Advanced Applied Econometrics
Homework 1
Due to: 17th April, 2025, 20.00

General information: solution should be submitted electronically (via email using the SGH email address) and contain two files: pdf with solution and do-file (or other code, e.g. R script) that allows to replicate results. Please title your mail *[SGH] Advanced Applied Econometrics. Homework 1.*

Exercise 1. Consider the autoregressive model of order 1:

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \tag{1}$$

where ε_t is the error term and ρ is less in modulus than one. Assume additionally that the error term is serially correlated, i.e., $\varepsilon_t = \Xi\varepsilon_{t-1} + \eta_t$, where η_t is i.d.d. Check whether the least square estimator of ρ is unbiased and consistent.

Exercise 2. Assume the linear relationship between y and a set of explanatory variables, i.e, $y = X\beta + \varepsilon$. Consider now the ridge regression estimator:

$$\hat{\beta}^R = \left(X'X + \lambda I\right)^{-1} X'y, \tag{2}$$

where λ is constant and $\lambda > 0$. Check whether $\hat{\beta}^R$ is consistent and unbiased.

Exercise 3. Acemoglu, Johnson, and Robinson (2001) study an effect of institutions on economic development. Read this article and consider the relationship between economic development and institutions:

$$\ln gdp = \beta_0 + \beta_1 avexpr + \varepsilon \tag{3}$$

where gdp is the PPP-adjusted GDP per capita in 1995, $avexpr$ is the average protection against expropriation risk and ε is the error term.

(i) What signs would you expect on the coefficients β_1. Why?

(ii) Dataset `AcemogluEtAl2001` consists of data that are used by Acemoglu, Johnson, and Robinson (2001). Using their dataset and applying standard least squares estimator estimate (3) and interpret estimates of β_1.

(iii) Discuss reliability of the above estimates. Do you think that $avexpr$ is endogenous variable? Why?

(iv) Discuss why the settlers mortality why could be an appropriate (exogenous and relevant) instrumental variables for the $avexpr$.

(v) Test the relevance of the chosen instrumental variable.

(vi) Run the Hausman test interpret its result.

(vii) Consider now additional instrumental variable, i.e., European settlements in 1900 which is denoted by $euro1900$. Discuss why this variable could be an appropriate (exogenous and relevant) instrumental variables for the $avexpr$.

(viii) Using both instrumental variable estimate (3). Discuss possible differences in estimates. Repeat points (v)-(vi).

(ix) Run overidentification test and interpret its results.

(x) Extend the baseline regression (3) by the absolute latitude (lat_abst)

$$\ln gdp = \beta_0 + \beta_1 avexpr + \beta_2 lat\_abst + \varepsilon \tag{4}$$

(xi) What signs would you expect on the coefficients β_2. Why?

(xii) Estimate (4) using the least squares estimator. Interpret estimates.

(xiii) Discuss exogeneity of lat_abst.

(xiv) Repeat IV estimation of (4) with both instrumental variables. Interpret estimates and compare them with previous results. Run the first stage, Hausman and overidentification test and interpret results.

(xv) Actualize database. Download recent data on GDP (per capita, PPP-adjusted) and some proxy of institutions. Discuss you choice of proxy of institutions and compare it with the previously exploited measure, i.e., *avexpr*. Describe data sources that you used. Merge more recent data with the existing dataset. And replicate above regressions, i.e., (3) and (4), applying the least square estimator and IV regression with two instrumental variables. Discuss results.

Exercise 4. Aggregate Production Function. Consider the following Cobb-Douglas production function:

$$\mathcal{PF}^{CD}: \quad Y_t = \Gamma_t K_t^{\alpha_1} L_t^{\alpha_2}, \tag{5}$$

where Y_t is the real value added (output), K_t is the capital stock, L_t denotes the labor input while Γ_t represents the technical change. Despite the nonlinear nature the considered Cobb-Douglas production function can be expressed as:

$$\mathcal{PF}^{CD}: \quad y_t = \gamma_t + \alpha_1 k_t + \alpha_2 l_t, \tag{6}$$

where $y_t = \ln(Y_t)$, $\gamma_t = \ln(\Gamma_t)$, $k_t = \ln(K_t)$ and $l_t = \ln(L_t)$.

