Asymptotic distribution of least square estimators for linear models with dependent errors

Emmanuel Caron [1] & Sophie Dede [2]

Directeurs de thèse: Bertrand Michel (LMJL), Jérôme Dedecker (MAP5)

31 Mai 2018

[1]Ecole Centrale Nantes, Laboratoire de Mathématiques Jean Leray UMR 6629, 1 Rue de la Noë, 44300 Nantes. **Email: emmanuel.caron@ec-nantes.fr**

[2]Lycée Stanislas, 22 Rue Notre-Dame-des-Champs, 75006 Paris. Email: dede.sophie@gmail.com

Sommaire

Sommaire

Linear Regression Model

Linear Regression model:

$$Y = X\beta + \epsilon,$$

- X is a fixed design, $[n \times p]$
- Y is a n random vector
- β is a p vector of unknown parameters
- ϵ are the errors, $\epsilon \in \mathbb{R}^n$.

Usual assumptions:

- the errors are i.i.d.
- $\mathbb{E}(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I_n$
- Sometimes, $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$

Least Square Estimator

$$\hat{\beta} = argmin_{\beta \in \mathbb{R}^p} \|Y - X\beta\|_2^2 = (X^t X)^{-1} X^t Y.$$

$\hat{Y} = X\hat{\beta}$: Orthogonal Projection of Y on $\mathcal{M}_X = Vect\{X_{.,1}, ..., X_{.,p}\}$

- $\mathbb{E}(\hat{\beta}) = \beta$ and $Cov(\hat{\beta}) = \sigma^2 (X^t X)^{-1}$
- Residual vector: $\hat{\epsilon} = Y - \hat{Y} = Y - X\hat{\beta} \in \mathcal{M}_X^{\perp}$
- $\hat{\sigma}^2 = \frac{\|\hat{\epsilon}\|_2^2}{n-p}$.

Distribution of the LSE:

- Gaussian Case: $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X^t X)^{-1})$
- Non-Gaussian Case: $D(n)(\hat{\beta} - \beta) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2 Q^{-1})$.

Goals and Plan

Main Goal : Remove the independence hypothesis and find results similar to the i.i.d. case.

Plan :

1. Hannan's Theorem (1973) [4]: convergence of the LSE in the stationary case under very mild conditions
2. Show that for a large class of designs, the asymptotic covariance matrix is as simple as the i.i.d. case
3. Estimation of the covariance matrix
4. Applications with Fisher's tests.

Stationarity

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. $(\epsilon_i)_{i\in\mathbb{Z}}$ is an error process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, supposed strictly stationary, with zero mean, and $\epsilon_0 \in \mathbb{L}^2$.

Definition : Strict Stationarity

A stochastic process $(\epsilon_i)_{i\in\mathbb{Z}}$ is said to be strictly stationary if the joint distributions of $(\epsilon_{t_1}, \ldots, \epsilon_{t_k})$ and $(\epsilon_{t_1+h}, \ldots, \epsilon_{t_k+h})$ are the same for all positive integers k and for all $t_1, \ldots, t_k, h \in \mathbb{Z}$.

Spectral density

Autocovariance function:

$$\gamma(k) = \mathrm{Cov}(\epsilon_m, \epsilon_{m+k}) = \mathbb{E}(\epsilon_m \epsilon_{m+k}).$$

Let f be the associated spectral density, $\lambda \in [-\pi, \pi]$:

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda,$$

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{ik\lambda}.$$

Sommaire

Hannan's condition on the error process

Stationary case: Hannan (1973) $\rightarrow$ Central Limit Theorem for the usual LSE $\hat{\beta}$, under very mild conditions.

