Universidad Carlos III de Madrid

César Alonso

ECONOMETRICS

Topic 8: AUTOCORRELATION

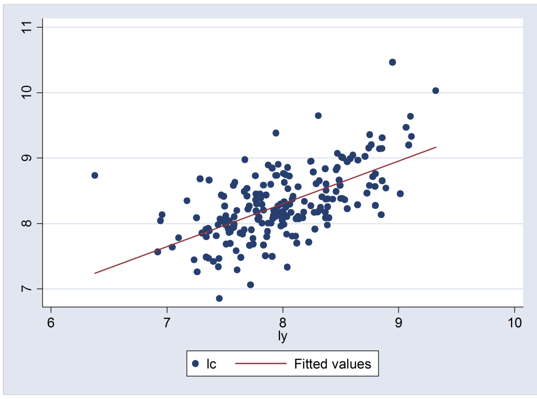
Contents

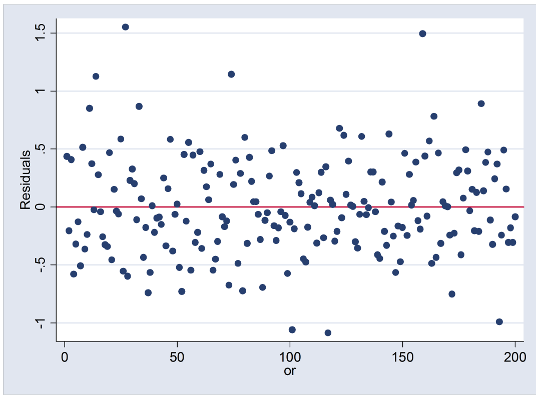
1 Introduction

- The term **autocorrelation** (or *serial correlation*) denotes those situations where realizations (observations) of the dependent variable are not independently drawn.
- This situation is very usual in the case of time series data.
- On the contrary, this situation does not happen in the case of cross-section data, for which indvidual units are independent each other.

Example (Cross section data):

- Cross section data about 200 Spanish households composed by couples with or without children, randomly drawn from the Spanish Consumer Expenditure Survey (EPF), 1990-91.
- Simple regression of the logarithm of per capita consumption on the logarithm of per capita disposable income.
- Left figure: logarithm of per capita consumption vs. logarithm of per capita income, and the corresponding OLS regression line.
- Right figure: residuals of the OLS linear regression vs. the order number of each observation (such order is arbitrary in the case of cross section data).
  - As residuals are the sample analogs of the population regression, if errors were independent each other we should not find any pattern.





- In the case of time series data (for which order of occurrence matters), autocorrelation is a frequent phenomenon, among other things, because of the **time dependence** associated with the inertia in economic data.

- It is quite plausible that **shocks o disturbances** affecting economic variables may show time dependence. *Examples:*

  - Consider the relationship between employment and inflation between 1950 and 1990.

    All unobserved factors which potentially can affect employment but are not included in such relationship are contained in the error term.

    Some of these unobserved factors can exhibit time dependence: particularly, **energy shocks** in the 1970s'.

    Such energy shocks can induce higher employment drop that that predicted by the model (i.e., a large positive error).

    As the effects of an energy shock are not damped fastly, we would expect errors in the next following years to be positive too.

    The magnitude of such errors due to the energy shock will decrease over time and, eventually, we will have errors closer to zero (even, negative) after some years.

  - Consider the relationship between stock index and growth.

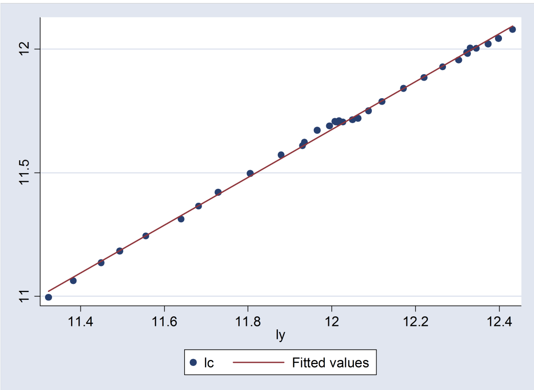
    Among the unobserved factors that can affect the stock index, we should mention the level of confidence of economic agents (which, in turn, is affected by unexpected news or surprises).

