Econometrics 1

1. For the simple linear regression $y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, \ldots, N$, the OLS estimator b for β has the variance $V\{b\} = \sigma^2/(N s_x^2)$ with variances σ^2 and s_x^2 of the error term and the regressor, respectively. To achieve an accuracy of the estimate b as high as possible, how has the design of the data collection be chosen?
2. Assume that the random variable ε_t follows the AR(1) process $\varepsilon_t = \varphi\varepsilon_{t-1} + v_t$ with $|\varphi| < 1$ and v_t being uncorrelated and having mean zero and variance σ^2. Show that the covariance $Cov\{\varepsilon_t, \varepsilon_{t-s}\}$ equals $\varphi^s \sigma^2 (1-\varphi^2)^{-1}$ for $s > 0$. Draw a schematic figure which shows the covariance $Cov\{\varepsilon_t, \varepsilon_{t-s}\}$ as a function of s for $\sigma^2 = 1$ and (a) $\varphi = 0.7$ and (b) $\varphi = -0.5$.
3. For estimating the parameters of the model $y_i = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$, observations for $i = 1, \ldots, N$ are available. Given that the regressors fulfill the relation $x_{i2} = h x_{i1}$ with a real number h, show that the rank of the (Nx3) matrix $X = (\ell, x_1, x_2)$ has a rank of at most 2; ℓ is a N-vector of ones, x_1 and x_2 are the N-vectors of observations of x_{i1} and x_{i2}, respectively.
4. The linear regression model $y_i = x_i'\beta + \varepsilon_i$, $i = 1, \ldots, N$ with K regressors is fitted to a set of data; you suspect that the error term variances are a function of variables $Z_2, \ldots, Z_p$. State the model for heteroskedasticity and the null hypothesis which are basis of the Breusch-Pagan test. Which probability distribution follows the test statistic of the Breusch-Pagan test?
5. The private consumption C_t of households is assumed to depend of the disposable income Y_t as indicated by $C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$. It is, moreover, assumed that $Y_t = C_t + I_t$, where I_t are exogenous investments; exogeneity means that $E\{I_t \varepsilon_t\} = 0$. Show that both Y_t and C_t are endogenous, i.e., that $E\{C_t \varepsilon_i\} \neq 0$ and $E\{Y_t \varepsilon_i\} \neq 0$.
6. Open the Ramanathan sample file “data3-7, Toyota station wagon repairs”, offered within the Gretl system. Perform the following analyses and interpret the results:
    a. Interpret the scatter plots of COST over (i) AGE and (ii) MILES. What do they suggest for the specification of a model for COST?
    b. Estimate the linear regression for COST with regressors AGE and MILES; show appropriate graphs of the residuals and report results of diagnostic tests suggested by a.); discuss possible remedies for at least two of the issues.
    c. Repeat b.) for the linear regression for log(COST) with regressors MILES and squared MILES, using heteroskedasticity-robust standard errors.
    d. Use both the White and the Breusch-Pagan test for heteroskedasticity; explain the results.
    e. Perform the Chow test for break (i) starting with observation 15 and (ii) starting with observation 43; compare the results with analogous ones for the regression of b.).
    f. Perform the PE test to test whether the regression of c.) is preferable to the regression for COST with regressors MILES and squared MILES (using heteroskedasticity-robust standard errors).
    g. Generate a dummy-variable D43 and estimate the regression for COST with regressor MILES and a term that make use of D43 in order to take into account the break starting with observation 43 (using heteroskedasticity-robust standard errors); compare the model based on suitable diagnostic tests with the model from b.).

    Document your analyses and interpretations in Gretl: save all relevant outputs, write your explanations and comments in a session file.