Violations of Classical Assumptions III: Serial Correlation

Marcio Santetti | Spring 2023

Table of contents

Introduction

Serial correlation, also known as **autocorrelation**, is a violation of CLRM **Assumption IV**, which states that observations of the error term are *uncorrelated* with each other. When this is no longer the case, values of the error term depend in some *systematic* way on observations from previous periods (in case of time-series and panel data), or on previous observations within the data set (for cross-section data).

Such problem is more frequent in *time-series data*, where the *order* of observations matters. Thus, in this lecture we take a break from cross-section to look at time-series data. When there is correlation within a regression's error term, our **inference** capabilities from it are undermined, although not causing **bias** to OLS coefficients.

In this lecture, we will see two "versions" of serial correlation: one that appears when the model is *correctly specified*, and another that is caused by a *specification error* such as omitted variables or incorrect functional form. Then, we look at its main consequences, and a few ways to deal with this problem. Regarding the latter point, we look at two *tests* to detect autocorrelation, and a statistical procedure to correct for this violation.

Pure serial correlation

Cases of "pure" serial correlation occur when CLRM Assumption IV is violated in a *correctly specified* model. Analytically, Assumption IV states that the *expected value* of the correlation coefficient (r) between two observations, i and j, is zero:

$$\mathbb{E}(r_{u_i, u_j}) = 0 \qquad \text{with } i \neq j$$

In case this condition is *not* satisfied, the model suffers from autocorrelation. Let us investigate the error term in more detail:

$$u_t = \rho u_{t-1} + e_t \tag{1}$$

Notice that we are now using t subscripts, denoting *time*. Here, *time* may be a day, a year, a quarter, etc. And the $t-1$ subscript simply means the value of a given variable in the *previous period*. Thus, if *t* is today, $t-1$ will be yesterday; if t is 2020, $t-1$ is 2019, and so on.

In equation (1), e_t is a classical (not serially correlated and normally distributed) stochastic residual term, and ρ is a parameter depicting the functional relationship between observations of the error term at time t and $t-1$, known as the *autocorrelation coefficient*.[1] As it should not be surprising, ρ is the *slope* coefficient of equation (1), and is thus calculated by $\hat{\rho} = Cov(u_t, u_{t-1})/Var(u_t)$.

[1] This term is the Greek letter *rho*.

The process described in (1) is known as a 1st *order Markov scheme*. Thus, ρ is a 1st *order autocorrelation coefficient*. As ρ approaches 1 (in absolute value), the higher the degree of serial correlation—that is, the higher the dependence of u_t on past observations. If $\rho > 1$, we have an *explosive* (unstable) trajectory.

It is also possible to model the dependency of the error term on higher-order autocorrelation coefficients, such as the following:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t \tag{2}$$

Equation (2) illustrates a 2nd order Markov scheme. We, however, will not go beyond the first-order case in our course. Just be

aware that the error term can have deeper levels of dependence with its previous observations.

Impure serial correlation

The "impure" version of serial correlation is more common to observe in practice, and is usually caused by a *specification* error, such as omitted variables and/or incorrect functional forms. The good news is that it can often be corrected through a better specification of the regression model.

Consider the following model, which we assume to be the "true" specification of y:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

And, instead, we estimate

$$y_t = \beta_0 + \beta_1 x_{1t} + u_t^*$$

where, as you may already expect, $u_t^* = u_t + \beta_2 x_{2t}$.

Now, u_t^* can be serially correlated **even if** u_t is not. The most plausible explanation for this fact is when x_{2t} *itself* is serially correlated. An independent variable being autocorrelated does not violate any classical assumption; however, when it is part of the population model and is, for some reason, omitted from our sample regression model, it is now part of the error term. Given the presence of a serially correlated component in u_t^*, the latter will be serially correlated, violating CLRM Assumption IV.

