

Serial Correlation

EC 339

Marcio Santetti

Spring 2023

Motivation

The road so far

Over the past weeks, we have learned:

- That *omitting* relevant variables from a model causes **bias**;
- That deterministic/strong stochastic *linear relationships* between two independent variables harm regression **standard errors**, and, therefore, OLS **inference**.

This week and the next, we turn our attention to the **residual** term, u .

- We begin by investigating what happens when observations within u share some sort of **linear relationship**.
- This problem is *extremely common* in time-series data, given that the **order** of observations matters, which is not true for cross-section data.

Pure serial correlation

Pure serial correlation

Recall **CLRM Assumption IV**:

"Observations of the error term are uncorrelated with each other."

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

In a **well-specified** model, autocorrelation can be characterized in the following way:

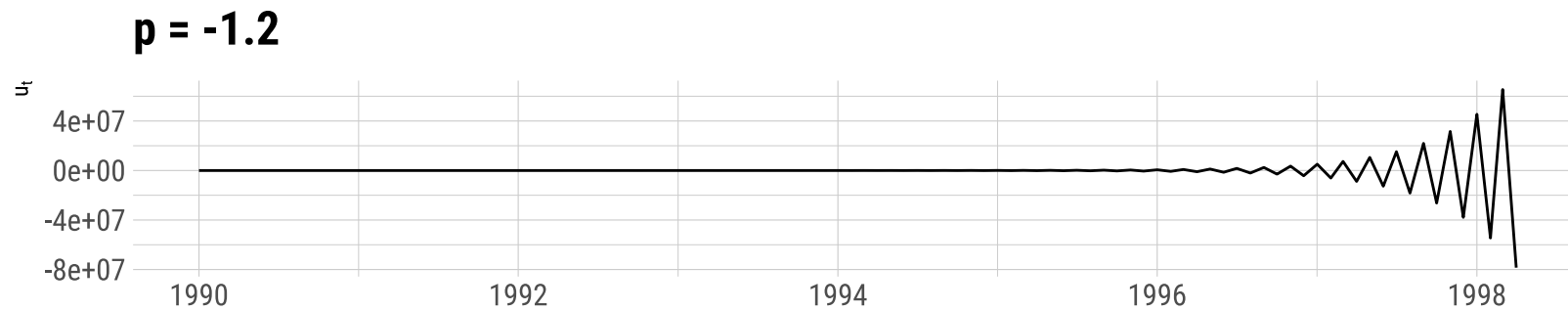
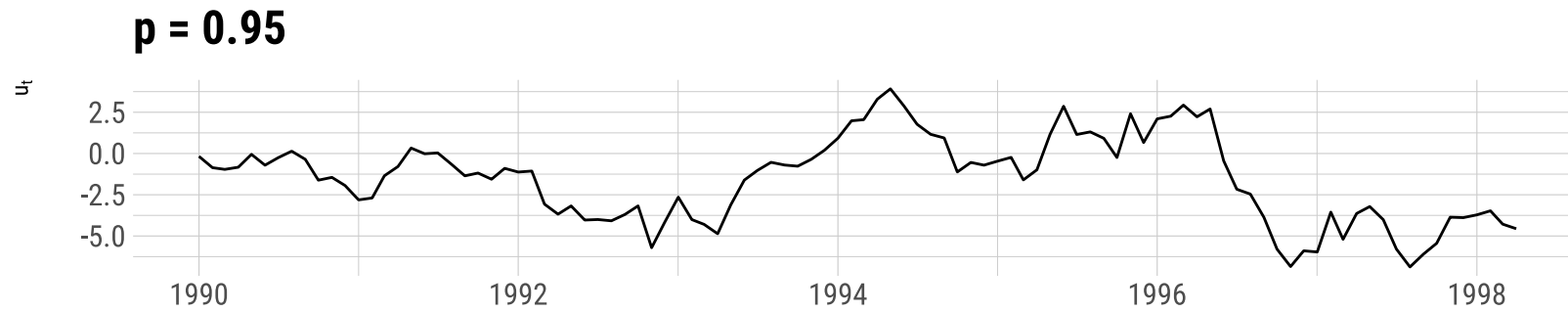
$$u_t = \rho u_{t-1} + e_t$$

where ρ is known as the *autocorrelation coefficient*.

