Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes.

D. Dacunha-Castelle* E. Gassiat†

November 1996

Abstract:
In this paper, we address to the problem of testing hypothesis using maximum likelihood statistics in non identifiable models, with application for model selection in situations where the parametrization for the larger model leads to non identifiability in the smaller model. We give two major applications: the case where the number of populations has to be tested in a mixture, and the case of stationary ARMA(p,q) processes where the order (p,q) has to be tested. We give the asymptotic distribution for the maximum (pseudo)-likelihood statistic when testing the order of the model. A locally conic parametrization appears to be a key tool for the problem, and allows to discover the deep similarity between the two problems.

A.M.S. Classification: 62 F 05, 62 H 30, 62 A 10.

Key Words: model selection, tests, non identifiable models, mixtures, ARMA processes.

Soumis aux Annals of Statistics.

*Laboratoire Modélisation Stochastique et Statistique, Université d'Orsay, Bat. 425, 91 405 Orsay
†Laboratoire Analyse et Probabilité, Université d'Evry-Val d'Essonne, Boulevard des Coquibus, 91 025 Evry

1 Introduction.

In this paper, we propose a general theory to deal with testing problems in non identifiable context. In case the model is a regular parametric model, the lack of identifiability leads in general to a degenerated Fisher information, so that the usual χ^2-theory does not apply. The usual proof for the χ^2-theory relies on an expansion of the likelihood till order 2, and then a maximization of the expansion. The major questions that now arise are the following:

- Question 1: Since the parameter is not identifiable, around which point can an expansion be made ?
- Question 2: In the optimization procedure, the inverse of the Fisher information appears, what to do since it is not invertible ?

Two famous examples of such situation are the test of the number of components in a mixture model, and the test of the order of an ARMA process. We solved completely the first example for testing one population against a mixture of populations in a previous work, see Dacunha-Castelle and Gassiat (1996). We develop here a general theory which applies to both problems. Let us first recall on the examples the lack of identifiability, where the difficulty of the problem comes from, and the general idea of our theory.
Let $\mathcal{F} = (f_\gamma)_{\gamma\in\Gamma}$ be a family of probability densities with respect to ν. Γ is a compact subset of $\mathbb{R}^k$ for some integer k. $\mathcal{G}_p$ is the set of all $p-$mixtures of densities of $\mathcal{F}$:

$$\begin{aligned}\mathcal{G}_p &= \{g_{\pi,\alpha} = \sum_{i=1}^{p} \pi_i \cdot f_{\gamma^i} \ : \ \pi = (\pi_1,\ldots,\pi_p),\ \alpha = (\gamma^1,\ldots,\gamma^p), \\ &\qquad \forall i = 1,\ldots,p,\ \gamma^i \in \Gamma,\ 0 \le \pi_i \le 1,\ \sum_{i=1}^{p} \pi_i = 1\}\end{aligned} \tag{1}$$

Obviously, the model is not identifiable for the parameters $\pi = (\pi_1,\ldots,\pi_p)$ and $\alpha = (\gamma^1,\ldots,\gamma^p)$. There exist mixtures g in $\mathcal{G}_p$ which have different representations $g_{\pi,\alpha}$ with different parameters π and α. For instance, we have for any permutation σ of the set $\{1,\ldots,p\}$:

$$\sum_{i=1}^{p} \pi_i \cdot f_{\gamma^i} = \sum_{i=1}^{p} \pi_{\sigma(i)} \cdot f_{\gamma^{\sigma(i)}}$$

Another example which may not be avoided by taking some quotient with respect to permutations is:

$$f_{\gamma^0} = \sum_{i=1}^{p} \pi_i \cdot f_{\gamma^0}$$

for any (π_i) such that $\pi_i \geq 0$ and $\sum_{i=1}^{p} \pi_i = 1$.
Since we solved part of the problem in Dacunha-Castelle and Gassiat (1996), we do not recall previous results concerning mixtures which were discussed there.

ARMA processes are given by the recurrence equation:

$$X_n + a_1 X_{n-1} + \ldots + a_p X_{n-p} = \epsilon_n + b_1 \epsilon_{n-1} + \ldots + b_q \epsilon_{n-q} \tag{2}$$

$(X_n)_{n\in\mathbb{N}}$ is a stationary process with $(\epsilon_n)_{n\in\mathbb{N}}$ as innovation process as soon as the complex polynomials $P(z) = 1 + \sum_{j=1}^{p} a_j z^j$ and $Q(z) = 1 + \sum_{j=1}^{q} b_j z^j$ do not have roots inside the complex unit disc and $(\epsilon_n)_{n\in\mathbb{N}}$ is a white noise. If $(\epsilon_n)_{n\in\mathbb{N}}$ is gaussian, then $(X_n)_{n\in\mathbb{N}}$ is a gaussian ARMA process. The spectral density f of such a process is given by:

$$f(x) = \frac{\sigma^2}{2\pi} |\frac{Q}{P}(e^{ix})|^2 \tag{3}$$

where σ^2 is the variance of the noise. Assume now that the true spectral density is

$$f_0(x) = \frac{\sigma_0^2}{2\pi} |\frac{Q_0}{P_0}(e^{ix})|^2 \tag{4}$$

with Q_0 of degree q_0, P_0 of degree p_0, and we want to test (p_0, q_0) against (p, q), where $p \geq p_0$ and $q \geq q_0$, $(p,q) \neq (p_0, q_0)$. The general model is that of stationary processes with spectral densities which have the form (3), where the degree of P is not larger than p and the degree of Q is not larger than q. As soon as $p > p_0$ and $q > q_0$, f_0 has in this model infinitely many representations, for instance by multiplying the representation (4) with the constant 1 written as the quotient of two identical polynomials. With this parametrization of the model, the information matrix for any parameter leading to f_0 has a kernel of dimension $\inf\{p - p_0, q - q_0\}$, see Theorem 3.3 in Azencott and Dacunha-Castelle (1986). The classical theory does not apply, the maximum likelihood statistic does not have a χ^2 asymptotic distribution. Alternative solutions for this testing problem may be given, using the correlation properties, see for instance Gill and Lewbel (1992). As test problems are intimately connected with estimation problems via confidence sets, another way to solve the problem could be the use of an estimator together with its asymptotic distribution. A now classical way to estimate the order of an ARMA process is by using some compensation term to the maximum likelihood statistic. Such idea has been introduced by Akaike (see Akaike (1970) and Akaike (1974)), with comparisons in Hannan (1980). The right choice of the compensating sequence requires a carefull investigation of the speed of convergence of the maximum likelihood statistic, which is studied for instance in Azencott and Dacunha-Castelle (1986). They use some quotient of the parameter space. Redner (1981) gives general ideas to use quotient spaces to prove the consistency (in the quotient space) of the maximum likelihood estimator in non identifiable situations. However, the precise asymptotic behavior of the estimator of the order is not known, so that it cannot be used for testing at a known asymptotic level. Another point of view is that of predictive stochastic complexity developed for example by Gerencser and Rissanen (). They give the asymptotic behaviour of three kinds of predictive stochastic complexities associated with ARMA processes, which give a computable criterion for model order estimation. However, this again does not lead to known asymptotics. Let us mention that, for testing the order using maximum likelihood, the computation of the asymptotic distribution was made by Hannan (1982) for the particular case $p_0 = q_0 = 0$ and $p = q = 1$. Hannan introduced a reparametrization of the model, which does not seem to be easily generalised to handle the general case. It was used again by Veres (1987) to find the asymptotic distribution for the case $p = p_0 + 1$ and $q = q_0 + 1$.

In this paper, we give the asymptotic distribution of the maximum likelihood statistic for *any* mixture model selection and for *any* ARMA model selection problem, so that this allows the construction of a test for the order at a known asymptotic level. To find this asymptotic distribution, we introduce a reparametrization of the model which we call a locally conic structure of the model. The general idea is that a first positive and real parameter θ contains some "distance" to the true model, this is the perturbation parameter, and a second parameter β is some direction of approach to the true model, or in other words the direction of the perturbation. A normalisation of the directional vector imposes the directional Fisher informations to be uniformly equal to 1. This gives an answer to question 2. β contains all the non identifiable part of the model. θ contains all the model order information, and is identifiable. θ may be consistently estimated: this gives an answer to question 1, the expansion will be done for θ near 0. Now, if $l_n(\theta, \beta)$ is the logarithm of the likelihood, the usual proof proceeds as follows (for i.i.d. observations $X_1, \ldots, X_n$) :

$$l_n(\theta, \beta) - l_n(\theta, 0) = \theta \sum_{i=1}^{n} h_\beta(X_i) - \frac{\theta^2}{2} \sum_{i=1}^{n} h_\beta(X_i)^2 + o_P(1)$$

Maximisation over θ leads to:

$$\sup_\theta l_n(\theta, \beta) = \frac{1}{2} \frac{(\sum_{i=1}^{n} h_\beta(X_i))^2}{\sum_{i=1}^{n} h_\beta(X_i)^2} + o_P(1)$$

Now, when maximizing over β, two problems appear. $\sum_{i=1}^{n} h_\beta(X_i)^2$ may take the value zero. There is a need of some uniform central limit theorem for a normalised variable. More importantly, **the remaining $o_P(1)$ has to be uniform over β !!!**

To our opinion, this last point has been overlooked in previous literature. In the most general cases of model selection for mixtures and ARMA processes, the expansion has to be done to an order bigger than 2, at least 4, leading to new asymptotics. The general situation seems to be the following. All directional models are LAN (locally asymptotically normal) models. The set $\mathcal{D}$, which is the union of all directional scores, may be defined as a subset of the unit sphere of some Hilbert space H. $\mathcal{D}$ has to be precompact, and its compactification leads to adding a new set $\tilde{\mathcal{D}}$, due to the non identifiability. Define ξ_d to be the (continuous) gaussian process indexed by $\mathcal{D}$ and with covariance the Hilbert product in H. Two different situations appear:

- $\tilde{\mathcal{D}}$ is small, the remaining terms in the expansion till order 2 are uniformly small, and the asymptotic distribution is

$$\sup_{d\in\mathcal{D}} \frac{1}{2}\xi_d^2 1_{\xi_d\geq 0}$$

  This is the case for testing one population against a simple mixture (see Dacunha-castelle and Gassiat(1996)), and for testing an ARMA(p_0, q_0) against an ARMA(p_0+1, q_0+1).

- $\tilde{\mathcal{D}}$ is not small, the expansion has to be done till order 4, and the asymptotic distribution is

$$\sup\left\{\sup_{d\in\mathcal{D}} \frac{1}{2}\xi_d^2 1_{\xi_d\geq 0};\ \sup_{d_1\in\mathcal{D}_1, d_2\in\mathcal{D}_2} \frac{1}{2}(\xi_{d_1}^2 + \xi_{d_2}^2 1_{\xi_{d_2}\geq 0})\right\}$$

  where $\mathcal{D}_1$ and $\mathcal{D}_2$ are two orthogonal subsets of $\tilde{\mathcal{D}}$. This is the case for testing q against p populations $(q < p)$ and for testing an ARMA(p_0, q_0) against an ARMA(p,q), $p_0 \leq p$, $q_0 \leq q$, $(p - p_0 + q - q_0) \geq 3$.

The paper is organized as follows: In section 2, we give the definition of a locally conic parametrization. Section 3 is devoted to the problem of testing q against p populations, and section 4 is devoted to the problem of testing an ARMA(p_0, q_0) against an ARMA(p,q). All technical proofs that are not essential for a comprehensive reading are given in section 5.

2 Locally conic models.

$X^{(n)} = (X_1, \ldots, X_n)$ is an n-dimensional real observation with distribution $P_0^{(n)}$ in a set $\mathcal{P}_n$, which is assumed to be dominated by some positive measure $\nu^{(n)}$.
We assume there exists a parametrization of all $\mathcal{P}_n$ through two parameters θ and β:

$(\theta, \beta) \in [0, M] \times \mathcal{B}$, M is a positive real number, $\mathcal{B}$ is a compact Polish space, and there exists a subset $\mathcal{T}$ of $[0, M] \times \mathcal{B}$ such that:

$$\forall n \in \mathbb{N}, \quad \mathcal{P}_n = \{P^{(n)}_{(\theta,\beta)}, (\theta, \beta) \in \mathcal{T}\}$$

where $[0, M] \times \mathcal{B}$ is endowed with the product topology of $\mathbb{R}$ and $\mathcal{B}$, and $\mathcal{T}$ has compact closure $\overline{\mathcal{T}}$.

The parametrization is assumed to be non identifiable in the parameter β for $\theta = 0$, but identifiable in the parameter θ at $\theta = 0$, that is:

$$P^{(n)}_{(\theta,\beta)} = P^{(n)}_0 \iff \theta = 0$$

which in particular implies:

$$\forall \beta \in \mathcal{B}, P^{(n)}_{(0,\beta)} = P^{(n)}_0$$

Define for any positive number c:

$$\mathcal{B}^c = \{\beta \in \mathcal{B} \ : \ \exists \theta \leq c, \ (\theta, \beta) \in \overline{\mathcal{T}}\}$$

For any β in $\mathcal{B}^c$, define:

$$\theta_\beta = \sup\{t \ : \ [0, t] \times \{\beta\} \subset \overline{\mathcal{T}}\}$$

Assume moreover:
$\forall \beta \in \mathcal{B}$, either $\theta_\beta > 0$, or there exists $c > 0$ such that $[0, c] \times \{\beta\} \cap \overline{\mathcal{T}}$ is empty.
Define now:

$$\tilde{\mathcal{B}} = \bigcap_{c>0} \mathcal{B}^c$$

The assumption says that it is impossible to find accumulation sequences of parameter leading to $\theta = 0$ with directions β where the submodel $(P^{(n)}_{(\theta,\beta)}, (\theta, \beta) \in \mathcal{T})_\theta$ (where β is fixed) is not defined in a rightneighborhood of 0. Moreover, $\tilde{\mathcal{B}}$ is then the set of all directions β for which the submodel approaches 0.

Such parametrization is called a locally conic parametrization. Models for which there exists a locally parametrization are called locally conic models.

