Estimation and Inference in Linear Models

Jesús Gonzalo

UC3M

October 30, 2017

References

- "Asymptotics for Linear Processes" (Phillips and Solo, 1992)
- "Testing the Autocorrelation Structure....." (Cumby and Huizinga, 1992)
- Brockwell, P.J. and R.A. Davis, Time Series: Theory and Methods. New York. SpringerVerlag, second edition, second printing 2009
- Hayashi, F., Econometrics. Princeton University Press, 2000.
- In my web page there are several Eviews applets related to this chapter.

Content

- Some asymptotic results
- Estimation and Inference on
  - μ
  - γ_k and ρ_k
  - the parameter of an AR(1) and of a MA(1)

Some Asymptotic Results

B-N decomposition

Beveridge Nelson (B-N) decomposition

$$X_t = C(L)\varepsilon_t = \left[C(1) + (1-L)\widetilde{C}(L)\right]\varepsilon_t$$

What is $\widetilde{C}(L)$?

$$\begin{aligned}
\sum_{j=0}^{\infty} c_j L^j &= \left(\sum_{j=0}^{\infty} c_j - \sum_{j=1}^{\infty} c_j\right) + \left(\sum_{j=1}^{\infty} c_j - \sum_{j=2}^{\infty} c_j\right) L + \left(\sum_{j=2}^{\infty} c_j - \sum_{j=3}^{\infty} c_j\right) L^2 + \dots \\
&= \sum_{j=0}^{\infty} c_j - \left(\sum_{j=1}^{\infty} c_j\right)(1-L) - \left(\sum_{j=2}^{\infty} c_j\right) L(1-L) - \left(\sum_{j=3}^{\infty} c_j\right) L^2(1-L) + \dots \\
&= \sum_{j=0}^{\infty} c_j + (1-L)\sum_{j=0}^{\infty} \tilde{c}_j L^j
\end{aligned}$$

with $\tilde{c}_j = -\sum_{s=j+1}^{\infty} c_s$.

Some Asymptotic Results

B-N decomposition

Lemma

Let $\widetilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$ and $\tilde{c}_j = \sum_{s=j+1}^{\infty} c_s$. Then,

$$\text{(a)} \quad \sum_{j=0}^{\infty} j^{\frac{1}{2}} |c_j| < \infty \quad \Longrightarrow \quad \sum_{j=0}^{\infty} \tilde{c}_j^2 < \infty \quad \left(\Longleftarrow \sum_{j=0}^{\infty} j^2 c_j^2 < \infty \right),$$

$$\Uparrow$$

$$\text{(b)} \quad \sum_{j=0}^{\infty} j |c_j| < \infty \quad \Longrightarrow \quad \sum_{j=0}^{\infty} |\tilde{c}_j| < \infty.$$

Some Asymptotic Results

Law of Large Numbers

Theorem (SLLN for m.d.s.)

Let $\{\varepsilon_t\}$ be a m.d.s. with

$$\mathbb{V}[\varepsilon_t] = \sigma_t^2 < \infty, \qquad \sum_{t=1}^{\infty} \frac{\sigma_t^2}{t^2} < \infty.$$

Then $\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \xrightarrow{a.s.} 0$ as $n \to \infty$.

Theorem (SLLN for Linear Processes)

If $X_t = C(L)\varepsilon_t$ with $\sum_{j=1}^{\infty} j^2 c_j^2 < \infty$ and $\{\varepsilon_t\}$ is a m.d.s. with

$$\sup_t \mathbb{E}[|\varepsilon_t|^2] < \infty,$$

then $\frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{a.s.} 0$ as $n \to \infty$.

Some Asymptotic Results

Law of Large Numbers

Proof (SLLN for Linear Processes)

$$X_t = C(1)\varepsilon_t + (1-L)\widetilde{C}(L)\varepsilon_t$$
$$\frac{1}{n}\sum_{t=1}^{n} X_t = C(1)\frac{\sum_{t=1}^{n}\varepsilon_t}{n} - \frac{1}{n}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_n\right),$$

where $\widetilde{\varepsilon}_j = \widetilde{C}(L)\varepsilon_j$. By the SLLN for m.d.s.