As $\rho \rightarrow |1|$, the higher the *degree* of serial correlation.

If $\rho > |1|$, we have an *explosive* trajectory.

Pure serial correlation



Impure serial correlation

Impure serial correlation

The "**impure**" version of serial correlation occurs in *misspecified models*.

Whenever the error term contains a relevant variable that has been omitted from the model, which in turn is **serially correlated** itself, we have a case of impure serial correlation.

- A simple *example*: suppose we are interested in a person's wealth over time. In case we omit their *credit score* measure, for instance, it will be part of the error term.
 - Do you believe one's credit score *today* is dependent on their *last year's* credit score?
 - If you *do*, then this omitted variable is affecting the error term, thus causing serial correlation, even if the error term, *by itself*, is not serially correlated.

Impure serial correlation

Recall what happens when we omit a relevant variable from a model:

- Suppose we have the "true" population model:

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

- And instead we estimate:

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

with $u_t^* = u_t + \beta_2 x_{2t}$.

In case x_2 is *serially correlated*, it will affect the residual term, which in turn *will be* serially correlated.

Consequences of serial correlation

Consequences of serial correlation

Firstly, autocorrelation **does not** cause **bias** to OLS estimates.

However, it affects OLS **standard errors**, undermining **inference** from OLS models.

- Since it usually **underestimates** SEs, we end up being *more likely to reject* null hypotheses, increasing the likelihood of *Type I error*.

This way, OLS is no longer **BLUE**.

- Why? Its **B** part is affected.
- **"Best"** refers to **minimum variance**, which is not achieved with serial correlation.

Dealing with serial correlation

Dealing with serial correlation

In addition to **visualizing** OLS residuals, there are several **tests** for serial correlation.

The most common ones are the **Durbin-Watson** and **Breusch-Godfrey** tests.

Moreover, we can use the **Cochrane-Orcutt** estimator to correct for serial correlation.

We will study these procedures through an **applied example**.

Okun's law

Okun's law illustrates the relationship between **unemployment** and **growth** in an economy over time.

In a very *basic* form, it can be expressed as follows:

$$u_t - u_{t-1} = -\gamma(g_t - g_n)$$

- where u_t and u_{t-1} are the unemployment rate at time t and $t - 1$, respectively;
- g_t is the output growth rate at time t , and g_n is the "normal" output growth rate, which can be assumed as constant.
- The γ coefficient measures this relationship. If the growth of output is *above* the normal rate, unemployment falls; a growth rate *below* the normal rate leads to an increase in unemployment.

Okun's law

We can rewrite Okun's law as:

$$\Delta u_t = -\gamma g_t$$

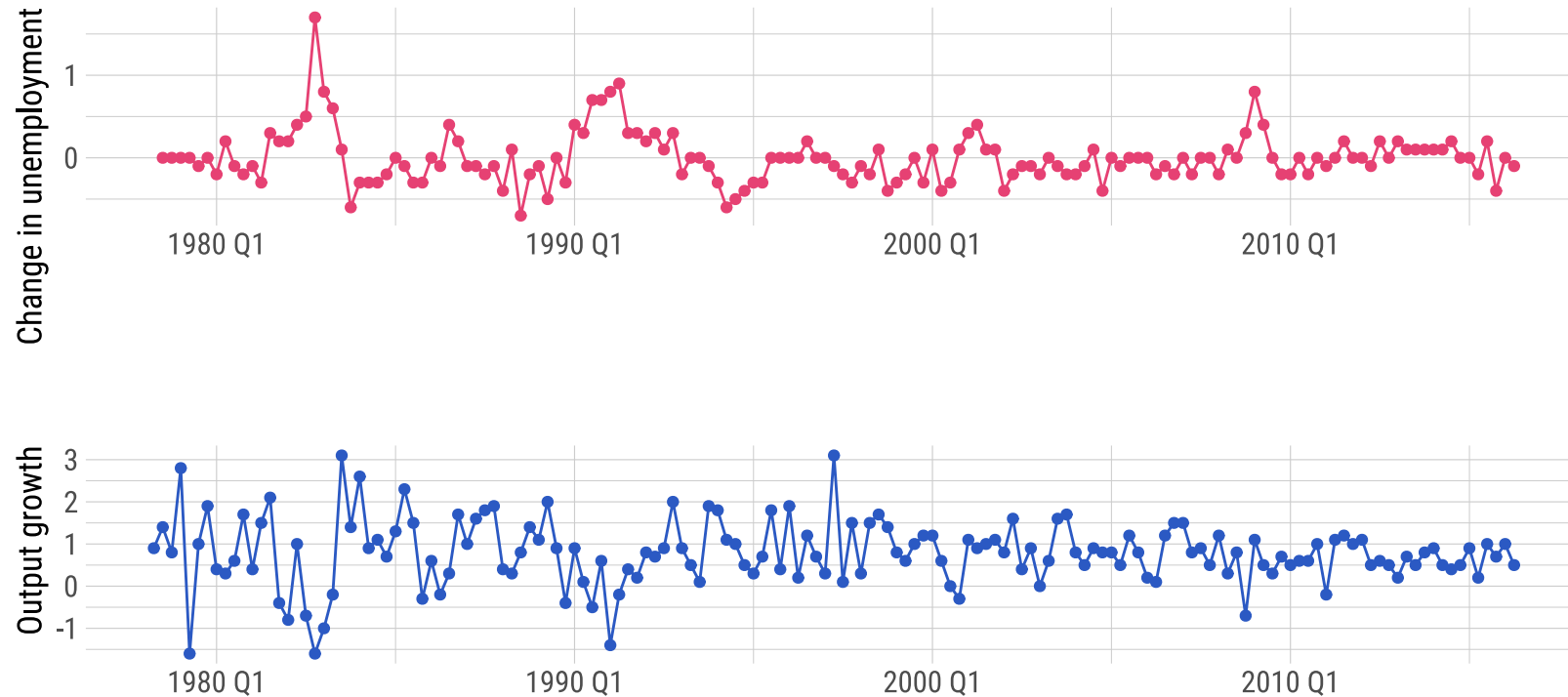
where Δu_t denotes the *change* in unemployment from $t - 1$ to t .

As an *econometric model*, we can write it as follows:

$$\Delta u_t = \beta_0 + \beta_1 g_t + \varepsilon_i$$

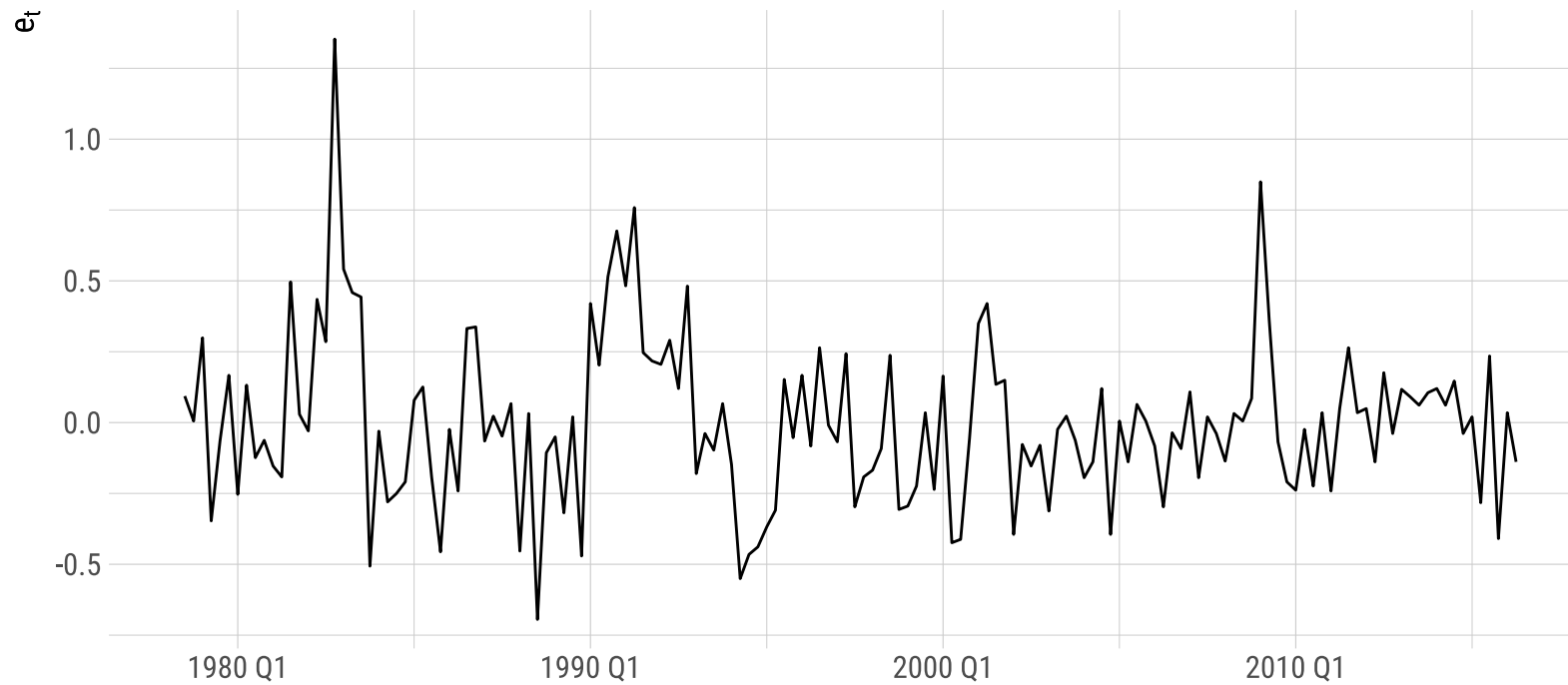
Let's throw some *data* in!

Okun's law: data for Australia (1978Q2–2016Q2):



Okun's law

A quick check at the model's **residuals**:



Does it look **autocorrelated**?

The Durbin-Watson test

The **Durbin-Watson** test for autocorrelation is used to test for **first-degree** serial correlation.

Provided that the regression model contains an *intercept* term (β_0) and has *no lagged* independent variable (e. g. , $x_{1,t-1}$), this test can be implemented.

$$d = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T (\varepsilon_t)^2}$$

with $0 \leq d \leq 4$.

It can be *approximated* by $2(1 - \hat{\rho})$.



The Durbin-Watson test

The **recipe** 🍳👨🍳:

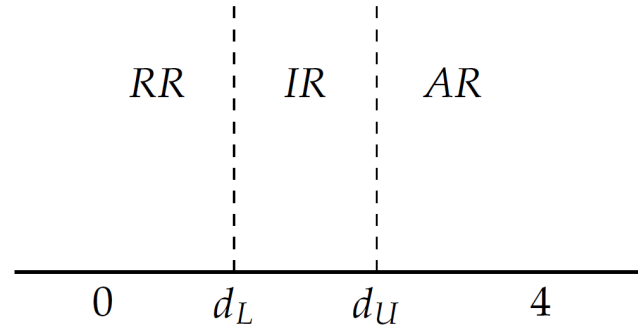
1. Estimate the regression model via OLS, storing its residuals;
2. Calculate the d test statistic;
3. Based on k , the number of slope coefficients, and on n , the sample size, consult the DW table for critical values.
4. The test's null hypothesis is of no serial correlation in the residuals. In case we reject H_0 , we have evidence of serial correlation.