3 Testing the number of populations in a mixture.

In this section, $X^{(n)}$ is an n-sample of a mixture of q populations, that is

$$P^{(n)}_0 = (g_0 \nu)^{\otimes n}$$

where g_0 is a mixture of q populations in the parametric family $(f_\gamma)_{\gamma\in\Gamma}$, $\Gamma \subset \mathbb{R}^k$:

$$g_0 = \sum_{l=1}^{q} \pi_l^0 f_{\gamma^{l,0}}$$

The general model is that of p mixtures $\mathcal{G}_p$ given by (1). We assume that $\mathcal{G}_p$ is identifiable in the weak following sense:

$$\sum_{l=1}^{p} \pi_l^0 f_{\gamma^{l,0}} = \sum_{l=1}^{p} \pi_l^1 f_{\gamma^{l,1}} \quad \nu \ a.e. \Longleftrightarrow \sum_{i=1}^{p} \pi_i^0 \cdot \delta_{\gamma_i^0} = \sum_{i=1}^{p} \pi_i^1 \cdot \delta_{\gamma_i^1}$$

as probability distributions on Γ. In other words, $\mathcal{G}_p$ is identifiable if the parameter is the mixing discrete probability distribution on Γ. Teicher (1965) or Yakowitz and Spragins (1968) give sufficient conditions for such weak identifiability, which hold for instance for finite mixtures of gaussian or gamma distributions.

The aim of this section is to derive the limiting distribution of the maximum likelihood statistic. Define for any g in $\mathcal{G}_p$:

$$l_n(g) = \sum_{i=1}^{n} \log g(X_i)$$

and the maximum likelihood statistic

$$T_n = \sup_{g\in\mathcal{G}_p} l_n(g)$$

We introduce the following locally conic parametrization, previously proposed by the authors in Dacunha-Castlle and Gassiat (1996). Define $\mathcal{B}_0$ the set of parameters

$$\beta = (\lambda_1, \ldots, \lambda_{p-q}, \gamma^1, \ldots, \gamma^{p-q}, \delta^1, \ldots, \delta^q, \rho_1, \ldots, \rho_q)$$

such that

$$\lambda_i \geq 0, \ \gamma^i \in \Gamma, \ i = 1, \ldots, p-q, \ \delta^l \in \mathbb{R}^k, \ \rho_l \in \mathbb{R}, \ l = 1, \ldots, q \tag{5}$$

$$\sum_{i=1}^{p-q} \lambda_i + \sum_{l=1}^{q} \rho_l = 0 \ \text{ and } \ \sum_{i=1}^{p-q} \lambda_i^2 + \sum_{l=1}^{q} \rho_l^2 + \sum_{l=1}^{q} \|\delta^l\|^2 = 1 \tag{6}$$

Let H be the Hilbert space $L^2(g_0\nu)$. Let then

$$N(\beta) = \| \sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{1}{g_0} \frac{\partial f_\gamma}{\partial \gamma_i}|_{\gamma=\gamma^{l,0}} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\gamma^i}}{g_0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^{l,0}}}{g_0} \|_H$$

We shall see later the interpretation of this quantity. For any β in $\mathcal{B}_0$ and any non negative θ such that for any integer $l = 1, \ldots, q$,
$\pi_l^0 + \rho_l \frac{\theta}{N(\beta)} \geq 0$, define the mixture:

$$g_{(\theta,\beta)} = \sum_{i=1}^{p-q} \lambda_i \frac{\theta}{N(\beta)} f_{\gamma^i} + \sum_{l=1}^{q} (\pi_l^0 + \rho_l \frac{\theta}{N(\beta)}) f_{\gamma^{l,0} + \frac{\theta}{N(\beta)} \delta^l} \tag{7}$$

Such parametrization may be viewed as a perturbation of g_0 in the following way: perturb the q mixture g_0 through a perturbation of the parameters $\gamma^{l,0}$ and the weights π_l^0, and add a perturbation as a $p-q$-mixture with weight tending to 0.
Such equation does not completely set a locally conic parametrization. Indeed, the equation (7) does not define unambiguously (θ,β) for a given mixture. For instance, different sets of parameters may give g_0. It is then important to define the set $\tilde{\mathcal{B}}$ such that $g_{(\theta,\beta)} = g_0 \Longleftrightarrow \theta = 0$, which is not an immediate consequence of the definition of $g_{(\theta,\beta)}$.

Let g be any p-mixture:

$$g = \sum_{i=1}^{p} \pi_i \cdot f_{\gamma^i}$$

To describe it through equation (7), one has to associate the parameters of g to those of g_0, that is to give a special order to the parameters. In other words: for any permutation σ of the set $\{1,\ldots,p\}$, we define the parameters θ_σ such that $g_{(\theta_\sigma,\beta_\sigma)} = g$. This leads to:

$$\beta_\sigma = (\lambda_{1,\sigma},\ldots,\lambda_{p-q,\sigma},\gamma^{1,\sigma},\ldots,\gamma^{p-q,\sigma},\delta^{1,\sigma},\ldots,\delta^{q,\sigma},\rho_{1,\sigma},\ldots,\rho_{q,\sigma})$$

with:

$$\forall i = 1,\ldots,p-q, \;\; \lambda_{i,\sigma}\cdot\theta_\sigma = \pi_{\sigma(i)}\cdot N(\beta_\sigma)$$
$$\forall i = 1,\ldots,p-q, \;\; \gamma^{i,\sigma} = \gamma^{\sigma(i)}$$
$$\forall i = 1,\ldots,q, \;\; \delta^{i,\sigma}\cdot\theta_\sigma = (\gamma^{\sigma(p-q+i)} - \gamma^{i,0})\cdot N(\beta_\sigma)$$
$$\forall i = 1,\ldots,q, \;\; \rho_{i,\sigma}\cdot\theta_\sigma = (\pi_{\sigma(p-q+i)} - \pi_i^0)\cdot N(\beta_\sigma)$$

It is easily seen that

$$\theta_\sigma = \| \sum_{l=1}^{q}\sum_{i=1}^{k}(\gamma_i^{\sigma(p-q+l)} - \gamma_i^{l,0})\frac{1}{g_0}\frac{\partial f_\gamma}{\partial\gamma_i}|_{\gamma=\gamma^{l,0}} + \sum_{i=1}^{p-q}\pi_{\sigma(i)}\frac{f_{\gamma^{\sigma(i)}}}{g_0} + \sum_{l=1}^{q}(\pi_{\sigma(p-q+l)} - \pi_l^0)\frac{f_{\gamma^{l,0}}}{g_0}\|_H$$

The system is not ambiguous on the scale of β_σ because of the normalizing condition (6).
The problem is then to choose between the permutations. The following choice is a good one. The idea is to associate step by step the nearest points γ^i involved in g to the set of points $\gamma^{l,0}$ involved in g_0. Look for:

$$\min_{l=1,\ldots,q,i=1,\ldots,p} \|\gamma^{l,0} - \gamma^i\|$$

It is attained for l_1 and i_1. Define then $\sigma(p-q+l_1) = i_1$. Look then for

$$\min_{l=1,\ldots,q,l\neq l_1,i=1,\ldots,p,i\neq i_1} \|\gamma^{l,0} - \gamma^i\|$$

It is attained for l_2 and i_2. Set then $\sigma(p-q+l_2)=i_2$. By induction, define in the same way $\sigma(p-q+l_j)=i_j$ for $j=1,\ldots,q$. In this algorithm, consider only points truly involved in g (eventually less than p points). Then complete the permutation σ in some ordered way. You then have defined a permutation $\sigma(g)$. The locally parametrization is then given by equation (7) with:

$$\mathcal{T}=\{(\theta,\beta_{\sigma(g)})\ :\ \ \theta\leq\theta_{\sigma(g)},\ \ g\in\mathcal{G}\}$$

This induces the set $\tilde{\mathcal{B}}$ as the intersection of all directions approaching 0 in $\mathcal{T}$.
Such parametrization is locally conic. An important point to notice, coming from the normalizing condition (6) is that:

$$\forall(\theta,\beta)\in\mathcal{T},\ \frac{\theta}{N(\beta)}\leq p+2q\sup_{\gamma\in\Gamma}\|\gamma\|^2 \tag{8}$$

so that θ tends to 0 as soon as $N(\beta)$ tends to 0.
Since we shall use partial derivatives of f_γ with respect to γ, we introduce some notations: $D^h_{i_1\ldots i_h}$ will be the h-th partial derivative operator with respect to $\gamma_{i_1}\ldots\gamma_{i_h}$. So that $D^h_{i_1\ldots i_h}f_\gamma$ will be the value of this partial derivative of f at point γ. We shall now need some more assumptions. Define $\mathcal{D}$ the subset of the unit sphere of H of functions of form

$$\frac{1}{N(\beta)}\left(\sum_{l=1}^{q}\pi_l^0\sum_{i=1}^{k}\delta_i^l\frac{D_i^1 f_{\gamma^{l,0}}}{g_0}+\sum_{i=1}^{p-q}\lambda_i\frac{f_{\gamma^i}}{g_0}+\sum_{l=1}^{q}\rho_l\frac{f_{\gamma^{l,0}}}{g_0}\right)$$

with β in $\tilde{\mathcal{B}}$. Define also ξ_d the gaussian process indexed by $\mathcal{D}$ with covariance the usual hilbertian product in H.
We shall use the following assumptions.

- **(P0)** There exists a function h in $L_1(g_0\nu)$ such that: $\forall\gamma\in\Gamma$, $|\log f_\gamma|\leq h$ ν-a.e. Moreover, f_γ possesses partial derivatives till order 5. For all $h\leq 5$, and all $i_1\ldots i_h$,

  $$\frac{D^h_{i_1\ldots i_h}f_{\gamma^0}}{g_0}\in L^3(g_0\nu)$$

  Moreover, there exist functions m_2, m_5 and a positive ϵ such that:

  $$\sup_{\|\gamma-\gamma^0\|\leq\epsilon}|\frac{D^2_{i_1 i_2}f_\gamma}{g_0}|\leq m_2\quad E_{g_0\nu}[m_2^3]<+\infty$$

  $$\sup_{\|\gamma-\gamma^0\|\leq\epsilon}|\frac{D^5_{i_1\ldots i_5}f_\gamma}{g_0}|\leq m_5\quad E_{g_0\nu}[m_5^3]<+\infty$$

- **(P1)** $\mathcal{D}$ is a Donsker class (see Van der Vart and Wellner (1996)), and ξ_d has continuous sample paths.

- **(P2)** For any integer p_1, p_2, such that $p_1 + p_2 \leq p - q$, for any set of distinct points γ^l, $l = 1, \dots, p_1$, distinct from any $\gamma^{l,0}$, the set of functions

$$((\frac{f_{\gamma^l}}{g_0})_{l=1,\dots,p_1}, (\frac{f_{\gamma^{l,0}}}{g_0})_{l=1,\dots,q}, (\frac{D_i^1 f_{\gamma^{l,0}}}{g_0})_{l=1,\dots,q,i=1,\dots,k}, (\frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0})_{l=1,\dots,p_2,i,j=1,\dots,k}$$

  is free in H.

To check **(P1)**, one may use metric entropy or bracketing entropy Theorems (see for instance Ossiander (1987) or Van der Vart and Wellner (1996)). Those entropy results may also be used to prove the continuity of the process (ξ_d). In fact, computation of entropy behavior is in general easy, since, as will be seen later, $\overline{\mathcal{D}}$ is a parametric set described with a finite number of parameters varying in a compact euclidian space.
Let now $\mathcal{D}_1$ be the subset of the unit sphere of H of functions of form

$$\sum_{l=1}^{q} \sum_{i=1}^{k} \lambda_{l,k} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}$$

and let $\mathcal{D}_2$ be the subset of the unit sphere of H of functions of form

$$\sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{q} \sum_{i=1}^{k} \lambda_{l,i} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{p_2} \sum_{i,j=1}^{k} \tau_l a_i a_j \frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0}$$

with $p_1 \leq p - q - 1$, $p_1 + p_2 \leq p - q$, $\mu_l \geq 0$, $\sum_{l=1}^{p_1} \mu_l + \sum_{l=1}^{q} \tilde{\rho}_l = 0$, $\tau_l \geq 0$, which are orthogonal to $\mathcal{D}_1$. We have the following asymptotic result:

Theorem 3.1 *Under the assumptions **(P0)**, **(P1)**, **(P2)**, $T_n - l_n(g_0)$ converges in distribution to the following variable:*

$$\sup \left\{ \sup_{d \in \mathcal{D}} \frac{1}{2} \xi_d^2 1_{\xi_d \geq 0}; \sup_{d_1 \in \mathcal{D}_1, d_2 \in \mathcal{D}_2} \frac{1}{2} (\xi_{d_1}^2 + \xi_{d_2}^2 1_{\xi_{d_2} \geq 0}) \right\}$$

Notice that $\mathcal{D}_1$ and $\mathcal{D}_2$ are subset of $\overline{\mathcal{D}}$, the (compact) closure of $\mathcal{D}$ in H. More precisely, $\overline{\mathcal{D}}$ is the subset of the unit sphere of H of functions of form

$$\sum_{l=1}^{p_1} \lambda_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{q} \sum_{i=1}^{k} \mu_{l,i} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{p_2} \sum_{i,j=1}^{k} \tau_l a_i a_j \frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0}$$

with $p_1 + p_2 \leq p - q$, $\lambda_l \geq 0$, $\sum_{l=1}^{p_1} \lambda_l + \sum_{l=1}^{q} \tilde{\rho}_l = 0$, $\tau_l \geq 0$.
The proof of Theorem 3.1 is in section 5.

