

**Problem 6: Paper discussion:
Suggested answers**

The paper “Legal Determinants of World Cup Success” had a number of econometric problems.

Comments related to violations of the Gauss-Markov assumptions (resulting in both biased coefficients and standard errors).

1) Endogeneity - omitted variable bias - simultaneity

Many of the right-hand side variables cannot be considered exogenous but are likely to be determined jointly with the dependent variable (by some unobservables that are captured by the error term). Hence the zero conditional mean assumption is violated. This is likely to bias the results and makes causal interpretation very difficult. The author obviously is aware of this, as can be inferred from the last sentence “maybe some other forces are at work...”. The paper is hence written as a critique to the use of several endogenous variables in other studies (e.g. those by La Porta et al. on “Legal Determinants of External Finance”), and warns us to be careful when interpreting the results of this type of cross-country regressions with endogenous variables in the right hand side.

2) Selection bias:

A particular subset of countries is used (those for which data are available), which is not a random draw of countries. Since the availability of data might very well be correlated with the left and right hand side variables, this is a concern, since it will lead to biased estimates. Obviously, just considering World Cup finalist also constitutes a selection bias in itself.

3) Linearity: “Antidirector rights” is an ordinal measure and it would hence be preferable to use dummies to distinguish different categories, rather than pretending there is a linear effect.

4) The no-perfect collinearity is obviously not violated, since the author was able to estimate the regression.

5) Heteroscedasticity might be a concern (e.g. because of more variation among countries of French origin compared to Scandinavian origin). While the authors use robust standard errors to correct for it, the small sample size makes us worry, because the validity of robust standard errors relies on the law of large numbers (p273).

Other concerns:

6) The high R^2 is not very meaningful; since with a small number of observations and a relatively large number of independent variables, it is not difficult to explain a lot of variation in the dataset.

7) There might be an issue of multicollinearity in the regression, causing some of the large standard errors. This might explain the differences in significance of coefficients between the 2 estimations. In general, the lack of robustness of the results should make us doubt the usefulness of any of the results.