Chapter 2

Misspecification Analysis II: Serial Correlation

Artur Silva Lopes, March 2022

ISEG–ULisboa

1 Introductory notes

In Econometrics, the expression "serial correlation problem" usually has a negative sense: it is the problem of serial correlation (or autocorrelation) of the errors.

GLS and FGLS estimation will not be addressed: GLS because it is usually irrealistic and FGLS because it rarely is a good solution. Grayham Mizon: "*autocorrelation correction: Don't!*"

2 Introduction

In the model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t, \;\; t = 1, 2, \ldots, T$ with $\mathsf{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$, the asssumption that

$$\mathsf{Var}(\mathbf{u}|\mathbf{X}) = \mathsf{E}(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \sigma^2\mathbf{I},$$

may not hold, and instead

$$\mathsf{Var}(\mathbf{u}|\mathbf{X}) = \mathsf{E}(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \mathbf{\Omega} = \sigma^2\mathbf{V} \neq \sigma^2\mathbf{I},$$

because the errors are serially correlated:

$$\exists t \neq s : \mathsf{Cov}(u_t, u_s) = \mathsf{E}(u_t u_s) \neq 0.$$

Assuming that they are stationary:

$$\mathbf{\Omega} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{t-3} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{t-3} & \cdots & 1 \end{bmatrix},$$

where γ_0 denotes the error variance (assumed as common), i.e., σ^2. The second matrix is the $\mathbf{V}$ matrix, $\mathbf{V} = [v_{ts}] = \left[\frac{\gamma_{|t-s|}}{\gamma_0}\right] = [\rho_{|t-s|}]$, i.e., the matrix of the autocorrelation coefficients.

3 Sources of “residual autocorrelation”

“Symptoms of residual autocorrelation”: what does this means?

According to a modern perspective, the errors of a correctly specified model should not be autocorrelated. Everything that is systematic, that has a pattern, must be in the main part of the model, must not be left to the “noise” component.

Symptoms of autocorrelation are often the outcome of:

a) errors of functional form specification;

b) omitted regressors, and in particular an insufficient dynamic specification.

Autocorrelation must not be seen as a problem intrinsic to the errors. It often is only the result of an inadequate specification.

4 The case of strict exogeneity

By assumption, $\mathsf{E}(u_t|\mathbf{X}) = 0, \forall t \Rightarrow \mathsf{Cov}(x_{tj}, u_s) = 0, \forall t, s, j$.

Purpose: to test a certain economic hypothesis (ex.: rational expectations hypothesis; MEH). The interest does not rely on analysing dynamic effects, or in getting a precise description of the economy, etc. . There is no freedom to change the specification of the model. The equation is fixed. It is not changeable.

1. **Unbiasdness.** The OLS estimator is still unbiased, $\mathsf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ (because $\mathsf{E}(\mathbf{u}|\mathbf{X}) = \mathbf{0}$).

2. **Covariance matrix.** But the covariance matrix is no longer the usual

$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\begin{aligned}
\text{Var}(\widehat{\beta}|\mathbf{X}) &= \text{E}[(\widehat{\beta}-\boldsymbol{\beta})(\widehat{\beta}-\boldsymbol{\beta})'|\mathbf{X}] \\
&= \text{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

3. **Inference.** Since $\text{Var}(\widehat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ was the basis for inference, usual methods cease to be valid. Therefore, incorrect inferences are more frequent than it is assumed.

   Example: $y_t = \beta_1 + \beta_2 x_t + u_t$, with $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$, $\epsilon_t \sim iid(0, \sigma^2)$ and $\bar{x} = 0$. Then:

$$\text{Var}(\widehat{\beta}_2|X) = \frac{\sigma_u^2}{TSS_x} + 2\frac{\sigma_u^2}{TSS_x^2}\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}\rho^j x_t x_{t+j},$$

   where TSS_x is the variation of the regressor, $\sum x_t^2$. The first term is $\text{Var}(\widehat{\beta}_2|\mathbf{X})$ when $\rho = 0$ (there is no autocorrelation, $u_t \equiv \epsilon_t$). Since the estimator of Var$(\widehat{\beta}_2|\mathbf{X})$ "forgets" the second term and the autocorrelation is usually positive (both of u_t and of x_t), generally the variance is underestimated. This underestimation can be large: $\widehat{\beta}_2$ appears to be much more precise than it really is.

