**IV Regression and 2SLS**

The data in **cigarette.dta** contains information related to the demand for cigarettes by state in the US. You are interested in estimating the cigarette demand equation where the dependent variable is the per capita log of packs sold **log(packpc)**, with explanatory the log of real per capita income **log(perinc)**, and log of real after tax price per pack **log(ravgprs)**.

The log of real after-tax price per pack **log(ravgprs)** is potentially endogenous, so that the following instruments were collected: average state sales tax (**rtaxso**), and cigarette specific taxes (**rtaxs**).

1. Generate the necessary variables for the demand equation of cigarettes.
2. Estimate the demand model using ordinary least squares. Interpret the estimated coefficients. Are the signs of the coefficients consistent with your expectations?
3. What is the potential source of endogeneity for **log(ravgprs)**? Explain why **rtaxso** and **rtaxs** are conceptually sound instrumental variable candidates for **log(ravgprs)**.
4. Using the **rtaxso** and **rtaxs** as the instrument-candidates, estimate the demand equation using 2SLS. Compare the results with the OLS and interpret the IV regression coefficients. What are the noticeable differences, if there are any?
5. Are the instruments relevant? Support your answer with the appropriate test result.
6. Are the instruments exogenous? Support your answer with the appropriate test result.
7. Is there evidence to say that log of real after-tax price per pack **log(ravgprs)** is indeed endogenous?