4 Testing the order of an ARMA process.

In this section we assume $X^{(n)}$ is an n-realization of a strictly stationary process with spectral density f_0 given by equation (4).
Recall that if X is an ARMA(p,q) process with spectral density f, the Fejer-Riesz canonical representation is:

$$f(e^{ix}) = \frac{\sigma^2}{2\pi} \cdot |\frac{Q}{P}(e^{ix})|^2 = \frac{\sigma^2}{2\pi} \cdot g(e^{ix})$$

where P is a polynomial with p roots of modulus strictly bigger then 1 and Q is a polynomial with q roots of modulus ≥ 1, $P(0) = Q(0) = 1$, and P and Q have real coefficients. We then define the parameter space $F(\rho, u)$ as the space of all spectral densities of the previous form with all poles and zeros $\geq 1+\rho$, and $0 < u \leq \sigma^2 \leq 1/u$. We shall use the maximum pseudo-likelihood statistic for testing the order. We set it now. For any integrable function h on the torus, define $\widehat{h}_k$ the Fourier coefficient:

$$\widehat{h}_k = \int_{-\pi}^{\pi} e^{-ikx} h(e^{ix}) \frac{dx}{2\pi}$$

Define also the Toeplitz operator of order n, T_n, as the operator which associates to each integrable function h on the torus the $n \times n$ Toeplitz matrix $T_n(h)$:

$$(T_n(h))_{i,j} = \widehat{h}_{i-j}, \quad i,j = 1, \ldots, n$$

Define for any continuous function v the periodogram

$$I_n(v) :=^T X^{(n)} \cdot T_n(v) \cdot X^{(n)} = \int_{-\pi}^{\pi} v(e^{ix}) |\sum_1^n X_k e^{ikx}|^2 \frac{dx}{2\pi}$$

In case the process is a gaussian process, the logarithm of the likelihood is $L_n(f)$ with:

$$-2L_n(f) = n \log 2\pi + \log \det T_n(f) +^T X^{(n)} [T_n(f)]^{-1} X^{(n)}$$

It is well known (see for instance Azencott and Dacunha-Castelle (1986)) that it is well approximated by the Whittle contrast function C_n, or pseudo likelihood, given by:

$$\begin{aligned} C_n(f) &= n \log 2\pi +^T X^{(n)} \cdot T_n(\frac{1}{f}) \cdot X^{(n)} + n \log \sigma^2 && (9) \\ &= n \log 2\pi + n \log \sigma^2 + I_n(\frac{1}{f}) && (10) \end{aligned}$$

which is a contrast function for the estimation of the parameters even if the process is not gaussian. As was explained in the introduction, for $p > p_0$ and $q > q_0$, the model is not identifiable. We now define a locally conic parametrization. Let f_0 have zeros

$1/u_i, i = 1, \ldots, q_0$ and poles $1/t_i, i = 1, \ldots, p_0$. We assume for simplicity that all poles are distinct and all zeros are distinct. In other case, the Theorem still holds with a similar proof. Define $r = \min(p - p_0, q - q_0)$. Suppose that $r = p - p_0$. The case where $r = q - q_0$ can be handled in the same manner. Let $s = q - q_0 - r$. Define for any non negative θ, and any

$$\beta = (\delta, \mu, \tau, c, \gamma, \nu)$$

with $\delta \in \mathbb{R}$, $\mu = (\mu_j)_{j=1,\ldots,q_0} \in \mathbb{C}^{q_0}$, $\tau = (\tau_j)_{j=1,\ldots,p_0} \in \mathbb{C}^{q_0}$, $c = (c_j)_{j=1,\ldots,r} \in \mathbb{C}^r$, $\gamma = (\gamma_j)_{j=1,\ldots,r} \in \mathbb{C}^r$, $\nu = (\nu_j)_{j=1\ldots,s} \in \mathbb{C}^s$, and such that

$$\|\mu\|^2 + \|\tau\|^2 + \|\gamma\|^2 + \|\nu\|^2 = 1$$

Define now

$$\begin{aligned} f_{(\theta,\beta)}(e^{ix}) \quad = \quad & \left(\frac{\sigma_0^2 + \frac{\theta}{N(\beta)}\delta}{2\pi}\right) \left|\frac{\prod_{i=1}^{q_0}(1 - (u_i + \frac{\theta}{N(\beta)}\mu_i)z)}{\prod_{i=1}^{p_0}(1 - (t_i + \frac{\theta}{N(\beta)}\tau_i)z)}\right|^2 \left|\prod_{i=1}^{r} \frac{(1 - (c_i + \epsilon_i\frac{\theta}{N(\beta)}\gamma_i)z)}{(1 - (c_i + (1 - \epsilon_i)\frac{\theta}{N(\beta)}\gamma_i)z)}\right|^2 \\ & |\prod_{i=1}^{s}(1 - (\frac{\theta}{N(\beta)}\nu_i)z)|^2 \end{aligned} \tag{11}$$

with $z = e^{ix}$, where, for $i = 1, \ldots, r$, $\epsilon_i = 1$ if $|c_i - t_i| < |c_i - u_i|$ and $\epsilon_i = 0$ if $|c_i - t_i| \geq |c_i - u_i|$, and with

$$\begin{aligned} N(\beta) \quad = \quad & \|\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}(\frac{\tau_i z}{1 - t_i z} + \frac{\overline{\tau_i}}{z - \overline{t_i}}) - \sum_{i=1}^{q_0}(\frac{\mu_i z}{1 - u_i z} + \frac{\overline{\mu_i}}{z - \overline{u_i}}) \\ & + \sum_{i=1}^{r}(\frac{(1 - 2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1 - 2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}) - \sum_{i=1}^{s}(\nu_i z + \frac{\overline{\nu_i}}{z})\|_H^2 \end{aligned}$$

where H is the Hilbert space $L^2([0, 2\pi]; \frac{dx}{2\pi})$. The parameters θ and β can be constrained to ly in a prescribed set in order that $f_{(\theta,\beta)}$ lies in $F(\rho, u)$ (in particular the polynomials have real coefficients), and in order that:

$$f_{(\theta,\beta)} = f_0 \Longleftrightarrow \theta = 0$$

This can be done for example by a choice of associations of the zeros and poles of f with those of f_0. Let f have zeros $1/v_j$, $j = 1, \ldots, q'$, $q_0 \leq q' \leq q$, and poles w_j, $j = 1, \ldots, p'$, $p_0 \leq p' \leq p$. Let ρ_1 be any permutation of $[1, \ldots, q']$ and ρ_2 any permutation of $[1, \ldots, p']$. Define then $P(z) = \prod(1 - w_{\rho_2(j)}z)$, $Q(z) = \prod(1 - v_{\rho_1(j)})$. The right choice of the permutations ρ_1 and ρ_2 will lead to the desired local identifiability without loosing infinite differentiability. For this choice, use the following rule: choose first any (maybe between several) $\rho_1(j)$ such that

$$|v_{\rho_1(j)} - u_j| = \inf_l \inf_\rho |v_{\rho_1(j)} - u_j|$$

Then iterate the rule for the $q_0 - 1$ remaining zeros of f_0. Do the same for the p_0 poles of f_0, so that ρ_2 is defined for p_0 points. It remains $p' - p_0$ poles and $q' - q_0$ zeros. Do the same coupling for r points relating r remaining poles to the r remaining poles. To end, complete τ_2 in some way. This gives an enumeration of poles and zeros so that polynomials P, Q may be built with this order, and then θ and β; The set $\mathcal{T}$ is then the set of all obtained (θ, β) when f runs over $F(\rho, u)$. Recall that, to have real parameters for the AR and MA polynomials, when one perturbation coefficient is associated to a zero or a pole which is not real, its conjugate is the perturbation coefficient associated to the conjugate zero or pole. This will always be the case. The number of real parameters is then $p + q + 1$. We then obviously have by construction that $f_{(\theta,\beta)} = f_0$ implies $\theta = 0$.
Notice that, in some sense, such construction chooses a particular member of the equivalent classes in the quotient space. Our locally conic parametrization is one way of thinking distances between quotient classes.
Notice also that, at least for small enough θ, the perturbation direction β may take any direction. In other words, the set of real parameters involved in $\tilde{\mathcal{B}}$ spans $\mathbb{R}^{p+q+1}$.
Another important remark is that we have for all (θ, β)

$$\frac{\theta}{N(\beta)} \leq 2(p_0 + q_0 + r + s)$$

We now define the derivative space $\mathcal{D}$ to be the subset of the unit sphere of H of functions of form

$$\begin{aligned}\frac{1}{N(\beta)} \quad & \Big(\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}\Big(\frac{\tau_i z}{1 - t_i z} + \frac{\overline{\tau_i}}{z - \overline{t_i}}\Big) - \sum_{i=1}^{q_0}\Big(\frac{\mu_i z}{1 - u_i z} + \frac{\overline{\mu_i}}{z - \overline{u_i}}\Big) \\ & + \sum_{i=1}^{r}\Big(\frac{(1-2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1-2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}\Big) - \sum_{i=1}^{s}\Big(\nu_i z + \frac{\overline{\nu_i}}{z}\Big)\Big)\end{aligned}$$

with $z = e^{ix}$ and with β in $\tilde{\mathcal{B}}$. Define also ξ_d the gaussian process indexed by $\mathcal{D}$ with covariance the usual hilbertian product in H. Let now $\mathcal{D}_1$ be the subset of the unit sphere of H of functions of form

$$\begin{aligned}& \frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}\left(\sum_{k=1}^{a_i} \tau_{i,k}\Big(\frac{z}{1 - t_i z}\Big)^{a_i} + \overline{\tau_{i,k}}\Big(\frac{1}{z - \overline{t_i}}\Big)^{a_i}\right) + \sum_{i=1}^{q_0}\left(\sum_{k=1}^{b_i} \mu_{i,k}\Big(\frac{z}{1 - u_i z}\Big)^{b_i} + \overline{\mu_{i,k}}\Big(\frac{1}{z - \overline{u_i}}\Big)^{b_i}\right) \\ & - \sum_{i=1}^{r_1}\Big(\frac{(1-2\epsilon_i)\gamma_i z}{1 - c_i z} + \frac{(1-2\epsilon_i)\overline{\gamma_i}}{z - \overline{c_i}}\Big) - \sum_{i=1}^{s}\Big(\nu_i z + \frac{\overline{\nu_i}}{z}\Big) \qquad (12)\end{aligned}$$

with $1 \leq a_i \leq 2$, $i = 1, \ldots, p_0$, $1 \leq b_i \leq 2$, $i = 1, \ldots, q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i > p_0 + q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0) \leq r$, and $r_1 \leq r - [\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0)]$.
Let also $\mathcal{D}_2$ be the subset of the unit sphere of H of functions orthogonal to $\mathcal{D}_1$ of

form (12) with $1 \leq a_i$, $i = 1, \ldots, p_0$, $1 \leq b_i$, $i = 1, \ldots, q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i > p_0 + q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0) \leq r$, and $r_1 \leq r - [\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0)]$.
Notice that $\mathcal{D}_1$ and $\mathcal{D}_2$ are subset of $\overline{\mathcal{D}}$, the (compact) closure of $\mathcal{D}$ in H. More precisely, $\overline{\mathcal{D}}$ is the subset of the unit sphere of H of functions of form (12) with $1 \leq a_i$, $i = 1, \ldots, p_0$, $1 \leq b_i$, $i = 1, \ldots, q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0) \leq r$, and $r_1 \leq r - [\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0 + q_0)]$. To identify the set $\overline{\mathcal{D}}$, one looks at the limit points of functions in $\mathcal{D}$ such that $N(\beta)$ tends to 0. This happens when at least one c_i tends to some t_i or u_i with corresponding γ_i tending to the corresponding τ_i, or to the corresponding μ_i.
Define now $U_n = \inf_{f \in F(\rho,u)} C_n(f)$. We have the following asymptotic result:

Theorem 4.1 *$U_n - C_n(f_0)$ converges in distribution to the following variable:*

$$-\sup\left\{\sup_{d\in\mathcal{D}} \frac{1}{2}\xi_d^2 1_{\xi_d \geq 0};\ \sup_{d_1\in\mathcal{D}_1, d_2\in\mathcal{D}_2} \frac{1}{2}(\xi_{d_1}^2 + \xi_{d_2}^2 1_{\xi_{d_2}\geq 0})\right\}$$

The proof of Theorem 4.1 is in section 5.

Remarks:

- An important point of the proof is the identification of $\overline{\mathcal{D}}$, that is the study of the limit points in $\mathcal{D}$ when $N(\beta)$ tends to 0. This is why the ϵ_i appears in (11), so that when $N(\beta)$ tends to 0, $f_{(\theta,\beta)}$ tends to f_0, and so that, due to the locally conic structure (coming from the choice of permutation):

$$\frac{\theta}{N(\beta)}|\mu_i|(1 + o(1)) \leq (|c_i - u_i|) \tag{13}$$

  for a c_i tending to u_i, for which $\epsilon_i = 0$, and

$$\frac{\theta}{N(\beta)}|\tau_i|(1 + o(1)) \leq (|c_i - t_i|) \tag{14}$$

  for a c_i tending to t_i, for which $\epsilon_i = 1$.

- In case $p_0 = q_0 = 0$ and $p = q = 1$, we recover the result of Hannan (1982).

- In case $p_0 + 1 = p$ and $q_0 + 1 = q$, and only in this case, $\mathcal{D}_2$ is empty so that we recover the asymptotic result of Veres (1987).