Therefore, when testing hypothesis about β_2, H_0 will be rejected (much) more frequently than it should when H_0 is true (much more than the $100\alpha\%$). The denominators of the t-statistics become deflated. Real size can be (much) larger than nominal size: there will be over-rejections of the true H_0 (or size distortions). Sometimes these are called spurious rejections.

Monte Carlo illustration. The DGP is

$$y_t = 2 + 1\, x_t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_{1t},\ |\rho| < 1, \quad \text{with}$$

$$x_t = \lambda x_{t-1} + \epsilon_{2t},\ \ |\lambda| < 1,$$

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim iid\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

A test of the true null hypothesis

$$H_0 : \beta_2 = 1, \quad vs. \quad H_1 : \beta_2 \neq 1,$$

is performed with $T = 50$. The MC results with $20,000$ replications are:

Percentual rejections (in %) for 5% nominal tests with $\lambda = 0.9$

ρ	0.0	0.2	0.4	0.6	0.8	1.0
% reje.	5.40	10.48	17.60	26.27	38.92	54.43

When there is no autocorrelation, estimated real size (5.4%) is very close to nominal, but with ρ growing rejection frequencies tend to grow very quickly, and when the errors have a unit root rejections exceed 50%.

The issue here is better defined than the one of heteroskedasticity: usual over-estimation of the OLS precision and over-rejection of true null hypotheses.

Note: spurious regressions are a particular case of this problem (more later).

4. **Consistency.** If

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right) = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right) = \mathbf{\Sigma}_{xx},$$

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{u}\right) = \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right) = \text{plim}(\bar{\mathbf{g}}) = \text{E}(\mathbf{g}_t) = \mathbf{0},$$

and if

$$\text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{V}\mathbf{X}\right) = \mathbf{Q}^*,$$

is a positive definite matrix, it can be shown that $\text{plim}(\widehat{\beta}) = \boldsymbol{\beta}$. What do the conditions require? Stationarity and ergodicity both of regressors and errors.

5. **Asymptotic normality.** Problem: it is not possible to employ a CLT for non-autocorrelated variables because

$$\frac{1}{\sqrt{T}}\mathbf{X}'\mathbf{u} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t u_t$$

is a sum of autocorrelated variables.

There is however a CLT that allows autocorrelation but requires stationary and ergodicity of regressors and errors: Gordin's CLT (see Hayashi, 2000).

Unless in extreme cases:

$$\widehat{\beta}|\mathbf{X} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}].$$

6. **Efficiency.** Gauss-Markov theorem is no longer valid: OLS is no longer BLUE.

   But GLS (BLUE) demands that the $\mathbf{V}$ matrix is known and EGLS or FGLS may not be more efficient than OLS.

7. **Robust inference. HAC.** Provided OLS is consistent, and provided its covariance matrix is consistently estimated, asymptotically valid inference is feasible.

With autocorrelation, estimation is more complex than with heteroskedasticity. The Newey and West estimator (1987) ou HAC (heteroskedastic and autocorrelation consistent) estimator is:

$$\widehat{\mathbf{S}}^* = \widehat{\mathbf{S}}_0 + \frac{1}{T}\sum_{j=1}^{l}\sum_{t=j+1}^{T}\omega_j e_t e_{t-j}(\mathbf{x}_t\mathbf{x}'_{t-j} + \mathbf{x}_{t-j}\mathbf{x}'_t)$$

with weights ω_j usually given by $\omega_j = 1 - \frac{j}{l+1}$, $j = 1, \ldots, l$ (Bartlett weights), to ensure that the matrix is, at least, semi-definite positive.

l is the lag truncation parameter, it is the maximum order considered for the autocorrelations of the errors; the order $l + 1$ is already considered as negligible. l is also called "bandwidth". If the errors follow a MA process this order is low. But the autocorrelation coefficients of an AR do not become zero. Loose rule: $l = 0.75\,T^{1/3}$. However, usually one should try several values for l, aiming to get robustness.

The weights ω_j are called the "kernel". In this case it is Bartlett's kernel. When using TSP (recall cha. 1): `GMM (HET/NOHET,NMA=L,INST=(regressors)) EQ;`

Monte Carlo example. The DGP is the same

$$y_t = 2 + 1\,x_t + u_t, \quad u_t = \rho u_{t-1} + \epsilon_{1t},\ |\rho| < 1, \quad \text{with}$$

$$x_t = \lambda x_{t-1} + \epsilon_{2t},\ \ |\lambda| < 1,$$

but $\lambda = 0.7$. Test of the true null $H_0 : \beta_1 = 1$ vs. $H_1 : \beta_2 \neq 1$ with:

- t_{OLS};
- t_{NW-5}, with a se given by the Newey-West matrix with $l = 5$;
- t_{NW-10} idem but with $l = 10$.