- $\forall j \in \mathbb{Z}$ and $\forall Z \in \mathbb{L}^2(\Omega)$: $P_j(Z) = \mathbb{E}(Z|\mathcal{F}_j) - \mathbb{E}(Z|\mathcal{F}_{j-1})$.
- **Hannan's condition** on the error process:

$$\sum_{i \in \mathbb{Z}} \|P_0(\epsilon_i)\|_{\mathbb{L}^2} < +\infty.$$

  This implies: $\sum_k |\gamma(k)| < +\infty$.
- Examples which verify Hannan's condition:
  - Linear Processes, functions of linear processes (Dedecker, Merlevède, Vólny (2007) [2])
  - Conditions à la Gordin ([2])
  - Framework of Wu (Wu (2005) [5])
  - Weakly dependent sequences (Dedecker-Prieur (2004) [3], Caron-Dede (2017) [1])

Hannan's conditions on the design

- Let $X_{.,j}$ be the column j of the matrix X, $j \in \{1, \ldots, p\}$:

$$d_j(n) = \|X_{.,j}\|_2 = \sqrt{\sum_{i=1}^{n} x_{i,j}^2},$$

  and let $D(n)$ be the diagonal matrix with diagonal term $d_j(n)$.
- **Conditions on the design**:

$$\forall j \in \{1, \ldots, p\}, \qquad \lim_{n\to\infty} d_j(n) = \infty,$$

$$\forall j \in \{1, \ldots, p\}, \qquad \lim_{n\to\infty} \frac{\sup_{1\leq i\leq n} |x_{i,j}|}{d_j(n)} = 0,$$

  and the following limits exist:

$$\forall j, l \in \{1, \ldots, p\}, \qquad \rho_{j,l}(k) = \lim_{n\to\infty} \sum_{m=1}^{n-k} \frac{x_{m,j}x_{m+k,l}}{d_j(n)d_l(n)}.$$

We define the $p \times p$ matrix $R(k)$:

$$R(k) = [\rho_{j,l}(k)] = \int_{-\pi}^{\pi} e^{ik\lambda} F_X(d\lambda),$$

where F_X is the spectral measure associated with the matrix $R(k)$.

Moreover, we suppose:

$$R(0) > 0.$$

Theorem (Hannan (1973) [4])

Under the previous conditions:

$$D(n)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, F^{-1}GF^{-1}),$$

and we have the convergence of second order moment:

$$\mathbb{E}\left(D(n)(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t D(n)^t\right) \xrightarrow[n\to\infty]{} F^{-1}GF^{-1},$$

with F and G the matrices:

$$F = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda),$$

$$G = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_X(d\lambda) \otimes f(\lambda).$$

Regular design

Definition (Regular design)

A fixed design X is called regular if, for any j, l in $\{1, \dots, p\}$, the coefficients $\rho_{j,l}(k)$ do not depend on k.

Interest:

- the asymptotic covariance matrix is easy to compute and similar to the i.i.d. case
- Not restrictive class (for instance Regularly varying sequence). Applications with Time Series.

Corollary

Under the assumptions of Hannan's Theorem, if moreover the design X is regular, then:

$$D(n)(\hat{\beta}-\beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \left(\sum_{k=-\infty}^{\infty} \gamma(k)\right) R(0)^{-1}\right),$$

and we have the convergence of the second order moment:

$$\mathbb{E}\left(D(n)(\hat{\beta}-\beta)(\hat{\beta}-\beta)^t D(n)^t\right) \xrightarrow[n\to\infty]{} \left(\sum_{k=-\infty}^{\infty} \gamma(k)\right) R(0)^{-1}.$$

For the i.i.d. case: $D(n)(\hat{\beta}-\beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2 Q^{-1}\right).$

Thus, to obtain confidence regions and tests for β, we need an estimator of $\sum_{k=-\infty}^{\infty} \gamma(k)$.

Sommaire

Spectral density estimate

Since $f(0) = 2\pi \sum_{k=-\infty}^{\infty} \gamma(k)$, we need an estimator of the spectral density.
Let us first consider a preliminary random function:

$$f_n(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq n-1} K\left(\frac{|k|}{c_n}\right) \hat{\gamma}_k e^{ik\lambda}, \quad \lambda \in [-\pi, \pi],$$

with:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{j=1}^{n-|k|} \epsilon_j \epsilon_{j+|k|}, \qquad 0 \leq |k| \leq (n-1).$$

K is the kernel:

$$\begin{cases} K(x) & = & 1 & if\ |x| \leq 1 \\ K(x) & = & 2 - |x| & if\ 1 \leq |x| \leq 2 \\ K(x) & = & 0 & if\ |x| > 2. \end{cases}$$

$c_n \xrightarrow[n \to \infty]{} \infty$ and $\frac{c_n}{n} \xrightarrow[n \to \infty]{} 0$.