As an example, consider that x_{2t} is *disposable income*. If you consider that your disposable income in 2020 is dependent on your disposable income from 2019, then this series is serially correlated. If it is contained in the error term of a regression where it should explicitly be present, then the error term will be serially correlated, thus violating CLRM Assumption IV.

Consequences of serial correlation

After we have learned the difference between "pure" and "impure" serial correlation, it is time to know its consequences in more detail. Firstly, autocorrelation **does not** cause bias to OLS estimates. Recall that impure serial correlation may be caused by omitted variables, and these may cause bias, but this is not due to serial correlation itself.

Secondly, the most important impact of serial correlation is to β estimates' *standard errors* (SEs). Autocorrelation tends to *underestimate* standard errors, leading to higher t-scores. This way, we are more likely to *reject* null hypotheses, even though these are correct (type I error).

Thirdly, a direct consequence of the latter point is that OLS will no longer be *BLUE*, given the lack of precision for its estimates' standard errors.

Dealing with serial correlation

Serial correlation affects the standard errors of OLS β estimates. Given that the nature of this problem arises from observations of the error term being dependent on previous observations, *reordering* a data set's observations is **not** an option for time-series data, since the ordering of observations matters for this type of data. In case this problem appears in cross-section data, reordering observations may be an option, but it is hard to

believe a particular ordering of observations will be the source of the issue.

More effective options concern a careful *specification* of the model and *statistical tests*. Regarding the first, a model's specification must occur in accordance to the underlying *theory*, as always highlighted when problems such as omitted variables and incorrect functional form are possible. Therefore, *thinking* before playing around with regression models is always a good option.

With respect to the second point, we will look at two statistical tests to detect serial correlation in a regression model: the **Durbin-Watson** and the **Breusch-Godfrey** tests. These procedures will help us in detecting this issue with more robustness, through hypothesis testing. In case these detect its presence, what do we do? We will look at a new procedure, the **Cochrane-Orcutt estimator**, which aims to correct fot this problem and deliver more precise standard errors.

We will apply these procedures in the context of Macroeconomics. More specifically, we will model the relationship between *inflation* and the *unemployment rate* for the Australian economy between 1987 and 2016.

An applied example: The Phillips curve

Contraty to the belief that *either* inflation or unemployment should guide which macroeconomic policy to implement at any given point in time, the seminal paper by A.W. Phillips, called *The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957* (1958), pointed out that the existence of an *inverse* relationship between changes in money wages and the unemployment rate for the British economy.

Many versions of the *Phillips curve* exist in the literature, with the most widely studied being the association between *price*

inflation and *changes* in the *unemployment rate*. Equation 3 illustrates this relationship:

$$\pi_t = \pi_t^E - \beta(unemp_t - unemp_{t-1}) \qquad (3)$$

where π_t is the inflation rate at time t, π_t^E comprehends the *expected* inflation for period t (assuming no change in unemployment), and $(unemp_t - unemp_{t-1})$ is the change in unemployment from time $t-1$ to t. The hypothesis is that falling changes in unemployment $(unemp_t - unemp_{t-1} < 0)$ puts a pressure on prices (through the wage channel), due to an excess demand for labor. On the other hand, if unemployment is rising $(unemp_t - unemp_{t-1} > 0)$, an excess supply of labor pushes the price level down. In case we set expectations as *constant* over time, we can estimate the following simple regression model:

$$\pi_t = \beta_0 - \beta_{1t}\Delta unemp_t + u_t \qquad (4)$$

where now the intercept captures the constant expected inflation. Moreover, $\Delta unemp_t$ is the change in the unemployment rate, and u_t is the error term.

