# The Half-sample Method for Testing Parametric Regressive and Autoregressive Models of Order 1

Jean Diebolt

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *The half-sample method for testing parametric regressive and autoregressive models of order 1.*

Jean Diebolt

**No 3418**

THÈME 4

Rapport
de recherche

ISSN 0249-6399

# The half-sample method for testing parametric regressive and autoregressive models of order 1.

Jean Diebolt

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet is2

Rapport de recherche n°3418 — — 22 pages

**Abstract:** The half-sample method has been introduced by Stephens (1978) for testing parametric models of distribution functions. In this paper, we present a similar method for testing the goodness-of-fit of linear or nonlinear regression or autoregression functions for parametric models of order 1, under minimal stationarity and ergodicity assumptions. Our procedure is based on a measure of the cumulated deviation process $\hat{A}_n$ between a weighted marked process of residuals and a parametric estimator of the cumulated conditional mean function (i.e. cumulated regression or autoregression function), under the null hypothesis $H_0$. We establish a functional limit theorem under $H_0$ for a variant $\hat{A}_n^{(\kappa)}$, $0 < \kappa \leq 1$, of the process $\hat{A}_n$. The half-sample method corresponds to $\kappa = 1/2$. We show that the limiting distribution of $\hat{A}_n^{(1/2)}$ under $H_0$ takes a very simple form. Several easily implemented goodness-of-fit tests can be based on this result. We provide simple conditions under which their power converges to 1 as the sample size goes to $\infty$. Finally, we investigate the asymptotic behavior of $\hat{A}_n^{(\kappa)}$ as $n \to \infty$ under sequences of $O(n^{-1/2})$ local alternatives. This allows us to compare the corresponding local powers of tests based on $\hat{A}_n^{(1/2)}$ and on $\hat{A}_n^{(1)}$.

**Key-words:** Autoregression ; Regression ; Goodness-of-fit ; Conditional mean squares ; Ergodic ; Half-sample ; Stationary ; Martingale difference array ; Nonlinear ; Nonparametric ; Local power ; Contiguity.

*(Résumé : tsvp)*

CNRS, Imag / LMC / Statistique et Modèles Stochastiques, BP 53, 38041 Grenoble cedex 09, France. Email : Jean.Diebolt@imag.fr

# La méthode du demi-échantillon pour tester des modèles paramétriques de régression et d'autorégression d'ordre 1.

**Résumé :** La méthode du demi-échantillon a été introduite par Stephens (1978) pour tester des modèles paramétriques de distributions. Dans cet article, nous présentons et étudions une méthode analogue pour tester l'adéquation de fonctions de régression et d'autorégression d'ordre 1, sous des hypothèses minimales de stationnarité et d'ergodicité. Nous utilisons le processus $\hat{A}_n$, qui cumule les différences entre les résidus et un estimateur paramétrique de la fonction de moyenne conditionnelle, ceci sous l'hypothèse nulle $H_0$. Nous établissons la loi limite sous $H_0$ d'une variante $\hat{A}_n^{(\kappa)}$, $0 < \kappa \leq 1$, du processus $\hat{A}_n$. La méthode du demi-échantillon correspond au cas $\kappa = 1/2$. Nous montrons que la loi limite de $\hat{A}_n^{(1/2)}$ sous $H_0$ admet une forme remarquablement simple. Des tests faciles à implémenter peuvent alors être envisagés. Nous étudions des conditions suffisantes pour que leur puissance tende vers 1 quand la taille de l'échantillon tend vers l'infini. Enfin, nous nous intéressons au comportement asymptotique de $\hat{A}_n^{(\kappa)}$ quand $n \to \infty$ sous des suites d'alternatives locales en $O(n^{-1/2})$. Ceci nous permet de comparer les puissances locales d'une classe de tests selon qu'ils sont calculés à partir de $\hat{A}_n^{(1/2)}$ ou de $\hat{A}_n^{(1)}$.

**Mots-clé :** Autorégression ; Régression ; Adéquation ; Moindres carrés conditionnels ; Ergodique ; Demi-échantillon ; Stationnaire; Tableau de différence de martingale ; Non linéaire ; Non paramétrique ; Puissance locale ; Contiguïté.

# 1 Introduction

The half-sample method has been introduced by Stephens (1978) in the context of goodness-of-fit tests for parametric models of distribution functions. In this paper, we present a new half-sample method for testing the goodness-of-fit of linear or nonlinear regression or autoregression functions for parametric homoscedastic models of order 1. We consider both parametric models of regression or autoregression functions for which the response variables $Y_t$ or $X_{t+1}$ are related to $X_t$ by a relation of the form

$$Y_t = m(X_t; \theta) + \sigma\varepsilon_t \qquad \text{or} \qquad X_{t+1} = m(X_t; \theta) + \sigma\varepsilon_{t+1}, \tag{1.1}$$

where $\{X_t\}$ is stationary and ergodic. We will occasionally denote $m(\cdot; \theta) = m_\theta$. We do not assume that the sequence $\{(X_t, \varepsilon_t)\}$ (or $\{(X_t, \varepsilon_{t+1})\}$, respectively) is stationary and ergodic. The rv's $\varepsilon_t$ are $(\mathcal{G}_t : t \geq 1)$-martingale differences with conditional variance 1, i.e.

$$E(\varepsilon_t | \mathcal{G}_{t-1}) = 0 \qquad \text{and} \qquad E\left(\varepsilon_t^2 \middle| \mathcal{G}_{t-1}\right) = 1 \tag{1.2}$$

for all $t \geq 1$, with $\mathcal{G}_0 = \sigma(X_1)$ (or $\mathcal{G}_0 = \sigma(X_0)$) and $\mathcal{G}_t = \sigma(X_1, \varepsilon_1, \ldots, X_t, \varepsilon_t, X_{t+1})$ (or $\mathcal{G}_t = \sigma(X_0, \varepsilon_1, \ldots, \varepsilon_t)$), $t \geq 1$, respectively, and $(\mathcal{G}_t : t \geq 1)$ the corresponding filtration.

We wish to test the goodness-of-fit of models of the form (1.1). In this situation, the null hypothesis $H_0$ asserts that the true conditional mean function actually belongs to some parametric model $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$. Many papers have been devoted to this question (see, e.g., Stute and González Manteiga, 1996, and references therein, and Zuber, 1996, 1997, for a survey). Here, we focus on testing procedures based on some measure of the deviation process

$$\hat{A}_n(x) = n^{-1/2} \sum_{t=1}^{n} \left(Y_t - m(X_t; \hat{\theta}_n)\right) I(X_t \leq x), \qquad x \in \mathbf{R}, \tag{1.3}$$

where $\hat{\theta}_n$ is some strongly consistent estimate of the true parameter $\theta_0$ under $H_0$. Such cumulated marked residual processes converge in distribution to Gaussian processes derived from Brownian motion. In contrast, analogous processes based on local nonparametric estimators of the regression or autoregression function (e.g., nearest neighbors estimators in Stute and González Manteiga, 1996) have no functional limiting distribution. The basic reason is that such local nonparametric estimators can be seen as derivatives of cumulated processes such as

$$n^{-1} \sum_{t=1}^{n} Y_t I(X_t \leq x) \qquad \text{or} \qquad n^{-1} \sum_{t=1}^{n} X_{t+1} I(X_t \leq x),$$

and Brownian motion has no functional derivative. This is the basic mathematical reason why workable goodness-of-fit tests should be based on cumulated marked residual processes such as (1.3) rather than on local nonparametric estimators.

An and Cheng (1991) have introduced a Kolmogorov-Smirnov type test based on (1.3) to check the linearity of autoregressive time series of order 1. They propose to replace $\hat{A}_n$ with the partial sums $p_n^{-1/2} \sum_{t=1}^{p_n} (X_{t+1} - m(X_t; \hat{\theta}_n))\, I(X_t \leq x)$, where $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$, to get rid of the influence of $\hat{\theta}_n$ on the limiting distribution. However, this method results in a dramatic loss of power, since it can only detect $O(p_n^{-1/2})$ local alternatives. Su and Wei (1991) have studied a similar Kolmogorov-Smirnov type test. They propose a simulation method to implement it, but mistakes (see Stute, 1997) invalidate their results. Cheng and Wu (1994) have considered a similar testing process for testing the goodness-of-fit of a parametric family of link functions in the context of Generalized Linear Models, where the $X_t$'s are assumed to be iid. They base their Kolmogorov-Smirnov type test on the maximum likelihood estimation of $\theta$. They provide no practical method for computing the fractiles of the limiting distribution that they obtain.

Following Diebolt (1990), Diebolt and Laïb (1994) have established a functional limit theorem for cumulated regressograms when the parameters are assumed known. McKeague and Zhang (1994) have established more general functional limit theorems for cumulated regressograms when the parameters are unknown, with $\hat{\theta}_n$ the least squares estimator (LSE). Diebolt and Laïb (1993, 1995), Diebolt (1995) (for iid $X_t$'s), and Diebolt and Ngatchou Wandji (1995, 1996, 1997) have considered the process (1.3) in the particular case where the parameters are assumed known. Stute (1997) and Zuber (1997) have established the limiting distribution of (1.3) in the regression setting when $\hat{\theta}_n$ is the LSE, whereas Ngatchou Wandji and Laïb (1998) have obtained similar results for autoregressive models.

However, practical computation of the fractiles of the limiting distribution of possible goodness-of-fit tests based on these results remains hardly tractable. This is the reason why a half-sample type approach can turn out to be of great practical interest. In this paper, we establish a functional limit theorem for a variant of the process (1.3), namely

$$\hat{A}_n^{(\kappa)}(x) \;=\; n^{-1/2} \sum_{t=1}^{n} \left( Y_t \;-\; m(X_t; \hat{\theta}_n^{(\kappa)}) \right) I\left(X_t \leq x\right), \qquad x \in \mathbf{R}, \tag{1.4}$$

under $H_0$, where $\{\hat{\theta}_n^{(\kappa)}\}$ denotes the (conditional) LSE of the $K$-dimensional parameter $\theta$ based on randomly selected observations in proportion $\kappa$, $0 < \kappa \leq 1$. The half-sample method corresponds to $\kappa = 1/2$. It turns out that for $\kappa = 1/2$, this process converges in distribution to the remarkable process $\sigma\, W(F(\cdot))$, with $W$ a standard Wiener process (Corollary 1 to Theorem 2). This simple time-changed Brownian motion is exactly the limiting process of the process

$$A_n(x) \;=\; n^{-1/2} \sum_{t=1}^{n} \left(Y_t \;-\; m_0(X_t)\right) I\left(X_t \leq x\right), \qquad x \in \mathbf{R}. \tag{1.5}$$

for testing $m = m_0$ against $m \neq m_0$, under the simple null hypothesis $H_0$. Corollary 1 is extremely interesting for practical use, since the distribution of the limiting process $\sigma\, W(F(\cdot))$ only depends on $\sigma$ and $F$ and does not depend on the model $\mathcal{M}$ to be tested. Moreover, the distributions of numerous functionals of $W$ are analytically known.