Corollary 4.2 *When $p = p_0 + 1$ and $q = q_0 + 1$, $U_n - C_n(f_0)$ converges in distribution to the following variable:*

$$-\sup_{d\in\mathcal{D}} \frac{1}{2}\xi_d^2 1_{\xi_d \geq 0}$$

5 Proofs.

5.1 Proof of Theorem 3.1.

First of all, define $\widehat{\theta}_\beta$ a maximizer of $l_n(\theta,\beta)$ for the fixed value of β. Recall the following result, which is proved in Dacunha-Castelle and Gassiat(1996) coming from **(M0)** and the fact that the parametrization is locally conic:

Proposition 5.1 *Define $\eta_n = \sup_{\beta\in\tilde{\mathcal{B}}} \widehat{\theta}_\beta$. Then η_n converges to 0 in probability as n tends to infinity.*

Using **(P0)**, $g_{(\theta,\beta)}$ is differentiable with respect to θ till order 5, and we have for all $2\le h\le 5$, if $g^{(h)}_{(\theta,\beta)}$ is the h-th derivative of g with respect to θ at point (θ,β):

$$\begin{aligned} g^{(h)}_{(\theta,\beta)} &= \frac{h}{N(\beta)^h}\sum_{i_1\dots i_{h-1}=1}^{k}\sum_{l=1}^{q}\rho_l\delta^l_{i_1}\dots\delta^l_{i_{h-1}}D^{h-1}_{i_1\dots i_{h-1}}f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l} \\ &+\frac{1}{N(\beta)^h}\sum_{i_1\dots i_h=1}^{k}\sum_{l=1}^{q}(\pi^0_l+\rho_l\frac{\theta}{N(\beta)})\delta^l_{i_1}\dots\delta^l_{i_h}D^h_{i_1\dots i_h}f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l}\end{aligned}$$

and also

$$g'_{(\theta,\beta)} = \frac{1}{N(\beta)}\left(\sum_{l=1}^{p-q}\lambda_l f_{\gamma^l} + \sum_{l=1}^{q}\rho_l f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l} + \sum_{i=1}^{k}\sum_{l=1}^{q}(\pi^0_l+\rho_l\frac{\theta}{N(\beta)})\delta^l_i D^1_i f_{\gamma^{l,0}+\frac{\theta}{N(\beta)}\delta^l}\right)$$

The proof of the Theorem follows the same lines as that of Theorems 4.2 and 4.3 in Dacunha-Castelle and Gassiat (1996). However, the parametrization is not exactly the same, and has more terms, so that we detail the proof. It will aslo rely on some control's Lemma that we set now.

Lemma 5.2 *Under **(P2)**, there exists a constant number a such that for β in $\tilde{\mathcal{B}}$:*

$$\frac{\sup_l \|\delta^l\|^2}{N(\beta)} \le a$$

Proof of Lemma 5.2.
First of all, using assumption **(P2)**, the hermitian matrix of all hilbertian scalar products involving functions in the free system is positive, so that it has a positive smallest eigenvalue σ, and the associated hermitian product is larger than σ multiplied by the usual scalar product in $\mathbb{R}^s$, $s=p+kq+p_2(k^2-1)$.
If $\frac{\|\delta^l\|^2}{N(\beta)}$ is unbounded, there exists a sequence β_n such that

$$\lim_{n\to+\infty}\frac{\|\delta^{l,n}\|}{N(\beta_n)} = +\infty$$

Since $\delta^{l,n}$ is bounded (see (6)), this implies that $N(\beta_n)$ tends to 0. Using assumption **(P2)**, this implies that $\delta^{l,n}$ tends to 0. Let now J_l be the set of indices i such that $\gamma^{i,n}$ tends to $\gamma^{l,0}$. We have:

$$\frac{\|\delta^{l,n}\|^2}{N(\beta_n)} \leq \frac{1}{\sigma} \frac{\|\delta^{l,n}\|^2}{(\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})+\delta^l\|^2+\frac{1}{4}(\sum_{j\in J_l}\lambda_j\|\gamma^{j,n}-\gamma^{l,0})\|^2)^2)^{1/2}(1+o(1))}$$

Now,

- If $\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})\| = o(\|\delta^l\|)$ or $\|\delta^l\| = o(\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})\|)$, obviously
$$\frac{\|\delta^{l,n}\|^2}{N(\beta_n)} \leq \frac{1}{\sigma}$$

- If now $\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0}) = -\delta^l(1+o(1))$ we have
$$\sum_{j\in J_l}\lambda_j\|\gamma^{j,n}-\gamma^{l,0}\|^2 \geq \frac{\|\sum_{j\in J_l}\lambda_j(\gamma^{j,n}-\gamma^{l,0})\|^2}{(\sum_{j\in J_l}\lambda_j)^2}$$
so that
$$\sum_{j\in J_l}\lambda_j\|\gamma^{j,n}-\gamma^{l,0}\|^2 \geq \frac{\|\delta^l\|^2(1+o(1))}{(\sum_{j\in J_l}\lambda_j)^2}$$
and so that again
$$\frac{\|\delta^{l,n}\|^2}{N(\beta_n)} \leq \frac{(\sum_{j\in J_l}\lambda_j)^2}{\sigma} \leq \frac{p-q}{\sigma}$$

and the Lemma follows.

The proof of Theorem 3.1 will rely on covering the domain in different regions. Let us look first at
$$A_n = \{\beta \ : \ \frac{\|\sup_l \delta^l\|}{N(\beta)^2} \leq \frac{1}{\eta_n^\alpha}\}$$
for some $\alpha < 3/4$. A first Lemma states the asymptotic distribution of the likelihood maximized on (θ,β) for $\beta \in A_n$.

Lemma 5.3 *Under the assumptions of Theorem 3.1,* $\sup_{\beta\in A_n} l_n(\widehat{\theta}_\beta,\beta) - l_n(0)$ *converges in distribution to*
$$\frac{1}{2}\sup_{d\in\mathcal{D}}(\xi_d)^2 \cdot 1_{\xi_d} \geq 0$$

Proof of lemma 5.3.
First, the following expansion holds for θ tending to 0:

$$l_n(\theta,\beta) - l_n(0) = \sum_{i=1}^n \frac{g_{(\theta,\beta)}-g_0}{g_0}(X_i) - \frac{1}{2}\sum_{i=1}^n\left(\frac{g_{(\theta,\beta)}-g_0}{g_0})^2(X_i) + O(\frac{g_{(\theta,\beta)}-g_0}{g_0}(X_i))^3\right) \tag{15}$$

Let us now write an expansion of $g_{(\theta,\beta)}$ till order 2:

$$g_{(\theta,\beta)}(x) = g_0(x) + \theta \cdot g'_{(0,\beta)}(x) + \frac{\theta^2}{2} \cdot g"_{(\theta*,\beta)}(x)$$

for a $\theta* \leq \theta$ and depending on x. Now as θ tends to 0:

$$g"_{(\theta*,\beta)}(x) = \frac{2}{N(\beta)^2} \sum_{l=1}^{q} \sum_{i=1}^{k} \rho_l \delta_i^l D_i^1 f_{\gamma^{l,0}}(x) + 0((\sup_l \frac{\|\delta^l\|^2}{N(\beta)^3})\theta m_2(x) g_0(x))$$

since δ^l is bounded and using **(P0)**. Write:

$$D_n(\beta) = \sum_{i=1}^{n} \frac{g'_{(0,\beta)}}{g_0}(X_i)$$

$$F_n^{i,l} = \sum_{i=1}^{n} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}(X_i)$$

Define also

$$u(\beta, i, l) = \langle \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}, \frac{g'_{(0,\beta)}}{g_0} \rangle_H$$

$$v(i, l; i', l') = \langle \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}, \frac{D_{i'}^1 f_{\gamma^{l',0}}}{g_0} \rangle_H$$

Notice that

$$\sum_{i=1}^{n} \left(\frac{g'_{(0,\beta)}}{g_0} \right)^2 (X_i) = n \cdot (1 + o(1))$$

$$\sum_{i=1}^{n} \left(\frac{g'_{(0,\beta)}}{g_0} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} \right) (X_i) = n u(\beta, i, l) \cdot (1 + o(1))$$

$$\sum_{i=1}^{n} \left(\frac{D_{i'}^1 f_{\gamma^{l',0}}}{g_0} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} \right) (X_i) = n v(i, l; i', l') \cdot (1 + o(1))$$

where the $o(1)$ are uniform in probability, thanks to **(P1)**. Let us now see what happens on A_n and for $\theta \leq 2\eta_n$. Applying lemma 5.2 we obtain for any $l = 1, \ldots, q$:

$$\frac{\|\delta^l\|^2}{N(\beta)^3} \theta \leq \eta_n^{1-4\alpha/3}$$

which goes to 0 since $\alpha < 3/4$. One can now prove:

$$\begin{aligned} l_n(\theta,\beta) - l_n(0) &= \theta D_n(\beta) + \sum_{l=1}^{q} \sum_{i=1}^{k} \rho_l \frac{\delta_i^l}{N(\beta)^2} F_n^{i,l} \theta^2 + o(n\theta^2) \\ &\quad - \frac{\theta^2}{2} n(1 + o(1)) - \sum_{l=1}^{q} \sum_{i=1}^{k} \rho_l \frac{\delta_i^l}{N(\beta)^2} \theta^3 n u(\beta, i, l)(1 + o(1)) \\ &\quad - \frac{\theta^4}{2} \sum_{l,l'=1}^{q} \sum_{i,i'=1}^{k} \rho_l \rho_{l'} \frac{\delta_i^l \delta_{i'}^{l'}}{N(\beta)^4} n v(l, i; l', i')(1 + o(1)) + o(n\theta^2) \end{aligned}$$

where all the $o(\cdot)$ are uniform in probability over β in A_n.Now,

$$\frac{\|\delta^l\|}{N(\beta)^2}\theta \leq \eta_n^{1-\alpha}$$

and since $(D_n(\beta), F_n^{i,l})_{i,l}/\sqrt{n}$ converges uniformly in distribution using **(P1)** we have easily

$$\frac{\delta_i^l}{N(\beta)^2}F_n^{i,l}\theta^2 = o(\theta D_n(\beta))$$

where the $o(\cdot)$ is uniform in probability over β in A_n. We finally get for β in A_n and for $\theta \leq 2\eta_n$:

$$l_n(\theta,\beta) - l_n(0) = \left(\theta D_n(\beta) - \frac{\theta^2}{2}n\right)(1+o(1))$$

where again the $o(\cdot)$ is uniform in probability over β in A_n. Since $\widehat{\theta}_\beta \leq \eta_n$ this obviously leads, by maximizing $\theta D_n(\beta) - \frac{\theta^2}{2}n$ to:

$$V_n(\beta) = \frac{1}{2}\frac{D_n(\beta)^2}{n}1_{D_n(\beta)\geq 0}(1+o(1))$$

for β in A_n and where the $o(\cdot)$ is uniform in probability over β in A_n. The conclusion of Lemma 5.3 follows using **(P1)** and the fact that the set of functions in $\mathcal{D}$ such that β is in $\cup_n A_n$ is exactly $\mathcal{D}$.

Let us now study what happens on

$$B_n = \{\beta \;:\; \exists l = 1,\ldots,q,\; \frac{\|\delta^l\|}{N(\beta)^2} \geq \frac{1}{\eta_n^\alpha}\}$$

First of all, notice that on B_n, $N(\beta)$ tends to 0, and using **(P2)** all δ^l tend to 0. We have as an immediate consequence of Lemma 5.2:

Lemma 5.4 *There exists a constant number M such that for β in B_n:*

$$N(\beta) \leq M\eta_n^{2\alpha/3}$$

and for any (i,l):

$$|\delta_i^l| \leq M\eta_n^{\alpha/3}$$

Proof of Lemma 5.4.
Define $\phi_{i,l} = \frac{\delta_i^l}{N(\beta)^2}$. We first have, using Lemma 5.2

$$N(\beta) \leq \frac{a^{1/3}}{|\phi_{i,l}|^{2/3}}$$

Now, on B_n, there exists (i,l) such that $|\phi_{i,l}| \geq 1/k\eta_n^\alpha$, and the first inequality of the Lemma follows. Now, using Lemma 5.2 we have for all (i,l)

$$|\delta_i^l| \leq \sqrt{aN(\beta)}$$

and the second inequality follows .
In other words, on B_n, $N(\beta)$ and all $|\delta_i^l|$ tend uniformly to 0.

We shall use again expansion (15), but the expansion for $g_{(\theta,\beta)}$ has now to be done till order 5:

$$g_{(\theta,\beta)}(x) = g_0(x) + \theta \cdot g'_{(0,\beta)}(x) + \sum_{i=2}^{4} \frac{\theta^i}{i!} \cdot g^{(i)}_{(0,\beta)}(x) + \frac{\theta^5}{5!} \cdot g^{(5)}_{(\theta*,\beta)}(x)$$

for a $\theta* \leq \theta$ and depending on x. The aim is now to prove that for $\theta \leq 2\eta_n$ and for $\beta \in B_n$ we have:

$$l_n(\theta,\beta) - l_n(0) = P_n(\theta,\beta)(1+o(1)) \tag{16}$$

where all the $o(\cdot)$ are uniform in probability over β in B_n and with $P_n(\theta,\beta)$ the polynomial of degree 4 in the variable θ:

$$\begin{aligned} P_n(\theta,\beta) &= \theta D_n(\beta) + \sum_{l=1}^{q} \theta^2 \left(\sum_{i=1}^{k} \rho_l \frac{\delta_i^l}{N(\beta)^2} F_n^{i,l} - \frac{n}{2} \right) - \theta^3 n \sum_{l=1}^{q} \sum_{i=1}^{k} \rho_l \frac{\delta_i^l}{N(\beta)^2} u(\beta,i,l) \\ &\quad - \frac{\theta^4}{2} n \sum_{l,l'=1}^{q} \sum_{i,i'=1}^{k} \rho_l \rho_{l'} \frac{\delta_i^l \delta_{i'}^{l'}}{N(\beta)^4} v(i,l;i',l') \\ &= \sum_{j=1}^{4} p_j(n,\beta)\theta^j \end{aligned}$$

From now on, any $o(1)$ will be uniform over $B_n(J_1,J_2)$. Since $|\delta_i^l|$ tends uniformly to 0, we have

$$\frac{\delta_{i_1}^l \dots \delta_{i_h}^l}{N(\beta)^h} = o\left(\frac{\delta_{i_1}^l \dots \delta_{i_{h-1}}^l}{N(\beta)^h} \right)$$

So that we can write for for $h \leq 4$:

$$g^{(h)}_{(\theta,\beta)} = \left(\frac{h}{N(\beta)^h} \sum_{i_1 \dots i_{h-1}=1}^{k} \sum_{l=1}^{q} \rho_l \delta_{i_1}^l \dots \delta_{i_{h-1}}^l D^{h-1}_{i_1 \dots i_{h-1}} f_{\gamma^{l,0} + \frac{\theta}{N(\beta)}\delta^l} \right) (1+o(1))$$