The next step is to estimate (4). Here, we use quarterly data for the Australian economy, for the 1987Q1–2016Q4 period, with a total of $n = 117$ observations :

$$\hat{\pi}_t = 0.73 - \underset{(0.21)}{0.398}\Delta unemp_t \qquad (5)$$

Notice that the sign of $\hat{\beta}_1$ confirms the Phillips curve's prediction of a *negative* association between π_t and $\Delta unemp_t$. However, this output does not inform anything about the error term. The first thing we can do is plot $\hat{u}_t$:
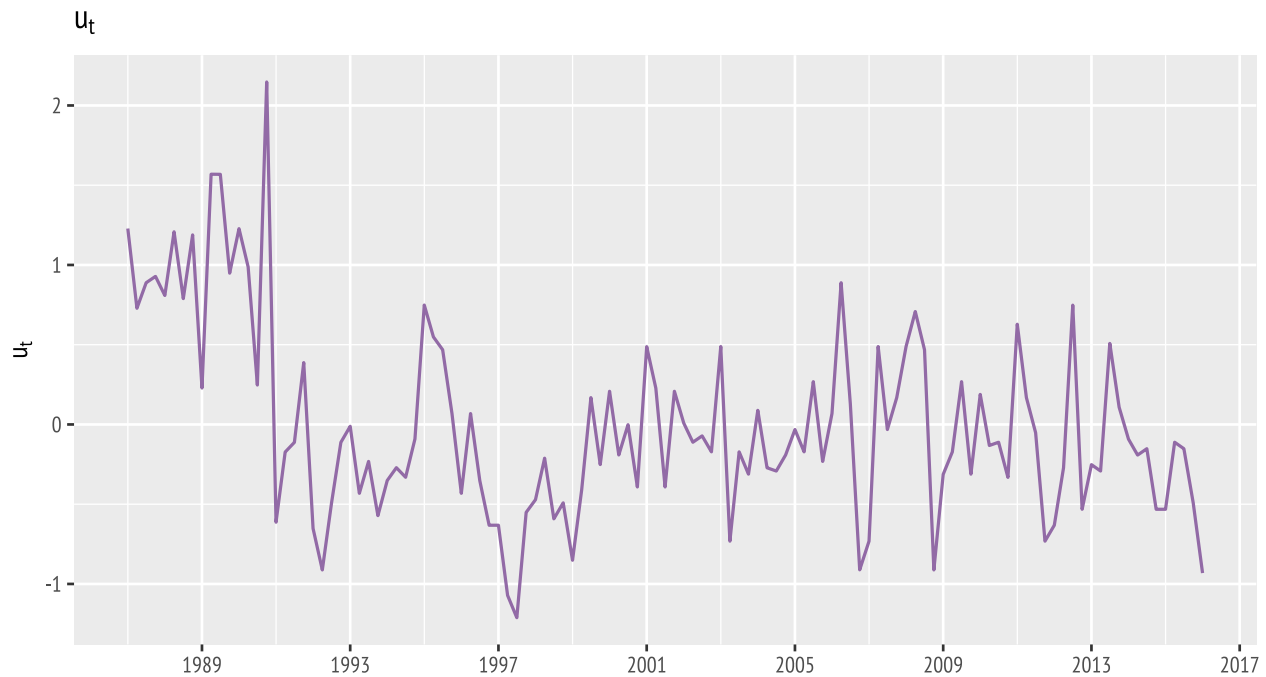
Figure 1: Model's residual term.

There seems to be some degree of dependence between past and present observations over time. For a more robust identification of autocorrelation, we can also look at the *Autocorrelation Function* (ACF) plot. The latter basically plots the values of $\hat{\rho}$, the autocorrelation coefficient from equation (1), across different *lags* (past values) of a variable. We look a it in Figure 2.

This plot shows the values for $\hat{\rho}$ from t (lag 0) until $t - 10$ (lag 10). When its values exceed the dashed bands, these are *statistically different* from zero (i.e., statistically significant) at $\alpha = 5\%$. Notice how significant the autocorrelation is across several lags.

