Lectures 18-19

Autocorrelation.

The assumption C.6 of Model C which is equivalent to the third Gauss-Markov condition, requires of the values of the disturbance term to have independent distributions, i.e. u_t is distributed independently of $u_{t'}$ for all $t \neq t'$. This assumption is usually satisfied for cross-sectional data models, while it frequently turns out to be violated when one deals with time series data: error terms for time periods not too far apart may be correlated. When C.6 assumption does not hold, it is said that the disturbance term is subject to autocorrelation, or serial correlation. It can be explained by the persistence of the behaviour of factors combined in the disturbance term in time. This lecture will analyze autocorrelation according to the following plan:

1. Reasons
2. Consequences
3. Detection
4. Remedial measures

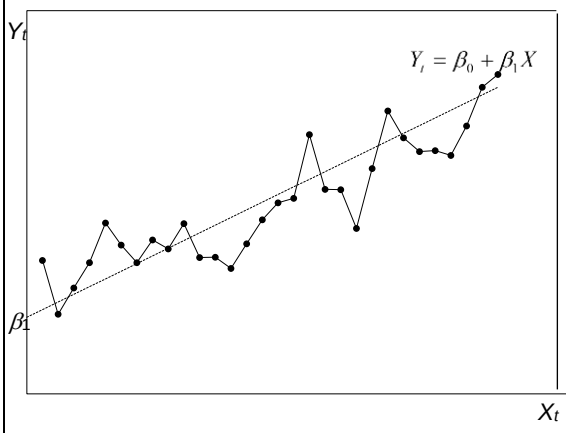
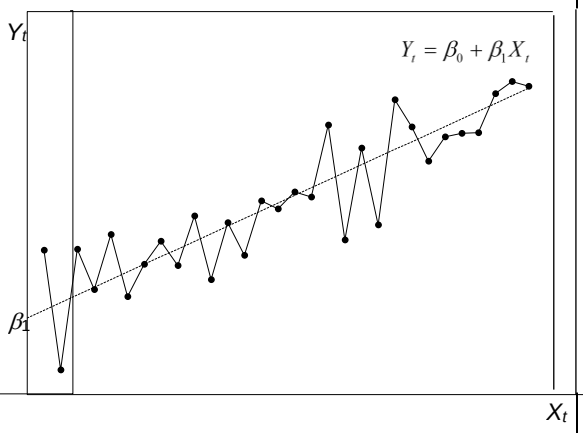
I. Reasons

In time-series models autocorrelation of the disturbances can arise from the following reasons:

1) Omitted variables: relevant factors omitted from the regression are correlated across periods. Since the disturbance term picks these missing factors up, the dependence of omitted factors (their persistence in time) translates into apparent autocorrelation of the disturbance term. For example, let the true model be a model with 2 explanatory variables:
$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + v_t$$
However, the second factor, X_{t2}, is not included into the estimated regression:
$$Y_t = \beta_0 + \beta_1 X_{t1} + u_t$$
Thus, the effect of X_{t2} is captured by the disturbance term u_t.Suppose that the omitted variable exhibits trends over time: X_{t2} depends on $X_{t-1,2}$, $X_{t-2,2}$… Therefore, as u_t combines these effects, the assumption of the serial independence of the disturbance term is violated
2) Misspecification of the functional form: for example, suppose that we use a linear specification for the relationship between Y and X instead of a quadratic one. Consider that X is growing over time. Thus, as in the linear model the disturbance term depends on X^2 that is increasing, the error term will also exhibit such growth, giving rise to autocorrelation.
3) Measurement errors: systematic errors (in the way how variables are measured) are reflected in the disturbance term that will behave systematically, showing a pattern. This can cause serial correlation.

Illustration:

Autocorrelation can be observed in scatter diagrams.