In section 2, we precise the models that we consider and list the assumptions that we need. We establish the main asymptotic properties of the (conditional) LSE $\hat{\theta}_n^{(\kappa)}$, $0 < \kappa \leq 1$, in the general case where the $\varepsilon_t$'s are not necessarily iid (Theorem 1). This extends the results of Mangeas and Yao (1996, 1997). Section 3 is devoted to the functional limiting distribution of $\hat{A}_n^{(\kappa)}$ under the null hypothesis $H_0$ (Theorem 2 and its corollary). In section 4, easily implemented goodness-of-fit tests based on Corollary 1 are discussed. In section 5, we examine the asymptotic behavior of $\hat{A}_n^{(\kappa)}$ under the alternative hypothesis $H_1$ that the true conditional mean function remains sufficiently far from the manifold $\mathcal{M}$, and provide simple conditions under which the power of such tests converges to 1 as the sample size goes to $\infty$ (Theorem 3). Finally, section 6 is devoted to the study of the asymptotic behavior of $\hat{A}_n^{(\kappa)}$ as $n \to \infty$ under a sequence of local alternatives $H_1^n$ of the form $m = m_{\theta_0} + n^{-1/2}\delta$. This allows us to compare the corresponding local powers of tests based on $\hat{A}_n^{(1/2)}$ (half-sample process) and on $\hat{A}_n^{(1)}$ (full-sample process) when testing for no effect. The proofs are postponed to section 7.

## 2 Models, assumptions and estimation

For simplicity, we establish our functional limit theorems for models $\mathcal{M}$ of the form

$$Y_t \;=\; m\left(X_t; \theta\right) \;+\; \sigma\, \varepsilon_t, \qquad t \geq 1, \tag{2.1}$$

under the assumptions that $\{X_t : t \geq 1\}$ is a stationary and ergodic process of real-valued rv's and that (1.2) holds. We do not assume that the sequence $\{(X_t, \varepsilon_t)\}$ is stationary and ergodic. We denote by $F$ the common repartition function of the $X_t$'s. The function $x \longrightarrow m(x;\,\theta)$ corresponds to the conditional mean of $Y_t$ given the $\sigma$-field $\mathcal{G}_{t-1}$, $E\left(Y_t|\,\mathcal{G}_{t-1}\right) = m(X_t;\,\theta)$ a.s. The autoregressive case follows by replacing $Y_t$ with $X_{t+1}$ and $\varepsilon_t$ with $\varepsilon_{t+1}$ in (2.1).

We assume that the model $\mathcal{M}$ is identifiable, i.e. for all $\theta$ and $\theta'$ in the compact subset $\Theta$ of $\mathbf{R}^K$, the almost everywhere equality of the functions $x \longrightarrow m(x;\,\theta)$ and $x \longrightarrow m(x;\,\theta')$ implies that $\theta = \theta'$. The parameter $\theta = (\theta_1, \ldots, \theta_K)^T$, $K \geq 1$, is considered as a $K \times 1$ matrix. The parameter space $\Theta$ is a compact subset of $\mathbf{R}^K$ with nonempty interior, $\mathrm{int}(\Theta)$. We denote by

$$\partial_k m(x;\,\theta) = \partial m(x;\,\theta)/\partial\theta_k \quad \text{and} \quad \partial_k m(x;\,\theta)|_{\theta=\zeta} = \partial m(x;\,\theta)/\partial\theta_k|_{\theta=\zeta}, \qquad k = 1, \ldots, K,$$

the partial derivative of $m(x;\,\theta)$ with respect to $\theta_k$ and the value of this partial derivative at some point $\theta = \zeta \in \mathrm{int}(\Theta)$, respectively. We denote by $\nabla_\theta m(x;\,\theta) = (\partial_1 m(x;\,\theta), \ldots, \partial_K m(x;\,\theta))^T$ the gradient of $m(x;\,\theta)$ with respect to $\theta$. The gradient $\nabla_\theta m(x;\,\theta) \in \mathbf{R}^K$ is considered as a $K \times 1$ matrix. In order to simplify the notation we will write $\nabla_\theta m(x;\,\theta) = \nabla m(x;\,\theta) = \nabla m_\theta(x)$ when no confusion can arise, and will denote $\nabla_\theta m_\theta(x)|_{\theta=\theta_0} = \nabla m_0(x)$, where $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,K})^T$ is the true value of the parameter.

Below we list the assumptions that we need. These assumptions concern the existence of conditional moments of sufficient order for the $\varepsilon_t$'s, regularity of the repartition function $F$ of $X_1$, regularity, smoothness and moment properties for the function $m(x;\,\theta)$. We suppose that the true value $\theta_0$ of the parameter under the null hypothesis $H_0$ that the observations are actually issued from the postulated model $\mathcal{M}$ is in the interior $\mathrm{int}(\Theta)$. We denote by $\|\cdot\|$ a norm on $\mathbf{R}^K$. We first assume that for all fixed $x$ the function $\theta \longrightarrow m(x;\,\theta)$ has continuous derivatives up to order 2, and for all $\theta$ the functions $x \longrightarrow m(x;\,\theta)$, $\partial_j m(x;\,\theta)$ and $\partial^2_{j\,j'} m(x;\,\theta) = \partial^2 m(x;\,\theta)/\partial\theta_j\partial\theta_{j'}$ are continuous.

**(A1)** There exists $\gamma > 0$ such that

$$c_{(\varepsilon,\,\gamma)} = \sup_{t\geq 1} E\left(|\varepsilon_t|^{2+\gamma}\middle|\,\mathcal{G}_{t-1}\right) < \infty.$$

**(A2)** The distribution function $F$ is continuous and increasing and there exists a real number $\beta \geq 2$ such that $E\left(|X_1|^\beta\right) = \int_{-\infty}^{\infty} |y|^\beta \, dF(y) < \infty$.

**(A3)** There exists $\gamma' > 0$ and a nondecreasing continuous function $\omega_0(\cdot)$ such that $\lim_{\delta\to 0} \omega_0(\delta) = \omega_0(0) = 0$ and

$$|m(x;\,\theta_2) - m(x;\,\theta_1)| \leq \omega_0\left(\|\theta_2 - \theta_1\|\right)\left(1 + |x|^{\beta/(2+\gamma')}\right) \tag{2.2}$$

for all $x \in \mathbf{R}$ and $(\theta_1,\,\theta_2) \in \Theta \times \Theta$. Moreover, there exists $c_{0,\,0}$ such that

$$|m\left(x;\,\theta_0\right)| \leq c_{0,\,0}\left(1 + |x|^{\beta/(2+\gamma')}\right) \ \text{ for all}\, x \in \mathbf{R}. \tag{2.3}$$

**(A4)** There exists a positive finite real number $r_0$ such that the closed ball $\bar{B}_0 = \bar{B}\left(\theta_0,\,r_0\right)$ is contained in $\mathrm{int}(\Theta)$, and there exists a nondecreasing continuous function $\omega_{0,\,2}(\cdot)$ such that $\lim_{\delta\to 0} \omega_{0,\,2}(\delta) = \omega_{0,\,2}(0) = 0$ and

$$\left|\partial^2_{j\,j'} m(x;\,\theta) - \partial^2_{j\,j'} m(x;\,\theta)\Big|_{\theta=\theta_0}\right| \leq \omega_{0,\,2}\left(\|\theta - \theta_0\|\right)\left(1 + |x|^{\beta/(2+\gamma')}\right) \tag{2.4}$$

for all $x$, all $\theta \in \bar{B}_0$ and all $j, j' = 1, \ldots, K$. Moreover, there exists $c_{0,\,2}$ such that

$$\left|\partial^2_{j\,j'} m(x;\,\theta)\Big|_{\theta=\theta_0}\right| \leq c_{0,\,2}\left(1 + |x|^{\beta/(2+\gamma')}\right) \ \text{ for all}\, x \in \mathbf{R}. \tag{2.5}$$

**(A5)** There exists $c_{0,1}$ such that for all $x$,

$$\|\nabla m_0(x)\| \;\le\; c_{0,1}\left(1 + |x|^{\beta/(2+\gamma')}\right). \tag{2.6}$$

**(A6)** The matrix

$$V_0 \;=\; \int_{-\infty}^{\infty} \nabla m_0 \nabla m_0^T \, dF \tag{2.7}$$

is definite positive.

**Remark 1.** *(i)* The assumption (A3) has a global nature, whereas (A4) is local ($\theta \in \bar{B}_0$) and (A5) is pointwise ($\theta = \theta_0$).

*(ii)* Since $\Theta$ is compact the continuous functions $\omega_0(\cdot)$ and $\omega_{0,2}(\cdot)$ are bounded, hence it follows from (A3)–(A4) that

$$|m(x;\theta_2) - m(x;\theta_1)| \le \text{cst}\left(1 + |x|^{\beta/(2+\gamma')}\right) \text{ for all } x \in \mathbf{R} \text{ and } (\theta_1, \theta_2) \in \Theta \times \Theta, \tag{2.8}$$

$$|m\,(x;\theta)| \;\le\; \text{cst}\left(1 + |x|^{\beta/(2+\gamma')}\right) \text{ for all } x \in \mathbf{R} \text{ and } \theta \in \Theta, \tag{2.9}$$

and

$$\left|\partial^2_{j\,j'} m(x;\theta)\right| \le \text{cst}\left(1 + |x|^{\beta/(2+\gamma')}\right) \text{ for all } x \in \mathbf{R},\ \theta \in \bar{B}_0 \text{ and } j, j' \;=\; 1,\ldots,K, \tag{2.10}$$

for some positive constants. Therefore, there exists a nonnegative finite measurable function $M(x)$, $x \in \mathbf{R}$, such that $\int_{-\infty}^{\infty} M\,dF \;<\; \infty$ and

$$\sup_{\theta \in \bar{B}_0} \left|\partial^2_{j\,j'} m\,(x;\theta)\right| \;\le\; M(x) \text{ for all } x \in \mathbf{R} \text{ and } j, j' \;=\; 1,\ldots,K. \tag{2.11}$$

*(iii)* It follows from the mean-value theorem and (2.10) that

$$\|\nabla m_\theta(x) - \nabla m_0(x)\| \le \text{cst}\left(1 + |x|^{\beta/(2+\gamma')}\right) \text{ for all } x \in \mathbf{R} \text{ and } \theta \in \bar{B}_0. \tag{2.12}$$

Therefore, in view of (A5),

$$\|\nabla m_\theta(x)\| \;\le\; \text{cst}\left(1 + |x|^{\beta/(2+\gamma')}\right) \text{ for all } x \in \mathbf{R} \text{ and } \theta \in \bar{B}_0. \tag{2.13}$$

*(iv)* Similarly, it follows from (2.13) that

$$|m\,(x;\theta) - m\,(x;\theta_0)| \;\le\; \text{cst}\;\|\theta - \theta_0\|\left(1 + |x|^{\beta/(2+\gamma')}\right) \tag{2.14}$$

for all $x \in \mathbf{R}$ and $\theta \in \bar{B}_0$.