In our context, $(\epsilon_i)_{i\in\{1,\ldots,n\}}$ is not observed. Only the residuals are available:

$$\hat{\epsilon}_i = Y_i - (x_i)^t\hat{\beta} = Y_i - \sum_{j=1}^{p} x_{i,j}\hat{\beta}_j,$$

because only the data Y and the design X are observed. Consequently, we consider the following estimator:

$$f_n^*(\lambda) = \frac{1}{2\pi}\sum_{|k|\leq n-1} K\left(\frac{|k|}{c_n}\right)\hat{\gamma}_k^* e^{ik\lambda}, \qquad \lambda\in[-\pi,\pi],$$

where:

$$\hat{\gamma}_k^* = \frac{1}{n}\sum_{j=1}^{n-|k|}\hat{\epsilon}_j\hat{\epsilon}_{j+|k|}, \qquad 0\leq |k|\leq (n-1).$$

Consistence

Theorem (Caron-Dede (2017), submitted)

Let c_n be a sequence of positive integers such that $c_n \to \infty$ as n tends to infinity, and:

$$c_n \mathbb{E}\left(|\epsilon_0|^2 \left(1 \wedge \frac{c_n}{n} |\epsilon_0|^2\right)\right) \xrightarrow[n\to\infty]{} 0.$$

Then, under the assumptions of Hannan's Theorem:

$$\sup_{\lambda \in [-\pi,\pi]} \|f_n^*(\lambda) - f(\lambda)\|_{\mathbb{L}^1} \xrightarrow[n\to\infty]{} 0.$$

Remark

If $\epsilon_0 \in \mathbb{L}^2$, then there exists $c_n \to \infty$ such that $c_n \mathbb{E}\left(|\epsilon_0|^2 \left(1 \wedge \frac{c_n}{n} |\epsilon_0|^2\right)\right) \xrightarrow[n\to\infty]{} 0$ holds.

Corollary

Corollary

Under the assumptions of Hannan's Theorem, if the design X is regular and if $f(0) > 0$, then:

$$\frac{R(0)^{\frac{1}{2}}}{\sqrt{2\pi f_n^*(0)}} D(n)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p),$$

where I_p is the $p \times p$ identity matrix.

Consequently, we can obtain confidence regions and tests for β in this dependent context.

Sommaire

Fisher Test

If the errors are i.i.d. Gaussian, the test statistic is:

$$F = \frac{1}{p - p_0} \times \frac{RSS_0 - RSS}{\hat{\sigma}_\epsilon^2}.$$

- p_0 is the dimension of the model under the H_0-hypothesis
- $RSS = \|\hat{\epsilon}\|_2^2$ (for the complete model)
- RSS_0 is the corresponding quantity under H_0
- $\hat{\sigma}_\epsilon^2 = \frac{RSS}{n-p}$

Under H_0:

$$F \overset{\mathcal{L}}{\sim} \mathcal{F}_{n-p}^{p-p_0}.$$

If the error process $(\epsilon_i)_{i\in\mathbb{Z}}$ is stationary, the usual Fischer tests can be corrected by replacing the estimator of $\sigma^2 = \mathbb{E}(\epsilon_0^2)$ by an estimator of: $\sum_k \gamma(k)$:

$$\tilde{F}_c = \frac{1}{p-p_0} \times \frac{RSS_0 - RSS}{2\pi f_n^*(0)},$$

where $f_n^*(\lambda) = \frac{1}{2\pi}\sum_{|k|\leq n-1} K\left(\frac{|k|}{c_n}\right)\hat{\gamma}_k^* e^{ik\lambda}$. Thanks to the previous results:

$$\tilde{F}_c \xrightarrow[n\to\infty]{\mathcal{L}} \frac{\chi^2(p-p_0)}{p-p_0}.$$

In practice, only a finite number of $\gamma(k)$ is estimated. For the simulations, to choose this number (called a_n) we shall use the graph of the empirical autocovariance of the residuals.