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t \xrightarrow{a.s.} 0.$$

Now we only have to prove that

$$\frac{1}{n}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_n\right) \xrightarrow{a.s.} 0.$$

Some Asymptotic Results

Law of Large Numbers

Proof (SLLN for Linear Processes) (cont.)

Take for instance $\frac{\widetilde{\varepsilon}_n}{n}$:

$$\sum_{n=1}^{\infty} P\left(\frac{|\widetilde{\varepsilon}_n|}{n} > \delta\right) < \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[\widetilde{\varepsilon}_n^2\right]}{n^2\delta^2} = \frac{1}{\delta^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\left(\sum_{j=0}^{\infty}\tilde{c}_j^2\right)k < \infty$$

by Markov's Inequality, where $k = \sup_t \mathbb{E}\left[|\varepsilon_t|^2\right] < \infty$ and $\sum_{j=0}^{\infty}\tilde{c}_j^2 < \infty$.
By Borel-Cantelli Lemma $\frac{\widetilde{\varepsilon}_n}{n} \xrightarrow{a.s.} 0$.

Reminder:

- Markov's Inequality $P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}\left[|X|^p\right]}{\varepsilon^p}$ if $\mathbb{E}\left[|X|^p\right] < \infty$ and $p > 0$.
- Borel-Cantelli Lemma $\sum_{n=1}^{\infty} P(|X_n(\omega) - X(\omega)| \geq \varepsilon) < \infty \implies X_n \xrightarrow{a.s.} X$.

Some Asymptotic Results

Central Limit Theorems

Theorem (CLT for m.d.s.)

Let $\{\varepsilon_t\}$ be a strictly stationary and ergodic m.d.s. with

$$\mathbb{V}[\varepsilon_t] = \sigma^2 < \infty.$$

Then

$$\frac{\sum_{t=1}^{n} \varepsilon_t}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Theorem (CLT for Linear processes)

If $X_t = C(L)\varepsilon_t$ with $\sum_{j=1}^{\infty} j|c_j| < \infty$ and $\{\varepsilon_t\}$ a strictly stationary and ergodic m.d.s. with $\mathbb{V}[\varepsilon_t] = \sigma^2 < \infty$, then

$$\sqrt{n}\frac{\sum_{t=1}^{n} X_t}{n} \xrightarrow{d} \mathcal{N}(0, \underbrace{C(1)^2\sigma^2}_{\text{Long run variance}})$$

Some Asymptotic Results

Central Limit Theorems

Proof (CLT for Linear processes)

$$\sqrt{n}\bar{X}_n = \frac{\sum_{t=1}^{n} X_t}{\sqrt{n}} = C(1)\frac{\sum_{t=1}^{n} \varepsilon_t}{\sqrt{n}} - \frac{1}{\sqrt{n}}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_n\right),$$

$$C(1)\frac{\sum_{t=1}^{n} \varepsilon_t}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, C(1)^2\sigma^2) \qquad \text{by CLT for m.d.s.}$$

We need to prove that $\frac{1}{\sqrt{n}}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_n\right) = o_p(1)$.
Notice that we cannot use now the Borel-Cantelli (WHY?) argument used to prove the SLLN. This can be done by proving $\mathbb{E}[|\widetilde{\varepsilon}_n|] < \infty$ and then applying Markov's inequality (HOW?).

$$\mathbb{E}[|\widetilde{\varepsilon}_n|] = \mathbb{E}\left[\left|\sum_{j=0}^{\infty} \widetilde{c}_j \varepsilon_{n-j}\right|\right] \leq \sum_{j=0}^{\infty} |\widetilde{c}_j|\, \mathbb{E}\left[|\varepsilon_{n-j}|\right] < \infty.$$