	Positive autocorrelation	**Negative autocorrelation**
Features:	Positive values of residuals tend to be followed by positive ones, and negative values by negative ones. Successive values tend to have the same sign.	Positive values of residuals tend to be followed by negative ones, and negative values by positive ones. Successive values tend to alternate in sign.
Graph:		
Frequency	Common in economic analysis. Explained by persistence of effects of omitted variables. For more observations (longer intervals) this effect is less perceived.	Uncommon in economic analysis. Sometimes this type of autocorrelation is induced by transformations of the original specification to make the model suitable for regression analysis.

Autocorrelation processes:

Consider a model: $Y_t = \beta_1 + \beta_2 X_t + u_t$. There are many forms of process to capture serial correlation:

1) First-order autoregressive autocorrelation: AR(1) – common type of autocorrelation described as:

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

It is autoregressive, because u_t depends on lagged values of itself. It is the first-order, because there is only one lag (dependence on its previous value). ε_t is innovation term at time t (white noise with zero mean and positive variance). For $\rho > 0$, the process is subject to positive autocorrelation. While for $\rho < 0$, there is negative autocorrelation.

For stationarity purposes we consider $|\rho| < 1$. Note that $E(u_t) = 0$.

Demonstration:

Lagging the process on time period: $u_{t-1} = \rho u_{t-2} + \varepsilon_{t-1}$. Substitute it for u_{t-1} into AR(1):

$$u_t = \rho^2 u_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$

Continuing to lag and substitute, we can express AR(1) in terms of innovation terms with diminishing weights: $\mathrm{E}(u_t) = \mathrm{E}(\varepsilon_t) + \rho\mathrm{E}(\varepsilon_{t-1}) + \rho^2\mathrm{E}(\varepsilon_{t-2}) + \rho^3\mathrm{E}(\varepsilon_{t-3}) + \ldots = 0$

By implementing Monte Carlo simulation, we can generate plots of residuals depending on the value of ρ. It shows that for $|\rho| < 0.3$, autocorrelation is weak and practically invisible. From $|\rho| > 0.3$, autocorrelation is beginning to be apparent. From $|\rho| > 0.6$, there is an obvious pattern in the behaviour of the disturbance term that is observed more frequently than we would expect as a matter of chance. When $|\rho|$ is around 0.9, autocorrelation is strong.

2) p[th] order autoregressive autocorrelation: AR(p)

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_p u_{t-p} + \varepsilon_t$$

Similarly, it is autoregressive but here it depends on lagged values of u_t up to the p[th] lag. It can also be demonstrated (by analogy) that if $|\rho| < 1$, then $E(u_t) = 0$.

3) q[th] order moving average autocorrelation: MA(q)

$$u_t = \lambda_0 \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \ldots + \lambda_q \varepsilon_{t-q}$$

The disturbance term is a linear combination of innovation terms up to the q[th] lag. Immediately, it follows that $E(u_t) = 0$.

4) ARMA(p,q) – general type of process:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \ldots + \lambda_q \varepsilon_{t-q}$$

It is a combination (aggregation) of AR(p) and MA(q). Hence, $E(u_t) = 0$

II. Consequences of autocorrelation:

The consequences of autocorrelation for OLS are similar to those of heteroscedasticity. In general, the regression coefficients remain unbiased, but OLS is inefficient because one can find an alternative regression technique that yields estimators with smaller variances. The other main consequence is that autocorrelation causes the standard errors to be estimated wrongly, often being biased downwards (for positive autocorrelation). Thus, *t*- and *F*-tests are no longer valid. Finally, although in general OLS estimates are unbiased, there is an important special case where they are biased (lagged dependent variable as a regressor).

Let's demonstrate some of these results for the case of the simple linear regression model:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Unbiasedness:

The estimated slope coefficient can be decomposed as follows:

$b_2 = \beta_2 + \sum_{t=1}^{T} a_t u_t$ where $a_t = \frac{X_t - \bar{X}}{\sum_{s=1}^{T}(X_s - \bar{X})^2}$

Taking expectation: $\mathrm{E}(b_2) = \beta_2 + \mathrm{E}(\sum_{t=1}^{T} a_t u_t) = \beta_2 + \sum_{t=1}^{T} \mathrm{E}(a_t u_t)$. If C.7 is satisfied, then a_t and u_t are distributed independently => expectation can be rewritten as $\sum_{t=1}^{T} \mathrm{E}(a_t u_t) = \sum_{t=1}^{T} \mathrm{E}(a_t)\mathrm{E}(u_t)$.