After these visual inspections, it is time to run *statistical tests* for serial correlation. Let us start with the Durbin-Watson test.
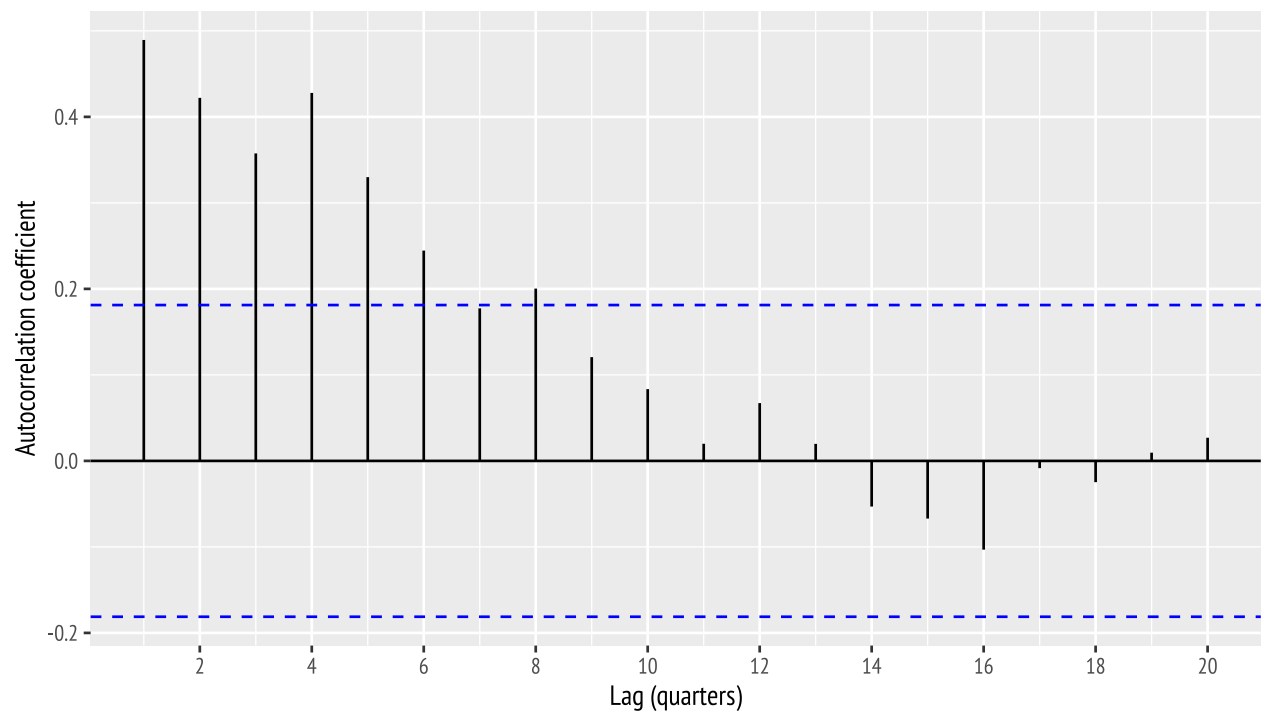
8

Figure 2: Autocorrelation function plot for the residual.

The Durbin-Watson *d* test

The Durbin-Watson (DW) test is implemented to determine whether there is **first-order** serial correlation in the error term, by examining the residuals of a model satisfying the following criteria:

- The regression model must have an intercept term (β_0);
- The model has no lagged independent variable (that is, having a variable such as $x_{i,t-1}$).

Its test statistic is computed with the following formula:

$$d = \sum_{t=2}^{T} (u_t - u_{t-1})^2 \Big/ \sum_{t=1}^{T} (u_t)^2$$

where *d*, the DW test statistc, lies between 0 and 4. When $d = 0$, there is evidence of *extremely positive* serial correlation (values of u_t will have the same sign as u_{t-1}); when $d \approx 4$, there is evidence of *extremely negative* serial correlation (values of u_t will have the opposite sign of u_{t-1}); and when $d \approx 2$, there is no evidence of serial correlation.

Fortunately, the value for *d* can be approximated by $2(1 - \hat{\rho})$, where $\hat{\rho}$ is the first-order autocorrelation coefficient, as in equation (1).