Sample size is $T = 100$. Objective: how good is the quality of HAC inference made with the NW estimator in a reasonably large sample?

Estimates of real size (in %) with tests with nominal 5% and $\lambda = 0.7$

ρ	0.0	0.2	0.4	0.6	0.8	1.0
t_{OLS}	4.87	8.76	14.39	21.76	30.67	38.84
t_{NW-5}	5.99	6.90	8.27	9.98	12.68	15.53
t_{NW-10}	6.25	6.90	7.75	8.90	10.86	12.64

a) the over-rejection problem becomes rather serious for OLS when $\rho > 0.2$;

b) the two statistcs based on NW estimation reduce the problem but they never remove it completely;

c) with $\rho = 0.8$ and $L = 10$ even the NW method is affected by significant size distortions;

d) it is very likely that samples with $T > 100$ are required for the NW method to work well;

e) if ρ is small there is no advantage making $l = 10$. But it suffices that $\rho = 0.4$ only for the test with $l = 10$ to become better than with $l = 5$.

Actually, the procedure needs large samples. Gordin's theorem:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{g}_t \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}),$$

where $\mathbf{S} = \sum_{j=-\infty}^{\infty} \mathbf{\Gamma}_j$, with

$$\mathbf{\Gamma}_j = \mathsf{E}[(\mathbf{g}_t - \boldsymbol{\mu})(\mathbf{g}_{t-j} - \boldsymbol{\mu})'] = \mathsf{E}(\mathbf{g}_t \mathbf{g}'_{t-j}),$$

is the long run covariance matrix of the process g_t. Since

$$\mathbf{S} = \lim_{T\to\infty} \mathsf{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{g}_t \right) = \lim_{T\to\infty} \mathsf{Var}(\sqrt{T}\bar{\mathbf{g}}),$$

to better understand the problems consider the case of a scalar $\mathbf{g}_t$:

$$
\begin{aligned}
\mathrm{Var}(\tfrac{1}{\sqrt{T}}\textstyle\sum_{t=1}^{T} g_t) &= \tfrac{1}{T}\mathrm{Var}(g_1 + g_2 + \ldots + g_T) \\
&= \tfrac{1}{T}[\mathrm{Cov}(g_1, g_1 + \ldots + g_T) + \mathrm{Cov}(g_2, g_1 + g_2 + \ldots g_T) + \ldots + \\
&\quad \mathrm{Cov}(g_T, g_1 + g_2 + \ldots + g_T)] \\
&= \tfrac{1}{T}[(\gamma_0 + \gamma_1 + \ldots + \gamma_{T-1}) + (\gamma_1 + \gamma_0 + \gamma_1 + \ldots + \gamma_{T-2}) + \ldots \\
&\quad (\gamma_{T-1} + \gamma_{T-2} + \ldots + \gamma_1 + \gamma_0)] \\
&= \tfrac{1}{T}[T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \ldots + 2\gamma_{T-1}] \\
&= \gamma_0 + 2\textstyle\sum_{j=1}^{T-1}(1 - \tfrac{j}{T})\gamma_j.
\end{aligned}
$$

$$
\lim_{T\to\infty} \mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t) = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j = \sum_{j=-\infty}^{\infty}\gamma_j.
$$

Now, how to estimate an infinite number of autocovariances? If we estimate only a few $\Rightarrow$ the estimator is inconsistent. If we estimate a large number $\Rightarrow$ the estimator is inefficient. A balance is achieved making l depend on T, but this can be insufficient.

Note: however, there are good parametric estimators (no need to follow the herd).