Define

$$E_n(\theta,\beta) = \sum_{i=1}^{n} \sum_{h=2}^{5} \frac{\theta^h}{(h-1)!} \sum_{i_1 \dots i_{h-1}=1}^{k} \sum_{l=1}^{q} \rho_l \frac{\delta_{i_1}^l \dots \delta_{i_{h-1}}^l}{N(\beta)^h} \frac{D^{h-1}_{i_1 \dots i_{h-1}} f_{\gamma^{l,0}}}{g_0}(X_i)$$

$$J_n(\theta,\beta) = \sum_{i=1}^n \frac{\sup \|\delta^l\|^5}{N(\beta)^5}\theta^5 \frac{m_5}{g_0}(X_i))$$

We may now write, using also **(P0)**:

$$\sum_{i=1}^n \frac{g_{(\theta,\beta)} - g_0}{g_0}(X_i) = \theta D_n(\beta) + E_n(\theta,\beta)(1+o(1)) + O(J_n(\theta,\beta))$$

Now, using expansion (15) together with the previous result leads to

$$\begin{aligned} l_n(\theta,\beta) - l_n(0) &= \theta D_n(\beta) + E_n(\theta,\beta)(1+o(1)) + O(J_n(\theta,\beta)) \\ &\quad -\frac{1}{2}\left(\theta D_n(\beta) + E_n(\theta,\beta)(1+o(1)) + O(J_n(\theta,\beta))\right)^2 \\ &\quad +O(\theta D_n(\beta) + E_n(\theta,\beta)(1+o(1)) + O(J_n(\theta,\beta)))^3 \end{aligned}$$

so that we obtain

$$\begin{aligned} l_n(\theta,\beta) - l_n(0) &= \theta D_n(\beta) + E_n(\theta,\beta) - \frac{1}{2}(\theta D_n(\beta) + E_n(\theta,\beta))^2 + R_n && (17) \\ &= P_n(\theta,\beta) + R_n && (18) \end{aligned}$$

where R_n is a sum of terms which are $o(Q_n(\theta,\beta))$, with $Q_n(\theta,\beta) = \sup_{1\leq j\leq 4} |p_j(n,\beta)\theta^j|$, plus terms which may be bounded with one of the following forms:

$$\frac{\theta^h \delta^l_{i_1}\dots\delta^l_{i_{h-1}}}{N(\beta)^h}\sum_{i=1}^n \frac{D^{h-1}_{i_1\dots i_{h-1}} f_{\gamma^{l,0}}}{g_0}(X_i) \quad h\geq 3$$

$$\frac{\theta^5\|\delta^l\|^5}{N(\beta)^5}n \ ; \ \frac{\theta^{h+1}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}}{N(\beta)^h}n \ ; \ \frac{\theta^{h+l}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}}{N(\beta)^{h+l}}n \quad with\ h,l\geq 3$$

$$\theta^3 n \ ; \ \frac{\theta^{h+2}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}}{N(\beta)^h}n \ ; \ \frac{\theta^{h+l+1}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}}{N(\beta)^{h+l}}n$$

$$\frac{\theta^{h+l+m}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}\delta^l_{k_1}\dots\delta^l_{k_{m-1}}}{N(\beta)^{h+l+m}}n \quad with\ h,l,m\geq 2$$

Now, since $\theta/N(\beta)$ and the δ^l_i are bounded, the first term in this list may be bounded by:

$$M\eta_n^{\alpha/3}\cdot\frac{\theta^2\delta^l_{i_1}}{N(\beta)^2}\sum_{i=1}^n\frac{D^1_i f_{\gamma^0}}{g_0}(X_i)$$

which is uniformly in probability

$$o(\frac{\theta^2\delta^l_{i_1}}{N(\beta)^2}F_n^{i_1,l})$$

Some of the other terms will be proven to be $o(n\theta^2)$ using Lemma 5.2, Lemma 5.4 and the fact that β is in B_n:

$$\frac{\theta^5\|\delta^l\|^5}{N(\beta)^5}n = O(n\theta^2\eta_n^{\alpha/3})$$

$$\frac{\theta^{h+1}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}}{N(\beta)^h}n = O(n\theta^2\eta^{\alpha(k-3)/3})\ k \geq 4$$

$$\frac{\theta^{h+l}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}}{N(\beta)^{h+l}}n = O(n\theta^2\eta^{\alpha(k+l-6)/3})\ \ k,l \geq 3$$

$$\theta^3 n = o(n\theta^2)$$

$$\frac{\theta^{h+2}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}}{N(\beta)^h}n = O(n\theta^2\eta_n^{(h-1)/3})$$

$$\frac{\theta^{h+l+1}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}}{N(\beta)^{h+l}}n = O(n\theta^2\eta_n^{\alpha(h+l-4/3))})\ \ k,l \geq 2$$

$$\frac{\theta^{h+l+m}\delta^l_{i_1}\dots\delta^l_{i_{h-1}}\delta^l_{j_1}\dots\delta^l_{j_{l-1}}\delta^l_{k_1}\dots\delta^l_{k_{m-1}}}{N(\beta)^{h+l+m}}n = O(n\theta^2\eta_n^{\alpha(h+l+m-7)/3})\ \ k+l+m \geq 8$$

The remaining terms may be proven to be $o(n\frac{\theta^4\delta^l_i\delta^l_j}{N(\beta)^4})$ for some i,j. They are:

$$\frac{n\theta^4\delta^l_i\delta^l_j}{N(\beta)^3} = O(N(\beta)\cdot n\frac{\theta^4\delta^l_i\delta^l_j}{N(\beta)^4})$$

$$\frac{n\theta^5\delta^l_i\delta^l_j}{N(\beta)^4} = O(\theta\cdot n\frac{\theta^4\delta^l_i\delta^l_j}{N(\beta)^4})$$

$$\frac{n\theta^6\delta^l_i\delta^l_j\delta^l_{j'}}{N(\beta)^6} = O(\eta_n^{\alpha/3}\cdot n\frac{\theta^4\delta^l_j}{N(\beta)^4})$$

$$\frac{n\theta^7\delta^l_i\delta^l_j\delta^l_{i'}\delta^l_{j'}}{N(\beta)^7} = O(\eta_n^{2\alpha/3}\cdot n\frac{\theta^4\delta^l_j}{N(\beta)^4})$$

So every term of R_n is $o(Q_n(\theta,\beta))$ uniformly in β. It is not possible now to conclude that (16) holds since we need $o(P_n(\theta,\beta))$ instead of $o(Q_n(\theta,\beta))$. It will be seen later that, at the optimizing value $(\widehat{\theta},\widehat{\beta})$, all terms in P_n have the same order and that $Q_n(\theta,\beta) = O(P_n(\widehat{\theta},\widehat{\beta}))$. Let us now prove that when at least one term $p_j(n,\beta)\theta^j$ of P_n tends to $+\infty$, then this implies that $P_n(\theta,\beta)$ is negative. Recall that $\phi_{i,l} = \frac{\delta^l_i}{N(\beta)^2}$. We can limit ourselves to subsequences; we omit the related subscripts for simplicity.
We shall need two technical Lemmas. The first one is a simple consequence of **(P2)**.

Lemma 5.5 *There exists $\tau > 0$ such that:*

$$\sum_{l,l'=1}^{q}\sum_{i,i'=1}^{k}\rho_l\rho_{l'}\phi_{i,l}\phi_{i',l'}nv(i,l;i',l') \geq \tau\sum_{l=1}^{q}\sum_{i=1}^{k}\rho_l^2\phi_{i,l}^2 nv(i,l;i,l)$$

Let now $J \subset \{1,\dots,q\}\times\{1,\dots,k\}$ be the set of indices (l,i) such that $\theta\phi_{i,l}$ is bounded, and $\tilde{J} \subset J$ be the set of indices (l,i) such that the accumulation point of $\theta\phi_{i,l}$ is non zero; let $\alpha_{i,l}$ be this accumulation point.

Lemma 5.6 *Suppose that $J = \{1, \dots, q\} \times \{1, \dots, k\}$. Then*

$$\forall j = 1, \dots, 4, \ \limsup_{n \to +\infty} p_j(n, \beta)\theta^j = O(1)$$

Moreover, there exists $m > 0$ such that at the maximizing value $P(|p_j(n, \widehat{\beta})\widehat{\theta}^j| \geq m)$ tends to 1.

Indeed, in this case we have

$$P_n(\theta, \beta) = \theta\sqrt{n}A_1(n, \beta) + (\theta\sqrt{n}A_2(n, \beta) - \frac{1}{2}n\theta^2) + \theta^2 nB_1(n, \beta) - \frac{1}{2}n\theta^2 B_2(n, \beta)$$

with $A_1 = O(1)$, $A_2 = O(1)$, $B_1 = O(1)$, $B_2 = O(1)$. Optimizing P_n leads to the maximum value:

$$\frac{1}{2}\frac{(A_1 + A_2)^2}{(B_2 - B_1 + 1)^2}$$

and Schwartz inequality leads to the result.
Now

- If $p_1(n, \beta)\theta$ tends to infinity, then since $D_n(\beta) = O(\sqrt{n})$, $p_1(n, \beta)\theta = o(n\theta^2)$ and $\sum_{l=1}^{q}\sum_{i=1}^{k}\rho_l\phi_{i,l}F_n^{i,l}\theta^2 = o(p_4(n, \beta)\theta^4)$ and $p_4(n, \beta) > 0$. So the only possibility to have $P_n > 0$ is that $|p_4(n, \beta)\theta^4| = o(|p_3(n, \beta)\theta^3|)$, with $|p_4(n, \beta)\theta^4|$ and $|p_3(n, \beta)\theta^3|$ tending to infinity. But this implies by Lemma 5.5 that no term $\theta\phi_{i,l}$ tends to infinity. So that, either $\tilde{J}$ is empty and $|p_3(n, \beta)\theta^3| = o(n\theta^2$, and P_n is negative for large enough n, or by Lemma 5.6, the maximum value of P_n is $O(1)$, doesn't tend to 0, and no term of P_n tends to 0.

- If $|p_2(n, \beta)\theta^2|$ tends to infinity, the same analysis holds.

- If $|p_3(n, \beta)\theta^3|$ tends to infinity, then in case there exists some leading term $\theta\phi_{i_0,l_0}$ tending to infinity, $|p_3(n, \beta)\theta^3| = o(|p_4(n, \beta)\theta^4|)$. In the other case, if J is empty, $|p_3(n, \beta)\theta^3| = o(n\theta^2)$ and P_n is negative. The case where J is not empty has been treated already.

In summary, if one of the $|p_j(n, \beta)\theta^j|$ tends to infinity, either the maximum value of P_n is negative for large enough n, or the maximum value of P_n is $O(1)$, doesn't tend to 0, and no term of P_n tends to 0. We may conclude that the supremum value of l_n is attained in the region where all terms of P_n are $O(1)$, where (16) holds.

Let us now look at the maximum value of $P_n(\theta, \beta)$ for β in B_n. Let $\phi = \frac{1}{N(\beta)^2}$.

Define, considering ϕ and β as different variables (in fact ϕ is a function of β):

$$\begin{aligned} Y_n(\beta,\phi,\theta) &= \theta D_n(\beta) - \frac{n}{2}\theta^2 + (\theta^2 \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l F_n^{i,l} - \theta^3 n \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l u(\beta,i,l)) \cdot \phi \\ &\quad - \frac{\theta^4}{2} n \sum_{l,l'=1}^{q} \sum_{i,i'=1}^{k} \rho_l \rho_{l'} \delta_i^l \delta_{i'}^{l'} v(i,l,i',l') \cdot \phi^2 \end{aligned}$$

We have $P_n(\theta,\beta) = Y_n(\beta,\phi,\theta)$, so that

$$\sup_{\theta \leq \eta_n, \beta \in B_n} P_n(\theta,\beta) \leq \sup_{\beta \in B_n} Z_n(\beta)$$

where $Z_n(\beta) = \sup_{\phi,\theta \geq 0} Y_n(\beta,\phi,\theta)$. Optimizing in ϕ leads to:

$$\phi = \frac{1}{n\theta^2} \frac{\sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l F_n^{i,l}}{\sum_{l,l'=1}^{q}\sum_{i,i'=1}^{k} \rho_l \rho_{l'} \delta_i^l \delta_{i'}^{l'} v(i,l,i',l')} - \frac{1}{\theta} \frac{\sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l u(\beta,i,l)}{\sum_{l,l'=1}^{q}\sum_{i,i'=1}^{k} \rho_l \rho_{l'} \delta_i^l \delta_{i'}^{l'} v(i,l,i',l')} \tag{19}$$

Define

$$F_n(\beta) = \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l F_n^{i,l}$$

$$S(\beta) = \sum_{l,l'=1}^{q}\sum_{i,i'=1}^{k} \rho_l \rho_{l'} \delta_i^l \delta_{i'}^{l'} v(i,l,i',l')$$

$$U(\beta) = \sum_{l=1}^{q}\sum_{i=1}^{k} \rho_l \delta_i^l u(\beta,i,l)$$

This leads to:

$$\begin{aligned} Y_n(\beta,\phi,\theta) &= \frac{1}{2n}\left(\frac{F_n(\beta)^2}{S(\beta)}\right) + \theta\left(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)}\right) \\ &\quad - \frac{n\theta^2}{2}\left(1 - \frac{U(\beta)^2}{S(\beta)}\right) \end{aligned}$$

which leads to the maximizing value for θ:

$$\theta = \frac{1}{n}\left(\frac{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)}}{1 - \frac{U(\beta)^2}{S(\beta)}}\right) 1_{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)} \geq 0}$$

The maximized value of Y_n is then:

$$\begin{aligned} Z_n(\beta) &= \frac{1}{2n}\left(\frac{(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)})^2}{1 - \frac{(U(\beta))^2}{S(\beta)}}\right) 1_{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)} \geq 0} \\ &\quad + \frac{1}{2n}\left(\frac{(F_n(\beta))^2}{S(\beta)}\right) \end{aligned}$$

Let us now study the behaviour of $Z_n(\beta)$ for $\beta \in B_n$.

