Introductory Econometrics

Based on the textbook by Wooldridge: *Introductory Econometrics: A Modern Approach*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

January 14, 2013

Outline

OLS with serially correlated errors

Presume that assumptions on uncorrelated errors are violated, i.e. $\mathrm{corr}(u_t, u_s) \neq 0$ for some s, t:

- As long as assumptions **TS.1–TS.3** are fulfilled, OLS remains unbiased (Remember that **TS.3** fails to hold in dynamic regressions);
- If $\mathbf{TS}.\mathbf{1}'$–$\mathbf{TS}.\mathbf{3}'$ hold, OLS remains consistent (Remember that $\mathbf{TS}.\mathbf{3}'$ holds in a well-specified dynamic regression);
- Assumption **TS.5** is violated, typically OLS is inefficient, and indicated variance estimates are incorrect.

Thus, in *static* regressions, 'autocorrelation' (correlated errors) has consequences similar to heteroskedasticity. In a dynamic regression model, however, autocorrelation usually makes OLS inconsistent.

An important element: AR(1) processes

A simple model for autocorrelation is provided by the first-order autoregressive or **AR(1)** process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t,$$

with ε_t uncorrelated and homoskedastic ('white noise') and $|\phi_1| < 1$. This process is stationary, and its autocovariances $C(h) = \phi_1^h \sigma_y^2$ decrease fast with increasing h.

For $\phi_1 = 0$, the AR(1) becomes uncorrelated white noise. The excluded case $\phi_1 = 1$ is the non-stationary random walk.

The model is easily generalized to higher orders AR(p). Autocorrelation may of course follow more complex patterns.

The prototypical example: autoregression with AR errors

Consider the dynamic regression model

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t,$$

with $|\beta_1| < 1$, and $u_t = \rho u_{t-1} + \varepsilon_t$ with $|\rho| < 1$ and ε_t serially uncorrelated.

With these assumptions, u_t and y_t are stationary. It is easily shown that

$$\hat{\beta}_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_+)(y_{t-1} - \bar{y})}{\sum_{t=1}^{n-1}(y_t - \bar{y})^2} \to \frac{\mathrm{cov}(y_t, y_{t-1})}{\mathrm{var}(y_t^2)} = \frac{\beta_1 + \rho}{1 + \beta_1 \rho} \neq \beta_1$$

Remarks on dynamic regression with AR errors

- These models violate **TS**.**3**′ due to *dynamic misspecification*. They are simply the wrong models for the data. In the example, inserting the regressor y_{t-2} leads to correct dynamic specification, to valid **TS**.**3**′ and **TS.5**, and to consistent OLS;
- Dynamic misspecification often but not always implies inconsistency. For example, $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ with $u_t = \rho u_{t-2} + \varepsilon_t$ fulfils **TS**.**3**′, and OLS becomes consistent.

It makes sense to aim at correct dynamic specification. This aim requires good tests for autocorrelation.

The Durbin-Watson test

Durbin and Watson (1950,1951) analyzed the small-sample distribution of the statistic

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

for OLS residuals $\hat{u}_t$ under the null of **TS.1–TS.5** with normal u_t. This statistic is sensitive to deviations from the null $\rho = 0$ in the sense of

$$y_t = X_t' \beta + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1,$$

with **TS.1–TS.4** valid also under the alternative $\rho \neq 0$.

Some properties of the Durbin-Watson statistic *DW*

It generally holds that $0 \leq DW \leq 4$, such that the distribution always has a bounded support on $[0, 4]$. Because of

$$DW \approx 2(1 - \hat{\rho}),$$

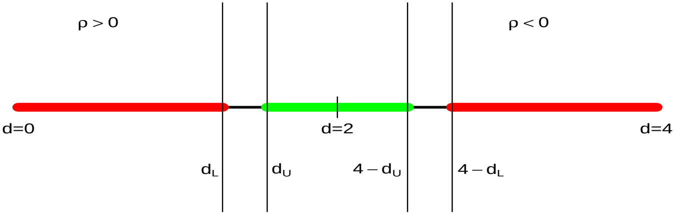
with $\hat{\rho}$ an estimate of ρ,

- $DW \in (0, 2)$ indicates positive autocorrelation $\rho > 0$;
- The ideal value $DW = 2$ indicates absence of autocorrelation;
- $DW \in (2, 4)$ indicates negative autocorrelation $\rho < 0$.

As $n \to \infty$, DW will converge to 2 under the null $\rho = 0$. With normality, one might expect tables of significance points depending on n and an *exact test*.

The distribution of DW under the null

Unfortunately, the DW test is not similar, its null distribution depends on more than just n, changes with the regressor matrix X. It is customary to consider tabulated upper and lower bounds d_U and d_L for the significance points.



Reject in the red intervals, do not reject in the green interval, no decision in the glacis in between.

More remarks on the Durbin-Watson test

- The *DW* null distribution differs for homogeneous regressions;
- The *DW* null distribution is entirely different in dynamic regressions, all tabulated d_U and d_L tables are invalid. The *DW* test should never be used nor even reported in dynamic regressions;
- The alternative model of the *DW* test, a stable first-order autoregression AR(1), is very special. If the test fails to reject, still there can be substantial autocorrelation and OLS may be inconsistent;
- Some contemporary textbooks discourage using the *DW* test. Still, the *DW* statistic is routinely reported in some econometric software.