Empirical example. This example is not typical.

```
                                   Equation   1
                                   ============
                    Method of estimation = Ordinary Least Squares
Dependent variable: DLC
Number of observations:  30
  Std. error of regression = .018050
                 R-squared = .652677
             Durbin-Watson = 1.73048 [.079,.454]
Breusch/Godfrey LM: AR/MA1 = .508995 [.476]
Breusch/Godfrey LM: AR/MA3 = 2.75500 [.431]
```

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	.018740	.460240E-02	4.07185	[.000]
DLR	.330562	.100103	3.30221	[.003]
DLS	.375609	.085867	4.37429	[.000]
DINF	.294196E-02	.853164E-03	3.44830	[.002]

GENERALIZED METHOD OF MOMENTS
==================================
WITH STARTING VALUES VIA:
NONLINEAR TWO STAGE LEAST SQUARES

EQUATIONS: EQ

INSTRUMENTS: C DLR DLS DINF

CONVERGENCE ACHIEVED AFTER 1 ITERATIONS

Number of observations = 30 E'PZ*E = .122894E-31

Parameter	Estimate	Standard Error	t-statistic	P-value
BEQ0	.018740	.447534E-02	4.18745	[.000]
BEQ1	.330562	.096896	3.41151	[.001]
BEQ2	.375609	.084745	4.43221	[.000]
BEQ3	.294196E-02	.793705E-03	3.70662	[.000]

Standard Errors computed from quadratic form of analytic first derivative
(Gauss)
(also robust to autocorrelation: NMA= 4, Kernel=Bartlett)

Notes:
a) there is no support to use HAC estimation; this is just to illustrate the estimation of the covariance matrix with the Newey-West method;
b) the bandwidth parameter was set equal to 4, which is a small but (more than) adequate value in this case, justified both by the small sample size and the absence of autocorrelation symptoms;
c) the `NOHET` option was chosen, i.e., robustness is searched only with respect to autocorrelation;
d) the estimated parameters are denoted with `BEQ0` to `BEQ3` and they are obviously the same as OLS; the corrected standard errors differ so slightly from the original that the robust t-statistics are also practically the same.

Often the evidence for positive first order autocorrelation is strong and it advisable to use one (or several) large value(s) for l. Often, the robust standard errors are much larger than the (OLS) originals, and a finding for significance is reverted, the robust t-statistic insufficiently large (in absolute value) to allow rejecting the null hypothesis.

5 The case of dynamic models

Objective: to provide a good description of the intertemporal relationships between variables. I assume that one of the regressors is y_{t-1}. Now we are free to change the specification of the model.

A new case of endogeneity: certain forms of serial correlation of the errors combined with the presence of at least one lagged dependent variable as regressor:

$$\mathsf{E}(\mathbf{g}_t) = \mathsf{E}(\mathbf{x}_t u_t) \neq 0 \Rightarrow \text{plim}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right) = \text{plim}\left(\frac{1}{T}\mathbf{X}'\mathbf{u}\right) \neq 0 \Rightarrow \text{plim}\widehat{\boldsymbol{\beta}} \neq \boldsymbol{\beta},$$

since the orthogonality condition fails, the OLS estimator of the regression coefficients (besides biased) is not (even) consistent.

Simple example. In the AR(1) stationary model

$$y_t = \beta y_{t-1} + u_t,\ |\beta| < 1, \text{ with } u_t \equiv \epsilon_t \sim iid(0, \sigma^2),$$

the regressor y_{t-1} is pre-determined and the OLS estimator of β is biased but consistent. Suppose now that

$$u_t = \rho u_{t-1} + \epsilon_t, \;\; |\rho| < 1.$$

The status of the regressor y_{t-1} changes:

$$\begin{aligned} \text{Cov}(y_{t-1}, u_t) &= \text{Cov}(\beta y_{t-2} + u_{t-1}, \rho u_{t-1} + \epsilon_t) \\ &= \beta\rho \, \text{Cov}(y_{t-2}, u_{t-1}) + \rho\sigma_u^2 \\ &= \beta\rho \, \text{Cov}(y_{t-1}, u_t) + \rho\sigma_u^2, \end{aligned}$$

Therefore,

$$\text{Cov}(y_{t-1}, u_t) = \frac{\rho\sigma_u^2}{1 - \beta\rho} \neq 0 \text{ (if } \;\rho \neq 0).$$

Hence,

$$\text{E}(y_{t-1}u_t) \neq 0 \Rightarrow \text{plim}(\widehat{\beta}) \neq \beta,$$

that is, the OLS estimator becomes inconsistent because the regressor turned into endogenous.

In summary: frequently, the status of lagged dependent variables depends upon the serial correlation properties of the errors. The solution cannot be "OLS + HAC". Why?

