ECONOMETRICS 1, Spring 2024
Bent E. Sørensen

Midterm Exam 2—April 1, 2024

Each sub-question in the following carries equal weight.

1. (20%) a) Explain (give the formulas) how the White heteroskedasticity-consistent variance estimator is calculated. You will get 7% if you assume that there is only one regressor to simplify the expressions (that is, assuming it it otherwise correct).
b) Explain in detail why this estimator gives consistent estimates of the variances (make clear the assumptions that you use).

2. (20%) Assume that you are looking at a standard linear model with 5 regressors (including the constant) and that you estimate the model over a period of length 100 and that you find the sum of squared residuals to be equal to 70.
Now you suspect that the last 10 periods are different and you therefore estimate the model using the first 90 observations where you obtain a sum of squared residuals equal to 40, and you estimate the model using the last 10 observations and obtain a sum of squared residuals equal to 20. Test whether the parameters of the last 10 periods are equal to the parameters of the first 90 periods. (Write down the test in terms of residuals.) (State the assumptions under which the test is valid).

3. (20%) Consider the following linear regression model:

$$y_i = b_0 + b_1 x_i + b_{2i}^* w_i + e_i \tag{1}$$

where the errors e_i are independently normally distributed with mean 0 and variance s_i^2. You somehow know that the parameter b_{2i}^* varies in the population as $b_2 + z_i$ (for some unobserved vector z with mean 0, with variance σ_z^2, and uncorrelated with e, x and w).
a) What is the mean and variance of the OLS-estimator of the coefficients b_1 and b_2 (for a sample of N observations)?
b) What would be an efficient estimator of the coefficients?
c) If z_i were observed, how would you get more precise estimates of b_1 and b_2?

4. (20%) Consider the regression model

$$y_i \ = \ \beta_0 + \epsilon_i \ ; \quad i = 1, ..., n \ ,$$

where ϵ_i are independent identically distributed normally distributed variables with variance σ^2.

a) (4%) Find the OLS estimator $\hat{\beta}_0$.
b) (10%) Derive the distribution of $\Sigma_{i=1}^{n}(y_i - \bar{y})^2$. (3% from stating what the distribution is.)
c) (6%) What (explaln why) is the distribution of $(\bar{y} - \beta_0)/\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2/n(n-1)}$. (This is not hard if you look at it quietly for a bit.)

5. (20%) Matlab question. You want to estimate the model

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \ .$$

You have data for y, X_1, X_2 and another variable W for 10000 individuals. W is a valid instrument for X_1. X_2 is exogenous.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% To do: Fill in the missing code.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Estimate the coefficients.

n = size(y,1);                              % Sample size.

X = [ones(n,1) X_1 X_2];                    % X matrix.

Z = XXXXXXXXXXXXXXXXXXXXXXX                   % Z matrix.

b_est_IV = XXXXXXXXXXXXXXXXX;                  % IV-Estimates of beta.

u = XXXXXXXXXXXXXXXXX;                      % Residuals.

var_est = XXXXXXXXXXXXXXXXX;                % Variance-Covariance Matrix.

SE = XXXXXXXXXXXXXXXXX;                     % Standard errors.

t = XXXXXXXXXXXXXXXXX;                      % t-statistics.
```