

Due Wednesday, March 25
(40 points)

This assignment is on multicollinearity problems, specification error and instrumental variables estimation. The necessary data come from two sources: the Excel files `dat3.11a.xls` and `dat303sb.xls` available at <http://u.arizona.edu/~rlo/> under econ 522a. Be sure to use matrix commands and attach the supporting computer print out to the completed assignment and make clear where your answers are shown.

The following information pertains to questions 1 & 2. Consider the following (original) model of monthly imports of the chemical barium chloride from China:

$$\ln(\text{chnimp})_t = \beta_0 + \beta_1 \ln(\text{chempi})_t + \beta_2 \ln(\text{gas})_t + \beta_3 \ln(\text{rtwex})_t + \beta_4 \text{time}_t + u_t, \\ t = 1, \dots, 131$$

where *chnimp* is the volume of barium chloride imported from China, *chempi* is an index of chemical production, *gas* is the volume of gasoline production, *rtwex* is an exchange rate index that measures the strength of the U.S. Dollar against other currencies, and *time* is a linear time trend variable that takes on the values 1 through 131 corresponding to each of the 131 months spanned by the data.

1. OLS estimation of the model seems to indicate that monthly imports of barium chloride is driven entirely by a pure time trend. This could be because chemical production, gasoline production, and the exchange rate have nothing to do with barium chloride imports. Or it could be that there is a multicollinearity problem. Evaluate the model for multicollinearity problems versus an irrelevant/redundant variables problem. Your evaluation should include but not necessarily be limited to the condition number and the variance inflation factor criteria.
2. Assume that the true (non constant term) parameter values for the model are given by $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$ and $\beta_4 = 0.01$.

- a. Suppose a researcher estimates the hypothesized model

$$\ln(\text{chnimp})_t = \beta_0 + \beta_1 \ln(\text{chempi})_t + \beta_2 \ln(\text{gas})_t + \beta_3 \ln(\text{rtwex})_t + \varepsilon_t$$

Calculate the biases for the *OLS* estimators of β_1 , β_2 and β_3 in the researcher's model.

- b. Suppose a researcher estimates the true model

$$\ln(\text{chnimp})_t = \beta_0 + \beta_4 \text{time}_t + u_t,$$

Would one expect that the true variance of the OLS estimator of β_4 in the original model is larger than the true variance of the OLS estimator of β_4 in the true model? Explain. Determine whether or not the estimated difference between these two variances is positive.

3. Consider the following model of labor supply for a random sample of workers:

$$\ln(hours_t) = \beta_0 + \beta_1 \ln(wage_t) + \beta_2 \ln(\widetilde{nly}_t) + \varepsilon_t, t = 1, \dots, 200$$

where *hours* is annual hours of work, *wage* is the hourly wage rate, and \widetilde{nly} is the correct but unobserved measure of non labor income. Assume that ε_t is normally distributed and satisfies all of the standard assumptions. A researcher has no choice but to use the observed measure nly_t in her regression but believes that $nly_t = \widetilde{nly}_t e^{v_t}$, where v_t is an unobserved normally distributed random error term that is independent of \widetilde{nly}_t and that satisfies all of the standard assumptions.

- a. Estimate the model by *IV* (instrumental variables) using $\ln(fsize)$ as an instrument for $\ln(nly)$, where *fsize* is family size. If the researcher is correct about the relationship between nly_t and \widetilde{nly}_t , describe the statistical properties of OLS estimation of the model when using nly_t as a regressor.
- b. Test for the endogeneity of $\ln(nly_t)$ at the 5% level of significance using the Wu version (“variable addition test”) of the Hausman test.
- c. Evaluate the case for whether or not $\ln(fsize)$ is a weak instrument.