Some Asymptotic Results

Central Limit Theorems

Example: AR(1)

$$X_t = \varphi X_{t-1} + \varepsilon_t$$

and $\{\varepsilon_t\}$ is a m.d.s. satisfying the conditions of the previous theorem.

$$C(1) = \frac{1}{1-\varphi}, \quad \text{then} \quad \sqrt{n}\bar{X}_n \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{(1-\varphi)^2}\right)$$

or

$$\sqrt{n}\bar{X}_n \xrightarrow{d} \mathcal{N}\left(0, \mathbb{V}[X_t]\frac{1+\varphi}{1-\varphi}\right)$$

because $\mathbb{V}[X_t] = \frac{\sigma^2}{1-\varphi^2}$.
Notice the problem when $\varphi = 1$.

Some Asymptotic Results

Sample Correlations

Let $\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}\left(X_t - \bar{X}_n\right)\left(X_{t+h} - \bar{X}_n\right)$ and $\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}$ for $0 \leq h \leq n-1$.

Theorem

If $\{X_t\}$ is the stationary process

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \{\varepsilon_t\} \sim iid(0, \sigma^2),$$

where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\mathbb{E}[\varepsilon_t^4] < \infty$, then for each $h \in \{1, 2, \dots\}$ we have

$$\sqrt{n}\,(\hat{\boldsymbol{\rho}}(h) - \boldsymbol{\rho}(h)) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{W}),$$

where

$$\hat{\boldsymbol{\rho}}(h) = \begin{bmatrix}\hat{\rho}(1) & \hat{\rho}(2) & \cdots & \hat{\rho}(h)\end{bmatrix}' \quad \text{and} \quad \boldsymbol{\rho}(h) = \begin{bmatrix}\rho(1) & \rho(2) & \cdots & \rho(h)\end{bmatrix}'$$

and $\boldsymbol{W}$ is the covariance matrix whose (i, j) element is given by Bartlett's formula.

Some Asymptotic Results

Sample Correlations

- The condition $\mathbb{E}[\varepsilon_t^4] < \infty$ can be relaxed at the expense of $\sum |j|\psi_j^2 < \infty$.
- Bartlett's formula:

$$w_{ij} = \sum_{k=-\infty}^{\infty} \Big\{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho(k)^2 - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \Big\}$$

- Simple algebra shows that

$$w_{ij} = \sum_{k=1}^{\infty} \Big\{ \rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k) \Big\} \Big\{ \rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k) \Big\}$$

and $w_{ii} = \sum_{k=1}^{\infty} (\rho_{i+k} + \rho_{i-k} - \rho_i \rho_k)^2$.

- Confidence intervals will be formed by

$$\pm C_\alpha \sqrt{\frac{w_{ii}}{n}} \qquad \text{with } C_\alpha = \Phi^{-1}\left(\frac{\alpha}{2}\right).$$

Some Asymptotic Results

Sample Correlations

Examples:

$$\bullet w_{ii} = 1 \quad \left(\pm 1.96 \frac{1}{\sqrt{n}} \text{ for } 95\%\right) \qquad \text{(WN)}$$

$$\bullet w_{ii} = 1 + 2\sum_{i=1}^{q} \rho^2(i), \qquad \text{for } i > q \qquad \text{(MA(q))}$$

$$\pm 1.96\sqrt{\frac{1+2\rho(1)^2}{n}}, \qquad \text{for } i > 1 \qquad \text{(MA(1))}$$

$$\bullet w_{ii} = \sum_{k=1}^{i} \phi^{2i}(\phi^{-k} - \phi^{k})^2 + \sum_{k=i+1}^{\infty} \phi^{2k}(\phi^{-i} - \phi^{i})^2 \qquad \text{(AR(1))}$$

$$= (1-\phi^{2i})(1+\phi^2)(1-\phi^2)^{-1} - 2i\phi^{2i}$$

$$\approx (1+\phi^2)/(1-\phi^2) \qquad \text{for } i \text{ large.}$$

Estimation and Inference of an AR(1)

Suppose $\{X_t\}$ is a stationary autoregressive process of order one satisfying

$$X_t = \varphi X_{t-1} + \varepsilon_t,$$

where $|\varphi| < 1$ and $\{\varepsilon_t\} \sim iid$, $\mathbb{V}[\varepsilon_t] = \sigma^2 < \infty$.