The Durbin-Watson test

```
okun_model %>%  
  dwtest()
```

```
#>  
#>      Durbin-Watson test  
#>  
#> data:  .  
#> DW = 1.331, p-value = 1.562e-05  
#> alternative hypothesis: true autocorrelation is greater than 0
```

The Durbin-Watson test decision regions

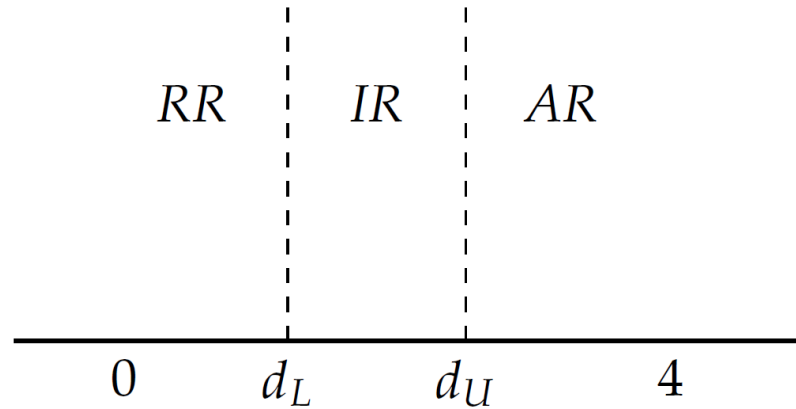


The Durbin-Watson test

```
. estat dwatson
```

```
Durbin-Watson d-statistic( 2, 152) = 1.330972
```

The Durbin-Watson test decision regions



The Breusch-Godfrey test

The **Breusch-Godfrey** test follows a similar procedure as the Durbin-Watson test's.

Its main *difference* involves the **auxiliary regression** estimated to find the autocorrelation coefficient, ρ . It must also include all **independent variables** from the original model.

$$LM = (n - q)R_{\hat{\epsilon}}^2$$

where n is the sample size from the original regression model;

q is the order of autocorrelation we wish to test for;

and $R_{\hat{\epsilon}}^2$ is the coefficient of determination from the auxiliary regression.

The Breusch-Godfrey test

```
. estat bgodfrey, lag(1) nomiss0
```

Breusch-Godfrey LM test for autocorrelation

lags(p)		chi2	df	Prob > chi2
1	1	18.154	1	0.0000

H0: no serial correlation

What is our *inference*?

The Cochrane-Orcutt estimator

From the two previous tests, we can infer that our Okun's law model suffers from *serial correlation*.

So what do we do?

The **Cochrane-Orcutt** procedure allows for the estimation of a *modified* version of the original regression model, allowing for serially *uncorrelated* residuals.

The Cochrane-Orcutt estimator

The **recipe** 🍳👨🍳:

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

The Cochrane-Orcutt estimator

Step 3: Transform the variables from the original regression into *quasi-differenced* terms, using $\hat{\rho}$.

For our Okun's law model, we have:

$$\tilde{g}_t = g_t - \hat{\rho}g_{t-1}$$

$$\widetilde{\Delta u}_t = \Delta u_t - \hat{\rho}\Delta u_{t-1}$$

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

$$\widetilde{\Delta u}_t = \tilde{\beta}_0 + \beta_1\tilde{g}_t + e_t$$

where $\tilde{\beta}_0 = (1 - \hat{\rho})\beta_0$.

The Cochrane-Orcutt estimator

```
. prais du g, corc
```

Cochrane-Orcutt AR(1) regression with iterated estimates

Source	SS	df	MS	Number of obs	=	151
-----+-----				F(1, 149)	=	0.19
Model	.0124954	1	.0124954	Prob > F	=	0.6622
Residual	9.71553772	149	.065204951	R-squared	=	0.0013
-----+-----				Adj R-squared	=	-0.0054
Total	9.72803312	150	.064853554	Root MSE	=	.25535