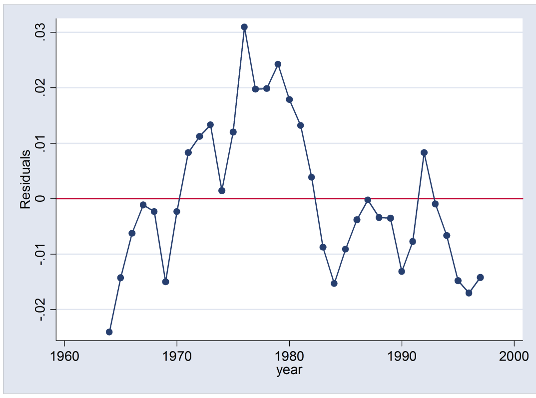
    Such unexpected news can lead the stock index below the value predicted by the model if agents' confidence is worsened.

    Presumably, the level of confidence of agents will exhibit inertia.

Example (Time series data):

- Annual data for the Spanish economy between 1964 and 1997 of aggregate consumption and GDP at market prices, in constant million euros of 1986.
- Simple regression of the logarithm of consumption on the logarithm of GDP.
- Left figure: logarithm of consumption vs. logarithm of GDP and the corresponding OLS regression line.
- Right figure: residuals of the OLS linear regression vs. the order number of each observation (i.e., the year, **which establishes a non-arbitrary order**).
  - As residuals are the sample analogs of the population regression of the relation between the log of consumption and the log of GDP, *if errors were independent each other, we should not find any pattern along time.*
  - However, we find periods when residuals are predominantly negative followed by other periods when residuals are predominantly positive.





- When observations are correlated each other, the OLS estimator can no longer be optimal, and it might be that the standard expression of its standard error can be inappropriate..

- For the sake of simplicity, we will consider the simple regression model, where we are conditioning on a single variable.

- To index observations, we will *use* t, s or $t-j$ *as subindices* –instead of i, h– to emphasize that we are considering time series.

2 El regression model with time series

- Consider a time series of (Y_t, X_t) $(t = 1, \ldots, T)$,
  i.e.: we observe T consecutive observations of Y_t and X_t.

- Suppose that all the classical regression assumptions, except that about independence among observations, are held.

- For a time series with T observations, we can write the model as

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (t = 1, \ldots, T)$$

- **Assunptions of the regression model with dependent observations**:

  1. Linearity in parameters $(Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t)$
  2.

     (i) $E\left(\varepsilon_t | X_1, \ldots, X_T\right) = E\left(\varepsilon_t | X_t\right) \quad \forall t$
     This assumption was always satisfied with independent observations (cross section data).
     But with time series data, this assumption is very restrictive,
     as it establishes that the conditional mean of the disturbance is only affected by the contemporaneous value of X.

     (ii) $E\left(\varepsilon_t | X_t\right) = 0 \quad \forall t$

  Then, assumptions 2.(i) and 2.(ii) together imply that

$$E\left(\varepsilon_t | X_1, \ldots, X_T\right) = E\left(\varepsilon_t | X_t\right) = 0 \quad \forall t$$

  Particularly, note that we are requiring the disturbance to be uncorrelated with past, present and future realizations of X.
  This implies that we are discarding the possibility that disturbances can affect future values of X.
  This implication, which entails **strict exogeneity** of X, is very restrictive.

It implies that, even though shocks or disturbances affect, by definition, Y, never affect X. (what seems unlikely)

Implications:

* $E(\varepsilon_t) = 0 \quad \forall t$ (by the law of iterated expectations).
* $C(X_t, \varepsilon_t) = 0 \quad \forall t$
* From assumptions 1. and 2.,

$$E(Y_t | X_1, \ldots, X_T) = E(Y_t | X_t) = \beta_0 + \beta_1 X_t$$

  i.e., the CEF is linear.

3.

(i) $V(\varepsilon_t | X_1, \ldots, X_T) = V(\varepsilon_t | X_t) = \sigma^2 \quad \forall t$
(**Conditional homoskedasticity**)

(ii) $C(\varepsilon_t, \varepsilon_{t-j} | X_1, \ldots, X_T) = C(\varepsilon_t, \varepsilon_{t-j} | X_t, X_{t-j}) = \sigma_{t,t-j} \quad \forall t, j \quad (j \neq 0)$
(**Conditional autocorrelation or conditional serial correlation**)
This assumption implies that the covariance between any two disturbances occurred in different periods (conditional to the realizations of X in those same periods) can differ from zero.