Objective: to find a dynamically complete model. The model $y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$ (where $\mathbf{x}_t'$ may contain lagged variables) is **dynamically complete** if

$$\mathsf{E}(y_t|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots) = \mathsf{E}(y_t|\mathbf{x}_t),$$

that is, if additional historical (past) information, not yet included in the model, is irrelevant; given the information that the model already contains (in $\mathbf{x}_t$), it does not add anything useful; it doesn't help to explain the behaviour of y_t besides the explanation that is already provided by $\mathbf{x}_t$.

It can be shown that if a model is dynamically complete, then its errors satisfy

$$\mathsf{E}(u_t|\mathbf{x}_t, u_{t-1}, \mathbf{x}_{t-1}, u_{t-2}, \ldots) = 0, \forall t,$$

which is the sufficient condition for $\mathbf{g}_t$ ($\mathbf{g}_t = \mathbf{x}_t u_t$) to be a m. d. s. (assumption H5' of the model with pre-determined regressors). Therefore, if a model is dynamically complete its errors are not autocorrelated since

$$\mathsf{E}(u_t|u_{t-1}, u_{t-2}, \ldots) = 0, \forall t.$$

Therefore, autocorrelation symptoms can be interpreted as indirect evidence that the model is not dynamically complete.

COMFAC. "LSE school": we achieve valid inferences only by eliminating specification errors, not by "correcting" the estimation method with an AR(1) (FGLS) for the errors. Mizon (1995): "autocorrelation correctors: Don't!". Otherwise, we may end up with an inadequate model and an inconsistent estimator.

"The story": someone starts with a static model and assumes that $u_t \sim iid$:

$$\mathcal{M}_0 : \; y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t, \;\; u_t \sim iid(0, \sigma_u^2),$$

where $\mathbf{x}_t$ contains only contemporaneous information. It is very likely that the model exhibits symptoms of serial correlation, that the practitioner tries to correct assuming an AR(1) for the errors, $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$, and estimates the model:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \rho(y_{t-1} - \mathbf{x}_{t-1}' \boldsymbol{\beta}) + \epsilon_t,$$

that is,

$$\mathcal{M}_1 : y_t - \rho y_{t-1} = (\mathbf{x}_t' - \rho \mathbf{x}_{t-1}') \boldsymbol{\beta} + \epsilon_t.$$

Denoting with γ the vector of coefficients of $\mathbf{x}_{t-1}$ in the unrestricted model, the model $\mathcal{M}_1$ imposes the non-linear restrictions

$$\boldsymbol{\gamma} = -\rho\,\boldsymbol{\beta}.$$

That is, $\mathcal{M}_1$ is the restricted version of the more general model

$$\mathcal{M}_2 : y_t = \mathbf{x}_t'\boldsymbol{\beta} + \rho y_{t-1} + \mathbf{x}_{t-1}'\gamma + \varepsilon_t,\ \varepsilon_t \sim iiid(0, \sigma^2),$$

where $\mathcal{M}_1$ is nested: $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{M}_2$.

The initial rejection of the no autocorrelation hypothesis may be due to the fact that $\mathcal{M}_2$ and not $\mathcal{M}_1$ is the DGP. Therefore, rather than "correcting" autocorrelation estimating $\mathcal{M}_1$, which implies untested restrictions and hence an estimator that can be inconsistent, it is better to first estimate $\mathcal{M}_2$ and test the restrictions of $\mathcal{M}_1$.

This test is the common factors test (COMFAC):

$$\mathcal{M}_1 : (1 - \rho L)y_t = (1 - \rho L)\mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t,$$

$$
\begin{aligned}
\mathcal{M}_2 : (1-\rho L) y_t &= \mathbf{x}_t' \boldsymbol{\beta} + L \mathbf{x}_t' \boldsymbol{\gamma} + \varepsilon_t \\
&= \mathbf{x}_t' (\boldsymbol{\beta} + \boldsymbol{\gamma} L) + \varepsilon_t,
\end{aligned}
$$

L denoting the lag operator $(L^p z_t = z_{t-p})$. In $\mathcal{M}_1$ but not in $\mathcal{M}_2$ the common factor $(1 - \rho L)$ is present. $\mathcal{M}_1$ imposes a restriction of one common factor in the autoregressive polynomials of y_t and $\mathbf{x}_t$.

One must begin with $\mathcal{M}_2$, the most general model. Afterwards the restrictions are tested and we must estimate $\mathcal{M}_1$ only in the case that data do not reject them .

GTS (*general-to-specific*): begin with a general model and test restrictions to simplify it, not the other way around. To act in reverse may leave us with an inadequate model and an inconsistent estimator.