The DW test is a bit different from other procedures, since it can sometimes be *inconclusive*. Thus, in addition to *rejecting* or *not rejecting* the null hypothesis, we can also have an inconclusive decision from this test.

Here are the steps of the DW test:

1. Obtain the OLS residuals from the regression model;
2. Calculate *d*;

3. Based on k, the number of slope coefficients, and on n, the sample size, consult the DW table (available in this week's module files) for critical values. The table provides *upper* and *lower* critical values, d_L and d_U, respectively;
4. The test's *null* hypothesis is of *no serial correlation* in the residuals. In case we *reject* H_0, we have evidence of (first-order) serial correlation.

Below, we illustrate the decision regions for the DW test. *RR* is the *rejection region*, while *AR* is the "acceptance" region. The area in the middle is where the test is inconclusive (*IR*). Notice that the d statistic lies between 0 and 4. When $d < d_L$, we *reject* H_0; when $d > d_U$, we *do not reject* H_0; and when $d_L \leqslant d \leqslant d_U$, the DW test is *inconclusive*.

The Durbin-Watson test decision regions

$$
\begin{array}{ccccc}
RR & \vdots & IR & \vdots & AR \\
\end{array}
$$

$$
\begin{array}{ccccc}
0 & d_L & & d_U & 4 \\
\end{array}
$$
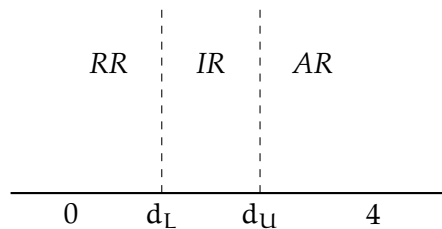
Let us compute d for the Phillips curve example. For the estimated model presented in equation (5), we extract the residuals, $\hat{u}_t$, and estimate the following *auxiliary regression*:

$$\hat{u}_t = \gamma_0 + \rho \hat{u}_{t-1} + e_t \tag{6}$$

where γ_0 is an intercept term. We want to estimate $\hat{\rho}$ from (6), which equals 0.5000583. Next, we compute the approximate value of the d statistic:

$$2(1 - \hat{\rho}) = 2(1 - 0.5000583) = 0.9998$$

Since we have only one dependent variable in our Phillips curve model, $k = 1$. With $n = 117$, the DW table gives us approximately $d_L = 165$ and $d_U = 169$, for a significance level of 5%.[2] Since our d statistic lies *below* the lower bound, it lies within the rejection region. Thus, we *reject* the null hypothesis of no serial correlation, and our Phillips curve model **violates** CLRM Assumption IV.

[2] Try to find these values in the Durbin-Watson table for critical values. In case you could not find the same values, do not worry. We will also use the table in the applied lecture.

The Breusch-Godfrey test

Another option to test for serial correlation is the *Breusch-Godfrey* (BG) test. Its main difference from the Durbin-watson test lies in the specification of the *auxiliary regression*. While equation (6) includes only the first lag of $\hat{u}_t$, the auxiliary regression for the BG test includes the *explanatory* variables of the regression model, allowing for a relationship between a control variable and the error term.

In addition, the BG procedure can be used to test for *higher-order* serial correlation in the error term, which is a limitation of the DW test. Here, however, we will stick to first-degree autocorrelation, so we can compare the results from boh tests.

The *test statistic* for the BG test is a *Lagrange Multiplier* (LM) type of test, which is simply

$$LM = (n - q)R^2_{\hat{u}}$$

where q denotes the order of serial correlation we wish to test for, and $R^2_{\hat{u}}$ is the R-squared coefficient from its auxiliary regression.

This test's *null hypothesis* is also of no serial correlation, and it follows a Chi-squared distribution with q degrees-of-freedom.

