ECOM-G314 Econometrics 1
Homework Assignment 4

This version of the assignment includes some additional links to the course material to help you find where a particular topic is being discussed. It should otherwise be identical to the original assignment.

1. (Adapted from Verbeek, Exercise 6.1) Consider the following linear regression model [35%]

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

We have sample of N independent observations, and assume that the error term $\varepsilon_i \sim NID(0, \sigma^2)$ and independent of all x_i. The density function of y_i (given x_i) is

$$f(y_i|x_i; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2}\right]$$

(a) Give an expression for the log-likelihood contribution of observation i, $\log L_i(\beta, \sigma^2)$. Explain why the likelihood function of the entire sample is given by

$$\log L(\beta, \sigma^2) = \sum_{i=1}^{N} \log L_i(\beta, \sigma^2).$$

Tips: The definition of independence (see e.g. B.4 in the textbook) will be helpful. Recall the basic properties of the logarithm (can you write $\log(a \cdot b)$ as a sum?).

(b) Determine the expressions for the two elements in $\partial \log L_i(\beta, \sigma^2)/\partial\beta$, where $\beta = (\beta_1, \beta_2)'$, and show that both have expectation zero for the true parameter values.

Tips: Note that (differentiable) functions of several variables can simply be differentiated component-wise. After differentiation, solve for ε_i from the model. Use the expectation of ε_i, and independence and its implications for expectations.

(c) Derive an expression for $\partial \log L_i(\beta, \sigma^2)/\partial\sigma^2$ and show that it also has expectation zero for the true parameter values.

Tips: As with (b).

(d) Show that $\partial^2 \log L_i(\beta, \sigma^2)/\partial\beta\partial\sigma^2 = \partial^2 \log L_i(\beta, \sigma^2)/\partial\sigma^2\partial\beta$, and show that it has expectation zero. What are the implications of this for the asymptotic covariance matrix of the ML estimator $(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)'$?

Tips: it may be useful to express both the partial derivatives in b) in matrix form. You can refer to the Schwarz-Clairaut theorem (differentiation order does not matter, $D_{x_1}D_{x_2}f = D_{x_2}D_{x_1}f$) if you don't want to differentiate the same things twice. Textbook pp. 192–193 and slide 7 in slides "Maximum likelihood estimation" will be helpful for the asymptotic properties.

(e) Present two ways to estimate the asymptotic covariance matrix of $(\hat{\beta}_1, \hat{\beta}_2)'$ and compare the two covariance matrix estimators.

Tips: read either the slides on "Maximum likelihood estimation" or the book chapter 6.1.2 carefully. For the comparison, consider whether there any similarities and differences between the estimators, either in limited samples or asymptotically.

2. (Verbeek, Exercise 6.1d) Consider the following linear regression model [15%]

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

We have sample of N independent observations, and assume that the error term $\varepsilon_i \sim NID(0, \sigma^2)$ and independent of all x_i. Suppose x_i is a dummy variable equal to 1 for the first N_1 observations and equal to 0 for $i = N_1 + 1, \ldots, N$. Derive the first-order conditions for the ML estimator. Show that the maximum likelihood estimators of β_1 and β_2 are

$$\hat{\beta}_1 = \frac{1}{N - N_1} \sum_{i=N_1+1}^{N} y_i \text{ and } \hat{\beta}_2 = \frac{1}{N_1} \sum_{i=1}^{N_1} y_i - \hat{\beta}_1,$$

respectively. What is the interpretation of these two estimators? What is the interpretation of the true parameter values β_1 and β_2?

Tips: you may find it useful to think about what is the sum of dummies over N observations. For interpretation of the values, it may be useful to think about conditional expectations, and how we interpret a dummy variable's coefficient in a regression setting.

3. Consider maximum likelihood estimation of the linear regression model [20%]

$$y_i = x_i'\beta + \varepsilon_i,$$

where $\varepsilon_i \sim NID(0, \sigma^2)$, discussed in the video lecture.

(a) Show that

$$s_i(\beta, \sigma^2) = \begin{pmatrix} \frac{\partial \log L_i(\beta, \sigma^2)}{\partial \beta} \\ \frac{\partial \log L_i(\beta, \sigma^2)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon_i x_i}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{1}{2}\frac{\varepsilon_i^2}{\sigma^4} \end{pmatrix}.$$

Tips: For the definition of the score vector (the gradient of the log-likelihood function), see slide 7 on "Maximum likelihood estimation" or chapter 6.1.2 in the textbook.