How to test these restrictions: using the "delta method" for non-linear restrictions.

For the AR(1) case TSP provides automatically a statistic when the model does not contain y_{t-1}: it is the Wald test, called "`Wald nonlin. AR1 vs. lags`".

Autocorrelation symptons + significant test statistic $\Rightarrow$ respecify the model introducing lags of all variables.

6 Tests for serial correlation

Menu:
a) t-test for the case of strictly exogenous regressors (review),
b) the Durbin-Watson test (drawbacks),
c) the Ljung-Box tests (review) and
d) the Breusch-Godfrey (and h-alt) tests.

Common basic idea: the unobserved errors are replaced by the residuals.

6.1 t-test for AR(1) with strict exogeneity

Assumed model: $u_t = \rho u_{t-1} + \epsilon_t, |\rho| < 1$ (AR(1)). Test of

$$H_0 : \rho = 0 \quad vs. \quad H_1 : \rho > 0 \text{ (or } H_1 : \rho \neq 0).$$

In the auxiliar regression $e_t = \rho e_{t-1} + v_t, \qquad t_\rho = \widehat{\rho}/se(\widehat{\rho}) \overset{a}{\sim} \mathcal{N}(0,1)$ under H_0.

Critical region: one-sided right or two-sided.

The test has also power for other forms of first order serial correlation.

6.2 Durbin-Watson test

The DW or d statistic was the only available to test serial correlation in the regression model for many years.

Shortcoming 1: model $u_t = \rho u_{t-1} + \epsilon_t$, $|\rho| < 1$ (AR(1)).

Shortcoming 2: it requires strict exogeneity. In particular, if y_{t-1} is a regressor the test becomes biased towards non-rejection, that is, powerless.

Shortcoming 3: u_t is assumed as normally distributed and the model must contain an intercept.

Statistic:

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

Shortcoming 4: the statistic does not have a small sample distribution independent from the $\mathbf{X}$ matrix. Even when $\text{Var}(u|\mathbf{X}) = \sigma^2 I$,

$$\text{Var}(e|\mathbf{X}) = \sigma^2 M.$$

Therefore:

a) even when $u_t \sim iid$ the residuals, e_t, will be serially correlated $(\mathbf{M} \neq \mathbf{I})$;

b) Since $\mathbf{M} = f(\mathbf{X})$, it is not possible to obtain a distribution independent from the particular $\mathbf{X}$.

Shortcoming 5: the solution for the previous problem is obtained at cost of an inconclusive region for the test.

Informally: values below 1 or 1.1 or 1.2 indicate (loosely) evidence for positive first order autocorrelation.

6.3 Ljung-Box tests

They are imported from time series literature. Let $\rho_1, \rho_2, \ldots, \rho_j$, denote the autocorrelation coefficients of the errors. The test is

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_p = 0, \quad vs. \quad H_1 : \exists \rho_j \neq 0, j = 1, \ldots, p.$$

Objective: to analyse the autocorrelations of all orders until order p (*portmanteau* test).

Denoting with r_j the empirical coefficients of the residuals, the Ljung-Box statistic

$$Q^* = T(T+2)\sum_{j=1}^{p} \frac{r_j^2}{T-j} \overset{a}{\sim} \chi^2_{(p)} \text{ under } H_0.$$

Choice of p:

a) if p is too low, the test is unable to detect autocorrelation at a high order;

b) if p is too high, tests may loose power because significant low order autocorrelations may become dissolved.

Order $p = 1$ is mandatory: (usually) it is the most recent past the most important in explaining the present.

Usually one more test is performed at the cost of raising the overall size above 5%. p depends on the data frequency:
a) annual: $p = 2$ or 3;
b) quarterly: $p = 4$ (is there unexplained seasonality?);
c) monthly: $p = 12$ (idem).

Godfrey (1998): in regression models the properties of these tests are not well known; hence it is better to use Breusch-Godfrey tests.

6.4 Breusch-Godfrey tests

These tests generalize Durbin's h-alt test for AR(1) errors when y_{t-1} is a regressor: $H_0 : \rho = 0$ $vs.$ $H_1 : \rho \neq 0$ in $u_t = \rho u_{t-1} + \epsilon_t$.

In the auxiliar regression $e_t = x_t'\alpha + \phi e_{t-1} + v_t, \quad h - alt = t_\phi \overset{a}{\sim} \mathcal{N}(0,1)$ under H_0.