Lemma 5.7 *Let $d(\beta)$ be the function in $\mathcal{D}$*

$$d(\beta) = \frac{1}{N(\beta)} \left(\sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\gamma^i}}{g_0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^{l,0}}}{g_0} \right)$$

On B_n, the set of possible accumulation points of $d(\beta)$ are the $d(\mu, \tilde{\rho}, \lambda, \tau, a)(\cdot)$ in the unit sphere of H:

$$\sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{q} \sum_{i=1}^{k} \lambda_{l,i} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{p_2} \sum_{i,j=1}^{k} \tau_l a_i a_j \frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0}$$

with $p_1 \leq p-q-1$, $p_1 + p_2 \leq p-q$, $\mu_l \geq 0$, $\sum_{l=1}^{p_1} \mu_l + \sum_{l=1}^{q} \rho_l = 0$, $\tau_l \geq 0$. On the subsequence,

$$d(\beta) = d(\mu, \tilde{\rho}, \lambda, \tau, a)(1 + o(1))$$

so that the same approximation holds for $u(\beta, i, l)$.

Proof of Lemma 5.7.
On B_n, $N(\beta)$ tends uniformly to 0. This implies that there exists an integer $p_1 \leq p-q-1$ such that for the indices (eventually reordered) $l \leq p_1$, the γ^l do not converge to any of the $\gamma^{l,0}$. For the other indices, let J_l be the (possibly empty but not all empty) set of indices m such that γ^m tends to $\gamma^{l,0}$. Writing a Taylor expansion, and keeping only the leading terms, we have

$$\begin{aligned} d(\beta) \;=\; & \Big(\frac{1}{N(\beta)}\Big(\sum_{l=1}^{p_1} \lambda_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \Big(\sum_{m \in J_l} \lambda_m + \rho_l\Big) \frac{f_{\gamma^{l,0}}}{g_0} \\ & + \sum_{l=1}^{q} \sum_{j=1}^{k} \Big(\sum_{m \in J_l} \lambda_m (\gamma_j^m - \gamma_j^{l,0}) + \pi_l^0 \delta_j^l\Big) \frac{D_j^1 f_{\gamma^{l,0}}}{g_0} \\ & + \sum_{l=1}^{p_2} \sum_{j,j'=1}^{k} \frac{1}{2} \Big(\sum_{m \in J_l} \lambda_m (\gamma_j^m - \gamma_j^{l,0})(\gamma_{j'}^m - \gamma_{j'}^{l,0})\Big) \frac{D_{jj'}^2 f_{\gamma^{l,0}}}{g_0}\Big)\Big)(1 + o(1)) \end{aligned}$$

for some p_2 with $p_1 + p_2 \leq p-q$. $N(\beta)$ has the same expansion. The sequences of coefficients

$$\frac{\lambda_l}{N(\beta)}, \quad \frac{\sum_{m \in J_l} \lambda_m + \rho_l}{N(\beta)}, \quad \frac{\sum_{m \in J_l} \lambda_m (\gamma_j^m - \gamma_j^{l,0}) + \pi_l^0 \delta_j^l}{N(\beta)}$$
$$\frac{\sum_{m \in J_l} \lambda_m (\gamma_j^m - \gamma_j^{l,0})(\gamma_{j'}^m - \gamma_{j'}^{l,0})}{N(\beta)}$$

are bounded. Let the accumulations points be respectively μ_l, $\tilde{\rho}_l$, $\lambda_{l,i}$ and $\tau_l a_i a_j$. The result follows.
We then have on B_n:

$$D_n(\beta) = \sum_{i=1}^{n} d(\mu, \tilde{\rho}, \lambda, \tau, a)(X_i)(1 + o(1)) \tag{20}$$

for some $d(\mu, \tilde{\rho}, \lambda, \tau, a)(\cdot)$. and also

$$u(\beta, i, l) = u(\mu, \tilde{\rho}, \lambda, \tau, a)(1 + o(1))$$

with obvious notations. We thus obtain that

$$\frac{1}{2n}\left(\frac{(F_n(\beta))^2}{S(\beta)}\right)$$

converges in distribution to $\frac{1}{2}\xi_{d_1}^2$ for some $d_1 \in \mathcal{D}_1$, and that jointly

$$\frac{1}{2n}\left(\frac{(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)})^2}{1 - \frac{(U(\beta))^2}{S(\beta)}}\right) 1_{D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)} \geq 0}$$

converges in distribution, uniformly in $(\mu, \tilde{\rho}, \lambda, \tau, a)$, to $\frac{1}{2}(\xi_{d_2})^2 \cdot 1_{\xi_{d_2}} \geq 0$ for some $d_2 \in \mathcal{D}_2$. Notice indeed that the variables $(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)}$ and $F_n(\beta)$ are uncorrelated. We finally have that $\sup_{\theta \leq \eta_n, \beta \in B_n} P_n(\theta, \beta)$ is upper bounded by the variable Z_n which converges in distribution to the variable

$$Z = \frac{1}{2}\sup_{d_1 \in \mathcal{D}_1}(\xi_{d_1})^2 + \frac{1}{2}\sup_{d_2 \in \mathcal{D}_2}(\xi_{d_2})^2 \cdot 1_{\xi_{d_2}} \geq 0$$

Let us now prove the lower bound. Let $\Lambda = (\mu, \tilde{\rho}, \lambda, \tau, a)$ be given such that $d_2(\Lambda) = d_2(\mu, \tilde{\rho}, \lambda, \tau, a)$ is in $\mathcal{D}_2$ with:

$$d_2(\Lambda) = \sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{q}\sum_{i=1}^{k} \lambda_{l,i} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{p_2}\sum_{i,j=1}^{k} \tau_l a_i a_j \frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0}$$

Let $\tilde{\Lambda} = (\tilde{\lambda}_{i,l})$, and $d_1(\tilde{\Lambda})$ in $\mathcal{D}_1$ given by

$$d_1(\tilde{\Lambda}) = \sum_{l=1}^{q}\sum_{i=1}^{k} \tilde{\lambda}_{l,i} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0}$$

For any sequence of ϵ tending to 0, define $\tilde{\beta}$ by:

$$\lambda_l = \mu_l \cdot \epsilon, \quad l = q+1, \ldots q + p_1$$

$$\lambda_l + \rho_l = \tilde{\rho}_l \cdot \epsilon, \quad l = 1, \ldots, q$$

if $p_1 + q \leq p - q$, and in case $\inf\{p_1 + q; p - q\} = p - q$,

$$\rho_l = \tilde{\rho}_l \cdot \epsilon, \quad l = p_1 + q + 1, \ldots, p - q\}$$

$$\lambda_l(\gamma_j^l - \gamma_j^{l,0}) + \pi_l^0 \delta_j^l = \lambda_{j,l} \cdot \epsilon, \quad l = 1, \ldots, q, \ j = 1, \ldots, k$$

$$\lambda_l(\gamma_i^l - \gamma_i^{l,0})(\gamma_j^l - \gamma_j^{l,0}) = \tau_l a_i a_j \cdot \epsilon, \quad l = 1, \ldots, p_2, \ i, j = 1, \ldots, k$$

$N(\tilde{\beta})$ has order ϵ. The parameters of $\tilde{\beta}$ are not uniquely determined by this set of equations. It is possible to find also a sequence of ζ tending to 0 such that:

$$\rho_l \delta_j^l = \tilde{\lambda}_{j,l} \cdot \zeta, \quad l = 1, \ldots, q, \ j = 1, \ldots, k$$

(just consider the case $a_i = 0$ and $a_i \neq 0$, to set an equation of order 2 in δ_j^l, and choose the speed ζ so that the equation has a solution for all j, l.) Choose now the speed ϵ of convergence to 0 of $N(\tilde{\beta})$ such that

$$\frac{1}{N(\tilde{\beta})^2} = \phi$$

with ϕ given in (19), where $\theta \leq \eta_n$. This is possible, since it is an homogeneous equation in ϵ. Since δ_j^l is at least of order ϵ and at most of order $\sqrt{\epsilon}$, we have that $\delta_j^l / N(\tilde{\beta})^2$ is at least of order $\frac{1}{\theta}$, so that

$$\tilde{\beta} \in B_n$$

In case

$$1_{D_n(\tilde{\beta}) - \frac{F_n(\tilde{\beta}) U(\tilde{\beta})}{S(\tilde{\beta})} \geq 0} = 1,$$

choose

$$\theta = \frac{1}{n} \left(\frac{D_n(\tilde{\beta}) - \frac{F_n(\tilde{\beta}) U(\tilde{\beta})}{S(\tilde{\beta})}}{1 - \frac{U(\tilde{\beta})^2}{S(\tilde{\beta})}} \right) 1_{D_n(\tilde{\beta}) - \frac{F_n(\tilde{\beta}) U(\tilde{\beta})}{S(\tilde{\beta})} \geq 0}$$

We have easily that $\theta / N(\tilde{\beta})$ is of order $n^{-1/4}$, so that we obtain that, using this particular $(\theta, \tilde{\beta})$,

$$\sup_{\theta \leq \eta_n, \beta \in B_n} P_n(\theta, \beta) \geq \sup_{\Lambda, \tilde{\Lambda}} Z_n(\tilde{\beta}) \tag{21}$$

(the supremum over Λ, $\tilde{\Lambda}$ is over all $\tilde{\beta}$ constructed from them.) In case

$$1_{D_n(\tilde{\beta}) - \frac{F_n(\tilde{\beta}) U(\tilde{\beta})}{S(\tilde{\beta})} \geq 0} = 0,$$

we obtain that

$$\sup_{\theta \leq \eta_n, \beta \in B_n} P_n(\theta, \beta) \geq \frac{1}{2n} \frac{F_n(\tilde{\beta})^2}{S(\tilde{\beta})} = Z_n(\tilde{\beta})$$

so that in all cases, (21) holds. Now, with those $\tilde{\beta}$ we have that

$$\frac{F_n(\tilde{\beta})}{\sqrt{S(\tilde{\beta})}} = \left(\sum_{i=1}^{n} d_1(\tilde{\Lambda})(X_i)\right)(1 + o(1))$$

and

$$\frac{(D_n(\beta) - \frac{F_n(\beta)U(\beta)}{S(\beta)})}{\sqrt{1-\frac{(U(\beta))^2}{S(\beta)}}} = (\sum_{i=1}^n d_2(\Lambda)(X_i))(1+o(1))$$

so that $\sup_{\Lambda,\tilde{\Lambda}} Z_n(\tilde{\beta})$ converges in distribution to

$$\frac{1}{2}\sup_{d_1\in\mathcal{D}_1}(\xi_{d_1})^2 + \frac{1}{2}\sup_{d_2\in\mathcal{D}_2}(\xi_{d_2})^2\cdot 1_{\xi_{d_2}\geq 0}$$

We then have proved the following Lemma:

Lemma 5.8 *Under the assumptions of Theorem 3.1,* $\sup_{\beta\in B_n} l_n(\widehat{\theta}_\beta,\beta) - l_n(0)$ *converges in distribution to*

$$\frac{1}{2}\sup_{d_1\in\mathcal{D}_1}(\xi_{d_1})^2 + \frac{1}{2}\sup_{d_2\in\mathcal{D}_2}(\xi_{d_2})^2\cdot 1_{\xi_{d_2}\geq 0}$$

which, together with Lemma 5.3, gives the Theorem.

5.2 Proof of Theorem 4.1.

First of all, define $\widehat{\theta}_\beta$ a minimizer of $C_n(f_{(\theta,\beta)})$ for the fixed value of β. We have, since the parametrization is locally conic, and using similar arguments as for Proposition 5.1:

Proposition 5.9 *Define* $\eta_n = \sup_{\beta\in\tilde{\mathcal{B}}}\widehat{\theta}_\beta$. *Then* η_n *converges to* 0 *in probability as* n *tends to infinity.*

We shall now study derivatives of $C_n(f_{(\theta,\beta)})$ with respect to θ for fixed β. Define

$$\begin{aligned} e_{(\theta,\beta)} &= \frac{1}{N(\beta)}\Big(\frac{\delta}{\sigma_0^2+\frac{\theta\delta}{N(\beta)}} + \sum_{i=1}^{p_0}\Big(\frac{\tau_i z}{1-(t_i+\frac{\theta}{N(\beta)}\tau_i)z} + \frac{\overline{\tau_i}}{z-(\overline{t_i}+\frac{\theta}{N(\beta)}\overline{\tau_i})}\Big) \\ &\quad -\sum_{i=1}^{q_0}\Big(\frac{\mu_i z}{1-(u_i+\frac{\theta}{N(\beta)}\mu_i)z} + \frac{\overline{\mu_i}}{z-(\overline{u_i}+\frac{\theta}{N(\beta)}\overline{\mu_i})}\Big) \\ &\quad -\sum_{i=1}^{r}\Big(\frac{\gamma_i z}{1-(c_i+\frac{\theta}{N(\beta)}\gamma_i)z} + \frac{\overline{\gamma_i}}{z-(\overline{c_i}+\frac{\theta}{N(\beta)}\overline{\gamma_i})}\Big) - \sum_{i=1}^{s}(\nu_i z+\frac{\overline{\nu_i}}{z})\Big) \end{aligned}$$

and $e^{(k)}_{(\theta,\beta)}$ to be the partial derivative of $e_{(\theta,\beta)}$ with respect to θ. $e_{(\theta,\beta)}$ is an element of $\mathcal{D}$. Let $C_n'(\theta,\beta)$ be the partial derivative of $C_n(f_{(\theta,\beta)})$ with respect to θ, and $C_n^{(k)}(\theta,\beta)$ the k-th partial derivative of $C_n(f_{(\theta,\beta)})$ with respect to θ. We have:

$$C_n'(\theta,\beta) = n\frac{\frac{\delta}{N(\beta)}}{\sigma_0^2+\theta\frac{\delta}{N(\beta)}} - I_n\left((\frac{e_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)$$

$$
\begin{aligned}
C_n^{"}(\theta,\beta) &= -n\frac{(\frac{\delta}{N(\beta)})^2}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^2} + I_n\left(\frac{e^2_{(\theta,\beta)} - e'_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)\\
C_n^{(3)}(\theta,\beta) &= n\frac{2(\frac{\delta}{N(\beta)})^3}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^3} + I_n\left(\frac{-e^3_{(\theta,\beta)} + 3e_{(\theta,\beta)}e'_{(\theta,\beta)} - e"_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)\\
C_n^{(4)}(\theta,\beta) &= -n\frac{6(\frac{\delta}{N(\beta)})^4}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^4} + I_n\left(\frac{e^4_{(\theta,\beta)} + 4e"_{(\theta,\beta)}e_{(\theta,\beta)} - 6e^2_{(\theta,\beta)}e'_{(\theta,\beta)} + 3(e'_{(\theta,\beta)})^2 - e^{(3)}_{(\theta,\beta)}}{f_{(\theta,\beta)}}\right)\\
C_n^{(5)}(\theta,\beta) &= n\frac{24(\frac{\delta}{N(\beta)})^5}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^5} + I_n(\frac{1}{f_{(\theta,\beta)}}(-e^5_{(\theta,\beta)} + 10e'_{(\theta,\beta)}e^3_{(\theta,\beta)} + 10e'_{(\theta,\beta)}e"_{(\theta,\beta)} - 15e_{(\theta,\beta)}(e'_{(\theta,\beta)})^2\\
&\quad -10e^2_{(\theta,\beta)}e"_{(\theta,\beta)} + 5e^{(3)}_{(\theta,\beta)}e_{(\theta,\beta)} - e^{(4)}_{(\theta,\beta)}))
\end{aligned}
$$