Then $X_t = \sum_{j=0}^{\infty} \varphi^j \varepsilon_{t-j}$ is strictly stationary and ergodic.

Estimation and Inference of an AR(1)

Let

$$\widehat{\varphi} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2} = \varphi + \frac{\sum_{t=2}^{n} X_{t-1}\varepsilon_t}{\sum_{t=2}^{n} X_{t-1}^2}.$$

Then

$$\sqrt{n}\,(\widehat{\varphi} - \varphi) = \frac{\frac{1}{\sqrt{n}} \sum_{t=2}^{n} X_{t-1}\varepsilon_t}{\frac{1}{n} \sum_{t=2}^{n} X_{t-1}^2}$$

and, by the Ergodic Theorem,

$$\frac{1}{n}\sum_{t=2}^{n} X_{t-1}^2 \xrightarrow{a.s.} \mathbb{E}[X_t^2] = \frac{\sigma^2}{1-\varphi^2}.$$

Estimation and Inference of an AR(1)

Observe that $Z_t = \varepsilon_t X_{t-1}$ is a m.d.s. w.r.t. $\mathcal{F}(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$,

$$\mathbb{E}[Z_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = X_{t-1}\,\mathbb{E}[\varepsilon_t] = 0.$$

$\{Z_t\}$ as a function of $\{\varepsilon_t\}, \{X_{t-1}\}$ is clearly strictly stationary and ergodic,

$$\mathbb{V}[\varepsilon_t X_{t-1}] = \mathbb{E}[\varepsilon_t^2 X_{t-1}^2] = \sigma^2\,\mathbb{E}[X_{t-1}^2] = \frac{\sigma^4}{1-\varphi^2} < \infty.$$

Applying the CLT for m.d.s. gives

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{n} X_{t-1}\varepsilon_t \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^4}{1-\varphi^2}\right).$$

Therefore,

$$\sqrt{n}(\widehat{\varphi}_n - \varphi) \xrightarrow{d} \mathcal{N}(0, 1-\varphi^2).$$

What if $\phi = 1$? Problem again.

Estimation and Inference of a MA(1)

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad |\theta| < 1, \quad \{\varepsilon_t\} \sim iid(0, \sigma^2).$$

- The first estimator can be obtained from $\rho_1 = \dfrac{\theta}{1+\theta^2}$.
- Estimating ρ_1 by

$$\widehat{\rho}_1 = \frac{\sum_{t=2}^{n}(Y_t - \bar{Y}_n)(Y_{t-1} - \bar{Y}_n)}{\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2}$$

we obtain

$$\widehat{\theta} = \begin{cases} \frac{1-\sqrt{1-4\hat{\rho}_1}}{2\hat{\rho}_1} & \text{if } 0 < |\widehat{\rho}_1| < 0.5, \\ -1 & \text{if } \widehat{\rho}_1 < -0.5, \\ 1 & \text{if } \widehat{\rho}_1 > 0.5, \\ 0 & \text{if } \widehat{\rho}_1 = 0. \end{cases}$$

Estimation and Inference of a MA(1)