If one or more regressors are not strictly exogenous, e_t is not a good estimate of u_t. The inclusion of $\mathbf{x}_t'$ aims to "clean" the residuals.

H_0 and H_1 can be the same as those of LB's test. But also

$H_0 : u_t$ **not autocorrel.** $vs. \quad H_1 : u_t \sim ARMA(m,q), \forall m,q : m+q=p.$

The procedure is similar to the h-alt test:

i) auxiliar regression:

$$e_t = x_t'\alpha + \phi_1 e_{t-1} + \ldots + \phi_p e_{t-p} + error_t,$$

ii) test $H_0 : \phi_1 = \phi_2 = \ldots = \phi_p = 0, \quad vs. \quad H_1 : \exists \phi_j \neq 0, j = 1, \ldots, p,$ with the usual LM statistic

$$BG(p) = TR_e^2 \overset{a}{\sim} \chi^2_{(p)} \text{ under } H_0,$$

where R_e^2 is the R^2 of the auxiliar regression.

7 Empirical example

The example illustrates mainly the presence of symptoms of autocorrelation resulting from a model with a poor dynamic specification.

Portuguese economy, quarterly data from 1977:1 to 1995:4, a money demand (static) equation

$$LMR_t = \beta_1 + \delta_1 Q_{t1} + \delta_2 Q_{t2} + \delta_2 Q_{t3} + \beta_2 t + \beta_2 LGDP_t + \beta_4 r_t + u_t,$$

LMR is logged money (M1), $Q_{tj}, j = 1, \ldots, 4$ are quarterly dummies, $LGDP$ is logged GDP and r is an interest rate on term deposits. It contains a deterministic trend term (t) but it is purely static.

Equation 1
=============

R-squared = .936480
LM het. test = .486913 [.485]
Durbin-Watson = .846703 ** [.000,.000]
Breusch/Godfrey LM: AR/MA1 = 33.5133 ** [.000]
Breusch/Godfrey LM: AR/MA4 = 34.6787 ** [.000]
Wald nonlin. AR1 vs. lags = 16.1423 ** [.000]

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	1.82332	1.08542	1.67983	[.098]
Q1	-.072107	.014877	-4.84672	[.000]
Q2	-.068193	.014631	-4.66077	[.000]
Q3	-.049859	.014505	-3.43744	[.001]
T	-.649085E-02	.103312E-02	-6.28277	[.000]
LGDP	1.02708	.146509	7.01035	[.000]
R	-2.46789	.112671	-21.9036	[.000]

The model fails completely in what concerns the absence of error autocorrelation: $DW < 1$, and BG(1) and BG(4) are highly significant, with p-values equal to 0.000. [The critical regions are $RC_{BG(1)} = \{BG(1) : BG(1) > 3.841\}$, $RC_{BG(4)} = \{BG(4) : BG(4) > 9.488\}$].

Note: the presence of the seasonal dummies, Q_{tj}, $j = 1, 2, 3$, does not ensure that seasonality is totally captured.

`Wald nonlin. AR1 vs. lags` rejects that the first order serial correlation problem could be solved with an estimation "correction" (FGLS).

Slight modification (in this case): introduction of the lagged dependent variable as regressor (LMR_{t-1}). Important: this is not a general procedure (one must start from a more general model). Purpose: to illustrate that the problem appears to be only one of a poor dynamic specification.

Both the h-alt (BG(1)) and the BG(4) statistics become clearly insignificant. We are not sure about the source of the problems. But empirical evidence clearly supports the argument of the insufficient dynamic specification.

Equation 2

============

R-squared = .976019

LM het. test = .251423 [.616]

Durbin-Watson = 1.95318 [.132,.760]

Durbin's h alt. = -.054748 [.956]

Breusch/Godfrey LM: AR/MA1 = .299737E-02 [.956]

Breusch/Godfrey LM: AR/MA4 = 1.45031 [.835]

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	1.16358	.681946	1.70626	[.093]
Q1	-.119098	.010311	-11.5504	[.000]
Q2	-.058601	.914345E-02	-6.40912	[.000]
Q3	-.038966	.907600E-02	-4.29334	[.000]
T	-.177309E-02	.782624E-03	-2.26557	[.027]
LMR(-1)	.623910	.059107	10.5556	[.000]
LPIB	.325642	.113377	2.87221	[.005]
R	-.954095	.160583	-5.94144	[.000]