Read each question carefully. Motivate and explain your answers clearly. The number of points for each question is indicated below the question. The maximum number of points is 32.

1. You are interested in testing a hypothesis regarding the coefficient β_1 in the linear regression

$$y = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3 + \beta_4 + e,$$

where $(y, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ are scalar random variables. However, your computational capacity is constrained so you can only estimate regressions with one explanatory variable and an intercept at a time. Discuss whether or not it is possible in that case to estimate the variance of e, $\sigma^2 = \mathrm{E}[e^2|\boldsymbol{x}]$, which you would need e.g. for hypothesis testing. If so, how would you go about estimating σ^2 in the best way?

(6 points)

2. Let the regression model of y on two scalar random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ (ignoring the intercept) be

$$y = \boldsymbol{x}'\beta + e = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + e,\ \mathrm{E}[\boldsymbol{x}e] = 0,\ \mathrm{E}[e^2|\boldsymbol{x}] = \sigma^2 < \infty.$$

You are interested in testing the hypothesis

$$H_0 : \frac{\beta_1}{\beta_2} = 1.$$

Derive the test-statistic to test the hypothesis H_0 using the asymptotic distribution of the parameter estimator. What is the appropriate reference distribution for the test statistic? You do not need to derive the variance of the parameter estimator, you can use

$$\mathrm{Var}[\widehat{\beta}] = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}.$$

(6 points)

3. Derive the variance of out-of-sample predictions from the heteroscedastic linear projection model

$$y = \boldsymbol{x}'\beta + e,\ \mathrm{E}[\boldsymbol{x}e] = 0,\ \mathrm{E}[e^2|\boldsymbol{x}] = \sigma^2(\boldsymbol{x})$$

at a prediction point $\boldsymbol{x}_0$ (assumed to be fixed).

(4 points)

4. Let the model be

$$\begin{aligned} y &= \beta \boldsymbol{x}^* + e \\ \boldsymbol{x}^* &= \gamma \boldsymbol{z} + v, \end{aligned}$$

where $\boldsymbol{x}$ is the endogenous (scalar) regressor, β the coefficient we wish to estimate, and $\boldsymbol{z}^*$ is the (valid, scalar) instrument. All variables are measured as deviations from their means and intercepts have been omitted. We have endogeneity, so $\mathrm{E}[\boldsymbol{x}^* e] \neq 0$ but $\mathrm{E}[\boldsymbol{z}e] = \mathrm{E}[\boldsymbol{z}v] = 0$. The errors have homoscedastic variances σ_e^2 and σ_v^2. The true endogenous regressor is not observed, but instead we observe one that is contaminated by "classical" measurement error. The observed regressor is

$$\boldsymbol{x} = \boldsymbol{x}^* + u,\ \mathrm{E}[\boldsymbol{x}^* u] = 0,\ \mathrm{E}[u^2|\boldsymbol{x}^*] = \sigma_u^2(\boldsymbol{x}).$$

Note that the error u is heteroscedastic. Suppose you have an iid sample consisting of n obervations on $y, \boldsymbol{x}, \boldsymbol{z}$. Derive the probability limits of the estimators of β and γ. Are they consistent?

(6 points)

5. Consider the full regression model:

$$y = \alpha t + \boldsymbol{x}'\beta + e$$

where y is the scalar dependent variable of dimension, t is a scalar "treatment" variable, taking the value of 1 for those receiving a treatment and 0 for all others, and $\boldsymbol{x}$ is a $k \times 1$ vector of control variables. Assume that $\mathrm{E}[e|t;x] = 0$, and that the error is homoskedastic. The parameter of interest is α. In order to estimate it we collect a sample $\{y_i; t_i; \boldsymbol{x}_i\}_{i=1}^{n}$.

   (a) Derive the LS estimators for α from the full model and from the reduced, "short" regression:

   $$y = \alpha^* t + u$$

   Are they consistent?

   (b) Derive the conditional variance of the OLS estimator for α when considering the full model and when using the reduced model. Provide an expression for a feasible estimator of the conditional variance of α in both cases, and compare their magnitude.

   (c) Suppose now that we believe that the omitted variable bias in the reduced model is equal to 0 (for instance because we conducted a randomized experiment such that t was randomly assigned to the individuals i in our sample).

      i. How can we test our belief? Define the null hypothesis, and provide a strategy to test it.

      ii. Discuss how you would choose your favourite regression model. (Hint: refer to the result of the test and to the precision of the estimation of α as discussed at point 5b.)

(10 points)