Hence:

$$F_c = \frac{1}{p - p_0} \times \frac{RSS_0 - RSS}{\hat{\gamma}_0^* + 2\sum_{k=1}^{a_n} \hat{\gamma}_k^*}.$$

Example: An autoregressive process

The process $(\epsilon_1, \ldots, \epsilon_n)$ is simulated, according to the AR(1) equation:

$$\epsilon_{k+1} = \frac{1}{2}(\epsilon_k + \eta_{k+1}).$$

- ϵ_1 is uniformly distributed over $[-\frac{1}{2}, \frac{1}{2}]$
- $(\eta_i)_{i \geq 2}$ is a sequence of i.i.d. random variables, independent of ϵ_1, such that $\mathbb{P}(\eta_i = -\frac{1}{2}) = \mathbb{P}(\eta_i = \frac{1}{2}) = \frac{1}{2}$
- Hannan's conditions are satisfied and the Fisher tests can be corrected.

Model

Model simulated:

$$Y_i = \beta_0 + \beta_1\sqrt{i} + \beta_2\log(i) + 10\epsilon_i, \quad \forall i \in \{1, ..., n\}$$

- H_0: $\beta_1 = \beta_2 = 0$ against H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$
- $\beta_0 = 3$
- Under H_0, the same Fisher test is carried out 2000 times. Then we look at the estimated level of the test for different choices of n and a_n. (we want an estimated level close to 5%).

- Case $\beta_1 = \beta_2 = 0$ and $a_n = 0$ (no correction):

n	500	1000	2000	3000	4000	5000
Estimated level	0.4435	0.4415	0.427	0.3925	0.397	0.4075

If $a_n = 0$, the estimated levels are too large. The test reject the null hypothesis too often.

As suggested by the graph of the estimated autocovariances, the choice $a_n = 4$ should give a better result for the estimated level.

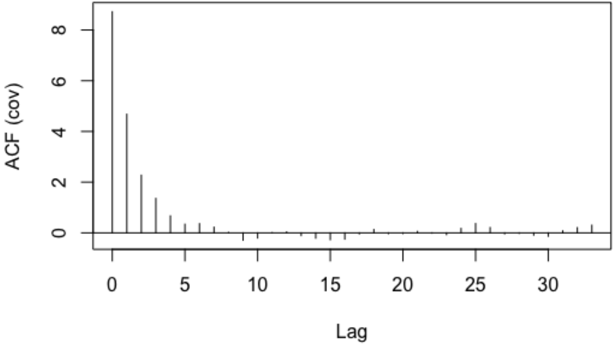

Figure : Empirical autocovariances, n = 2000.

- Case $\beta_1 = \beta_2 = 0$, $a_n = 4$:

n	500	1000	2000	3000	4000	5000
Estimated level	0.106	0.1	0.078	0.072	0.077	0.068

  - $a_n = 4$ works well. For $a_n = 4$ and $n = 5000$, the estimated level is around 0.07
  - If $n = 10000$, it is around 5%. Asymptotically, it converges to 0.05.

Then, under H_1, we study the estimated power of the test:

- Case $\beta_1 = 0$, $\beta_2 = 0.2$, $a_n = 4$:

n	500	1000	2000	3000	4000	5000
Estimated power	0.2505	0.317	0.4965	0.6005	0.725	0.801

The estimated power increases with the size of the samples, and it is around 0.8 as soon as $n = 5000$.

Perspectives

- To generalize these results in case where the design X is random
- To develop a data driven criterion for the coefficient a_n
- Package R for the applications of these results

Thank you for your attention !

E. Caron and S. Dede.
Asymptotic distribution of least square estimators for linear models with dependent errors : regular designs.
eprint arXiv:1710.05963, Oct. 2017.

J. Dedecker, F. Merlevède, and D. Volnỳ.
On the weak invariance principle for non-adapted sequences under projective criteria.
Journal of Theoretical Probability, 20(4):971–1004, 2007.

J. Dedecker and C. Prieur.
New dependence coefficients. examples and applications to statistics.
Probability Theory and Related Fields, 132(2):203–236, 2005.

E. J. Hannan.
Central limit theorems for time series regression.
Probability theory and related fields, 26(2):157–170, 1973.

W. B. Wu.
Nonlinear system theory: Another look at dependence.

Proceedings of the National Academy of Sciences of the United States of America, 102(40):14150–14154, 2005.