- More efficient estimator can be obtained by LS or by ML.
- By LS:

$$\varepsilon_t = -\theta\varepsilon_{t-1} + Y_t,$$

$$Y_t = -\sum_{j=1}^{t-1}(-\theta)^j Y_{t-j} - (-\theta)^t\varepsilon_0 + \varepsilon_t = f_t(Y;\theta,\varepsilon_0) + \varepsilon_t,$$

where

$$f_1(Y;\theta,\varepsilon_0) = \theta\varepsilon_0,$$

$$f_t(Y;\theta,\varepsilon_0) = -\sum_{j=1}^{t-1}(-\theta)^j Y_{t-j} - (-\theta)^t\varepsilon_0.$$

Estimation and Inference of a MA(1)

LS for $Y_t = f_t(Y;\theta,\varepsilon_0) + \varepsilon_t$ is inside the classical NLS set-up that you have learned in previous econometric courses. In fact because we can obtain an initial consistent estimator of θ^0, we will obtain our asymptotic results from a one-step minimization procedure (one-step Gauss-Newton estimator).

Let's assume we have an initial estimator $\tilde{\theta}$ satisfying

$$\tilde{\theta} - \theta = o_p(n^{-\frac{1}{4}}) \quad \text{and} \quad \tilde{\varepsilon}_0 = O_p(1).$$

The one-step Gauss-Newton estimator of θ is obtained by regressing

$$\varepsilon_t(Y;\tilde{\theta}) = Y_t - f_t(Y;\tilde{\theta},\tilde{\varepsilon}_0) = \sum_{j=0}^{t-1}(-\tilde{\theta})^j Y_{t-j} + (-\tilde{\theta})^t \tilde{\varepsilon}_0$$

on the first derivative of $f_t(Y;\theta,\varepsilon_0)$ evaluated at $\theta = \tilde{\theta}$.

Estimation and Inference of a MA(1)

That derivative is

$$W_t(Y;\tilde{\theta}) = \begin{cases} \tilde{\varepsilon}_0 & \text{for } t = 1, \\ \sum_{j=1}^{t-1} j(-\tilde{\theta})^{j-1} Y_{t-j} + t(-\tilde{\theta})^{t-1}\tilde{\varepsilon}_0 & \text{for } t = 2, 3, \ldots, n. \end{cases}$$

Regressing $\varepsilon_t(Y;\tilde{\theta})$ on $W_t(Y;\tilde{\theta})$ we obtain an estimator of $\theta - \tilde{\theta}$. The improved estimator of θ is then

$$\widehat{\theta} = \tilde{\theta} + \Delta\widehat{\theta}, \qquad (*)$$

where

$$\Delta\widehat{\theta} = \left(\sum_{t=1}^{n} W_t(Y;\tilde{\theta})^2 \right)^{-1} \sum_{t=1}^{n} \varepsilon_t(Y;\tilde{\theta}) W_t(Y;\tilde{\theta}).$$

Estimation and Inference of a MA(1)

Theorem (Asymptotic normality for MA(1)

Let $Y_t = \varepsilon_t + \theta^0 \varepsilon_{t-1}$, where $|\theta^0| < 1$ and $\{\varepsilon_t\} \sim iid(0, \sigma^2)$ with $\mathbb{E}[|\varepsilon_t|^{2+r}] < L < \infty$ for some $r > 0$.

Let $\tilde{\varepsilon}_0$ and $\tilde{\theta}$ be initial estimators satisfying $\tilde{\varepsilon}_0 = O_p(1)$, $\tilde{\theta} - \theta = o_p(n^{-\frac{1}{4}})$, and $|\tilde{\theta}| < 1$. Then

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \xrightarrow{d} \mathcal{N}\left(0, 1 - (\theta^0)^2\right),$$

where $\widehat{\theta}$ is defined in (*). Also, $\widehat{\sigma}^2 \xrightarrow{p} (\sigma^0)^2$, where σ^0 is the true value of σ and $\widehat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(Y; \widehat{\theta})$.

Again a problem with unity ($\theta^0 = 1$).