Since $\mathrm{E}(u_t) = 0$ independent of whether the disturbance term is subject to autocorrelation:

$$\mathrm{E}(b_2) = \beta_2 + \sum_{t=1}^{T} \mathrm{E}(a_t)\mathrm{E}(u_t) = \beta_2 + 0 = \beta_2$$ – unbiased.

Inefficiency:

We will not show this result analytically. Nevertheless, it can be mentioned that the proof of the Gauss-Markov theorem that established efficiency relies on the assumption of no autocorrelation. Since it is violated, OLS estimates are no longer BLUE and are hence inefficient.

Special case: The OLS estimators are biased and inconsistent for the model with the lagged dependent variable as a regressor with the disturbance term subject to autocorrelation:

Demonstration: $\begin{cases} Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{cases} <=> Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + \rho u_{t-1} + \varepsilon_t$

Lagging the model by one time period: $Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}$. It is evident that Y_{t-1} depends on u_{t-1}. Thus, assumption C.7 is violated, because the disturbance term affects Y_{t-1} in the observation for Y_t.

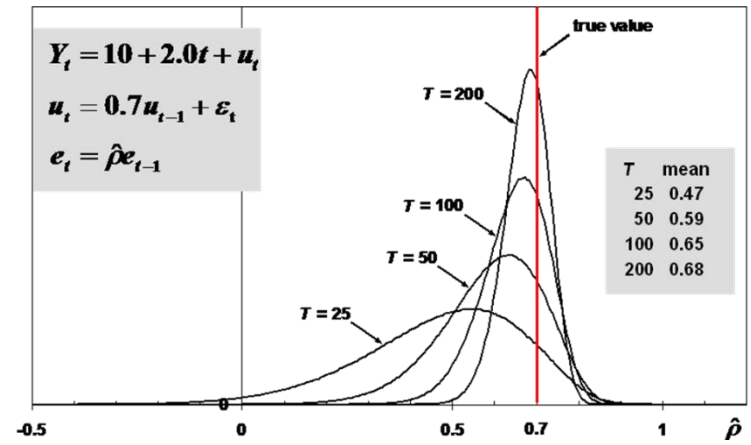
III. Detection: tests for autocorrelation:

1. **Residual plot:** a graph of the estimated residuals e_t against time.

Since the disturbance term is theoretical concept, it is non-observable. Residuals are not the values of the disturbance term but their behaviour is very much similar to the behaviour of the disturbance term. Therefore, one way to detect autocorrelation is to look at residuals: it is likely that

if the disturbance term is subject to autocorrelation, then the residuals will be subject to a similar pattern of autocorrelation.

For example, if the disturbance term follows the AR(1) process, then provided that the sample size is large enough (condition for consistency is satisfied), the regression parameters will converge to their true values and residuals will be very close to the values of the disturbance term. Hence, a regression of e_t on e_{t-1} can be sufficient, at least in large samples. It can be illustrated with the Monte Carlo experiment shown on the graph below.



The true model is generated as AR(1) with $\rho = 0.7$. On the horizontal axis it shows the estimated value of this coefficient when e_t is regressed on e_{t-1}. For a small sample size the distribution of ρ is skewed to the left and heavily biased downwards. However, as the sample size increases, the bias becomes sufficiently small and it is clear that the mean value is approaching to the true value of ρ.

