Computational Methods in Econometrics: Assignment 1

By Group 15

Sebastiaan van Dijen	1234567
Rick van der Ploeg	1234567
Lanlan Hou	2801069
Wa Chak Sou	2796840

School of Business and Economics

26^{th} September 2025

INTRODUCTION

In this assignment, we are going to work with a data set which contains the values of time series regression model

$$y_t = \alpha + x_{t,1}\beta_1 + x_{t,2}\beta_2 + \varepsilon_t = x_t'\beta$$

where

$$y_t = (y_1, \ldots, y_t), \quad x_t' = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{t,1} & x_{t,2} \end{bmatrix}, \quad \beta = (\alpha, \beta_1, \beta_2), \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

for $t = 1, ..., n$ observations. y_t represents contains data on yearly methane emissions in the Netherlands from 1968 to 2022; $x_{t,i}$ contains explanatory variables while $x_{t,1}$ is the produced cow milk (both in 1,000 metric tonnes) and $x_{t,2}$ is the population of cows (in 1,000) in the same year t. ε_t is the error terms where comes from a normal distribution with zero mean and unknown variance σ_ε^2. 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. We would like to test whether the assumption that $\text{Cov}(\varepsilon_t, \varepsilon_{t_1}) = 0$ holds for all $t = 2, ..., n$. The Durbin-Watson test is being used with the null hypothesis $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. This test defines $d = 2(1 - \rho)$ and uses the test 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 $H_1 : d \neq 2$. The test statistic $\hat{d}$ always ranges between [0, 4] and it has a non-symmetric rejection region $R_d = (0, c_1) \cup (c_2, 4)$. Using Monte Carlo Simulation, we are able to approximate the rejection region.

EXERCISE 1

a. Pivotal statistic $\hat{d}$ under the null hypothesis

In the Durbin-Watson test, uses the statistic $\hat{d}$:

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

Since $\hat{\varepsilon} = M_X\varepsilon$, then $\hat{\varepsilon}_t = r_t\varepsilon$:

$$= \frac{\sum_{t=2}^{n}(r_t\varepsilon - r_{t-1}\varepsilon)^2}{\sum_{t=2}^{n}(r_{t-1}\varepsilon)^2}$$

Rewrite $\varepsilon = \sigma u$:

$$= \frac{\sum_{t=2}^{n}(r_t\sigma u - r_{t-1}\sigma u)^2}{\sum_{t=2}^{n}(r_{t-1}\sigma u)^2}$$
$$= \frac{\sum_{t=2}^{n}(r_t - r_{t-1})^2}{\sum_{t=2}^{n}(r_{t-1})^2}$$

For hypothesis testing, it is possible to use Monte Carlo simulation if the test statistic is pivot. If the distribution of test statistic does not depend on a nuisance parameter, where $G_n(\cdot, F) = G_n(\cdot)$ for any $F \in H_0$, the rejection probabilities can be calculated. In Durbin-Watson test, using the test statistic $\hat{d}$, which is a pivotal statistic under the null hypothesis. As a result, Monte Carlo simulation can be proceed to approximate the critical values that define the corresponding rejection region.

EXERCISE 2

a. Approximated rejection region for $\alpha = 0.1$ and Monte-Carlo p-value

Using the y_t and x_t' as mentioned in the introduction, we ran a regression of the transformed methane data on an intercept, milk production, and population. Using OLS estimator, the estimated $\hat{\beta}$ can be computed by $\hat{\beta} = (X^T X)^{-1} X^T y$. Moreover, $\hat{y}_t$ can be computed with $\hat{\beta}$ by $\hat{y}_t = X\hat{\beta}$ and the estimated residuals $\hat{\varepsilon}_t$ can be estimated by $\hat{\varepsilon}_t = y_t - \hat{y}_t$. To assess the presence of serial correlation in the residuals, we computed the Durbin–Watson (DW) test statistic. The observed value of the test statistic is

$$\hat{d} = 1.342.$$

Since the null distribution of the DW statistic is non-standard, we employed a Monte Carlo procedure with $B = 9{,}999$ replications to approximate the critical region of the test. The estimated quantiles of the simulated distribution were

$$c_1^* = 1.556, \qquad c_2^* = 2.457,$$

corresponding to the 0.05 and 0.95 quantiles, respectively, for a two-sided test with significance level $\alpha = 0.1$.

Because the observed test statistic $d = 1.342 < c_1^* = 1.541$, we reject the null hypothesis of no serial correlation. The Monte Carlo p-value is defined as

$$p_{\mathrm{MC}}(y) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\left\{T^{(b)} \leq \hat{d}\right\},$$

where $T^{(b)}$ are the simulated test statistics under the null hypothesis. In our case, this evaluates to

$$p_{\mathrm{MC}}(y) = 0.013,$$

indicating strong evidence against the null.

b. Approximated $(1 - \alpha)$ confidence interval for ρ

To construct an approximate $(1-\alpha)$ -confidence interval for ρ, we can directly use approximated rejection region obtained in the Durbin-Watson test by inverting. The idea is to rewrite the rejection region as an event for which we do not reject.