Estimation and Inference of a MA(1)

Hannan-Rissanen algorithm

$$\widehat{\varepsilon}_t = Y_t - LinearProjection[Y_t | Y_{t-1}, Y_{t-2}, \dots],$$
$$Y_t = \theta \widehat{\varepsilon}_{t-1} + u_t$$
$$\Longrightarrow \widehat{\theta} = \frac{\sum_{t=2}^{n} Y_t \widehat{\varepsilon}_{t-1}}{\sum_{t=2}^{n} \widehat{\varepsilon}_{t-1}^2}.$$

Check the performance of this alternative estimator (in my web page there is an Eviews applet for this).

Testing for Autocorrelation

Box-Pierce with the true errors

Basic idea: use the correlations

$$\widehat{\rho}_i = \frac{\sum_{t=i+1}^{n} Y_t Y_{t-i}}{\sum_{t=i+1}^{n} Y_{t-i}^2}$$

to test

$$H_0: \quad Y_t = \varepsilon_t \text{ white noise.}$$

By the CLT on correlations we have

$$\sqrt{n}(\underset{k\times 1}{\hat{\boldsymbol{\rho}}_k} - \underset{k\times 1}{\boldsymbol{\rho}_k}) \xrightarrow{d} \mathcal{N}(0, \underset{k\times k}{W}).$$

Under H_0,

$$\sqrt{n}\hat{\boldsymbol{\rho}}_k \xrightarrow{d} \mathcal{N}(0, \underset{k\times k}{I}).$$

Therefore,

$$n\sum_{j=1}^{k}(\hat{\rho}_j)^2 \xrightarrow{d} \chi_k^2 \equiv Q \quad \text{- Box-Pierce statistic.}$$

Testing for Autocorrelation

Box-Pierce with the true errors

- If we had fitted an ARMA(p,q), the number of degrees of freedom would have been $k-(p+q)$.
- This test is testing

$$
\begin{aligned}
H_0 &: \mathbb{C}\mathrm{ov}(y_t, y_{t-j}) = 0 \quad j = 1, \ldots, k, \\
H_1 &: \mathbb{C}\mathrm{ov}(y_t, y_{t-j}) \neq 0 \quad \text{for at least one } j.
\end{aligned}
$$

- A modification

$$Q^* = n(n+2)\sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{n-j}$$

is the Ljung-Box statistic.

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

- A more realistic case is to assume that we have a model

$$y_t = x_t'\beta + \varepsilon_t, \qquad t = 1, \ldots, n$$

  and we want to test if the errors from this model are white noise.
- We do not observe the errors but the residuals, $\hat{\varepsilon}_t$.

$$\begin{cases} \hat{\rho}_j \equiv & \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \\ \hat{\gamma}_j \equiv & \frac{1}{n}\sum_{t=j+1}^{n} \varepsilon_t \varepsilon_{t-j} \end{cases} \qquad \text{(Before)}$$

$$\begin{cases} \hat{\rho}_j \equiv & \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \\ \hat{\gamma}_j \equiv & \frac{1}{n}\sum_{t=j+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} \end{cases} \qquad \text{(Now)}$$

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

- Is it right to use $\hat{\rho}_j$ (calculated from the residuals) and the residual-based Q statistics derived from $\{\hat{\rho}_j\}$ for testing for serial correlation?

  Yes, but only if the regressors are strictly exogenous:

  $$\mathbb{E}[\varepsilon_i | \boldsymbol{X}] = 0, \qquad i = 1, \ldots, n.$$

- This assumption is too strong for time series data. The strict exogeneity assumption implies that for any regressor k, $\mathbb{E}[x_{jk}\varepsilon_i] = 0, \forall i, j$, not only $i = j$.

- For instance, an AR(1) model:

  $$\begin{aligned} y_i &= \beta y_{t-i} + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \\ \mathbb{E}[y_i \varepsilon_i] &= \beta\, \mathbb{E}[y_{i-1}\varepsilon_i] + \mathbb{E}[\varepsilon_i^2] \\ &= \mathbb{E}[\varepsilon_i^2] \neq 0. \end{aligned}$$

  So, the regressor is not orthogonal to the **past** error term (y_i is the regressor for observation $i + 1$).