2. Breusch-Godfrey test:

The idea of the Breusch-Godfrey test is that in order to control the effects of any endogeneity on the residuals we should also include all explanatory variables from the original model into the residuals regression. However, the theory is complex relating to maximum-likelihood estimation. Therefore, several asymptotically-equivalent versions of the test have been proposed. We will look at the Lagrange Multiplier version of the test. It allows to identify serial correlation not only of the first order but higher orders as well. Note that it is valid only for large samples.

Consider a linear regression:

$$Y_t = \beta_1 + \sum_{j=2}^{k} \beta_j X_{jt} + u_t$$

The disturbance term might follow an AR(p) process: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \varepsilon_t$.

Steps:

1) Estimate the initial regression model by OLS and calculate its residuals, e_t;
2) Estimate $e_t = \gamma_1 + \sum_{j=2}^{k} \gamma_j X_{jt} + \sum_{s=1}^{p} \rho_s e_{t-s}$ (so called auxiliary equation) using the $T - p$ observations: $p + 1$ through T;
3) Calculate LM statistics: $(T - p) \cdot R_e^2$ where R_e^2 is the R-squared from the auxiliary regression. $T - p$ is the actual number of observations. LM statistics is distributed as $\chi^2(p)$ when testing for p[th] order autocorrelation.
4) H_0: no autocorrelation
   H_1: not H_0
   Perform Chi-square test: reject the null hypothesis of zero autocorrelation in favour of the alternative if $(T - p) \cdot R_e^2 > \chi^2_{\alpha\% \, sign \, level}(p)$ and not reject otherwise. This test is valid only asymptotically.

Note that, alternatively, for $p = 1$, simple *t*-test on coefficient e_{t-1} can be used with asymptotically validity. For $p > 1$, *t*-test version becomes *F*-test on the lagged residuals comparing RSS_{UR} for the auxiliary regression with RSS_R for the same specification without the residual terms, i.e. $H_0: \rho_1 = \rho_2 = \cdots = \rho_p$

$H_1: at\ least\ one\ \rho \neq 0$

Moreover, this test is asymptotically valid for MA(p) autocorrelation.

3. Durbin-Watson test:

This test is used for the detection of AR(1) autocorrelation. The null hypothesis states that in the equation $u_i = \rho \cdot u_{i-1} + \varepsilon_i$ the true value of $\rho = 0$. The following Durbin-Watson statistics is calculated from estimated residuals:

$$DW = d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

It is shown that $0 < d < 4$.

With a strong positive residual autocorrelation, $DW \to 0$;

With a strong negative residual autocorrelation, $DW \to 4$;

When $\rho = 0$, $DW \to 2$.

In practice it is convenient to use the approximate formula of the Durbin-Watson statistics (for large samples):

$DW = \frac{\sum_t {e_t}^2 - 2\sum_t e_t e_{t-1} + \sum_t e_{t-1}^2}{\sum_t e_t^2} \approx 2 \cdot (1 - \frac{\sum_t e_t e_{t-1}}{\sum_t e_t^2})$, since $\sum_t e_t^2$ is approximately equal to $\sum_t e_{t-1}^2$.

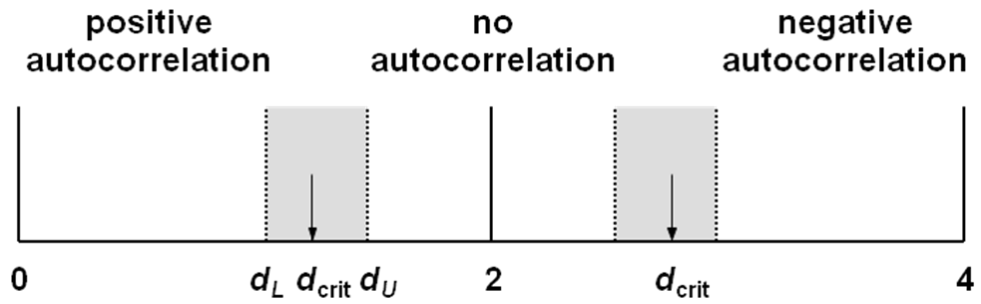
Further, as $\sum_t e_t^2 \approx \sqrt{\sum_t e_t^2} \cdot \sqrt{\sum_t e_{t-1}^2}$ and the sample correlation coefficient between e_t and e_{t-1} is given by $r_{t,t-1} = \frac{\sum_t e_t e_{t-1}}{\sqrt{\sum_t e_t^2 \cdot \sqrt{\sum_t e_{t-1}^2}}}$, the Durbin-Watson statistic can be approximated by the following expression:

$$DW \approx 2 \cdot (1 - \frac{\sum_t e_t e_{t-1}}{\sum_t e_t^2}) \approx 2 \cdot (1 - r_{t,t-1})$$

.

The distribution of the Durbin-Watson statistics depends not only on *k* (the number of explanatory variables) and *n* (the numbers of observations), but also on the sample values explanatory variables. Therefore the exact critical value of the statistic for any given significance level will be different for different samples. Durbin and Watson showed that the distribution of d is bounded by two limiting distributions.

The decision rule is described as follows:



From the table of the Durbin-Watson statistics we take upper and lower bounds of d, given the number of observations in the sample and the number of explanatory variables for the chosen significance level.

Test for positive autocorrelation:

$H_0: \rho = 0$

$H_1: \rho > 0.$

Reject H_0 if $d \leq d_L$. If $d \geq d_U$, we cannot reject H_0. If $d_L < d < d_U$, the test is inconclusive.

Test for negative autocorrelation:

$H_0: \rho = 0$

$H_1: \rho < 0$

use $4 - d$. This is done when d is greater than 2.

If $4 - d \leq d_L$, there is significant negative autocorrelation. If $4 - d \geq d_U$, there is evidence of no negative autocorrelation. The test is inconclusive for $d_L < 4 - d < d_U$.

Main features and remarks of the Durbin-Watson test:

1) It detects only first-order autocorrelation;
2) It cannot be applied if the model contains lagged dependent variable (the statistics is biased towards 2 for such case);
3) It cannot be applied if the model does not contain constant term;
4) There is uncertainty zone (test is inconclusive);
5) It is appropriate for finite samples and is provided directly by standard packages.

Extension: Durbin's test for detection of AR(1) autocorrelation in the case of the lagged dependent variable used as a regressor.

Durbin proposed ***h*-test:**

$$h = \hat{\rho}\sqrt{\frac{n}{1 - ns^2_{b_{Y(-1)}}}}$$

where

$\hat{\rho}$ is an estimate of ρ in the AR(1) process. Since the test is only valid for large samples, $d \to 2 - 2\hat{\rho}$. Hence, the estimator is then $\hat{\rho} = 1 - 0.5d$;

$s^2_{b_{Y(-1)}}$ is an estimate of the variance of the coefficient of the lagged dependent variable;

n is an actual number of observations.

This statistics is distributed as a normal variable with zero mean and unit variance (asymptotically). Hence, standard *Z*-test is performed. The problem is that $ns^2_{b_{Y(-1)}}$ can be greater than, or close to 1 (it usually happens when sample size is not very large). Then *h*-statistics cannot be computed.

IV. Remedial measures:

If the test for detection of autocorrelation results in the rejection of the null hypothesis, it is necessary to do the following:

1) Try to improve the model specification
2) If the final specification is considered correct but the null hypothesis is still rejected, the disturbance term is subject to autocorrelation. It can be eliminated by means of an **autoregressive transformation**.