Recall that for each observation i, x_i is typically a **vector** of regressors (not a scalar).

(b) Show that

$$s_i(\beta,\sigma^2)s_i(\beta,\sigma^2)' = \begin{pmatrix} \frac{\varepsilon_i^2 x_i x_i'}{\sigma^4} & -\frac{\varepsilon_i x_i}{2\sigma^4} + \frac{\varepsilon_i^3 x_i}{2\sigma^6} \\ -\frac{\varepsilon_i x_i'}{2\sigma^4} + \frac{\varepsilon_i^3 x_i'}{2\sigma^6} & \frac{1}{4\sigma^4} - \frac{2\varepsilon_i^2}{4\sigma^6} + \frac{\varepsilon_i^4}{4\sigma^8} \end{pmatrix}.$$

Tips: Can you get the results through a straightforward matrix multiplication?

(c) Show that conditional on x_i

$$I_i(\beta,\sigma^2) = E[s_i(\beta,\sigma^2)s_i(\beta,\sigma^2)'] = \begin{pmatrix} \frac{1}{\sigma^2}x_i x_i' & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Tips: Expectation and variance of ε_i, plus independence.

You may take as given that a for random variable distributed as $NID(0,\sigma^2)$, the third moment is 0 and the fourth moment is $3\sigma^4$.

4. Consider the simple linear regression model [30%]

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

Let z_i be an independently normally distributed random variable with mean zero and variance unity ($z_i \sim NID(0,1)$), while $x_i = \delta z_i + \eta_i$, where $\eta_i \sim NID(0,1)$. Moreover, the error term $\varepsilon_i = \rho\eta_i$.

(a) Show that $\text{cov}(x_i, \varepsilon_i) = \rho$. Under which conditions is the regressor x_i endogenous?

Tips: independence and its implications for expectations.

(b) Show that $\text{cov}(z_i.\varepsilon_i) = 0$ and $\text{cov}(z_i, x_i) = \delta$. Under which conditions is z_i a valid instrument for x_i?

(c) Let $\beta_1 = 0.0$, $\beta_2 = 1.0$ and $\rho = 0.99$. Consider for four values of δ, $\delta = 0$, $\delta = 0.1$, $\delta = 0.5$ and $\delta = 1.0$. For each δ, generate $S = 1000$ samples of size $N = 100$ from the regression model, and for each generated sample, compute the IV estimate of β_2 and the t-test statistic for $H_0 : \beta_2 = 1.0$ against $H_1 : \beta_2 \neq 1.0$.

Because 1.0 is the true value of β_2, H_0 should be rejected in 5% of the replications in the t-test conducted at the 5% level of significance (the nominal size of the test). Compute the rejection rate of the test, i.e., find the proportion of the replications where the absolute value of the t-test statistic exceeds the critical value (1.96). How does the rejection rate and the distribution of the estimator vary with δ? [Hint: If the model is estimated using the `ivreg()` function of the `ivreg` package in R and the result is stored in `iv1`, the IV estimate of β_2 is obtained as `coef(iv1)[2]` and the covariance matrix estimator of the OLS estimator as `vcov(iv1)`.]

(d) According to the rule of thumb discussed in the video lecture, a weak instrument problem is unlikely if the first-stage F-statistic of the test of the significance of the coefficient of z_i exceeds 10 in the linear regression of x_i on a constant and z_i. Repeat (c), but compute the proportion of replications where the rule of thumb suggests that z_i is a weak instrument for x_i. How well does the rule of thumb seem to work in detecting the weak instrument problem according to you simulations? [Hint: If the model is estimated using the `ivreg()` function of the `ivreg` package in R and the result is stored in `iv1`, the first-stage F-statistic is obtained as `summary(iv1)$diagnostics[7]`.]

(e) Repeat (c) in the case of two instruments: $z_{i1} \sim NID(0,1)$, $z_{i2} \sim NID(0,1)$, and $x_i = \delta z_{i1} + 0 \cdot z_{i2} + \eta_i$. Compare the results to those obtained in (c) with one instrument.

(f) Repeat (e), but instead of the t-test, assess the performance of the test of over-identifying restrictions. In other words, compute the proportion of the replications where the p-value of the test of over-identifying restrictions does not exceed 5%. [Hint: If the model is estimated using the `ivreg()` function of the `ivreg` package in R and the result is stored in `iv1`, `summary(iv1)$diagnostics[12]` gives the p-value of the test of over-identifying restricions.]