Computational Methods in Econometrics: Assignment 1

In this assignment, you will use a Monte Carlo testing procedure to test for autocorrelation in regression residuals. The starting point of interest is the parametric regression model

$$y_t = x_t'\beta + \varepsilon_t \qquad\qquad \text{where } \varepsilon_1, \ldots, \varepsilon_n \sim \text{NID}(0, \sigma_\varepsilon^2).$$

Suppose that y_t and x_t are part of a time series, that is they are measured over time for $t = 1, \ldots, n$. For example, imagine that y_t represents methane emissions in the Netherlands in a given year t, while x_t contains influencing factors such as the number of cows, the population or the liters of produced cow milk in the same year. We can observe these quantities over a number of years and obtain time series $y = (y_1, ..., y_n)$ and $x_i = (x_{i,1}, ..., x_{i,n})$, for $i = 1, ..., k$, in this way. With time series data, it is possible that the value of a variable observed in the current time period will be similar to its value in the previous period. If the actual data generating process contains intertemporal dependence which our regression model does not capture, then typically, this results in the innovations being correlated.

To make sure that we are not in this situation, we would like to test whether the assumption $\mathbb{C}\text{ov}(\varepsilon_t, \varepsilon_{t-1}) = 0$ holds for all $t = 2, ..., n$. One way to do this is to start with the larger model

$$y_t = x_t'\beta + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + \nu_t, \qquad \nu_1, \ldots, \nu_n \sim \text{NID}(0, \sigma_\nu^2).$$

We proceed by testing H_0: $\rho = 0$ versus $H_1 : \rho \neq 0$ using the Durbin-Watson test. This test defines $d = 2(1 - \rho)$ and uses the statistic

$$\hat{d} = \frac{\sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=2}^{n}\hat{\varepsilon}_{t-1}^2}$$

to test H_0: $d = 2$ versus $H_1 : d \neq 2$. In the above equation, $\hat{\varepsilon}$ denote the OLS regression residuals which can be obtained by $\hat{\varepsilon} = (I_n - X(X'X)^{-1}X')y = M_X y = M_X \varepsilon$, where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$ and $M = [x_1 \ldots x_k]$. Unfortunately, the Durbin-Watson test statistic has a difficult distribution to derive, which makes the theoretical method to find rejection probabilities impossible. What we know is that the test statistic always ranges between 0 and 4 (why?) and it has a non-symmetric rejection region $R_d = (0, c_1) \cup (c_2, 4)$. The question is now, how to obtain c_1 and c_2. Luckily, we know how the computer can help here!

However, before we can use Monte Carlo methods, we need to show the following.

Exercise 1.

a. Show that $\hat{d}$ is a pivotal statistic under the null hypothesis. Why do we need to show this?

Next, load the data from the file called "DataAssign1.csv". It contains data on yearly methane emissions in the Netherlands from 1968 to 2022, as well as yearly cow milk production (both in 1,000 metric tonnes) and population (in 1,000). Since the data contains trends, it is customary to work with first differences $\Delta y_t = y_t - y_{t-1}$. Calculate first differences for all three variables before you proceed with the next exercise.

Exercise 2.

a. Run a regression of the (transformed) methane data on an intercept, milk production and population. Obtain the Durbin-Watson test statistic after your regression. What can we conclude from this value? To answer this, perform a Monte Carlo test for $B = 9,999$ and report the approximated rejection region using Monte Carlo approximations c_1^* and c_2^* for $\alpha = 0.1$. Also, derive and report $p_{MC}(y)$, the Monte-Carlo p-value.

b. Invert the Durbin-Watson test with approximated rejection region to find an approximate $(1 - \alpha)$-confidence interval for ρ. Note that you will need to use the test statistic for general values d_0 under the null which is given by $\hat{d} - (d_0 - 2)$ for the test H_0: $d = d_0$ versus H_1: $d \neq d_0$.

Finally, we would like to perform a small Monte Carlo simulation study of our test procedure and confidence interval obtained in the previous exercise.

Exercise 3.

a. Generate data under the null $\rho = 0$. For the data generating process (DGP), you can use the X matrix as well as the estimated coefficients $\hat{\beta}$ from Exercise 2. Use an estimate of the error variance $\hat{\sigma}_\varepsilon^2$, obtained from your regression, as variance for your simulated error terms. Run the theoretical Durbin-Watson test with your approximated rejection region. Report the empirical size. Does the approximated confidence interval cover the true ρ? Repeat this for $M = 10,000$ times. Comment on your results. Are they in line with your expectations?

b. Generate data under the alternative using $\rho = 0.4$. For the error variance σ_ν^2, you can use $\hat{\sigma}_\varepsilon^2(1-\rho^2)$. Run the theoretical Durbin-Watson test with your approximated rejection region for $M = 10,000$ times and report the approximate power. Finally, compute how often your confidence interval covers the true ρ. Discuss your results. Do they align with your expectations?