Review: Linear Regression and Instrumental Variable

This exercise is based on Carstensen, K. and E. Gundlach (2006), "The primacy of institutions reconsidered: Direct income effects of malaria prevalence. *The World Bank Economic Review*, 20(3), 309-339.

"Some recent empirical studies deny any direct effect of geography on development and conclude that institutions dominate all other potential determinants of development. An alternative view emphasizes that geographic factors such as disease ecology, as proxied by the prevalence of malaria, may have a large negative effect on income, independent of the quality of a country's institutions."

"In line with previous empirical studies, the following cross-country regression equation is used to estimate the relative effects of institutional quality ($INSTITUTIONS$) and malaria prevalence ($MALARIA$) on economic development, which here is measured by the logarithm of GDP per capita ($\ln GDPC$):"

$$\ln\ GDPC_i = \beta_1 + \beta_2 INSTITUTIONS_i + \beta_3 MALARIA_i + \varepsilon_i \tag{1}$$

$INSTITUTIONS$ and $MALARIA$ can be proxied by different variables from the data set. Begin with *gadp* as proxy for $INSTITUTIONS$ and *malrisk* as proxy for $MALARIA$. Additional information regarding the variables from the data set and their definitions can be found in the appendix of the paper.

Questions

(a) Estimate equation (1) by OLS. Report the estimated parameters and their standard errors. Are the (signs of the) estimated parameters plausible?

(b) You note that *malrisk* might be endogenous. Therefore, you decide to use *maleco* as an instrument for *malrisk*. What conditions must be met for *maleco* to be a strong/valid instrument?

(c) Estimate the equation by 2SLS. Report the estimated parameters and their standard errors.

(d) Estimate the first-stage regression for *malrisk* and compute the relevant first-stage F statistic. Does the result confirm or violate one or more of the conditions from task (b)?

(e) You find that *gadp* is also an endogenous regressor and decide to use *lnmort* as corresponding instrument variable. Now, estimate the equation by 2SLS using two instrument variables for the two regressors. Is the estimated income effect of malaria prevalence significant? Compare the results with those from (a).

(f) Can you statistically test whether the instruments are exogenous in this IV setting? Briefly explain.

(g) Finally, try to reproduce the results from Carstensen and Gundlach (2006) in table 1, column 1. They have excluded countries with less than 1 million inhabitants (*popmill* = 1) and countries that depend mainly on the oil industry (*oil* = 1) from their data set.
Hint: Carstensen and Gundlach (2006) employ different proxies for $INSTITUTIONS$ and $MALARIA$ and use a degrees-of-freedom adjustment to account for the small sample size.

(h) What is the main reason for the different point estimates in (e) and (g)?