(i) Using dataset `USMacro.dta` estimate underlying structural parameters of the Cobb-Douglas production function (6). Assume temporarily that there is no technical change, i.e., $\gamma_t = const$. Interpret obtained estimates.

(ii) Modify assumption about technical change. Consider now the case that $\gamma_t = \gamma_0 + \gamma t$. Interpret this assumption. Next, estimate the underlying parameters of the Cobb-Douglas production function (6) and compare with previous results.

(iii) Interpret the R^2 for both regressions.

(iv) Are explanatory variables collinear?

(v) For both regression analyze whether the error term is normally distributed.

(vi) For both regressions analyze whether the error term is serially correlated. Interpret economically this property and, based on your economic knowledge, discuss whether the error term should be autocorrelated.

(vii) For both models, i.e., with and without technical change, apply the method of estimation that account for serial correlation of the error term. Discuss the differences in comparison to the previous points.

(viii) The key feature of the Cobb-Douglas production function is related to returns to scale. The returns to scale could be: (i) constant when $\alpha_1 + \alpha_2 = 1$, (ii) decreasing if $\alpha_1 + \alpha_2 < 1$, and (iii increasing when $\alpha_1 + \alpha_2 > 1$. Based on the previous results discuss what are the returns to scale in the US economy. Use an appropriate statistical test.

(ix) Consider now the more general production function, i.e. , the Constant Elasticity of Substitution production function (henceafter CES):

$$\mathcal{PF}^{CES}: \quad Y_t = \Gamma_t \left[\pi_0 \left(\Gamma_t^K K_t \right)^{\frac{\sigma-1}{\sigma}} + (1-\pi_0) \left(\Gamma_t^L L_t \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{7}$$

where σ is the elasticity of substitution between labor and capital, Γ_t^K and Γ_t^L is the capital- and labor-augmenting technical change, respectively. It is worth to note that if $\sigma = 1$ then the CES production function (7) takes the Cobb-Douglas form. In the associated literature, the key parameters of the CES production function (7) are estimated based on variation in relative prices of factors of production. In this vein, the standard optimization yields the following relationship between the logged labor share (ls_t) and the logged labor productivity (yl_t) and labor-augmenting technical change (Γ_t^L):

$$ls_t = \beta_0 + \frac{1-\sigma}{\sigma} \left[yl_t + \ln\left(\Gamma_t^L\right) \right], \tag{8}$$

where β_0 is the constant term. Estimate the parameters (8) by imposing assumption that $\Gamma_t^L = \exp(\gamma^L t)$. Test the serial correlation of the error term and, if it is necessary, account for this property in estimation of (8). What is the empirical value of the σ (Hint: the equation (8) is linear and thus σ will be nonlinear function of obtained parameters. Use the command `nlcom` to get also variance of key estimate.)? Is the assumption about Cobb-Douglas production function satisfied? Interpret estimates on γ^L.

(x) The elasticity between labor and capital could be also estimated based on data about capital share.Then,

$$cs_t = \eta_0 + \frac{1-\sigma}{\sigma}\left[yk_t + \ln\left(\Gamma_t^K\right)\right], \tag{9}$$

where η_0 is the constant term, cs_t is the logged cpaital share, yk_t denotes the capital productivity. Calculate the capital share by assuming that the labor and capital share sum to unity and that the sample average of the labor share is 0.66. Next, assume that $\Gamma_t^K = \exp(\gamma^K t)$ and estimate the underlying parameters of the CES production function based on (9). What is the empirical range of the σ ? Is the assumption about the Cobb-Douglas production function satisfied? Interpret estimate on γ^K. How the results are different from the previous point in which the variation in labor share was used to identify the σ.

(xi) Consider now the system estimation of the CES production function that combines equations (8) and (9). Discuss the problem of the parameters identification. Estimate the underlying parameters, i.e., σ, γ_k and γ_l. Compare results with previous points. Do the system estimates support and assumption of the Cobb-Douglas case?

(xii) Discuss potential sources of endogeneity in estimating the underlying parameters. Are they are the same for the Cobb-Douglas and the CES production function?

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.