In the Durbin–Watson test, it has a non-symmetric rejection region:

$$R_d = (0, c_1) \cup (c_2, 4)$$

which implies that $\mathbb{P}(T_n \in R_d) \leq \alpha$. To obtain the confidence interval we do not reject

$$\mathbb{P}(T_n \notin R_d) = 1 - \mathbb{P}(T_n \in R_d) \geq 1 - \alpha$$

Additionally, rewrite the probability as

$$\mathbb{P}(T_n \notin R_d) = c_1 \leq T_n \leq c_2$$

using the new test statistic which is given by $T_n = \hat{d} - (d_0 - 2)$, and for general values d_0 under the null, $d_0 = 2(1-\rho)$

$$\begin{gathered}
c_1 \leq T_n \leq c_2 \\
c_1 \leq \hat{d} - (d_0 - 2) \leq c_2 \\
c_1 \leq \hat{d} - (2(1-\rho) - 2) \leq c_2 \\
c_1 \leq \hat{d} - 2 + 2\rho \leq c_2 \\
c_1 \leq \hat{d} + 2\rho \leq c_2 \\
c_1 - \hat{d} \leq 2\rho \leq c_2 - \hat{d} \\
\frac{c_1 - \hat{d}}{2} \leq \rho \leq \frac{c_2 - \hat{d}}{2}
\end{gathered}$$

As a result, substituting the c_1^* and c_2^* obtained in Exercise 2a, the approximate $(1-\alpha)$-confidence interval for ρ is

$$\begin{aligned}
CI &= \left[\tfrac{1.541-1.342}{2}, \tfrac{2.454-1.342}{2}\right] \\
&= \left[\tfrac{0.199}{2}, \tfrac{1.112}{2}\right] \\
&= [0.107, 0.557]
\end{aligned}$$

By inverting the Durbin–Watson test with approximated rejection region, the approximate $(1-\alpha)$-confidence interval for ρ is $[0.107, 0.557]$.

EXERCISE 3

a. Theoretical Durbin-Watson test with approximated rejection region

We simulate $M = 10,000$ samples under the null hypothesis $\rho = 0$, using the estimated coefficients $\hat{\beta}$, design matrix X, and residual variance $\hat{\sigma_\epsilon}$ from Exercise 2. For every time, the new data are generated as following: $y^{(s)} = X\hat{\beta} + \epsilon^{(s)}, \quad \epsilon^{(s)} \sim \mathcal{N}(0, \hat{\sigma}_\epsilon^2 I_n)$

We then estimate the regression again, compute the Durbin-Watson statistic, and test H_0 using the Monte Carlo critical values(c_1^*, c_2^*). The corresponding confidence interval for ρ is obtained by test inversion.

The empirical size of the test is close to the nominal 0.1. The confidence interval covers the true $\rho = 0$ in the coverage rate of percentage for the simulations.

The result shows that the Durbin-Watson test maintains the size well under the null, and the confidence interval covers 90% of the time.

b. Under the alternative using $\rho = 0.4$

In this case, we repeat the simulation under the alternative using $\rho = 0.4$. To do one simulation, we start by generating $v_1, ..., v_n \sim \text{NID}(0, \sigma_v^2)$ for $n = 10000$, where $\sigma_v^2 = \hat{\sigma_\epsilon^2}(1 - \rho^2) = \hat{\sigma_\epsilon^2}(1 - (0.4)^2)) = 0.84\hat{\sigma_\epsilon^2}$. Then we can compute the errors by: $\epsilon_t = 0.4\epsilon_{t-1} + v_t$. We use this ϵ to calculate $y_t = x_t'\beta + \epsilon_t$, after which we fit a regression using OLS on y_t and X. Then we can compute the estimated residuals by $\hat{\epsilon_t} = y_t - \hat{y_t}$, where $\hat{y_t}$ is obtained from the regression. With everything that we calculated until now, we can calculate $\hat{d}$, using the formula from the introduction. After this we check whether $\hat{d} \in R_d = (0, c_1) \cup (c_2, 4) = (0, 1.541) \cup (2.454, 4)$, where we got c_1 and c_2 from exercise 2. If this is true we denote a rejection; otherwise we do not. After this, we calculate the confidence interval by CI $= \left[\frac{c_1 - \hat{d}}{2}, \frac{c_2 - \hat{d}}{2}\right] = \left[\frac{1.541 - \hat{d}}{2}, \frac{2.454 - \hat{d}}{2}\right]$ and check whether ρ falls in this interval. If this is true we denote that it falls in the interval, otherwise we do not.

After doing this simulation $M = 10000$ times, we can calculate the approximated power by approximated power $= \frac{\text{amount of rejections}}{M}$ and we can calculate how much the interval covers ρ by coverage of $\rho = \frac{\text{amount of times } \rho \text{ was in the confidence interval}}{M}$. After doing 10000 simulations, we got the following results: approximated power ≈ 0.7032 and coverage of $\rho \approx 0.8222$.