In this paper, we consider the estimator $\hat{\theta}_n^{(\kappa)}$ derived from the mean squares estimator $\hat{\theta}_n$ for nonlinear models of regression (or autoregression) functions,

$$\hat{\theta}_n \;=\; \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \left(Y_t - m(X_t;\theta)\right)^2,$$

studied, e.g., by Klimko and Nelson (1978), Lai (1994), McKeague and Zhang (1994), and Mangeas and Yao (1996, 1997) for stationary processes and by Duflo, Senoussi and Touati (1990, 1991) in the general linear regression or autoregression context. Under the assumptions (A1)–(A6), which are essentially similar to conditions [S], [M] and [N] of Mangeas and Yao (1997) (see also McKeague and Zhang [1994, conditions B]), this sequence of estimators is strongly consistent and satisfies

$$\xi_n \;=\; n^{1/2}\left(\hat{\theta}_n \,-\, \theta_0\right) \;=\; n^{-1/2}\,\sigma\,V_0^{-1}\sum_{t=1}^{n}\nabla m_0(X_t)\,\varepsilon_t \;+\; o_P(1) \quad \text{as } n\to\infty,$$

where the $K\times K$ symmetric matrix $V_0=\int_{-\infty}^{\infty}\nabla m_0 \nabla m_0^T\,dF$ is finite and definite positive, and $o_P(1)\in\mathbf{R}^K$ converges in probability to 0 as $n\to\infty$. Therefore, $\xi_n$ converges in distribution to a $K$-dimensional normal rv $\xi$ with mean 0 and variance matrix $\sigma^2 V_0$. We consider a real number $0<\kappa\le 1$ and a sequence $\{b_t\,:\,t\ge 1\}$ of iid Bernoulli($\kappa$) rv's (i.e. $b_t=1$ with probability $\kappa$ and $b_t=0$ with probability $1-\kappa$). We assume that the sequence $\{b_t\,:\,t\ge 1\}$ is independent of the $\sigma$-field $\mathcal{G}=\sigma\left(\cup_{t=1}^{\infty}\mathcal{G}_t\right)$ and set $\mathcal{H}_t=\sigma(X_1,\varepsilon_1,b_1\ldots,X_t,\varepsilon_t,b_t,X_{t+1})$. Letting $Q_n=\sum_{t=1}^{n}b_t$, we extract a $Q_n$-subsample $\{X_{i_1},\,\ldots,\,X_{i_{Q_n}}\}$ randomly as follows. For $t=1,\ldots,n$, we keep $X_t$ when $b_t=1$ and delete it when $b_t=0$. Let $\hat{\theta}_n^{(\kappa)}$ denote the mean squares estimate of $\theta$ based on this subsample, i.e.

$$\hat{\theta}_n^{(\kappa)} \;=\; \arg\min_{\theta\in\Theta}\sum_{t=1}^{n} b_t\,\left(Y_t \,-\, m(X_t;\,\theta)\right)^2. \tag{2.15}$$

**Theorem 1** *Under the assumptions (A1)–(A6), $\hat{\theta}_n^{(\kappa)}$ converges to $\theta_0$ a.s. as $n\to\infty$, and*

$$\xi_n^{(\kappa)} \;=\; n^{1/2}\left(\hat{\theta}_n^{(\kappa)} \,-\, \theta_0\right) \;=\; \kappa^{-1}\,V_0^{-1}\,n^{-1/2}\,\sigma\sum_{t=1}^{n} b_t\,\nabla m_0(X_t)\,\varepsilon_t \;+\; o_P(1) \tag{2.16}$$

*converges in distribution to a Gaussian rv $\xi^{(\kappa)}$ with mean 0 and variance matrix $\kappa^{-1}\sigma^2 V_0^{-1}$.*

# 3 Functional limit theorems

We wish to test the null hypothesis $H_0$ that the true regression function $m$ belongs to the manifold $\mathcal{M}=\{m_\theta\,:\,\theta\in\Theta\}$ representing some parametric model, i.e. we wish to test the null hypothesis $H_0$ that $E(Y_t\,|\,\mathcal{G}_{t-1})$ has the form $m(X_t;\,\theta)$ for some unknown value $\theta_0\in\Theta$. We introduce the process

$$\hat{A}_n^{(\kappa)}(x) \;=\; n^{-1/2}\sum_{t=1}^{n}\left(Y_t \,-\, m\left(X_t;\,\hat{\theta}_n^{(\kappa)}\right)\right)\,I\left(X_t\le x\right), \qquad x\in\mathbf{R}. \tag{3.1}$$

Under $H_0$, this process $\hat{A}_n^{\kappa}(\cdot)$ takes the form

$$\hat{B}_n^{(\kappa)}(x) \;=\; B_n(x) \;-\; n^{-1/2}\sum_{t=1}^{n}\left(m\left(X_t;\,\hat{\theta}_n^{(\kappa)}\right) \,-\, m\left(X_t;\,\theta_0\right)\right)\,I\left(X_t\le x\right), \qquad x\in\mathbf{R}, \tag{3.2}$$

where

$$B_n(x) \;=\; n^{-1/2}\,\sigma\sum_{t=1}^{n}\varepsilon_t\,I\left(X_t\le x\right), \qquad x\in\mathbf{R}. \tag{3.3}$$

**Theorem 2** *Under the assumptions (A1)–(A6) and under $H_0$, the processes $\hat{B}_n^{(\kappa)}(\cdot)$ converge in distribution to the centered Gaussian process*

$$\hat{B}^{(\kappa)}(x) \;=\; B(x) \;-\; g_0(x)^T \xi^{(\kappa)}, \qquad x \in \mathbf{R}, \tag{3.4}$$

*where*

$$g_0(x) \;=\; \int_{-\infty}^{x} \nabla_0 m \, dF, \qquad x \in \mathbf{R}, \tag{3.5}$$

*$\xi^{(\kappa)}$ is Gaussian with mean 0 and variance matrix $\kappa^{-1}\sigma^2 V_0^{-1}$, and the rv's $\hat{\xi}^{(\kappa)}$ and $B(x)$ have covariance column matrix*

$$h_0(x) \;=\; E\left(\xi^{(\kappa)} \, B(x)\right) \;=\; \sigma \, V_0^{-1} \, g_0(x). \tag{3.6}$$

*The covariance function of $\hat{B}^{(\kappa)}(\cdot)$ is*

$$E\left(\hat{B}^{(\kappa)}(x_1)\hat{B}^{(\kappa)}(x_2)\right) \;=\; \sigma^2 \,\left[F\left(x_1 \wedge x_2\right) \;+\; \left(\kappa^{-1} \;-\; 2\right) g_0(x_1)^T \, V_0^{-1} \, g_0(x_2)\right]. \tag{3.7}$$

If in addition $\kappa = 1/2$, then we find the following simple and remarkable result, which is the analogous in the present setting of Stephen (1978)'s result.

**Corollary 1** *If $\kappa \;=\; 1/2$, then under the assumptions (A1)–(A6) and under $H_0$, the processes $\hat{B}_n^{(1/2)}(\cdot)$ converge in distribution to the centered Gaussian process*

$$\hat{B}^{(1/2)}(x) \;=\; \sigma \, W\left(F(x)\right), \tag{3.8}$$

*where $W$ is a standard Wiener process.*

# 4 Possible tests based on these results

First, two possible tests based on Corollary 1 of Theorem 2 are a Kolmogorov-Smirnov type test and a Cramér-von Mises type test. The limiting distributions under $H_0$ of these test statistics are given by

$$\sup_{x \in \mathbf{R}} \left| \hat{A}_n^{(1/2)}(x) \right| \;\longrightarrow\; \sigma \sup_{u \in [0,\,1]} |W(u)|$$

and

$$\int_{-\infty}^{\infty} \left| \hat{A}_n^{(1/2)}(x) \right|^2 w\left(\mathbf{F}_n(x)\right) \, d\mathbf{F}_n(x) \;\longrightarrow\; \sigma^2 \int_0^1 W^2\left(u\right) w\left(u\right) \, du,$$

where $w$ is a weight function and $\mathbf{F}_n$ is the empirical distribution function of the sample. When $\sigma$ is unknown, it can be replaced with an a.s. convergent estimate $\hat{\sigma}_n$.

Another possibility is to make use of the Karhunen-Loève expansion of the Gaussian process $Z^{(\kappa)}(\cdot) = \sigma^{-1} \, (\hat{B}^{(\kappa)} \circ F^{-1})(\cdot)$ defined on the unit interval [0, 1],

$$Z^{(\kappa)} \;=\; \sum_{j=1}^{\infty} \left(\lambda_j^{(\kappa)}\right)^{1/2} \, Z_j^{(\kappa)} \, \phi_j^{(\kappa)}, \tag{4.1}$$

where $\lambda_1^{(\kappa)} \geq \lambda_2^{(\kappa)} \geq ... \geq 0$ are the eigenvalues of the covariance operator of $Z^{(\kappa)}$ on $L^2[0,1]$, the functions $\phi_1^{(\kappa)}$, $\phi_2^{(\kappa)}$, ... are an orthonormal basis of eigenfunctions of this operator, and under $H_0$ the rv's

$$Z_j^{(\kappa)} \;=\; \left(\lambda_j^{(\kappa)}\right)^{-1/2} \left(\int_0^1 Z^{(\kappa)}(u)\,\phi_j^{(\kappa)}(u)\,du\right) \quad \text{are iid normal } \mathcal{N}(0,1). \tag{4.2}$$

The proposed test statistics has the form

$$\mathcal{T}_n^{(\kappa,\,J)} \;=\; \sum_{j=1}^{J} \left(Z_{n,\,j}^{(\kappa)}\right)^2 \tag{4.3}$$

for some moderate $J > 1$, where

$$Z_{n,\,j}^{(\kappa)} \;=\; \left(\lambda_j^{(\kappa)}\right)^{-1/2} \sigma^{-1} \int_{-\infty}^{\infty} \hat{A}_n^{(\kappa)}(y)\,\phi_j^{(\kappa)}\left(F(y)\right)\,dF(y), \tag{4.4}$$

converges in distribution under $H_0$ to

$$\mathcal{T}^{(\kappa,\,J)} \;=\; \sum_{j=1}^{J} \left(Z_j^{(\kappa)}\right)^2, \tag{4.5}$$

a chi-square with $J$ degrees of freedom. When $F$ is unknown in (4.4), it can be replaced with $\mathbf{F}_n$. For $\kappa = 1/2$, $Z^{(1/2)}$ is a standard Wiener process $W$ on the unit interval [0, 1],

$$\lambda_j^{(1/2)} \;=\; (j\,-\,1/2)^{-2}\pi^{-2} \quad\text{and}\quad \phi_j^{(1/2)}(u) = 2^{1/2}\sin\left[(j\,-\,1/2)\pi u\right], \quad u\in[0,\,1],\; j\geq 1. \tag{4.6}$$

# 5 Behavior of the processes under a fixed alternative

In this section, we study the power of possible tests which can be derived from Corollary 1. We assume that the alternative hypothesis holds, namely that

$$Y_t \;=\; m_1(X_t) \,+\, \sigma\,\varepsilon_t, \quad t\geq 1, \tag{5.1}$$

where the true regression (or autoregression) function $m_1(\cdot)$ *cannot be written* in the form $m_\theta$. We denote $\|\psi\|_\infty = \sup_{x\in\mathbf{R}}|\psi(x)|$.

