments are realized. We concentrate our practice on increasing output in hedge funds and alternative investments. Additional information regarding Xicon Economics can be found at www.xiconeconomics.com; additional information regarding Xicon Squared, LP can be found by visiting https://www.xiconeconomics.com/hedge-funds.



XICON economics