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

$$
\begin{aligned}
\hat{\gamma}_j \equiv & \frac{1}{n} \sum_{t=j+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} \\
= & \frac{1}{n} \sum_{t=j+1}^{n} \left[\varepsilon_t - x_t'(\hat{\beta} - \beta) \right] \left[\varepsilon_{t-j} - x_{t-j}'(\hat{\beta} - \beta) \right] \\
= & \gamma_j - \frac{1}{n} \sum_{t=j+1}^{n} (x_{t-j}\varepsilon_t + x_t \varepsilon_{t-j})'(\hat{\beta} - \beta) + (\hat{\beta} - \beta) \left(\frac{1}{n} \sum_{t=j+1}^{n} x_t x_{t-j}' \right) (\hat{\beta} - \beta)
\end{aligned}
$$

- If $\mathbb{E}[x_t \varepsilon_{t-j}], \mathbb{E}[x_{t-j} \varepsilon_t]$, and $\mathbb{E}[x_t x_{t-j}']$ are all finite, then because $\hat{\beta} - \beta \xrightarrow{p} 0$, we have
$$\hat{\gamma}_j - \gamma_j \xrightarrow{p} 0.$$
- However $\sqrt{n}(\hat{\gamma}_j - \gamma_j) \not\to 0$

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

$$\bullet\sqrt{n}\hat{\gamma}_j = \sqrt{n}\gamma_j - \overbrace{\frac{1}{n}\sum_{t=j+1}^{n}(x_{t-j}\varepsilon_t + x_t\varepsilon_{t-j})'}^{?}\overbrace{\sqrt{n}(\hat{\beta}-\beta)}^{O_p(1)}$$

$$\underbrace{+\underbrace{\sqrt{n}(\hat{\beta}-\beta)'}_{Op(1)}\underbrace{\left(\frac{1}{n}\sum_{t=j+1}^{n}x_t x_{t-j}'\right)}_{O_p(1)}\underbrace{(\hat{\beta}-\beta)}_{o_p(1)}}_{o_p(1)}$$

$$\bullet\frac{1}{n}\sum_{t=j+1}^{n}(x_{t-j}\varepsilon_t + x_t\varepsilon_{t-j}) \xrightarrow{p} \underbrace{\mathbb{E}[x_{t-j}\varepsilon_t] + \mathbb{E}[x_t\varepsilon_{t-j}]}_{\substack{=0 \\ \text{if all the regressors} \\ \text{are strictly exogenous}}}$$

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

- What if the regressors are predetermined but not strictly exogenous?
- Predetermined regressors:

$$\mathbb{E}[x_{ik}\varepsilon_i] = 0 \quad i = 1, \dots, n, \; k = 1, \dots, K.$$

- The AR(1) satisfies this assumption.
- When the regressors are not strictly exogenous, we need to modify the Q statistic to restore its asymptotic distribution. For this purpose we will impose two restrictions
  - Stronger form of predeterminedness

$$\mathbb{E}[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, x_t, x_{t-1}, \dots] = 0.$$

  - Stronger form of homoskedasticity

$$\mathbb{E}[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, x_t, x_{t-1}, \dots] = \sigma^2 > 0.$$

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

Proposition (testing for serial correlation with predetermined regressors)

Suppose $\underset{1\times 1}{y_t} = \underset{1\times K}{x_t'} \underset{K\times 1}{\beta} + \underset{1\times 1}{\varepsilon_t}, t = 1, 2, \ldots n$, with

- $\{y_t, x_t'\}$ jointly stationary and ergodic,
- $\mathbb{E}[x_t x_t'] = \Sigma_{xx}$ full rank,
- $\mathbb{E}[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, x_t, x_{t-1}, \ldots] = 0$,
- $\mathbb{E}[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, x_t, x_{t-1}, \ldots] = \sigma^2 > 0$.