**Theorem 3** *We assume that the alternative hypothesis $H_1$ holds, i.e. the stationary process $\{X_t\}$ satisfies*

$$Y_t \;=\; m_1(X_t) \,+\, \sigma\,\varepsilon_t, \quad t\geq 1,$$

*where the regression (autoregression) function $m_1(\cdot)$ cannot be written in the form $m_\theta$, $\theta\in\Theta$, the martingale differences $\varepsilon_t$ satisfy (A1), and the stationary distribution function $F$ of the process $\{X_t\}$ is continuous and increasing. Moreover, we assume that $\int_{-\infty}^{\infty}|m_1|\;dF < \infty$ and that (A1)–(A3) hold. Let $N$ denote a seminorm on the Skorokhod space $D[\mathbf{R}]$, such that $N(\psi)\leq cst\;\|\psi\|_\infty$ for all $\psi\in D[\mathbf{R}]$. If*

$$N\left(\int_{-\infty}^{\bullet}(m_1 - m_\theta)\;dF\right) \;>\; 0 \quad \textit{for all } \theta\in\Theta, \tag{5.2}$$

*then*

$$N\left(\hat{A}_n^{(\kappa)}\right) \;\longrightarrow\; \infty \quad \textit{a.s. as } n\to\infty. \tag{5.3}$$

# 6 Behavior of the processes under local alternatives

## 6.1 Theory

In this section, we investigate the local power of tests based on the processes $\hat{A}_n^{(\kappa)}$. To this end, we first establish a functional limit theorem (Theorem 4 below) analogous to Theorem 1, under the sequence of local alternatives $H_1^n$ that

$$Y_t \;=\; m\,(X_t;\,\theta_0) \;+\; n^{-1/2}\delta\,(X_t) \;+\; \sigma\,\varepsilon_t \quad \text{or} \quad X_{t+1} \;=\; m\,(X_t;\,\theta_0) \;+\; n^{-1/2}\delta\,(X_t) \;+\; \sigma\,\varepsilon_{t+1} \tag{6.1}$$

for some fixed $\theta_0 \in \text{int}\Theta$ and some function $\delta$. Then, we compare the local power of the $\chi^2_J$ test defined in (4.1)–(4.5) when $\kappa = 1$ (full-sample testing process) and $\kappa = 1/2$ (half-sample testing process). Finally, we compute and compare the local powers of this test for $\kappa = 1$ and $\kappa = 1/2$ in the important particular case of testing for no effect.

The square root $p^{1/2}$ of the density $p$ is said to be differentiable in quadratic mean with respect to Lebesgue measure (e.g., Pollard, 1997) if there exists $q \in L^2(\mathbf{R}, dx)$ such that

$$\lim_{h\to 0}\, h^{-2} \int_{-\infty}^{\infty} \Big[p^{1/2}\,(x\,+\,h) \;-\; p^{1/2}\,(x) \;-\; h\,q\,(x)\Big]^2\, dx \;=\; 0. \tag{6.2}$$

If moreover $p(x) > 0$ for all $x$, we denote $q = \dot{p}/2\sqrt{p}$.

**Theorem 4** *Suppose that (A1)–(A6) hold, and one of the following assumptions is in force.*

*(1) In the regression case,*

$$|\delta(x)| \;\le\; c_{(\delta,\,\gamma')}\Big(1\;+\;|x|^{\beta/(2\,+\,\gamma')}\Big) \quad \textit{for all } x. \tag{6.3}$$

*(2) In the autoregression case, the function $\delta$ is bounded and the $\varepsilon_t$'s are iid with mean 0 and variance 1. Their common distribution has a positive density p differentiable in quadratic mean with respect to Lebesgue measure, and $\int_{-\infty}^{\infty} \phi_p(x)^{2+\gamma'}\, p(x)\, dx < \infty$, with $\phi_p = \dot{p}/p$.*

*Then under $H_1^n$ (see (6.1)), $\hat{A}_n^{(\kappa)}$ converges in distribution to the Gaussian process*

$$\tilde{B}^{(\kappa)}(x) \;=\; B(x) \;-\; g_0^T(x)\xi^{(\kappa)} \;+\; \int_{-\infty}^{x} \delta\, dF \;-\; g_0^T(x)\,\mu_0, \tag{6.4}$$

*where*

$$\mu_0 \;=\; V_0^{-1} \int_{-\infty}^{\infty} \delta\, \nabla m_0\, dF. \tag{6.5}$$

## 6.2 Comparisons for the $\chi^2_J$-test

Here, we now compare the local powers of the tests based on (4.1)–(4.5) for $\kappa = 1$ and $\kappa = 1/2$. We assume that the function $\delta$ is orthogonal to the tangent space of the model at $\theta_0$, i.e.,

$$\mu_0 \;=\; \int_{-\infty}^{\infty} \delta\, \nabla m_0\, dF \;=\; 0. \tag{6.6}$$

We then have

$$\begin{aligned} \hat{B}_\delta^{(\kappa)}(x) \;&=\; B(x) \;-\; g_0^T(x)\xi^{(\kappa)} \;+\; \int_{-\infty}^{x} \delta\, dF \\ &=\; \hat{B}^{(\kappa)}(x) \;+\; \int_{-\infty}^{x} \delta\, dF \end{aligned} \tag{6.7}$$

and

$$\begin{aligned}
Z_\delta^{(\kappa)}(u) &= \sigma^{-1}\left(\hat{B}_\delta^{(\kappa)} \circ F^{-1}\right)(u) \\
&= \sigma^{-1}\left(B \circ F^{-1}\right)(u) \;-\; \sigma^{-1}\, g_0^T\left(F^{-1}(u)\right)\xi^{(\kappa)} \;+\; \sigma^{-1}\int_{-\infty}^{F^{-1}(u)} \delta\, dF \\
&= Z^{(\kappa)}(u) \;+\; \sigma^{-1}\int_0^u \delta^\star(v)\, dv, \qquad (6.8)
\end{aligned}$$

where $\delta^\star = \delta \circ F^{-1}$. Under the sequence of alternatives $H_1^n$, the rv's

$$\begin{aligned}
Z_{\delta,\,j}^{(\kappa)} &= Z_j^{(\kappa)} \;+\; \left(\lambda_j^{(\kappa)}\right)^{-1/2}\sigma^{-1}\int_0^1\left(\int_0^u \delta^\star(v)\,dv\right)\phi_j^{(\kappa)}(u)\,du \\
&= Z_j^{(\kappa)} \;+\; d_j^{(\kappa)} \qquad \text{are iid normal } \mathcal{N}(0,\, d_j^{(\kappa)}). \qquad (6.9)
\end{aligned}$$

Therefore, the $\chi_J^2$ noncentrality term is $\sum_{j=1}^J \left(d_j^{(\kappa)}\right)^2$. Under the sequence of alternatives $H_1^n$ the process $Z^{(1/2)}$ has the same distribution as

$$W \;+\; \sigma^{-1}\int_0^\bullet \delta^\star(v)\,dv,$$

and under the sequence of alternatives $H_1^n$, the rv's

$$\begin{aligned}
Z_{\delta,\,j}^{(1/2)} &= Z_j^{(1/2)} \;+\; (j \,-\, 1/2)\,\pi\,\sigma^{-1}\int_0^1\left(\int_0^u \delta^\star(v)\,dv\right)\sqrt{2}\,\sin\left[(j \,-\, 1/2)\,\pi\, u\right]du \\
&= Z_j^{(1/2)} \;+\; d_j^{(1/2)} \qquad \text{are iid normal } \mathcal{N}(0,\, d_j^{(1/2)}). \qquad (6.10)
\end{aligned}$$

The corresponding $\chi_J^2$ noncentrality term is then $\sum_{j=1}^J \left(d_j^{(1/2)}\right)^2$. An integration by parts yields

$$d_j^{(1/2)} \;=\; \sigma^{-1}\int_0^1 \delta^\star(u)\,\sqrt{2}\,\cos\left[(j \,-\, 1/2)\,\pi\, u\right]du. \qquad (6.11)$$

## 6.3 Example: testing for no effect

In this case $m_\theta(x) = \text{cst} = \theta$ and $\partial m_\theta(x)/\partial\theta \;=\; 1$ for all $x$. The condition $\mu_0 = 0$ reads

$$\int_{-\infty}^{\infty} \delta\, dF \;=\; \int_0^1 \delta^\star(v)\,dv \;=\; 0, \qquad (6.12)$$

the process $Z^{(1)}$ is a Brownian bridge, $\lambda_j^{(1)} = j^{-2}\pi^{-2}$ and $\phi_j^{(1)}(u) = \sqrt{2}\sin(j\pi u)$. An integration by parts yields

$$d_j^{(1)} \;=\; \sigma^{-1}\int_0^1 \delta^\star(u)\,\sqrt{2}\,\cos(j\,\pi\, u)\,du. \qquad (6.13)$$

We wish to compare $\left(d_j^{(1)}\right)^2$ with $\left(d_j^{(1/2)}\right)^2$ in the special case where

$$\int_0^u \delta^\star(v)\,dv \;=\; k^{-1}\,\phi_k^{(1)}(u) \;=\; \sqrt{2}\,\sin(k\,\pi\, u). \qquad (6.14)$$

The condition (6.12) is satisfied, and $\delta^\star(u) = \pi\,\sqrt{2}\,\cos(k\pi u)$. In this case,

$$\left(d_k^{(1)}\right)^2 \;=\; \sigma^{-2}\left(\lambda_k^{(1)}\right)^{-1} \;=\; \sigma^{-2}\,\pi^2 \quad \text{and} \quad d_j^{(1)} \;=\; 0 \;\text{ for }\; j \neq k, \qquad (6.15)$$

whereas for $k = 1, \ldots, J$,

$$\sum_{j=1}^{J} \left( d_j^{(1/2)} \right)^2 = \sigma^{-2} \sum_{j=1}^{J} \frac{4\,(j - 1/2)^2}{\left[(j - 1/2)^2 - k^2\right]^2}. \tag{6.16}$$

Table I displays the values of the $\chi_J^2$ noncentrality terms for $\kappa = 1/2$ in the special case where $\sigma = 1$ and $\delta^\star(u) = \pi\sqrt{2}\cos(k\pi u)$, i.e. $\int_0^u \delta^\star(v)\,dv = k^{-1}\phi_k^{(1)}(u)$, $k = 1, \ldots, 10$, with $J = 10, 20$ and $100$. These values are to be compared with the corresponding common value of the $\chi_J^2$ noncentrality terms for $\kappa = 1$, namely $\pi^2 = 9.869605$. This example shows that even for the comparatively small value $J = 10$, the local powers of the $\chi_J^2$ tests corresponding to $\kappa = 1/2$ (half-sample estimation) and $\kappa = 1$ (full-sample estimation) against local alternatives of the form $\mathrm{cst}\, n^{-1/2}\cos(k\pi u)$ are close for the 8 first frequencies $k = 1, \ldots, 8$, and still comparable for the frequency $k = 9$. For the frequency $k = 10$, this local power is significantly smaller for $\kappa = 1/2$ than for $\kappa = 1$.

| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ | $k=6$ | $k=7$ | $k=8$ | $k=9$ | $k=10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $J=10$ | 9.467 | 9.459 | 9.444 | 9.420 | 9.384 | 9.328 | 9.233 | 9.050 | 8.559 | 4.389 |
| $J=20$ | 9.669 | 9.668 | 9.667 | 9.664 | 9.661 | 9.657 | 9.651 | 9.645 | 9.637 | 9.627 |
| $J=100$ | 9.830 | 9.830 | 9.830 | 9.830 | 9.830 | 9.830 | 9.829 | 9.829 | 9.829 | 9.829 |

Table 1: Values of $\sum_{j=1}^{J} \left( d_j^{(1/2)} \right)^2$, $J = 10, 20$ and $100$, for $\sigma = 1$ and $k = 1, \ldots, 10$.