We also have for $k \geq 1$:

$$
\begin{aligned}
(-1)^k e^{(k)}_{(\theta,\beta)} &= \frac{1}{N(\beta)^{k+1}}[\sum_{i=1}^{p_0}((\frac{\tau_i z}{1-(t_i+\frac{\theta}{N(\beta)}\tau_i)z})^{k+1} + (\frac{\overline{\tau_i}}{z-(\overline{t_i}+\frac{\theta}{N(\beta)}\overline{\tau_i})})^{k+1})\\
&\quad -\sum_{i=1}^{q_0}((\frac{\mu_i z}{1-(u_i+\frac{\theta}{N(\beta)}\mu_i)z})^{k+1} + (\frac{\overline{\mu_i}}{z-(\overline{u_i}+\frac{\theta}{N(\beta)}\overline{\mu_i})})^{k+1})\\
&\quad -\sum_{i=1}^{r}((\frac{\gamma_i z}{1-(c_i+\frac{\theta}{N(\beta)}\gamma_i)z)^{k+1}} + (\frac{\overline{\gamma_i}}{z-(\overline{c_i}+\frac{\theta}{N(\beta)}\overline{\gamma_i})})^{k+1}) + \frac{k!(\frac{\delta}{N(\beta)})^{k+1}}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^{k+1}}]
\end{aligned}
$$

So that we may write

$$C_n^{(3)}(\theta,\beta) = n\frac{2(\frac{\delta}{N(\beta)})^3}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^3} + \frac{1}{N(\beta)^2}I_n\left(K_{(\theta,\beta)}\right)$$

$$C_n^{(4)}(\theta,\beta) = -n\frac{6(\frac{\delta}{N(\beta)})^4}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^4} + \frac{1}{N(\beta)^4}I_n\left(S_{(\theta,\beta)}\right)$$

$$C_n^{(5)}(\theta,\beta) = n\frac{24(\frac{\delta}{N(\beta)})^5}{(\sigma_0^2+\theta\frac{\delta}{N(\beta)})^5} + \frac{1}{N(\beta)^5}I_n\left(T_{(\theta,\beta)}\right)$$

We shall make use of a Lemma which is a consequence of Theorem 2.5 of Dahlhaus (1988). We prove it below.

Let $\mathcal{G}$ be the subset of H:

$$\mathcal{G} = \left\{\frac{e_{(0,\beta)}}{f_0}, \frac{N(\beta)^2 e'_{(0,\beta)}}{f_0}, \frac{e^2_{(0,\beta)}}{f_0}, K_{(0,\beta)} \;:\; \beta \in \tilde{\mathcal{B}}\right\} \cup \left\{T_{(\theta,\beta)}, S_{(\theta,\beta)} \;:\; (\theta,\beta) \in \mathcal{T}\right\}$$

Define also for any function g in $\mathcal{G}$

$$I(g) = \frac{1}{2\pi}\int_0^{2\pi} g(\lambda)f(\lambda)d\lambda$$

Let $\mathcal{X}$ be the set of bounded real functions on $\mathcal{G}$, equipped with the metric generated by the uniform norm $\|x\| = \sup|x(g)|$.

Lemma 5.10 *Let $E_n(g) = \frac{1}{\sqrt{n}} I_n(g) - \sqrt{n} I(g)$, where g is in $\mathcal{G}$. Let $(W(g))_{g \in \mathcal{G}}$ be the centered gaussian process with covariance the scalar product in the Hilbert space H. Then the empirical spectral process $(E_n(g))_{g \in \mathcal{G}}$ converges weakly on $\mathcal{X}$ to $(W(g))_{g \in \mathcal{G}}$.*

As for mixtures, the proof relies on separating $\mathcal{T}$ in two regions. Define again $\eta_n = \sup_\beta \widehat{\theta}_\beta$, which tends to 0 in probability. Let

$$A_n = \{(\theta, \beta) \in \mathcal{T} \ : \ \frac{\theta}{N(\beta)^2} \leq \eta_n^\alpha, \theta \leq 2\eta_n\}$$

$$B_n = \{(\theta, \beta) \in \mathcal{T} \ : \ \frac{\theta}{N(\beta)^2} \geq \eta_n^\alpha, \theta \leq 2\eta_n\}$$

for some $\alpha \in]0, 1[$.
Let us first study $C_n(\theta, \beta)$ on A_n. Taylor expansion till order 4 with integral remaining term, together with Lemma 5.10 lead to:

$$\begin{aligned} C_n(\theta, \beta) - C_n(0) &= -\theta\sqrt{n} E_n(\frac{e_{(0,\beta)}}{f_0})(1 + o(1)) + \frac{1}{2} n\theta^2(1 + o(1)) \\ &\quad - \frac{\theta^2}{2N(\beta)^2}\sqrt{n} E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})(1 + o(1)) \\ &\quad + \frac{\theta^3}{6N(\beta)^2} na(1 + o(1)) + \frac{\theta^4}{24N(\beta)^4} nb(1 + o(1)) \end{aligned}$$

where a and b are real numbers, and the $o(1)$ are uniform in probability over A_n. Now, on A_n we have

$$\frac{\theta^2}{2N(\beta)^2}\sqrt{n} E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0}) = O(\theta\sqrt{n} E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})\eta_n^\alpha) = o(1)$$

using Lemma 5.10, and

$$\frac{\theta^3}{6N(\beta)^2} na = O(n\theta^2\eta_n^{2\alpha}) = o(n\theta^2)$$

where the $o(.)$ are uniform in probability, so that we have

$$C_n(\theta, \beta) - C_n(0) = -\theta\sqrt{n} E_n(\frac{e_{(0,\beta)}}{f_0})(1 + o(1)) + \frac{1}{2} n\theta^2(1 + o(1))$$

When minimizing over θ this leads to

$$-\frac{1}{2} \frac{E_n(\frac{e_{(0,\beta)}}{f_0})^2}{n} 1_{E_n(\frac{e_{(0,\beta)}}{f_0}) \geq 0}(1 + o(1))$$

The set of functions $\frac{e_{(0,\beta)}}{f_0}$ is exactly $\mathcal{D}$ so that we have:

Lemma 5.11 *$\inf_{(\theta,\beta)\in A_n} C_n(\theta,\beta) - C_n(f_0)$ converges in distribution to the following variable:*

$$-\sup_{d\in\mathcal{D}} \frac{1}{2}\xi_d^2 1_{\xi_d\geq 0}$$

Let us now study what happens on B_n. As for the mixtures, a key point of the proof will be a control Lemma to be able to stop Taylor's expansion and to have uniformly $o(1)$ remaining terms. We set it below. We first state a Lemma to describe what happens when $N(\beta)$ tends to 0.

Lemma 5.12 *On B_n, any accumulation point of $e_{(0,\beta)}/f_0$ has the form*

$$\frac{1}{f_0} e(\delta,\tau,\mu,\gamma,c,\nu,a,b)$$

which is a point in the unit sphere of H, with

$$\begin{aligned} e(\delta,\tau,\mu,\gamma,c,\nu,a,b) &= \frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}\left(\sum_{k=1}^{a_i} \tau_{i,k}(\frac{z}{1-t_i z})^{a_i} + \overline{\tau_{i,k}}(\frac{1}{z-\overline{t_i}})^{a_i}\right) \\ &+ \sum_{i=1}^{q_0}\left(\sum_{k=1}^{b_i} \mu_{i,k}(\frac{z}{1-u_i z})^{b_i} + \overline{\mu_{i,k}}(\frac{1}{z-\overline{u_i}})^{b_i}\right) \\ &- \sum_{i=1}^{r_1}(\frac{\gamma_i z}{1-c_i z} + \frac{\overline{\gamma_i}}{z-\overline{c_i}}) - \sum_{i=1}^{s}(\nu_i z + \frac{\overline{\nu_i}}{z}) \end{aligned}$$

with $1\leq a_i$, $i=1,\ldots,p_0$, $1\leq b_i$, $i=1,\ldots,q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i > p_0+q_0$, $\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0+q_0) \leq r$, and $r_1 \leq r - [\sum_1^{p_0} a_i + \sum_1^{q_0} b_i - (p_0+q_0)]$.

Proof of Lemma 5.12.

Let us look at a precise expansion when $N(\beta)$ tends to 0. Let for any $i=1,\ldots,q_0$ $U(i)$ be the set of indices j such that c_j tends to u_i, and $c_j = u_i + \alpha_j(u_i)$, and for any $i=1,\ldots,p_0$ $T(i)$ be the set of indices j such that c_j tends to t_i, and $c_j = t_i + \alpha_j(t_i)$. Let also J be the complementary set of the union of all $U(i)$ and $T(i)$ in $\{1,\ldots,r\}$. We then have:

$$\begin{aligned} e_{(0,\beta)} &= \frac{1}{N(\beta)}[\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}(\frac{(\tau_i - \sum_{j\in T(i)}\gamma_j)z}{1-t_i z} + \frac{\overline{\tau_i} - \sum_{j\in T(i)}\overline{\gamma_j}}{z-\overline{t_i}}) \\ &+ \sum_{i=1}^{p_0}(\sum_{j\in T(i)}\gamma_j \sum_{h\geq 2}\frac{\alpha_j^{h-1}z^h}{(1-t_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1}}{(z-\overline{t_i})^h}) \\ &- \sum_{i=1}^{q_0}(\frac{(\mu_i - \sum_{j\in U(i)}\gamma_j)z}{1-u_i z} + \frac{\overline{\mu_i} - \sum_{j\in U(i)}\overline{\gamma_j}}{z-\overline{u_i}}) \\ &- \sum_{i=1}^{p_0}(\sum_{j\in U(i)}\gamma_j \sum_{h\geq 2}\frac{\alpha_j^{h-1}z^h}{(1-u_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1})}{(z-\overline{u_i})^h}) \\ &- \sum_{i\in J}(\frac{\gamma_i z}{1-(c_i+\frac{\theta}{N(\beta)}\gamma_i)z} + \frac{\overline{\gamma_i}}{z-(\overline{c_i}+\frac{\theta}{N(\beta)}\overline{\gamma_i})}) - \sum_{i=1}^{s}(\nu_i z + \frac{\overline{\nu_i}}{z})) \end{aligned}$$

and a similar expansion holds for $N(\beta)$, which then appears as a quadratic function of the perturbation parameters. The result comes then from looking at adherence values of the parameters

$$\frac{1}{N(\beta)}\frac{\delta}{\sigma_0^2} \quad \frac{(\tau_i - \sum_{j\in T(i)}\gamma_j)}{N(\beta)} \quad \sum_{j\in T(i)} \frac{\gamma_j \alpha_j^{h-1}}{N(\beta)}$$

$$\frac{(\mu_i - \sum_{j\in U(i)}\gamma_j)}{N(\beta)} \quad \sum_{j\in U(i)} \frac{\gamma_j \alpha_j^{h-1}}{N(\beta)}$$

and

$$\frac{\gamma_i}{N(\beta)},\ i\in J, \quad \frac{\nu_i}{N(\beta)}, i=1,\dots,s$$

Let us now state the control Lemma.

Lemma 5.13 *On B_n we have*

$$N(\beta) \leq 2\eta_n^{(1-\alpha)/2}$$

and

$$\frac{\theta}{N(\beta)} \leq M\eta_n^{(1-\alpha)/2r}$$

for some constant number M.

First of all, since on B_n, $\theta \leq 2\eta_n$, we have

$$N(\beta)^2 \leq 2\eta_n^{(1-\alpha)}$$

and the first inequality follows. Now, let us study what happens when $N(\beta)$ tends to 0. There must be at least one c_i tending to some u_i with corresponding γ_i tending to corresponding μ_i, or (and) some c_i tending to some t_i with corresponding γ_i tending to corresponding τ_i. In each case, we have (13) or (14). Now, looking at the expansion of $N(\beta)$ near 0, it appears that the leading term is at least of order $\min_i(|c_i-u_i|^r, |c_i-t_i|^r)$, (recall that $r=\min(p-p_0, q-q_0)$) so that

$$\frac{\theta}{N(\beta)} = 0(N(\beta)^{1/r})$$

which, when combined with the first inequality, leads to the second inequality. .

