Question 1.

Consider the following linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + u_i$$

where

$$\mathbb{E}\left(u_i\right) = 0, \ \mathbb{E}\left(x_i u_i\right) \neq 0, \ and \ \mathbb{E}\left(w_i u_i\right) = 0$$

so that the regressor x_i is endogenous. Assume on top of that that we do not observe w_i. As we still need to estimate the effect of x on the conditional expectation of y, we consider the regression

$$y_i = \beta_1 + \beta_2 x_i + v_i$$

1. Why is it reasonable to believe that regression (2) as endogenous?
   Let z_i be an instrumental variable exogenous in regression (1).

2. Why the candidate set of instruments that we shall use for an IV estimation of (2) is $Z_i = (1, z_i)'$ (before we check whether z_i is a valid instrument or not)?

3. Give the analytical expression of $\hat{\beta}_{2,iv}$, the IV estimator of β_2. (Hint: Use the IV estimator formula for just-identified models)
   Calculate the probability limit of $\hat{\beta}_{2,iv}$ in the following cases:
   i) w_i and z_i are not correlated.
   ii) w_i and z_i are correlated.

4. From 3.i) and 3.2) why is each of the conditions $\mathbb{C}ov\left(z_i, x_i\right) \neq 0$ and $\mathbb{C}ov\left(z_i w_i\right) = 0$ important?

Question 2.

In this exercise, we want to determine the effect of the wage on labor supply. To address this question, we use a sample composed of 4165 US workers. We consider a linear model in which the dependent variable *weeks* is the number of weeks worked in the year. Explanatory variables are the log of the monthly wage, denoted $log(wage)$, the number of years of schooling ($educ$), an indicator variable that equals 1 if the wage is set by a union contract ($union$), an indicator variable that equals 1 if the person is a female ($female$).

   The model has been first estimated by OLS. Results are reported in Column (1) of Table 1.

1. What is the impact of the wage on labor supply?

2. We suspect the variable $log(wage)$ to be endogenous. Give one possible reason for the potential endogeneity of $log(wage)$. Explain briefly.

3. Give (in words) the two characteristics that a valid instrument should possess.

4. A two-stages least squares estimation is performed, using the following instruments: an indicator variable that equals 1 if the individual works in a manufacturing industry ($industry$) and an indicator variable equal to 1 if the individual lives in an urban area ($urban$). Discuss briefly whether the two potential instruments are likely to be valid and why.

5. Estimation results are reported in Column (2) of Table 1. What is the impact of the wage on labor supply according to TSLS (2SLS) results?

6. To test overidentifying restrictions, a Sargan (or J) test is performed. The value of the test statistic is 1.051. Give its behavior under the null hypothesis, the formal decision rule, and your conclusion based on the above results.

7. Obtain the Hausman t-test statistic based. Does it suggest you should use the OLS results or the IV results ?

Table 1: Estimation Results: OLS and TSLS

Dependent variable: weeks

	(1) OLS	(2) TSLS
(Intercept)	44.7665	30.7044
	(1.2153)	(4.9997)
log(wage)	0.7326	3.1518
	(0.1972)	(0.8572)
educ	-0.1532	-0.3200
	(0.0321)	(0.0661)
union	-1.9960	-2.1940
	(0.1701)	(0.1860)
female	-1.3498	-0.2378
	(0.2642)	(0.4679)
N	4165	4165

Note: standard errors are in parentheses