## References


**[1** ] AN, H. Z., AND CHENG, B. (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review* **59**, 287–307.

**[2** ] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

**[3** ] CHENG, K. F., AND WU, J. W. (1994). Testing goodness-of-fit for a parametric family of link functions. *Journal of the American Statistical Association* **89**, 657–664.

**[4** ] DIEBOLT, J. (1990). Testing the functions defining a nonlinear autoregressive time series. *Stochastic Processes and their Applications* **36**, 85–106.

**[5** ] DIEBOLT, J., AND LAÏB, N. (1993). Un principe d'invariance faible pour un test non-paramétrique de régression non linéaire avec données dépendantes. *Publications de l'ISUP* (nouvelle série) **37**, 1–32.

**[6** ] DIEBOLT, J., AND LAÏB, N. (1994). A weak invariance principle for cumulated functionals of the regressogram estimator with dependent data. *Journal of Nonparametric Statistics* **4**, 71–87.

**[7** ] DIEBOLT, J., AND LAÏB, N. (1995). Nonparametric tests for correlation or autoregression models under mixing conditions. *Notes aux Comptes Rendus de l'Académie des Sciences de Paris* série I, **320**, 1135–1139.

**[8** ] DIEBOLT, J. (1995). A nonparametric test for regression functions: Asymptotic theory. *Journal of Statistical Planning and Inference* **44**, 1–17.

**[9** ] DIEBOLT, J., AND NGATCHOU WANDJI, J. (1995). Un test non-paramétrique pour un modèle bilinéaire diagonal d'ordre un. *Notes aux Comptes Rendus de l'Académie des Sciences de Paris* série I, **320**, 727–730.

**[10** ] DIEBOLT, J., AND NGATCHOU WANDJI, J. (1996). A nonparametric test for general first-order autoregressive models with a mixed term. *Notes aux Comptes Rendus de l'Académie des Sciences de Paris* série I, **322**, 577–582.

**[11** ] DIEBOLT, J., AND NGATCHOU WANDJI, J. (1997). A nonparametric test for generalized first-order autoregressive models. *Scandinavian Journal of Statistics* **24**, 241–257.

**[12** ] DUFLO, M., SENOUSSI, R., AND TOUATI, A. (1990). Sur la loi des grands nombres pour les martingales vectorielles et l'estimateur des moindres carrés d'un modèle de régression. *Annales de l'I.H.P.* série B, **26**, 549–566.

**[13** ] DUFLO, M., SENOUSSI, R., AND TOUATI, A. (1991). Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Annales de l'I.H.P.* série B, **27**, 1–25.

**[14** ] HALL, P., AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Probability and Mathematical Statistics, Academic Press, New York.

**[15** ] HALLIN, M. (1996). Eléments de la théorie asymptotique des expériences statistiques. In *Inférence non paramétrique. Les statistiques de rangs* (J.-J.Droesbeke and J. Fine, Eds.), pp. 129–166. Association pour la Statistique et ses Utilisations. Editions of Brussels University, Brussels, and Editions Ellipses, Paris.

**[16** ] KLIMKO, L. A., AND NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics* **6**, 3, 629–642.

**[17** ] Kreis, J.-P. (1987). On adaptative estimation in stationary ARMA processes. *The Annals of Statistics* **1**, 112–133.

**[18** ] Lai, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression. *The Annals of Statistics* **22**, 1917–1930.

**[19** ] Le Cam, L., and Yang, G. L.(1990). *Asymptotics in Statistics. Some Basic Concepts.* Springer series in Statistics. Springer-Verlag, New York.

**[20** ] Mangeas, M., and Yao, J. F. (1996). Sur l'estimateur des moindres carrés d'un modèle autorégressif fonctionnel. *Notes aux Comptes Rendus de l'Académie des Sciences de Paris* série I, **324**.

**[21** ] Mangeas, M., and Yao, J. F. (1997). Sur l'estimateur des moindres carrés d'un modèle autorégressif fonctionnel. Technical report of the SAMOS # **67**, University of Paris-I, France. Submitted.

**[22** ] McKeague, I. W., and Zhang, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *The Annals of Statistics* **22**, 495–514.

**[23** ] Ngatchou Wandji, J., and Laïb, N. (1998). Limiting distribution of weighted processes of residuals. Application to testing time series. Technical report of the University of Caen, France. Submitted.

**[24** ] Pollard, D. (1997). Another look at differentiability in quadratic mean. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, Eds.) Research papers in Probability and Statistics. Springer-Verlag, New York.

**[25** ] Stephens, M. A. (1978). On the half-sample method for goodness-of-fit. *Journal of the Royal Statistical Society* **40**, 64–70.

**[26** ] Stute, W., and González Manteiga, W. (1996). NN goodness-of-fit tests for linear models. *Journal of Statistical Planning and Inference* **53**, 75–92.

**[27** ] Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics* **25**, 613–641.

**[28** ] Swensen, A. R. W. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16**, 54 – 70.

**[29** ] Su, J. Q., and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association* **86**, 420–426.

**[30** ] Zuber, J. (1996). Quelques tests non paramétriques d'absence d'effet (Testing for no effect in regression). Technical report # **96.2**, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

**[31** ] Zuber, J. (1997). Tester en régression des modèles paramétriques avec estimation des paramètres, cas multidimensionnel (Testing parametric regression models, multivariate case). Technical report # **97.1**, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

# 7 Appendix: Proofs of the results

## 7.1 Sketch of the proof of Theorem 1

The proof follows as in Mangeas and Yao (1997), with the help of the CLT in subsection 7.5. The basic ingredient is that for all $\psi \in L^{2+\gamma'}(dF)$ the sequence of partial sums $\sum_{t=1}^{n} b_t \, \psi(X_t) \, \varepsilon_t$ is a martingale with increasing process $\kappa n \int_{-\infty}^{\infty} \psi^2 \, dF$ with respect to the filtration $(\mathcal{H}_t : t \geq 1)$, since $\{b_t : t \geq 1\}$ is assumed independent of $\mathcal{G}$. We have

$$\sum_{t=1}^{n} b_t \, (Y_t \, - \, m(X_t; \theta)) \, \nabla m \, (X_t; \theta) \; = \; 0 \quad \text{ for } \theta \; = \; \hat{\theta}_n^{(\kappa)}. \tag{7.1}$$

Therefore, by Taylor expanding $m(X_t;\,\theta)$ and $\nabla m(X_t;\,\theta)$ around $\theta_0$ we obtain that

$$n^{-1/2}\,\sigma\sum_{t=1}^{n} b_t\,\varepsilon_t\nabla m_0(X_t) \;=\; n^{-1}\sum_{t=1}^{n} b_t\,\nabla m_0(X_t)\nabla m_0^T(X_t)\,\xi_n^{(\kappa)} \;+\; o_P(1). \tag{7.2}$$

Finally, $n^{-1}\sum_{t=1}^{n} b_t\nabla m_0(X_t)\nabla m_0^T(X_t)$ and the variance matrix of $n^{-1/2}\sum_{t=1}^{n} b_t\nabla m_0(X_t)\varepsilon_t$ both converge to $\kappa V_0$. ■

## 7.2 Proof of Theorem 2 and Corollary 1

*Sketch of the proof of Theorem 2 :* The proof is basically similar to the proofs concerning the limiting distribution of $\hat{B}_n(\cdot)$ in Zuber (1996) (who tests for no effect in the regression case with iid noise), Zuber (1997) (who treats the regression case with iid noise), Stute (1997) (who mainly treats the regression case with iid noise and regression function of the form $m(x;\,\theta) \;=\; \sum_{j=1}^{k}\theta_j g_j(x)$), and Ngatchou Wandji and Laïb (1998) (who treat the autoregression case with stationary ergodic $\{X_t\}$ and martingale difference noise $\{\varepsilon_t\}$). Therefore, we give a very short account of the proof, only insisting on the modifications related to the substitution of $\hat{\theta}_n$ with $\hat{\theta}_n^{(\kappa)}$. The process $\hat{B}_n^{(\kappa)}$ can be uniformly approximated by

$$B_n(x) \;-\; n^{-1}\sum_{t=1}^{n}\nabla m_0^T\,(X_t)\;\xi_n^{(\kappa)}\,I\,(X_t\le x)\,,$$

and the latter expression can in turn be uniformly approximated by $B_n(x) \;-\; g_0^T(x)\xi_n^{(\kappa)}$. These uniform approximations result from:

(i) A second order Taylor expansion of $m(X_t;\,\theta)$ around $\theta_0$ and assumptions (A4)–(A5).

(ii) The following extension of the version of Dini's Lemma which is used to extablish the Glivenko-Cantelli Theorem (e.g., Billingsley [1968, page 103]).

**Lemma 1** *Let $\mathbf{F}_n(x) \;=\; n^{-1}\sum_{t=1}^{n} I\,(X_t\le x)$ denote the empirical distribution function of the n-sample $\{X_1,\ldots,X_n\}$ of the stationary ergodic process $\{X_t\}$ with continuous and increasing stationary distribution function $F$. For each function $\Lambda\,:\,\mathbf{R}\,\longrightarrow\,\mathbf{R}^p$ such that $\int_{-\infty}^{\infty}\|\Lambda\|\;dF\;<\;\infty$, the processes $\int_{-\infty}^{x}\Lambda\;\,d\mathbf{F}_n$ converge a.s. to $\int_{-\infty}^{x}\Lambda\,dF$ uniformly in $x$.*

*Proof of Lemma 1 :* It is enough to prove this result in the case $p=1$. By splitting $\Lambda$ into the difference of nonnegative functions $\Lambda^+-\Lambda^-$, we may suppose without loss of generality that $\Lambda\ge 0$. Then the function

$$u\in[0,\,1]\;\;\longrightarrow\;\;\int_0^u\,\left(\Lambda\circ F^{-1}\right)(v)\,dv\;=\;\int_{-\infty}^{F^{-1}(u)}\,\Lambda\,dF$$

is continuous and nondecreasing. Moreover, by the stationarity and ergodicity of the sequence $\{X_t\}$ and integrability of $\Lambda$ with respect to $dF$, the nondecreasing functions $\int_0^u(\Lambda\circ F^{-1})\,d\mathbf{H}_n$, where $\mathbf{H}_n(\cdot)$ denotes the empirical distribution function of $\{U_1=F(X_1),\ldots,U_n=F(X_n)\}$, converge pointwise to the continuous function $\int_0^u(\Lambda\circ F^{-1})(v)\,dv$ a.s. for $u$ in the compact interval $[0,\,1]$. By making use of the above-mentioned version of Dini's Lemma it results that this convergence is uniform a.s. ■

This lemma is used to prove that $\int_{-\infty}^{x}M\,d\mathbf{F}_n$ and $\int_{-\infty}^{x}\nabla m_0\,d\mathbf{F}_n$ converge to $\int_{-\infty}^{x}M\,dF$ and $\int_{-\infty}^{x}\nabla m_0\,dF$ uniformly in $x$, respectively, since by (A4)–(A5) (see (2.11), $\int_{-\infty}^{\infty}M\,dF<\infty$ and $\int_{-\infty}^{\infty}\|\nabla m_0\|\,dF<\infty$.