Then

$$\sqrt{n}\hat{\gamma} \xrightarrow{d} \mathcal{N}(0, \sigma^4(I_p - \Phi)), \qquad \sqrt{n}\hat{\rho} \xrightarrow{d} \mathcal{N}(0, I_p - \Phi),$$

where

$$\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p)', \qquad \hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_p)'$$

and ϕ_{jk}, (j, k) element of $p \times p$ matrix Φ, is given by

$$\phi_{jk} = \mathbb{E}[x_t \varepsilon_{t-j}]' \, \mathbb{E}[x_t x_t']^{-1} \, \mathbb{E}[x_t \varepsilon_{t-k}] / \sigma^2.$$

Testing for Autocorrelation

Box-Pierce with sample autocorrelation calculated from residuals

Proof

Hayashi (p. 165)

- By the Ergodic theorem, matrix Φ is consistently estimated by its sample counterpart $\hat{\Phi} \equiv (\hat{\phi}_{j,k})$ with

$$\hat{\phi}_{j,k} \equiv \hat{\mu}_j' S_{xx}^{-1} \hat{\mu}_k / s^2, \qquad j, k = 1, 2, \ldots, p,$$

where

$$s^2 \equiv \frac{1}{n-K} \sum_{t=1}^{n} \hat{\varepsilon}_t^2, \quad \hat{\mu}_j \equiv \frac{1}{n} \sum_{t=j+1}^{n} x_t \hat{\varepsilon}_{t-j}.$$

- From last proposition we have

$$\text{modified Box-Pierce } Q \equiv n \hat{\rho}' (I_p - \hat{\Phi})^{-1} \hat{\rho} \xrightarrow{d} \chi_p^2.$$

Testing for Autocorrelation

An Auxiliary Regression-Based Test:

- Regress $\hat{\varepsilon}_t$ on $x_t, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \ldots, \hat{\varepsilon}_{t-p}$
- $$pF \sim \chi^2_p,$$
  where F stands for the F statistic for the hypothesis that the p coefficients of $\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \ldots, \hat{\varepsilon}_{t-p}$ are all zero
- Also $nR^2 \sim \chi^2_p$.

Appendix

(Cumby & Huizinga, 1992)

- Let
$$y_t = X_t\delta + \epsilon_t (t = 1, ..., T) \qquad and \qquad \hat{\epsilon}_t = y_t - X_t d.$$
- By a mean value theorem
$$\sqrt{T}\hat{r} = \sqrt{T}r + \frac{\partial r}{\partial \delta}\sqrt{T}(d - \delta).$$
- Proposition:
$$\sqrt{T}\hat{r} \sim N(0, V_{\hat{r}}) \qquad with \qquad V_{\hat{r}} = V_r + BV_d B^{'} + CD^{'}B^{'} + BCD^{'}.$$
- In general having to estimate the residuals will affect the asymptotic distribution of their sample autocorrelations. One special case is when X only contains lagged dependent variables and the errors are conditionally homokedasticity. In this case the Box-Pierce statistic $Q_s = T\hat{r}^{'}\hat{r}$ will asymptotically as a chi-squared random variable with $s - k$ degrees of freedom (s=number of correlations, and r= number of lags in the regression).

Appendix
(Cumby & Huizinga, 1992)

- Ljung (1986) investigates how large s must be before $Q \approx \chi^2_{s-k}$. In samples of 50 or 100 observations, $s \geq 10$ is sufficient for all AR(1) models and $s \geq 2$ is sufficient for AR(1) models with $|\varphi| < 0.9$.

Appendix

Distributions of empirical sizes

$X_t = 0.8X_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n, \quad \{\varepsilon_t\} \sim ii\mathcal{N}(0, 1)$

- Simulating values of Q for $p = 2, \ldots, 20$ and comparing with χ^2_{p-1}