Eliminating AR(1) autocorrelation: one explanatory variable

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

The regression is also valid for $t-1$: $Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$. Multiply this equation by ρ to get:

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{t-1} + \rho u_{t-1}$$

Subtract it from the original equation:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + u_t - \rho u_{t-1}$$

As $u_t - \rho u_{t-1} = \varepsilon_t$ for which all Gauss-Markov conditions are satisfied, the problem of the autocorrelated disturbance term is now eliminated. The model becomes:

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \beta_2 \rho X_{t-1} + \varepsilon_t$$

It is ADL(1,1) model of the form $Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \varepsilon_t$ with a non-linear restriction: the coefficient of X_{t-1} is minus the product of the coefficients of X_t and Y_t, i.e. $\lambda_4 = -\lambda_2 \cdot \lambda_3$. Therefore, a non-linear estimation technique is used. For example, in EViews it is done by the following command: **NLS Y=C(1)*(1-C(2))+C(2)*Y(-1)+C(3)*X-C(2)*C(3)*X(-1).**

Eliminating AR(1) autocorrelation: two explanatory variables

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

Using the same procedure as before: lagging the original specification, multiplying it by ρ and then subtracting the derived equation from the initial regression, we get:

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{2t-1} + \beta_3 \rho X_{3t-1} + \rho u_{t-1}$$
$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + u_t - \rho u_{t-1}$$

So, the model is transformed into ADL(1,1) of the form $Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_{2t} + \lambda_4 X_{2t-1} + \lambda_5 X_{3t} + \lambda_6 X_{3t-1} + \varepsilon_t$ with 2 non-linear restrictions:

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \beta_3 X_{3t} - \beta_3 \rho X_{3t-1} + \varepsilon_t$$

Restrictions: $\begin{cases} \lambda_4 = -\lambda_2 \cdot \lambda_3 \\ \lambda_6 = -\lambda_2 \cdot \lambda_5 \end{cases}$

There are several ways to estimate this regression in EViews:

1) Y=C(1)*(1-C(2))+C(2)*Y(-1)+C(3)*X2-C(2)*C(3)*X2(-1)+C(4)*X3-C(2)*C(4)*X3(-1);
2) Add AR(1) to the list of explanatory variables in the initial regression. Note that coefficients of lagged explanatory variables not presented in the original specification (X_{2t-1} andX_{3t-1} here) are not reported. For higher orders of autocorrelation add AR(1), AR(2), …, up to AR(p), where p is expected order of autocorrelation to deal with.

Cochrane-Orcutt iterative process:

In early days of computing, non-linear techniques was not so simple and widely applicable. Therefore, some other methods were used. The Cochrane–Orcutt iterative procedure requires the transformation of the regression model to a form in which the OLS procedure is applicable.

Consider a simple linear regression model $Y_t = \beta_1 + \beta_2 X_t + u_t$ where the disturbance term follows AR(1) process $u_t = \rho u_{t-1} + \varepsilon_t$.

This transformation is done in several steps:

1) Estimate the original equation by OLS and compute its residuals e_t;
2) Estimate the first-order serial correlation coefficient (ρ) by regressing e_t on e_{t-1}
3) Transform the variables as follows:

$$\tilde{Y}_t = Y_t - \rho Y_{t-1}$$
$$\tilde{X}_t = X_t - \rho X_{t-1}$$

$\beta_1' = \beta_1(1-\rho)$;

4) Regress $\tilde{Y}_t$ on $\tilde{X}_t$ to obtain revised estimates b_1 and b_2;
5) Plug estimated b_1 and b_2 into the original regression, and then obtain a new set of estimates

for residuals. Go back and repeat step 2).

This iterative procedure can be stopped when the estimates of ρ from two successive iterations differ by no more than some predetermined values, such as 0.001. The final value of ρ is then used to get estimates for transformed regression.

Making the autoregressive transformation, we lose the first observation. To avoid this, the **Price–Winsten correction** is applied: on the 4-th step of the Cochrane-Orcutt procedure the first observation multiplied by $\sqrt{1-\rho^2}$ is added to the transformed observations [2; T].

The introduction of this multiplier is explained by the need to deal with the heteroscedasticity problem, which arises due to the following effect:

$u_t = \rho \cdot u_{t-1} + \varepsilon_t \Rightarrow \sigma_u^2 = var(u_t) = var(\rho \cdot u_{t-1} + \varepsilon_t) = \rho^2\, var(u_{t-1}) + 2\, cov(u_{t-1}, \varepsilon_t) + var(\varepsilon_t) = \rho^2\sigma_u^2 + \sigma_\varepsilon^2$

$\Rightarrow \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$.