Therefore, the functional limit as $n\;\rightarrow\;\infty$ of the sequence of processes $\hat{B}_n^{(\kappa)}(\cdot)$ is the same as that of $B_n(\cdot)-g_0(\cdot)^T\xi_n^{(\kappa)}$. Using Theorem 1 of Ngatchou Wandji and Laïb (1998) (which relies on a version of the CLT for martingale difference arrays, see, e.g., Hall and Heyde [1980, Corollary 3.1, pages 58–59] and subsection 7.5 below) it follows that the $\mathbf{R}^{1+K}$-valued $(\mathcal{H}_t)$-martingale array $(B_n(x),\, \xi_n^{(\kappa)})$ converges in distribution to a normal rv with mean 0 and variance matrix

$$\begin{pmatrix} \sigma^2 F(x) & h_0^T(x) \\ h_0(x) & \kappa^{-1}\sigma^2 V_0 \end{pmatrix}$$

for each $x$, whereas the corresponding finite-dimensional distributions also converge to multivariate normal distributions. It remains essentially to establish the tightness of the sequence $\{B_n(\cdot)\}$. We make use of the quantile transformation in order to deal with rv's uniformly distributed over the unit interval [0, 1], and functions and processes defined on [0, 1], as in the proof of Lemma 1. Let $B_n^\star = B_n \circ F^{-1}$. By an extension of Billingsley (1968) used, e.g., in McKeague and Zhang [1994, page 507], it is enough to prove that we have, for some $\rho > 2$,

$$E\left(|B_n^\star(u_2) \;-\; B_n^\star(u_1)|^\rho\right) \;\leq\; \mu_B^{\rho/2}\left([u_1,\, u_2]\right) \;+\; o(1) \quad \text{as } n \to \infty$$

uniformly in $0 \leq u_1 < u_2 \leq 1$, for some nonnegative continuous measure $\mu_B(\cdot)$. This can be done using Rosenthal's inequality (e.g., Hall and Heyde [1980, Theorem 2.12, pages 23–24]). For more details, see subsection 7.6 below. It follows that the process

$$B_n(x) \;-\; g_0(x)\xi_n^{(\kappa)} \;=\; \left(B_n(x),\, \xi_n^{(\kappa)}\right)(1,\, -g_0(x))^T$$

converges in distribution to the Gaussian process $B(\cdot) - g_0(\cdot)\xi^{(\kappa)}$ with covariance function

$$\sigma^2 F\left(x_1 \wedge x_2\right) \;-\; E\left(B(x_1) g_0(x_1)^T \xi^{(\kappa)}\right) \;-\; E\left(B(x_2) g_0(x_2)^T \xi^{(\kappa)}\right) \;+\; \kappa^{-1}\sigma^2 g_0(x_1)^T V_0^{-1} g_0(x_2).$$

■

## 7.3 Proof of Theorem 3

For simplicity we will denote $\nu_\theta = m_1 - m_\theta$. The process $\hat{A}_n^{(\kappa)}$ can be written in the form

$$\begin{aligned} \hat{A}_n^{(\kappa)}(x) &= B_n(x) \;-\; n^{-1/2} \sum_{t=1}^{n} \left(m\left(X_t;\, \hat{\theta}_n^{(\kappa)}\right) \;-\; m_1\left(X_t\right)\right) I\left(X_t \leq x\right) \\ &= B_n(x) \;+\; n^{1/2} \int_{-\infty}^{x} \nu_{\hat{\theta}_n^{(\kappa)}}\, d\mathbf{F}_n. \end{aligned} \tag{7.3}$$

By assumptions (A2)–(A3) we have for all $x$ that

$$\begin{aligned} \left| \int_{-\infty}^{x} \nu_{\theta_1}\, d\mathbf{F}_n \;-\; \int_{-\infty}^{x} \nu_{\theta_2}\, d\mathbf{F}_n \right| &\leq \omega_0\left(\|\theta_1 \;-\; \theta_2\|\right) \int_{-\infty}^{x} \left(1 \;+\; |y|^{\beta/2}\right) d\mathbf{F}_n(y) \\ &\leq \omega_0\left(\|\theta_1 \;-\; \theta_2\|\right) \int_{-\infty}^{\infty} \left(1 \;+\; |y|^{\beta/2}\right) d\mathbf{F}_n(y) \\ &\longrightarrow \omega_0\left(\|\theta_1 \;-\; \theta_2\|\right) \int_{-\infty}^{\infty} \left(1 \;+\; |y|^{\beta/2}\right) dF(y) \end{aligned} \tag{7.4}$$

almost surely. Since $N(\cdot) \leq \text{cst}\, \|\cdot\|_\infty$, setting

$$L_n \;=\; \int_{-\infty}^{\infty} \left(1 \;+\; |y|^{\beta/2}\right) d\mathbf{F}_n(y) \quad \text{and} \quad L \;=\; \int_{-\infty}^{\infty} \left(1 \;+\; |y|^{\beta/2}\right) dF(y) \;<\infty,$$

it follows that

$$N\left(\int_{-\infty}^{\bullet} \nu_{\theta_1}\, d\mathbf{F}_n \;-\; \int_{-\infty}^{\bullet} \nu_{\theta_2}\, d\mathbf{F}_n\right) \;\;\le\;\; \omega_0\left(\|\theta_1 \,-\, \theta_2\|\right) \,\mathrm{cst}\, L_n$$
$$\longrightarrow \;\; \omega_0\left(\|\theta_1 \,-\, \theta_2\|\right) \,\mathrm{cst}\, L \;\text{ a.s.} \tag{7.5}$$

Therefore, $\theta \longrightarrow N(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n)$ and $\theta \longrightarrow N(\int_{-\infty}^{\bullet} \nu_\theta\, dF)$ are continuous. By compacity of $\Theta$ and making use of Lemma 1 we obtain that

$$\lim_{n\to\infty}\, \sup_{\theta\in\Theta}\, N\left(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n \;-\; \int_{-\infty}^{\bullet} \nu_\theta\, dF\right) \;=\; 0 \quad \text{a.s.} \tag{7.6}$$

Finally,

$$\lim_{n\to\infty}\, \inf_{\theta\in\Theta}\, N\left(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n\right) \;=\; \inf_{\theta\in\Theta}\, N\left(\int_{-\infty}^{\bullet} \nu_\theta\, dF\right) \quad \text{a.s.} \tag{7.7}$$

If $\inf_{\theta\in\Theta} N(\int_{-\infty}^{\bullet} \nu_\theta\, dF) \;>\; 0$ then a.s. $\inf_{\theta\in\Theta} N(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n) \;>\; 0$ for all $n$ large enough. Suppose that $\inf_{\theta\in\Theta} N(\int_{-\infty}^{\bullet} \nu_\theta\, dF) > 0$. For all $n$ large enough,

$$\begin{aligned}
N\left(B_n \;+\; n^{1/2}\int_{-\infty}^{\bullet} \nu_{\hat\theta_n^{(\kappa)}}\, d\mathbf{F}_n\right) \;&\ge\; \left| n^{1/2}\, N\left(\int_{-\infty}^{\bullet} \nu_{\hat\theta_n^{(\kappa)}}\, d\mathbf{F}_n\right) \;-\; N\left(B_n\right)\right| \\
&=\; n^{1/2}\, N\left(\int_{-\infty}^{\bullet} \nu_{\hat\theta_n^{(\kappa)}}\, d\mathbf{F}_n\right) \;-\; N\left(B_n\right) \\
&\ge\; n^{1/2}\, N\left(\int_{-\infty}^{\bullet} \nu_{\hat\theta_n^{(\kappa)}}\, d\mathbf{F}_n\right) \;-\; N\left(B_n\right) \\
&\ge\; n^{1/2}\, \inf_{\theta\in\Theta}\, N\left(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n\right) \;-\; N\left(B_n\right),
\end{aligned} \tag{7.8}$$

hence

$$P\left\{N\left(B_n \;+\; n^{1/2}\int_{-\infty}^{\bullet} \nu_{\hat\theta_n^{(\kappa)}}\, d\mathbf{F}_n\right) \;\le\; q_\alpha\right\} \;\;\le\;\; P\left\{N\left(B_n\right) \;\ge\; n^{1/2}\, \inf_{\theta\in\Theta}\, N\left(\int_{-\infty}^{\bullet} \nu_\theta\, d\mathbf{F}_n\right) \;-\; q_\alpha\right\},$$

which is asymptotic to $P\left\{\, N(B) \;\ge\; n^{1/2}\, \inf_{\theta\in\Theta}\, N(\int_{-\infty}^{\bullet} \nu_\theta\, dF)\,\right\}$ as $n \to \infty$. ∎

## 7.4 Proof of Theorem 4

(i) Regression case. Under $H_{1,\,n}$, the process $\hat A_n^{(\kappa)}$ takes the form

$$\hat A_n^{(\kappa)}(x) \;=\; \hat B_n^{(\kappa)}(x) \;+\; n^{-1}\sum_{t=1}^{n} \delta\left(X_t\right) I\left(X_t \le x\right), \qquad x \in \mathbf{R}, \tag{7.9}$$

where $\hat B_n^{(\kappa)}$ is defined in (3.2)–(3.3). In the regression case, the only modification is in the asymptotic distribution of $\hat\theta_n^{(\kappa)}$ under $H_{1,\,n}$. An inspection of the first part of the proof of Mangeas and Yao (1997) shows that under $H_{1,\,n}$, $\hat\theta_n^{(\kappa)}$ still converges a.s. to $\theta_0$. Moreover, by Taylor expanding $m(X_t;\,\theta)$ and $\nabla m(X_t;\,\theta)$ around $\theta_0$ we now obtain that

$$n^{-1}\sum_{t=1}^{n} b_t\, \nabla m_0\left(X_t\right) \nabla m_0^T\left(X_t\right) \xi_n^{(\kappa)} \;=\; n^{-1/2}\sigma\sum_{t=1}^{n} b_t\, \varepsilon_t\, \nabla m_0\left(X_t\right) \;+\; n^{-1}\sum_{t=1}^{n} b_t\, \delta\left(X_t\right) \nabla m_0\left(X_t\right) \;+\; o_P(1).$$

Since $n^{-1}\sum_{t=1}^n b_t\,\delta(X_t)\,\nabla m_0(X_t)$ converges a.s. to $\kappa\int_{-\infty}^{\infty}\delta\,\nabla m_0\,dF$ and $\sum_{t=1}^n b_t\,\nabla m_0(X_t)\nabla m_0^T(X_t)$ converges a.s. to the matrix $\kappa V_0$, it follows that

$$\xi_n^{(\kappa)} \;=\; n^{1/2}\left(\hat{\theta}_n^{(\kappa)} \,-\, \theta_0\right) \quad \text{converges in distribution to} \quad \xi^{(\kappa)} \,+\, \mu_0$$

under $H_{1,\,n}$.