(iii) $\sigma_{t,t-j} = \gamma_j \quad \forall t, j \quad (j \neq 0)$
This assumption implies that the covariance between any two disturbances occurred in different periods (conditional to the realizations of X in those same periods) depends only on the time length j between both periods, but not on the particular period t.
In other words: the conditional covariance between disturbances occurred in 1980 and 1985 is the same as the conditional covariance between disturbances occurred in 1990 and 1995.
This assumption relies on the **stationarity (in covariance) condition**, which establishes that all first- and second- order conditional moments (means, variances and covariances) do not depend on the period of reference.
Actually, stationarity condition is already implicit in assumption 2. an in

assumption 3.(i).

Intuitively, stationarity entails that the relationship between variables is relatively stable along time. Otherwise, the parameters characterizing such relationship would vary along time, precluding us to infer how changes in a variable affect the mean value of another variable.

Thus, assumptions 3.(i), 3.(ii) and 3.(iii) together establish that

$$V\left(\varepsilon_t \mid X_1, \ldots, X_T\right) = V\left(\varepsilon_t \mid X_t\right) = \sigma^2 \quad \forall t$$
$$C\left(\varepsilon_t, \varepsilon_{t-j} \mid X_1, \ldots, X_T\right) = C\left(\varepsilon_t, \varepsilon_{t-j} \mid X_t, X_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$$

Implications (by the law of iterated expectations):

$$V\left(\varepsilon_t\right) = \sigma^2 \quad \forall t$$
$$C\left(\varepsilon_t, \varepsilon_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$$

3 Consequences on OLS estimation

- What are the implications of autocorrelation on the properties of the OLS linear regression of Y on X, i.e., $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \widehat{X}$?

- Recall that, for a given sample, the OLS slope can be expressed as

$$\widehat{\beta}_1 = \sum_t c_t Y_t$$

where, defining the explanatory variable in deviations with respect to its mean, $x_s = X_s - \overline{X}$,

$$c_t = x_t / \sum_s x_s^2$$

- Besides, we know that

$$\sum_t c_t = 0, \quad \sum_t c_t X_t = 1, \quad \sum_t c_t^2 = 1/\sum_s x_s^2$$

Thus, $\widehat{\beta}$ has conditional mean:

$$\begin{aligned} E\left(\widehat{\beta}_1 \middle| X_1, \ldots, X_T\right) &= \sum_t c_t E\left(Y_t | X_t\right) = \sum_t c_t \left(\beta_0 + \beta_1 X_t\right) \\ &= \beta_0 \sum_t c_t + \beta_1 \sum_t c_t X_t = \beta_1 \end{aligned}$$

Consequently, with the earlier assumptions, the OLS estimator remains **unbiased**.

- Moreover, it can be easily checked that

$$\operatorname{p\lim} \widehat{\beta}_1 = \beta_1,$$

so that the OLS estimator $\widehat{\beta}_1$ is a **consistent** estimator of β_1. To see this, notice that

$$\widehat{\beta}_1 = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t Y_t}{\frac{1}{T}\sum_{s=1}^{T} x_s^2} = \frac{S_{XY}}{S_X^2}.$$

$$\beta_1 = \frac{C(X_t, Y_t)}{V(X_t)}$$

As S_{XY}, S_X^2 are consistent estimators of $C(X_t, Y_t)$, $V(X_t)$, respectively, the ratio between S_{XY} and S_X^2 estimates consistently β_1.

- **Nevertheless**, the variance of the estimator is now

$$V\left(\widehat{\beta}_1 \middle| X_1, \ldots, X_T\right) = \sum_j \sum_t c_t c_{t-j} \sigma_{t,t-j} = \sigma^2 \sum_t c_t^2 + 2 \sum_j \sum_{t>j} c_t c_{t-j} \sigma_{t,t-j}$$

and, therefore, since $c_t c_{t-j} = x_t x_{t-j} \Big/ \left(\sum_s x_s^2\right)^2 = x_t x_{t-j} \Big/ (TS_X^2)^2$,

$$\begin{aligned} V\left(\widehat{\beta}_1\right) &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \sum_{t>j} x_t x_{t-j} \sigma_{t,t-j}\right] \\ &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \sum_{t>j} x_t x_{t-j} \gamma_j\right] \\ &= \sigma^2 E\left(\frac{1}{TS_X^2}\right) + 2E\left[\frac{1}{(TS_X^2)^2} \sum_j \gamma_j \sum_{t>j} x_t x_{t-j}\right] \end{aligned}$$