The LM test for autocorrelation

The (approximate) LM test for autocorrelation is due to Breusch and Godfrey (1978). It proceeds in the following steps:

1. Estimate $y_t = X_t\beta + u_t$ by OLS and save residuals $\hat{u}_t$;
2. Estimate the *auxiliary regression* $(t = p+1, \ldots, n)$

$$\hat{u}_t = \gamma_1 x_{t,1} + \ldots + \gamma_k x_{t,k} + \delta_1 \hat{u}_{t-1} + \ldots + \delta_p \hat{u}_{t-p} + \mathrm{error}$$

3. Calculate $LM = (n-p)R^2$ from the auxiliary regression. Under the null of no autocorrelation, LM is distributed as $\chi^2(p)$ in large samples.

This (asymptotic) test works for static and dynamic regressions. The choice of p is done by the user, $p = 1$ corresponds vaguely to the DW test (which is invalid in dynamic regressions).

Variants of the autocorrelation tests

WOOLDRIDGE considers the following variants of the DW and Breusch-Godfrey LM tests:

- Instead of the *DW* statistic, use the empirical correlation of $\hat{u}_t$ and $\hat{u}_{t-1}$;
- Instead of $LM = (n-p)R^2$, use the F statistic on the lagged residuals;
- Omit the main regressors from the auxiliary regression.

These variants can be used, but they offer no advantage over the commonly used tests. Other variants (Durbin's h test, Q test due to LJUNG AND BOX) have less power than the LM test.

What if autocorrelation is detected?

Several responses are conceivable:

- Search for a correct dynamic specification, using lags of the dependent and of the explanatory variables as potential regressors (recommended);
- In analogy to weighted least squares, transform the original model such that the Gauss-Markov assumptions are fulfilled (feasible GLS): severe problems in dynamic regression models, may not lead to full efficiency;
- Stick to the original specification and to OLS, modify standard errors: does not work in dynamic regression, usually not satisfactory.

GLS for AR(1) errors

Presume the following model is correct:

$$y_t = X_t'\beta + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

with ε_t white noise and $|\rho| < 1$. Then, a transformed problem can be shown to fulfil Gauss-Markov conditions:

$$\begin{aligned}
\tilde{y}_t &= y_t - \rho y_{t-1}, t = 2, \ldots, n, \\
\tilde{x}_{t,j} &= x_{t,j} - \rho x_{t-1,j}, t = 2, \ldots, n, j = 1, \ldots, k, \\
\tilde{y}_1 &= \sqrt{1-\rho^2} y_1, \quad \tilde{x}_{1,j} = \sqrt{1-\rho^2} x_{1,j}, \\
\tilde{y}_t &= \tilde{X}_t \beta + \tilde{u}_t, t = 1, \ldots, n,
\end{aligned}$$

and OLS applied to this problem, i.e. GLS, is efficient.

The algebra of GLS

The transformation (due to Prais and Winsten) can be viewed as multiplication by a matrix $\tilde{y} = Ay$ and $\tilde{X} = AX$, with

$$A = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 \\ 0 & -\rho & 1 & 0 \\ 0 & 0 & -\rho & 1 \end{bmatrix},$$

such that

$$\mathrm{E}\tilde{u}\tilde{u}' = \sigma_\varepsilon^2 I_n,$$

and the GLS estimate $\tilde{\beta}$ becomes

$$\begin{aligned} \tilde{\beta} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \\ &= (X'A'AX)^{-1}X'A'Ay. \end{aligned}$$

Feasible GLS with AR(1) errors

The coefficients ρ are usually unknown. So, they must be estimated. An obvious suggestion is to run a regression on OLS residuals $\hat{u}$,

$$\hat{u}_t = \rho_0 + \rho \hat{u}_{t-1} + \text{error},$$

and to plug in the corresponding $\hat{\rho}$ for ρ in the transformations. This **fGLS** (feasible GLS) method can be shown to achieve the efficiency of GLS in large samples for static regressions and correctly specified autocorrelation. In dynamic regressions, GLS is not BLUE and not even unbiased, the procedure becomes unreliable.

Variants of fGLS with autocorrelated errors

- The method as described can be iterated: new residuals from improved estimates yield new ρ estimates etc. Not much is gained by iteration;
- Omitting the awkward first transformed observation and running fGLS on $n-1$ transformed observations is called the procedure of COCHRANE AND ORCUTT;
- The method can be generalized to higher-order AR models for errors or even other models, such as MA (moving average) or ARMA. Not really recommended.

Corrected standard errors: the basic idea

If OLS is used in the 'GLS model'

$$y_t = X_t'\beta + u_t, \quad \mathrm{E}uu' = \Omega,$$

OLS will be inefficient. The variance of $\hat{\beta}$ will not be $\sigma^2(X'X)^{-1}$ but rather

$$\mathrm{var}\hat{\beta} = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Even if this looks complex, why not plug in some estimate for Ω, evaluate this formula via computer, and use the diagonal of the resulting matrix for standard errors?

Parametric and nonparametric corrections

A parametric correction would for example estimate some ρ based on an AR(1) assumption, and fill Ω by $\omega_{i,j} = \hat{\sigma}_u^2 \hat{\rho}^{|i-j|}$.

A nonparametric correction would plug in autocovariance estimates, downweighted with larger distance from the diagonal (Bartlett or Newey-West for simple linear weighting).

Corrected standard errors with OLS is less popular for autocorrelation than for heteroskedasticity, although both follow from the same White-Eicker idea. The autocorrelation and the heteroskedasticity corrections can even be combined.