(ii) AUTOREGRESSION CASE WITH IID $\varepsilon_t$'S. The assumption that $\delta$ is bounded basically ensures that the homogeneous Markov chain $\{X_t\}$ remains ergodic. Its stationary distribution $F = F_{\delta,\,n}$ now depends on the additional perturbation term $n^{-1/2}\delta(\cdot)$. Therefore, we need a contiguity result to proceed. However, we are not in a standard situation where the parameter is shifted. On the contrary, here the parameter has a fixed value, $\theta_0$. In view of (6.1), the shift is rather *functional* than parametric. More precisely, the shift

$$m_{\theta_0} \;\longrightarrow\; m_{\theta_0} \,+\, n^{-1/2}\,\sigma^{-1}\,\delta$$

is effective only if $\delta$ is not in the tangent space of the model $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ at $\theta_0$, i.e. if $\delta$ is not in the subspace spanned by the functions $\partial_j m_0$, $j = 1, \ldots, K$. Typically, we will choose $\delta$ *orthogonal* in some sense to this tangent space. Since the idea of likelihood ratio has no precise meaning in the present setting, we have to somehow revisit the usual application scheme of mean-square differentiability, LAN and contiguity theory. We begin with a change of measure and the related Girsanov type formula for (not necessarily Gaussian) dicrete-time processes. By equation (6.1), $X_t$ is $\sigma(X_0, \varepsilon_1, \ldots, \varepsilon_t)$-measurable for $t \geq 1$.

**Lemma 2** *Let*

$$a_0 \;=\; cst \;\; and \;\; a_t \;=\; a_t\,(e_1, \ldots, e_t)\,, \quad t \geq 1, \tag{7.10}$$

*denote real-valued measurable functions defined on the product space* $\mathbf{R}^{\mathbf{N}^*}$ *equipped with its cylindrical* $\sigma$*-field. Let* $\Pi_n$ *denote the product measure*

$$\Pi_n\,(de_1, \ldots, de_n) \;=\; p\,(e_1) \ldots p\,(e_n)\; de_1 \ldots de_n.$$

*and* $\Pi_{a,\,n}$ *denote*

$$\Pi_{a,\,n}\,(de_1, \ldots, de_n) \;=\; \exp\,(\Lambda_n)\,\Pi_n\,(de_1, \ldots, de_n)\,,$$

*where*

$$\Lambda_n \;=\; \sum_{t=1}^n\,[\ln p\,(e_t \,-\, a_{t-1}\,(e_1, \ldots, e_{t-1})) \,-\, \ln p\,(e_t)]\,. \tag{7.11}$$

*Let* $T^{(n)}$ *denote the measurable transformation*

$$T^{(n)}\,(e_1, \ldots, e_n) \;=\; (e_1 \,+\, a_0, \ldots, e_n \,+\, a_{n-1}\,(e_1, \ldots, e_{n-1}))\,. \tag{7.12}$$

*Then,* $(e_1, \ldots, e_n)$ *has the same distribution under* $\Pi_{a,\,n}$ *as* $T^{(n)}\,(e_1, \ldots, e_n)$ *under* $\Pi_n$.

*Proof of Lemma 2* : It is essentially an application of Fubini's Theorem. Consider bounded measurable functions $h_1, h_2$ and $h(e_1, e_2) = h_1(e_1)\;h_2(e_2)$. We have

$$\begin{aligned}
&\int\!\!\int h_1(e_1)\,h_2(e_2)\,\Pi_{a,\,2}(de_1,\,de_2)\\
&= \int\!\!\int h_1(e_1)\,h_2(e_2)\,p(e_1 \,-\, a_0)\,p(e_2 \,-\, a_1(e_1))\,de_1 de_2\\
&= \int h_1(e_1)\,p(e_1 \,-\, a_0)\left(\int h_2(e_2)\,p(e_2 \,-\, a_1(e_1))\,de_2\right)\,de_1
\end{aligned}$$

$$
\begin{aligned}
&= \int h_1(e_1)\, p(e_1 - a_0) \left( \int h_2(e_2' + a_1(e_1))\, p(e_2')\, de_2' \right) de_1 \\
&= \int h_1(e_1' + a_0) p(e_1') \left( \int h_2(e_2' + a_1(e_1'))\, p(e_2')\, de_2' \right) de_1' \\
&= \int\int h_1(e_1' + a_0)\, h_2(e_2' + a_1(e_1' + a_0))\, p(e_1')\, p(e_2')\, de_2'\, de_1' \\
&= \int\int h_1(e_1 + a_0)\, h_2(e_2 + a_1(e_1))\, \Pi_2(de_1, de_2) \\
&= \int\int h\,(e_1 + a_0,\, e_2 + a_1(e_1 + a_0))\, \Pi_2(de_1, de_2).
\end{aligned} \tag{7.13}
$$

We have a similar sequence of equalities for $h(e_1, \ldots, e_n) = h_1(e_1) \ldots h_n(e_n)$, where $h_1, \ldots, h_n$ are bounded measurable functions. The result follows by density. ■

As a consequence, if we set

$$
dP_{\delta,\, n} \;=\; \exp\left(\Lambda_{\delta,\, n}\right) dP_n, \tag{7.14}
$$

where

$$
\Lambda_{\delta,\, n} \;=\; \sum_{t=0}^{n} \left[ \ln p\left(\varepsilon_{t+1} \,-\, n^{-1/2}\, \sigma^{-1}\, \delta\,(X_t)\right) \,-\, \ln p\,(\varepsilon_{t+1}) \right], \tag{7.15}
$$

then

$$
\begin{aligned}
&P_{\delta,\, n}\left\{(\varepsilon_1, \ldots, \varepsilon_{n+1}) \in \mathbf{A}_n\right\} \\
&\quad = \; P_n\left\{ \left(\varepsilon_1 + n^{-1/2}\, \sigma^{-1}\, \delta\,(X_0), \ldots, \varepsilon_{n+1} + n^{-1/2}\, \sigma^{-1}\, \delta\,(X_n)\right) \in \mathbf{A}_n \right\}.
\end{aligned} \tag{7.16}
$$

Therefore, the process $\{X_t\}$ defined by (6.1) has the same distribution under the probability measure $P_n$ as the process $\{X_t\}$ defined by (1.1) has under the probability measure $P_{\delta,\,n}$. Set $h_{t,\,n} = n^{-1/2}\sigma^{-1}\delta(X_t)$. Under the regularity conditions and (i)–(iii), $E(\phi_p(\varepsilon_t)) = 0$ for all $t \geq 1$ and the martingale array

$$
n^{-1/2} \sum_{t=1}^{n} \delta\,(X_t)\, \phi_p\,(\varepsilon_{t+1})
$$

satisfies the CLT (see subsection 7.5) and converges in distribution to $\mathcal{N}(0, I_p \int_{-\infty}^{\infty} (\delta/\sigma)^2\, dF)$, where $I_p$ denotes the Fisher information $\int_{-\infty}^{\infty} \phi_p(x)^2\, p(x)\, dx$. Moreover (e.g., Swensen, 1985, Kreis, 1987, Hallin, 1996 and Pollard, 1997), we have for all $\tau$ that

$$
\begin{aligned}
\Lambda_{\tau\delta,\, n} \;&=\; \sum_{t=0}^{n} \left[\ln p\,(\varepsilon_{t+1} - \tau h_{t,\, n}) \,-\, \ln p\,(\varepsilon_{t+1}]\right) \\
&=\; -\,\tau \sum_{t=0}^{n} h_{t,\, n}\, \phi_p\,(\varepsilon_{t+1}) \,+\, \frac{\tau^2\, \lambda^2}{2} \,+\, o_P(1) \\
&=\; -\,\tau\, \sigma^{-1}\, n^{-1/2} \sum_{t=0}^{n} \delta\,(X_t)\; \phi_p\,(\varepsilon_{t+1}) \,+\, \frac{\tau^2\, \lambda^2}{2} \,+\, o_P(1) \\
&\longrightarrow\; \mathcal{N}\left(-\frac{\tau^2\, \lambda^2}{2},\, \tau^2\, \lambda^2\right),
\end{aligned} \tag{7.17}
$$

where

$$
\lambda^2 \;=\; I_p \int_{-\infty}^{\infty} (\delta/\sigma)^2\, dF \;>\; 0. \tag{7.18}
$$

By Le Cam's first lemma (e.g., Le Cam and Yang [1990, page 20] or Hallin [1996, page 150]), this implies that the two sequences of probability distributions $\{P_{\delta,n}\}$ and $\{P_n\}$ are contiguous. Now, denote by $\psi^{(\kappa)}(x) = (\psi_1^{(\kappa)}(x), \ldots, \psi_d^{(\kappa)}(x))^T$ an $\mathbf{R}^d$-valued function such that

$$\int_{-\infty}^{\infty} \left|\psi_j^{(\kappa)}\right|^{2+\gamma'} dF \;<\; \infty \;\text{ for all } j = 1, \ldots, d$$

and let

$$\Psi_n^{(\kappa)} \;=\; n^{-1/2} \sum_{t=1}^{n} b_t\, \psi^{(\kappa)}\,(X_t)\, \varepsilon_{t+1}. \tag{7.19}$$

**Lemma 3** *Suppose that (A1)–(A6) and the assumptions of Theorem 4 (2) hold. Then under* $H_0$, $(\Lambda_n,\, \Psi_n^{(\kappa)})$ *converges in distribution to* $\mathcal{N}((-\lambda^2/2,\, 0)^T;\, \Gamma^{(\kappa)})$, *where*

$$\Gamma^{(\kappa)} \;=\; \begin{pmatrix} \Gamma_{1,1}^{(\kappa)} & \Gamma_{1,2}^{(\kappa)\,T} \\ \Gamma_{1,2}^{(\kappa)} & \Gamma_{2,2}^{(\kappa)} \end{pmatrix},$$

*with* $\Gamma_{1,1}^{(\kappa)} = \lambda^2$,

$$\Gamma_{1,2}^{(\kappa)} \;=\; \sigma^{-1}\, \kappa \int_{-\infty}^{\infty} \delta\, \psi^{(\kappa)}\, dF \;\; \textit{and} \;\; \Gamma_{2,2}^{(\kappa)} \;=\; \kappa \int_{-\infty}^{\infty} \psi^{(\kappa)} \psi^{(\kappa)\,T}\, dF.$$

