Review of: "Decoding the Correlation Coefficient: A Window into Association, Fit, and Prediction in Linear Bivariate Relationships"

Hrishikesh Vinod[1]

1 Fordham University

Potential competing interests: No potential competing interests to declare.

comments for *Qeios* by H. D. Vinod on Hosseini's Decoding the Correlation Coefficient: A Window into Association, Fit, and Prediction in Linear Bivariate Relationships

This is an introductory paper about the simplest bivariate linear regression Y=a+bX, where r(X, Y) is the simple correlation coefficient between the two variables. The regression coefficients are "a," the intercept, and "b," the slope. The coefficient of determination R2 (R-squared) is identical if we regress Y on X or flip the model and regress X on Y. However, the identical R2 values upon flipping hold only under linearity and do not hold if the relation is nonlinear (non-parametric kernel). The simple correlation r(X, Y) is the square root of R2, where the square root sign is that of the covariance of X and Y, cov(X, Y) only in bivariate linear regressions.

If we standardize X and Y such that each has mean zero and unit standard deviation (SDx=1, SDy=1), then the intercept "a" disappears from the standardized data model, and the slope "b" becomes the correlation r(X, Y). Such a conversion of the slope is an algebraic fact. Hence verify that the slope is: b=r (SDy/SDx) as correctly stated by the author. However, some comments regarding forecasting are incorrect.

The author correctly acknowledges the potential implications of correlations for causality. However, causality from passively observed data is a profound subject that requires a study of kernel regressions using flipped models and three criteria. A free R software package called **generalCorr** has simple R commands to compute causal directions and strengths between variables, even allowing for control variables. For theory, see Vinod (2017), "Generalized Correlation and Kernel Causality with Applications in Development Economics," in *Communications in Statistics - Simulation and Computation Volume 46, issue 6, 2017,* DOI link: http://dx.doi.org/10.1080/03610918.2015.1122048

There are dozens of published applications of Vinod's causality methods, including Allen & Hooper (2018). Generalized correlation measures of causality and forecasts of the VIX using nonlinear models. *Sustainability* 10 (8), 2695. https://doi.org/10.3390/su10082695. https://www.mdpi.com/2071-1050/10/8/2695

A more recent application of Vinod's methods is in Fonseca, et al. (2023). Assessing causality among topics and sentiments: The case of the G20 discussion on Twitter. *Journal of Information Science*, pages 1-16. https://doi.org/10.1177/01655515231160034

The assertions in Hosseini's paper are not intended to be applicable to even a slightly more general linear model where we have two regressors (X and Z), Y=a+bX+cZ. The new slope coefficients (b and c) then become so-called *partial* correlation coefficients adjusted by suitable ratios of standard deviations. See an open-access paper, Vinod (2022), "Kernel regression coefficients for practical significance," Journal of Risk and Financial Management 15(1), pp.1-13. https://doi.org/10.3390/jrfm15010032