Example It is interesting to check the how the variance is obtained, explicitly, in a simple case. For $T = 3$, we have $\widehat{\beta}_1 = c_1 Y_1 + c_2 Y_2 + c_3 Y_3$, so that

$$\begin{aligned} V\left(\widehat{\beta}_1 \middle| X_1, X_2, X_3\right) &= c_1^2 \sigma_{11} + c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} \\ &+ c_2 c_1 \sigma_{21} + c_2^2 \sigma_{22} + c_2 c_3 \sigma_{23} \\ &+ c_3 c_1 \sigma_{31} + c_3 c_2 \sigma_{32} + c_3^2 \sigma_{33} \end{aligned}$$

Since $\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma^2$ and $\sigma_{12} = \sigma_{21} = \gamma_1$, etc., the expression becomes

$$\begin{aligned} V\left(\widehat{\beta}_1 \middle| X_1, X_2, X_3\right) &= \sigma^2 \left(c_1^2 + c_2^2 + c_3^2\right) + 2\left(c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} + c_2 c_3 \sigma_{23}\right) \\ &= \frac{\sigma^2}{TS_X^2} + 2\left(c_1 c_2 \sigma_{12} + c_1 c_3 \sigma_{13} + c_2 c_3 \sigma_{23}\right) \\ &= \frac{\sigma^2}{TS_X^2} + 2\left[\left(c_1 c_2 + c_2 c_3\right) \gamma_1 + c_1 c_3 \gamma_2\right]. \end{aligned}$$

- The fact that, in general, $\gamma_j \neq 0$ $(j \neq 0)$ makes that $V\left(\widehat{\beta}_1 \middle| X_1, \ldots, X_T\right) \neq \sigma^2 / \left(TS_X^2\right)$, so the variance differs from the standard expression.

- The computer package ignores that the model has changed, so it wll continue calculating the values

$$\begin{aligned} s^2 &= \textstyle\sum_{t=1}^T e_t^2 \Big/ (T-2) \\ s_{\widehat{\beta}_1}^2 &= s^2 / \left(TS_X^2\right) \end{aligned}$$

- Clearly, the standard estimator of the variance does not provide a proper estimate of $\sigma^2_{\widehat{\beta}_1}$, and the usual calculation of the standard error is inappropriate.
  Consequently, the usual confidence intervals and hypotheses test statistics are also wrong.

- **Efficiency**

  OLS, in this context, is **not** the Best Linear Unbiased Estimator.

  (Because it ignores the autocorrelation structure between errors).

4 Inference robust to autocorrelation: Newey-West variance estimator

- The rationale to obtain estimators of the variance of the OLS estimator that are robust to autocorrelation is similar to the one proposed by Eicker-White for conditional heteroskedasticity.

- In this respect, the natural way to estimate the additional term

$$\sum_j \sum_{t>j} x_t x_{t-j} \gamma_j = \sum_j \gamma_j \sum_{t>j} x_t x_{t-j}$$

would be, denoting the model residuals as $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$, $(t = 1, \ldots, T)$:

$$\sum_{j=1}^{J} \frac{1}{T-j} \sum_{t\ (t>j)} x_t x_{t-j} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j}.$$

and, using the assumption $C\left(\varepsilon_t, \varepsilon_{t-j}\right) = \gamma_j \quad \forall t, j \quad (j \neq 0)$, can be simplified to

$$\sum_{j=1}^{J} \frac{1}{T-j} \left(\sum_{t\ (t>j)} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j} \right) \left(\sum_{t\ (t>j)} x_t x_{t-j} \right).$$

- There appears the problem that, in principle, the sum across j should account from 1 to ∞ in order to capture all the possible non-zero covariances, but in practice, it is limited by the sample size T.

  In fact, the choice of J is arbitrary, a usual criterion being

$$J = int\left[(T)^{1/4}\right]$$

  where $\mathrm{int}(w)$ is the function that yields the integer of w.

  In practice, the choice is subject to the data frequency and the number of available observations.

  - Annual data: J between 1 and 3.
  - Quarterly data: J between 4 and 12.
  - Monthly data: J between 12 and36.

- The major problem of the resulting estimator

$$\widetilde{V}\left(\widehat{\beta}_1\right)=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\sum_{j=1}^{J}\frac{1}{T-j}\left(\sum_{t\ (t>j)}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}\right)\left(\sum_{t\ (t>j)}x_tx_{t-j}\right)$$

is that, for a given sample, it is not ensured that the estimation of $\widetilde{V}\left(\widehat{\beta}_1\right)$ is positive.

- Newey and West (1987) proposed a consistent estimator that circumvents this problem.