*Proof of Lemma 3* : This is a direct consequence of the CLT for arrays of martingales (subsection 7.5) and of the equalities

$$\Lambda_{\delta,n} \;=\; -\,\sigma^{-1}\, n^{-1/2} \sum_{t=1}^{n} \delta\,(X_t)\; \phi_p\,(\varepsilon_{t+1}) \;+\; \lambda^2/2 \;+\; o_P(1)$$

and $E(\varepsilon_t \phi_p(\varepsilon_t)) = -1$ for all $t$. ■

Since

$$\xi_n^{(\kappa)} \;=\; \kappa^{-1}\, V_0^{-1}\, n^{-1/2}\, \sigma \sum_{t=1}^{n} b_t\, \nabla m_0(X_t)\, \varepsilon_{t+1} \;+\; o_P(1),$$

and taking $\psi^{(\kappa)}(x) = \kappa^{-1} V_0^{-1} \sigma \nabla m_0(x)$ in Lemma 3, it follows that $(\Lambda_n, \xi_n^{(\kappa)})$ converges in distribution to $\mathcal{N}(\mu_0, \Gamma^{(\kappa)})$ under $H_0$, where

$$\mu_0 \;=\; \Gamma_{1,2}^{(\kappa)} \;=\; V_0^{-1} \int_{-\infty}^{\infty} \delta\, \nabla m_0\, dF \tag{7.20}$$

does not depend on $\kappa$, and $\Gamma_{2,2}^{(\kappa)} \;=\; \kappa^{-1}\, \sigma^2\, V_0^{-1}$. It then follows from Le Cam's third lemma (e.g., Hallin [1996, page 151] or Pollard, 1997) that under $H_1^n$, $\xi_n^{(\kappa)}$ converges in distribution to $\mathcal{N}(\mu_0, \kappa^{-1}\sigma^2 V_0^{-1})$. Finally, to prove that the sequence of processes $\{B_n\}$ converges under $H_1^n$ to the Gaussian process $B + \int_{-\infty}^{\bullet} \delta\, dF$, it suffices to study the finite-dimensional distributions and the tightness of $B_n$ under $H_1^n$. This can be done by using Lemmas 2–3 with $\kappa = 1$. ■

### 7.5 CLT for an array of martingales

The conditions required by Hall and Heyde [1980, Corollary 3.1, pages 58–59] for having a CLT for an array of martingales are expressed in terms of the corresponding martingale difference array

$$\eta_{n,t} \;=\; n^{-1/2}\, b_t\, \ell\,(X_t)\; \varepsilon_t, \qquad 1 \le t \le n, \tag{7.21}$$

where $\int_{-\infty}^{\infty} |\ell|^{2+\gamma'}\, dF < \infty$. The first condition is a conditional Lindeberg condition. The second condition involves the limit as $n \to \infty$ of the sums from $t = 1$ to $t = n$ of the conditional variances $E(\eta^2_{n,t} \,|\, \mathcal{H}_{t-1})$. We only examine the technical proof of the conditional Lindeberg condition. This condition has the following form. For all $\beta > 0$,

$$\sum_{t=1}^{n} E\left(\eta^2_{n,t} I\left\{|\eta_{n,t}| \;>\; \beta\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \;\to\; 0 \quad \text{a.s. as } n \to \infty.$$

In order to check this condition we have to prove for all $\beta > 0$ that

$$n^{-1} \sum_{t=1}^{n} E\left(b_t^2 \ell^2\,(X_t)\, \varepsilon_t^2 I\left\{|\ell\,(X_t)\,|\,|\varepsilon_t| \;>\; n^{1/2}\beta\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \;\to\; 0 \quad \text{a.s. as } n \to \infty. \tag{7.22}$$

To prove this result, we make use of a conditional version of Hölder's inequality. Let $a > 1$ and $b > 1$ be conjugate exponents, $a^{-1} + b^{-1} = 1$, such that $2a \le 2 + \gamma$, implying that

$$\sup_{t \ge 1} E\left(|\varepsilon_t|^{2a} \,\middle|\, \mathcal{H}_{t-1}\right) \;<\; \infty,$$

in view of (A1). We have

$$\begin{aligned}
& E\left(b_t^2\, \ell^2\,(X_t)\, \varepsilon_t^2 I\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \\
& \le\; \ell^2\,(X_t)\; E^{1/a}\left(|\varepsilon_t|^{2a} \,\middle|\, \mathcal{H}_{t-1}\right)\, E^{1/b}\left(I\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \\
& \le\; \text{cst}\, \ell^2\,(X_t)\; E^{1/b}\left(I\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right),
\end{aligned} \tag{7.23}$$

since $0 \le b_t \le 1$. Everything then boils down to the following sequence of inequalities:

$$\begin{aligned}
E\left(I\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \;&=\; P\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2} \,\middle|\, \mathcal{H}_{t-1}\right\} \\
&\le\; \beta^{-2}\, n^{-1}\, E\left(b_t^2\, \ell^2\,(X_t)\, \varepsilon_t^2 \,\middle|\, \mathcal{H}_{t-1}\right) \\
&=\; \beta^{-2}\, n^{-1}\, \kappa\, \ell^2\,(X_t)\, E\left(\varepsilon_t^2 \,\middle|\, \mathcal{H}_{t-1}\right) \\
&\le\; \text{cst}\; n^{-1}\, \ell^2\,(X_t)
\end{aligned} \tag{7.24}$$

by independence of $b_t$ and $\mathcal{G}$. Thus,

$$E^{1/b}\left(I\left\{|b_t\, \ell\,(X_t)|\; |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \;\;\le\;\; \text{cst}\; n^{-1/b}\; |\ell\,(X_t)|^{2/b} \tag{7.25}$$

and

$$\begin{aligned}
n^{-1} \sum_{t=1}^{n} E\left(b_t^2\, \ell^2\,(X_t)\, \varepsilon_t^2 I\left\{|b_t\, \ell\,(X_t)|\, |\varepsilon_t| > \beta n^{1/2}\right\} \,\middle|\, \mathcal{H}_{t-1}\right) \;\;&\le\;\; \text{cst}\, n^{-1/b}\, n^{-1} \sum_{t=1}^{n} |\ell\,(X_t)|^{2\,+\,2/b} \\
&\longrightarrow\; 0 \quad \text{a.s. as } n \to \infty,
\end{aligned} \tag{7.26}$$

since $n^{-1} \sum_{t=1}^{n} |\ell(X_t)|^{2\,+\,2/b}$ converges a.s. to $\int_{-\infty}^{\infty} |\ell|^{2\,+\,2/b}\, dF$ and this integral is finite provided that $1 < a \le 1 + \gamma/2$ has been chosen so close to 1 that $2 + 2/b \le 2 + \gamma'$. ■

## 7.6 Tightness of $\{B_n\}$

Consider a process of the form

$$\Psi_n(u) \;=\; n^{-1/2}\,\psi\,(U_t)\;I\,(U_t \le u)\;\varepsilon_t, \tag{7.27}$$

with $\int_{-\infty}^{\infty} |\psi|^{2+\gamma''} < \infty$, $0 < \gamma'' \le \gamma$. We have (Rosenthal's inequality)

$$\begin{aligned} E\,(|\Psi_n(u_2) \;-\; \Psi_n(u_1)|^{\rho}) \;&=\; E\left(\left|\sum_{t=1}^{n} n^{-1/2}\psi\,(U_t)\,I\,(u_1 < U_t \le u_2)\,\varepsilon_t\right|^{\rho}\right) \\ &\le\; c_{Ros}\,E\left[\left(n^{-1}\sum_{t=1}^{n}\psi^2\,(U_t)\,I\,(u_1 < U_t \le u_2)\right)^{\rho/2}\right] \\ &+\; c_{Ros}\sum_{t=1}^{n} E\left[n^{-\rho/2}\,|\psi\,(U_t)|^{\rho}\,I\,(u_1 < U_t \le u_2)\,|\varepsilon_t|^{\rho}\right]. \end{aligned} \tag{7.28}$$

We now select $2 \;<\; \rho \;\le\; 2 \;+\; \gamma''$. We then have

$$\begin{aligned} E\,(|\psi\,(U_t)|^{\rho}\,I\,(u_1 < U_t \le u_2)\,|\varepsilon_t|^{\rho}) \;&=\; E\,(\,E\,(\,|\psi\,(U_t)|^{\rho}\,I\,(u_1 < U_t \le u_2)\,|\varepsilon_t|^{\rho} \mid \mathcal{H}_{t-1})\,) \\ &=\; E\,(\,|\psi\,(U_t)|^{\rho}\,I\,(u_1 < U_t \le u_2)\,E\,(\,|\varepsilon_t|^{\rho} \mid \mathcal{H}_{t-1})\,) \\ &\le\; \text{cst}\int_0^1 |\psi\,(u)|^{\rho}\,du \;<\; \infty. \end{aligned} \tag{7.29}$$

Moreover, since $\rho > 2$ it follows that $n^{1-\rho/2} \;=\; o(1)$. Therefore, everything boils down to determining a suitable upper bound for

$$n^{-\rho/2}E\left[\left(\sum_{t=1}^{n}\psi^2\,(U_t)\,I\,(u_1 < U_t \le u_2)\right)^{\rho/2}\right] \tag{7.30}$$

with $\rho/2 > 1$. Let us denote

$$\mathbf{H}_n^{\psi}(u) \;=\; n^{-1}\sum_{t=1}^{n}\psi^2\,(U_t)\,I\,(U_t \le u)\,, \qquad u \in [0,\,1]. \tag{7.31}$$

The expectation (7.30) can be written as

$$\begin{aligned} &E\left(\left|\,\mathbf{H}_n^{\psi}\,(u_2) \;-\; \mathbf{H}_n^{\psi}\,(u_1)\,\right|^{\rho/2}\right) \\ &=\; E\left(\left|\,\mu^{\psi}\,([u_1,\,u_2]) \;+\; \left[\left(\mathbf{H}_n^{\psi} - H^{\psi}\right)(u_2) \;-\; \left(\mathbf{H}_n^{\psi} - H^{\psi}\right)(u_1)\right]\,\right|^{\rho/2}\right) \\ &\le\; E\left(\left|\,\mu^{\psi}\,([u_1,\,u_2]) \;+\; R_n\,\right|^{\rho/2}\right) \\ &\le\; c_{\rho/2}\,\left|\,\mu^{\psi}\,([u_1,\,u_2])\,\right|^{\rho/2} \;+\; c_{\rho/2}\,E\left(R_n^{\rho/2}\right), \end{aligned} \tag{7.32}$$

where

$$\mu^{\psi}\,(A) \;=\; \int_A \psi^2\,(v)\;dv \quad \text{and} \quad H^{\psi}\,(u) \;=\; \int_0^u \psi^2\,(v)\;dv \tag{7.33}$$

are well-defined and $H^{\psi}$ is continuous since $\int_{-\infty}^{\infty}\psi^2\,dF < \infty$. We have made use of the fact that for all for any $q > 1$ there exists a positive constant $c_q$ such that for all positive $w$ and $w'$ we have $(w + w')^q \le c_q\,(w^q + w'^{\,q})$. Moreover,

$$R_n \;=\; \sup_{u\in[0,\,1]}\left|\mathbf{H}_n^{\psi}(u) \;-\; H^{\psi}(u)\right|$$

is bounded by 1 and converges to 0 by ergodicity and Dini's Lemma, as in Lemma 1. Finally, taking $\mu_{\Psi} =$ cst $\mu^{\psi}$ yields the result. ∎

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399