Outline Solutions

Q1: A good answer could contain:

(a) The coefficient on *lpris* is an elasticity because the model is log-log. A 1% increase in prisoners p.100,00inh. is estimated to increase the number of crimes p.c. by 0.57 percent holding the number of police officers p. 100,000inh. constant. The coefficient is strongly significantly different from 0.
The variable *lpris* is endogenous if it is correlated with the error term of the model.
Possible sources of endogeneity in this model are: 1) model misspecification by choosing the wrong functional form. 2) Omitted relevant variables 3) Simultaneity between crimes and prison population 4) measurement error in the regressors.

(b) The F-test version of the RESET test for model specification is conducted. It tests for omitted polynomials and interaction terms. A good reply would state the details of the test, including the null, the steps of the test, the statistic and rejection rule. The results provide strong evidence of model misspecification.

(c) The model in Table 1D is the model in Table 1B plus the additional regressor *lincpc*. It is obvious that the coefficients change when lincpc is included. A good reply would state to that the results in Table 1B suffer from omitted variable bias. A very good reply would write the omitted variable bias and explain that it is non-zero if the omitted variable plays a role and is correlated with the included regressors. Compare lecture slides “endogeneity” pp. 5-6.

(d) The regression in Table 1E is another structural equation of a two simultaneous equations system. The other one is 1B. This reflects the simultaneity of the two equations as crime and prison population likely depend on each other. This invalidates both regression results because of endogeneity of these variables. A good answer would work this out formally as on lecture slides “simultaneous equation models” pp. 12-15.

(e) A good answer would state the 2sls estimation approach (compare lecture notes endogeneity pp. 27-29) and would discuss validity of the two IVs. While the first condition (partial relationship with lpris) can be tested for (with a F-test for joint significance after the regression in Table 1E), the second requirement (no correlation with the error of the structural equation for lcrim) cannot be tested. One could try to reason that overcrowding litigation does not directly play a role for crimes, but it is unclear whether there is anything else omitted in the model for lcrim that is correlated with the IVs. If one knew that one of the two IVs was valid, one could conduct the Sargan test for overidentification for IV validity.

(f) The regression based Hausman Test for endogeneity is conducted. A good reply would outline and explain the test in detail. Compare lecture notes endogeneity pp. 41. Because the residuals are significant at the 5% level in the model for *lcrim*, there is some evidence of endogeneity in the model (provided the instruments were valid as this is required by the test).