  To proceed, we must fix J to an integer number of periods after which the time dependence between errors is zero or negligible. The proposed estimator would be

$$s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\sum_{j=1}^{J}\left(\frac{j}{J+1}\right)\frac{1}{T-j}\left(\sum_{t\ (t>j)}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}\right)\left(\sum_{t\ (t>j)}x_tx_{t-j}\right)$$

  Examples:

  - $J=1$

$$s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}\left[\frac{1}{2}\left(\frac{1}{T-1}\sum_{t=2}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}\right)\left(\sum_{t=2}^{T}x_tx_{t-1}\right)\right].$$

  - $J=2$

$$\begin{aligned}s^2_{\widehat{\beta}_1}=\frac{s^2}{TS_X^2}+\frac{2}{\left(TS_X^2\right)^2}&\left[\frac{1}{2}\left(\frac{1}{T-1}\sum_{t=2}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}\right)\left(\sum_{t=2}^{T}x_tx_{t-1}\right)\right.\\&\left.+\frac{2}{3}\left(\frac{1}{T-2}\sum_{t=3}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-2}\right)\left(\sum_{t=3}^{T}x_tx_{t-2}\right)\right]\end{aligned}$$

- Newey and West proved this estimator to be consistent for arbitrary choices of J as long a J increases with sample size T (intuitively, the longer the available data history, the more flexible we can be about the autocorrelation pattern).

- in practice, most econometric packages include the option of autocorrelation-robust standard errors with OLS and IV/2SLS estimation.

  Some of these programs, like E-Views, take a default value of J (i.e., the maximum number of periods to calculate non-zero covariances).

- Besides, with time series data **heteroskedasticity and autocorrelation can appear together**. There exist a variacne estimator that is robust to both situations. In this respect, most econometric packages include this option.

5 Autocorrelation tests

- Suppose you suspect about the presence of autocorrelation, and want to assess such possibility given your sample data.

- In addition, suppose that if there is autocorrelation, it is first-order autocorrelation, i.e., $E(\varepsilon_t \varepsilon_{t-1}) \neq 0$. This would imply that, if we observed the time series of ε_t $(t = 1, \ldots, T)$, and run the regression

$$\varepsilon_t = \theta_1 \varepsilon_{t-1} + v_t,$$

  the null hypothesis of no autocorrelation $H_0 : E(\varepsilon_t \varepsilon_{t-1}) = 0$ would be equivalente to $H_0 : \theta_1 = 0$.
  (Since $E(\varepsilon_t) = 0$ for any t, in principle the linear projection of ε_t on ε_{t-1} does not require constant).

- However, we do not observe the errors ε_t. Instead, we observe the model residuals, $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$, which are the sample analogs of ε_t.

- Running the OLS linear regression of $\widehat{\varepsilon}_t$ on $\widehat{\varepsilon}_{t-1}$ without constant, the econometric package will yield a slope estimate

$$\widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1} \widehat{\varepsilon}_t}{\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1}^2}$$

  and a estandard error of the estimated slope $s_{\widehat{\rho}}$.
  To test $H_0 : \theta = 0$, it can be proved that $t_\theta = \widehat{\rho}/s_{\widehat{\rho}}$ has an approximate distribution $N(0, 1)$ under the null.
  A large magnitude of t_θ in absolute value will provide evidence against the null, $\theta = 0$ and in favor of the existence of first-order autocorrelation.

- The most popular first-order autocorrelation test (reported by most econometric packages) is the *Durbin-Watson test d*.

  This test is relarted with the first-order autocorrelation statistic that we have shown in the form,

  $$d \simeq 2(1 - \widehat{\rho}).$$

  However, the Durbin-Watson test has the inconvenient that its critical values depend on the sample size.

  Besides, its interpretation is counterintuitive, as *small* values of the Durbin-Watson statistic point out *high and positive* autocorrelation, and vice versa.

- In practice, we can test for the existence of higher order autocorrelation. For example, if we wanted to test for autocorrelation up to order q, we can run the regression

  $$\widehat{\varepsilon}_t = \theta_0 + \theta_1 \widehat{\varepsilon}_{t-1} + \theta_2 \widehat{\varepsilon}_{t-2} + \cdots + \theta_q \widehat{\varepsilon}_{t-q} + v_t,$$

  where $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t = Y_t - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_t\right)$ are the OLS residuals, and test the null $H_0 : \theta_1 = \theta_2 = \cdots = \theta_q = 0$.

  Intuitively, this test is, in practice, a regression test.