On B_n, we shall write Taylor expansion till order 5 again with integral remaining term, so that, when using Lemma 5.10, we obtain:

$$\begin{aligned} C_n(\theta,\beta) - C_n(0) &= -\theta\sqrt{n}E_n(\frac{e_{(0,\beta)}}{f_0})(1+o(1)) + \frac{1}{2}n\theta^2(1+o(1)) \\ &\quad -\frac{\theta^2}{2N(\beta)^2}\sqrt{n}E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})(1+o(1)) \\ &\quad +\frac{\theta^3}{6N(\beta)^2}na(1+o(1)) + \frac{\theta^4}{24N(\beta)^4}nb(1+o(1)) + O(\frac{\theta^5}{N(\beta)^5}n)(1+o(1)) \end{aligned}$$

where the $o(1)$ are in probability uniform over B_n,

$$a = I(K_{(0,\beta)}) + \frac{2(\frac{\delta}{N(\beta)})^3}{\sigma_0^6}$$

and

$$b = I(S_{(0,\beta)}) + \frac{6(\frac{\delta}{N(\beta)})^4}{\sigma_0^8}$$

All $o(1)$ will be now uniform over B_n using Lemma 5.13. We have

$$\frac{\theta^5}{N(\beta)^5}n = o(\frac{\theta^4}{N(\beta)^4}n)$$

Notice that the functions $\frac{e'_{(0,\beta)}}{f_0}$, after normalization, are in $\mathcal{D}_1$.
Notice also that on B_n we have, for some $e(\delta,\tau,\mu,\gamma,c,\nu,a,b)$ as in Lemma 5.12,

$$\frac{e_{(0,\beta)}}{f_0} = \frac{1}{f_0}e(\delta,\tau,\mu,\gamma,c,\nu,a,b)(1+o(1)) \tag{22}$$

Notice also that on B_n since $N(\beta)$ tends uniformly to 0 we have:

$$a = 3I(\frac{e_{(0,\beta)}N(\beta)^2 e'_{(0,\beta)}}{f_0})(1+o(1))$$

and

$$b = 3I(\frac{(N(\beta)^2 e'_{(0,\beta)})^2}{f_0})(1+o(1))$$

Define now $B_n(\delta,\tau,\mu,\gamma,c,\nu,a,b)$ the subset of B_n where (22) holds. We now minimize $C_n(\theta,\beta) - C_n(0)$ over $B_n(\delta,\tau,\mu,\gamma,c,\nu,a,b)$, the reason why it is sufficient is the same as that for the mixtures by a careful study of the leading terms in the expansion, and the optimum value in the expansion. For $\phi = 1/N(\beta)^2$ the polynomial to be minimized is

$$P(\theta,\phi) = -\theta\sqrt{n}W_1^n + +\frac{1}{2}n\theta^2 - \frac{\theta^2}{2}\phi\sqrt{n}W_2^n + \frac{\theta^3}{2}n\phi C_{12} + \frac{\theta^4}{8}n\phi^2 C_{22}$$

with

$$W_1^n = E_n(\frac{e_{(0,\beta)}}{f_0}) \ \ W_2^n = E_n(\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0})$$

and, up to a factor $1+o(1)$, C_{12} is the covariance of W_1^n and W_2^n, C_{22} the variance of W_2^n, W_1^n being of unit variance.
We shall minimize it over ϕ and then θ. This leads to a minimum value, which is reachable for some β in $B_n(\delta,\tau,\mu,\gamma,c,\nu,a,b)$ giving the minimizing value of ϕ.
Minimizing over ϕ leads to

$$\phi = \frac{1}{C_{22}}\left(\frac{2W_2^n}{\sqrt{n}\theta^2} - \frac{2C_{12}}{\theta}\right)$$

with minimum value of the polynomial

$$-\frac{1}{2}\frac{(W_2^n)^2}{C_{22}} - \theta\sqrt{n}\left(W_1^n - W_2^n\frac{C_{12}}{C_{22}}\right) + \frac{\theta^2}{2}\left(1 - \frac{C_{12}^2}{C_{22}^2}\right)$$

When minimizing over θ, this leads to

$$\theta = \frac{1}{\sqrt{n}}\frac{W_1^n - W_2^n\frac{C_{12}}{C_{22}}}{1 - \frac{C_{12}^2}{C_{22}^2}} 1_{W_1^n - W_2^n\frac{C_{12}}{C_{22}} \geq 0}$$

with minimum value

$$-\frac{1}{2}\left(\frac{(W_2^n)^2}{C_{22}} + \frac{(W_1^n - W_2^n\frac{C_{12}}{C_{22}})^2}{1 - \frac{C_{12}^2}{C_{22}^2}} 1_{W_1^n - W_2^n\frac{C_{12}}{C_{22}} \geq 0}\right)(1 + o(1))$$

We just have to verify that at the optimizing value, $\theta/N(\beta)^2 \geq \eta_n^\alpha$, which is indeed the case since $\theta\phi$ does not tend to 0.
Letting now n tend to infinity and minimizing over all $B_n(\delta, \tau, \mu, \gamma, c, \nu, a, b)$ leads to

Lemma 5.14 $\inf_{(\theta,\beta)\in A_n} C_n(\theta,\beta) - C_n(f_0)$ *converges in distribution to the following variable:*

$$-\sup_{d_1\in\mathcal{D}_1, d_2\in\mathcal{D}_2} \frac{1}{2}(\xi_{d_1}^2 + \xi_{d_2}^2 1_{\xi_{d_2}\geq 0})$$

Lemma 5.11 and Lemma 5.14 give Theorem 4.1.

Proof of Lemma 5.10.

The proof proceeds by a verification of the assumptions used in Theorem 2.5 of Dahlhaus (1988), that is his assumption 2.1.
Assumption (a) is verified since the process is an ARMA process and the spectral density has infinitely many derivatives, all bounded.
(b) is verified since the tapering is the constant 1.
Let us now verify (c).
$\mathcal{G}$ is a permissible subset of H (in the sense of Pollard (1984) Appendix C) since it is a parametric class of functions which is pointwise continuous in the interior, and may be approached by sequences of parameters on the boundary.
Let us recall that for all x:

$$m_O = \frac{1}{2\pi u}\left(\frac{(1 - \frac{1}{1+\rho})^{q_0}}{2^{p_0}}\right)^2 \leq |f_0(x)| \leq \frac{u}{2\pi}\left(\frac{2^{q_0}}{(1 - \frac{1}{1+\rho})^{p_0}}\right)^2 = M_0$$

It is then enough to prove that the functions in $\mathcal{G}$ are uniformly pointwise bounded, and to verify the entropy condition.

Moreover, since the functions $S_{\theta,\beta)}$, $T_{\theta,\beta)}$, $\frac{N(\beta)^2 e'_{(0,\beta)}}{f_0}$ are bounded functions of bounded parameters, which are continuous both pointwise and in H, with the square of the norm which is a quadratic function of some of the parameters, so that the entropy condition is verified. it is enough to verify the conditions for the set of functions

$$\frac{N(\beta)^2 e^3_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e'_{(0,\beta)} e_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0}$$

It is again enough to verify that the functions

$$\frac{e_{(0,\beta)}}{f_0},\ \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0}$$

are uniformly pointwise bounded, and that the set of such functions verify the entropy condition. This in turns implies the entropy condition for the whole set of functions. To see that they are bounded, recall the expansion when $N(\beta)$ tends to 0:

$$\begin{aligned}
e_{(0,\beta)} &= \frac{1}{N(\beta)}[\frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0}(\frac{(\tau_i - \sum_{j\in T(i)}\gamma_j)z}{1-t_i z} + \frac{\overline{\tau_i} - \sum_{j\in T(i)}\overline{\gamma_j}}{z-(\overline{t_i}})\\
&+ \sum_{i=1}^{p_0}(\sum_{j\in T(i)}\gamma_j\sum_{h\geq 2}\frac{\alpha_j^{h-1}z^h}{(1-t_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1}}{(z-\overline{t_i})^h})\\
&- \sum_{i=1}^{q_0}(\frac{(\mu_i - \sum_{j\in U(i)}\gamma_j)z}{1-u_i z} + \frac{\overline{\mu_i} - \sum_{j\in U(i)}\overline{\gamma_j}}{z-\overline{u_i}})\\
&- \sum_{i=1}^{p_0}(\sum_{j\in U(i)}\gamma_j\sum_{h\geq 2}\frac{\alpha_j^{h-1})z^h}{(1-u_i z)^h} + \overline{\gamma_j}\sum_{h\geq 2}\frac{\overline{\alpha_j}^{h-1})}{(z-\overline{u_i})^h})\\
&- \sum_{i\in J}(\frac{\gamma_i z}{1-(c_i + \frac{\theta}{N(\beta)}\gamma_i)z} + \frac{\overline{\gamma_i}}{z-(\overline{c_i} + \frac{\theta}{N(\beta)}\overline{\gamma_i})}) - \sum_{i=1}^{s}(\nu_i z + \frac{\overline{\nu_i}}{z}))
\end{aligned}$$

and also

$$\begin{aligned}
N(\beta)^2 e"_{(0,\beta)} &= \frac{1}{N(\beta)}[\frac{2\delta^3}{\sigma_0^6} + \sum_{i=1}^{p_0}(\frac{\tau_i^3 z^3}{(1-t_i z)^3} - \sum_{j\in T(i)}\frac{\gamma_j^3 z^3}{(1-t_i z)^3}(\sum_{h\geq 0}(\frac{\alpha_j z}{(1-t_i z))^h})^3\\
&+ \frac{\overline{\tau_i}^3 z^3}{(1-\overline{t_i}z)^3} - \sum_{j\in T(i)}\frac{\overline{\gamma_j}^3 z^3}{(z-\overline{t_i})^3}(\sum_{h\geq 0}(\frac{\overline{\alpha_j}}{(z-\overline{t_i})})^h)^3)\\
&- \sum_{i=1}^{q_0}(\frac{(\mu_i^3 z^3}{(1-u_i z)^3} - \sum_{j\in U(i)}\frac{\gamma_j^3 z^3}{(1-u_i z)^3}(\sum_{h\geq 0}(\frac{\alpha_j z}{(1-u_i z)})^h)^3\\
&- \frac{\overline{\tau_i}^3 z^3}{(1-\overline{u_i}z)^3} - \sum_{j\in U(i)}\frac{\overline{\gamma_j}^3 z^3}{(z-\overline{u_i})^3}(\sum_{h\geq 0}(\frac{\overline{\alpha_j}}{(z-\overline{u_i})})^h)^3)\\
&- \sum_{i\in J}(\frac{\gamma_i^3 z^3}{(1-c_i z)^3} + \frac{\overline{\gamma_i}^3}{(z-\overline{c_i})^3})
\end{aligned}$$

Looking at the leading terms in the expansions, we obtain that for all β:

$$\frac{e_{(0,\beta)}}{f_0} \leq \frac{1}{m_f}\left(u + (1 + 2(r + p_0 + q_0)\frac{1}{\rho})^r) + 2s\right)$$

$$\frac{N(\beta)^2 e"_{(0,\beta)}}{f_0} \leq \frac{1}{m_f}\left(2u^3 + (1 + 2(r + p_0 + q_0)\frac{1}{\rho})^{3r})\right)$$

The entropy condition is the following. Let $\mathcal{N}(\epsilon)$ be the number of balls of diameter ϵ in H necessary for covering the set of functions. We have to verify that

$$\int_0^1 \left[\log \frac{\mathcal{N}(\epsilon)^2}{\epsilon}\right]^2 d\epsilon < +\infty$$

The previous expansions allow to find that for the set

$$\left\{\frac{e_{(0,\beta)}}{f_0}, \frac{N(\beta)^2 e"_{(0,\beta)}}{f_0} \; : \; \beta \in \tilde{\mathcal{B}}\right\}$$

the norm square is a quadratic function of at most K bounded parameters, with $K = r(1 + 2s + 2p_0 + 2q_0)$, so that we have

$$\mathcal{N}(\epsilon) = O(\frac{1}{\epsilon^K})$$

and the condition holds.

References

Akaike, H.(1970). Statistical predictor identification. *Ann. Inst. Statis. Math.*, **22** 202–217.

Akaike, H.(1974). A new look at the statistical model identification. *IEEE Trans. Aut. Cont.* **AC-19** 716–723.

Azencott R, Dacunha-Castelle D.(1986). *Series of irregular observations forescasting and model building.* Springer.

Dacunha-Castelle D, Gassiat E.(1996). Testing in locally conic models. submitted to *ESAIM Prob. and stat.*.

Dahlhaus, R.(1988). Empirical processes and their applications to time series analysis. *Stochastic Processes and their Applications* **30** 69–83.

Gerencser, L., Rissanen, J. Asymptotics of predictive stochastic complexity. In *New Directions in Time Series Analysis*, Springer Verlag. Brillinger, D., Caines ,P., Geweke, J., Parzen, E. Rosenblatt, M. Taqqu, M. eds.

Gill, L., Lewbel, A.(1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and Arma models. *J. Amer. Statist. Assoc.* **87** 766–776.

Hannan, J.(1980) The estimation of the order of an Arma process. *Annals of Stat.* **8** 5:1071–1081.

Hannan, E.J.(1982) Testing for autocorrelation and Akaike's criterion. In *Essays in Statistical Science*, 403-412. Gani, J.M., Hannan, E.J. eds.

Ossiander, M.(1987). A central limit theorem under metric entropy with l^2 bracketing. *Annals of Prob.* **15** 897–919.

Pollard, D.(1984). *Convergence of Stochastic Processes.* Springer Verlag.

Redner, R. (1981). Note on the consistency of the maximum likelihood estimate for non-identifiable distributions. *Annals of Stat.*, **9** 225–228.

Teicher, H.(1965). Identifiability of finite mixtures. *Annals of Math. Stat.* **36** 423–439.

Van der Vart, A. W., Wellner, J.A.(1996). *Empirical Processes.* Springer Verlag.

Veres, S.(1987). Asymptotic distribution of likelihood ratios for overparamtrized Arma processes. *J. Time Series Anal.* **8** 345–357.

Yakowitz, S.J., Spragins, J.D.(1968). On the identifiability of finite mixtures. *Annals of Math. Stat.* **39** 209–214.