Applying the BG test to our Phillips curve example, we estimate the following auxiliary regression:

$$\hat{u}_t = \gamma_0 + \rho\hat{u}_{t-1} + \gamma_1 \Delta unemp_t + e_t \qquad (7)$$

This regression returns an R-squared of 0.255. Since we are testing for first-degree autocorrelation, q = 1. Therefore,

$$LM = (117 - 1)0.255 = 29.6$$

With $\alpha = 5\%$, we once again *reject* the null hypothesis.

The Cochrane-Orcutt estimator

Okay, so our Phillips curve model suffers from serial correlation. Even though the coefficient on $\Delta unemp_t$ has the expected sign, we can no longer *trust* the model's *standard errors*. This fact undermines our *inference,* since the standard errors reflect the *precision* of our β estimates. What to do, then?

A nice option would be to estimate our model such that the error term is no longer serially correlated. The good news is that the **Cochrane-Orcutt** procedure allows for that, using many of the estimations we have already done here. It works along the following steps:

1. Estimate the original regression model via OLS;
2. Store the residuals, $\hat{u}_t$, from step 1;
3. Estimate a first-order Markov scheme for $\hat{u}_t$, storing $\hat{\rho}$;
4. Transform the variables from the original regression into *quasi-differenced* terms, using $\hat{\rho}$;

5. Re-estimate the model via OLS using the quasi-differenced variables from step 4.

Let's break down the above steps. From 1–3, there should be no surprises, since we have already done that a few pages ago. For our Phillips curve mode, $\hat{\rho} = 0.5000583$. Now, we can proceed to step 4.

We will *transform* the dependent and independent variables by *incorporating* the autocorrelation coefficient $\hat{\rho}$ into the story. Now, instead of working with π_t and $\Delta unemp_t$, we will use *quasi-differenced* versions of them. These are defined by:

$$\tilde{\pi}_t = \pi_t - \hat{\rho}\pi_{t-1} \tag{7}$$

$$\widetilde{\Delta unemp}_t = \Delta unemp_t - \hat{\rho}\Delta unemp_{t-1} \tag{8}$$

Equations (7) and (8) are, then, quasi-differences of the original dependent and independent variable, respectively. We will use them to re-estimate the Phillips curve, then correcting for serial correlation:

$$\tilde{\pi}_t = \tilde{\beta}_0 + \beta_1 \widetilde{\Delta unemp}_t + e_t \tag{9}$$

where $\tilde{\beta}_0 = (1 - \hat{\rho})\beta_0$, and e_t is the error term from equation (6), which is assumed to be a classical error term, with no serial correlation.

By estimating (9), we obtain:

$$\hat{\tilde{\pi}}_t = 0.702 - \underset{(0.2086)}{0.383} \widetilde{\Delta unemp}_t \tag{10}$$

Equation (10) is computed using the Cochrane-Orcutt estimator, and it delivers consistent standard errors. Let us compare equations (5) and (10), that is, the original estimated model and the one computed via Cochrane-Orcutt:

- **Original model**: $\hat{\beta}_1 = -0.398$; $SE(\hat{\beta}_1) = 0.21$
- **Cochrane-Orcutt model**: $\hat{\beta}_1 = -0.383$; $SE(\hat{\beta}_1) = 0.2086$

Notice that the estimated coefficients remain close to each other, since bias is not a problem. The *standard errors* are different, and even though it does not look as a huge change, let us check the Durbin-Watson d statistics for each model:

- **Original model**: d = 0.9998
- **Cochrane-Orcutt model**: d = 2.24

Now, we *do not reject* the DW test's null hypothesis for the model using the Cochrane-Orcutt procedure. In a nutshell, we have preserved the sign and estimate from the original regression, but now we have **reliable** standard errors, as well as a serially *uncorrelated* error term.

Okay, that was a lot of new content. When we move on to do these tests and procedures in practice, we need not manually run all these steps. The software has commands that will do everything for us, we just need to interpret the output. However, we *must* know what is happening behind all the tests and commands the computer runs for us, so we do not lose sight of what we are actually doing. We will look at more examples, if there are any loose ends in your mind, so never worry.