Therefore, if we added the first observation without multiplying it by $\sqrt{1-\rho^2}$, the variance of the disturbance term would be equal to σ_u^2, while in the other observations, which have been subject to the Cochrane-Orcutt transformation, this variance is equal to σ_ε^2.

Common factor test:

We have shown that for a linear regression model with the disturbance term subject to AR(1) process, the model can be transformed into ADL(1,1) with non-linear restriction imposed for estimation purposes, i.e.

Restricted model (transformed AR(1)): $Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_{2t} - \beta_2\rho X_{2t-1} + \beta_3 X_{3t} - \beta_3\rho X_{3t-1} + \varepsilon_t$

This model can be considered as a special case of a more general model involving the same variables, i.e.

Unrestricted ADL(1,1) model: $Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_{2t} + \lambda_4 X_{2t-1} + \lambda_5 X_{3t} + \lambda_6 X_{3t-1} + \varepsilon_t$

Restrictions embodied in the AR(1) process: $\begin{cases}\lambda_4 = -\lambda_2 \cdot \lambda_3 \\ \lambda_6 = -\lambda_2 \cdot \lambda_5\end{cases}$

The common factor test is used to differentiate between these 2 cases. It involves a comparison of RSS_R and RSS_U, the residual sums of squares in the restricted and unrestricted specifications.

H_0: Restricted model should be used: $\begin{cases}\lambda_4 = -\lambda_2 \cdot \lambda_3 \\ \lambda_6 = -\lambda_2 \cdot \lambda_5\end{cases}$ are valid;

H_1: Unrestricted ADL(1,1) model should be used.

Under the null hypothesis that the restrictions are valid, the test statistic has a χ^2 (chi- square) distribution with degrees of freedom equal to the number of restrictions (2 in this case).Note that it is a large-sample test.

Test statistics: $\chi^2 = n \log\frac{RSS_R}{RSS_U} \overset{asympt}{\underset{H_0}{\sim}} \chi^2(d.f. = number\ of\ restrictions)$, where

n is the actual number of observations in the regression (after adjusting endpoints);

log is the natural logarithm.

Dynamic model specification:

Initially, in order not to deal with a poorly specified model for which tests can appear to be satisfactory, even though it is misspecified (high risk of making Type II error), one should adopt a general-to-specific approach. It means that one should start with a model that is sufficiently general to avoid potential problems of under specification, and then see if it is possible to simplify it by testing restrictions on parameters. For example, in our case we should make the following steps for models with dynamic specification:

1) Estimate the model with all the lagged variables (unrestricted ADL(1,1):
$$Y_t = \lambda_0 + \lambda_1 Y_{t-1} + \lambda_2 X_{2t} + \lambda_3 X_{2t-1} + \lambda_4 X_{3t} + \lambda_5 X_{3t-1} + \varepsilon_t;$$
2) Test whether the lagged variables individually and as a group do not have significant explanatory power, i.e. $\lambda_1 = \lambda_3 = \lambda_5 = 0$. If it is not rejected, then the model can be simplified to the static case: $Y_t = \lambda_0 + \lambda_2 X_{2t} + \lambda_4 X_{3t} + \varepsilon_t$;
3) If the lagged variables do have significant explanatory power, we could perform a common factor test and see if we could simplify the model to an AR(1) specification, i.e. $\begin{cases} \lambda_3 = -\lambda_1 \cdot \lambda_2 \\ \lambda_5 = -\lambda_1 \cdot \lambda_4 \end{cases}$. If the test shows that unrestricted version of ADL(1,1) should be used, then perform described tests detecting autocorrelation;
4) Test for a model with a lagged dependent variable, i.e. $\lambda_3 = \lambda_5 = 0$.