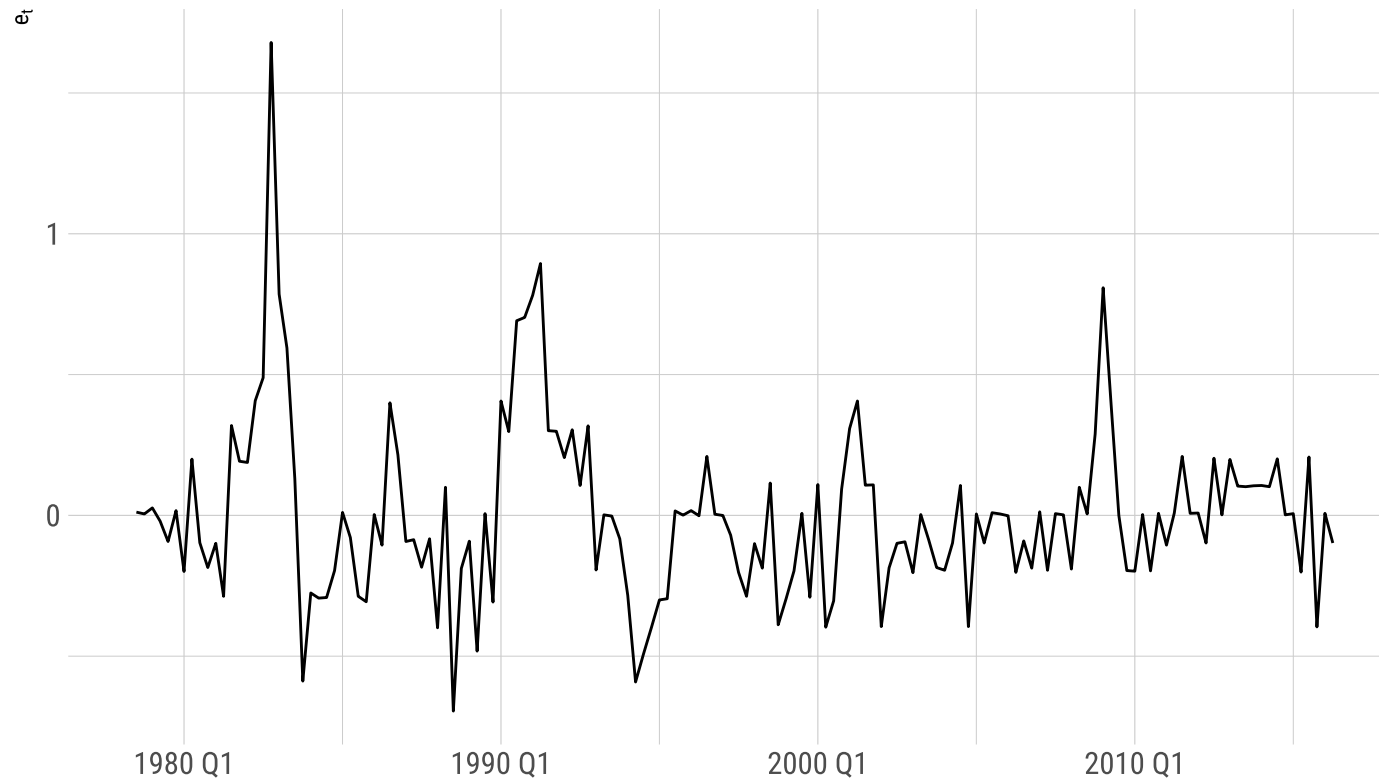
du	Coefficient	Std. err.	t	P> t	[95% conf. interval]
-----+-----					
g	-.010738	.0245295	-0.44	0.662	-.0592086 .0377326
_cons	.0035386	.0510639	0.07	0.945	-.0973644 .1044416
-----+-----					
rho	.5612189				

Durbin-Watson statistic (original) = 1.330972

Durbin-Watson statistic (transformed) = 2.270438

The Cochrane-Orcutt estimator

Now, the residuals from the **Cochrane-Orcutt** procedure:



Next time: Serial correlation in practice