Computing and estimating information matrices of weak ARMA models

Yacouba Boubacar Mainassara, Michel Carbon, Christian Francq

Computing and estimating information matrices of weak ARMA models

Y. Boubacar Mainassara[a], M. Carbon[b], C. Francq[a]

[a] *Université Lille III, EQUIPPE-GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France.*
[b] *Université Rennes 2 et Ensai*

Abstract

Numerous time series admit *weak* autoregressive-moving average (ARMA) representations, in which the errors are uncorrelated but not necessarily independent nor martingale differences. The statistical inference of this general class of models requires the estimation of generalized Fisher information matrices. We give analytic expressions and propose consistent estimators of these matrices, at any point of the parameter space. Our results are illustrated by means of Monte Carlo experiments and by analyzing the dynamics of daily returns and squared daily returns of financial series.

Key words: Asymptotic relative efficiency (ARE), Bahadur's slope, Information matrices, Lagrange Multiplier test, Nonlinear processes, Wald test, Weak ARMA models.

1. Introduction

The class of the standard ARMA models with independent errors is often judged too restrictive by practitioners, because they are inadequate for time series exhibiting a nonlinear behavior. Even when the independence assumption is relaxed and the errors are only supposed to be martingale differences, the ARMA models remain often unrealistic because such models postulate that the best predictor is a linear function of the past values.

Email addresses: mailto:yacouba.boubacarmainassara@univ-lille3.fr (Y. Boubacar Mainassara), mailto:carbon.michel@yahoo.ca (M. Carbon), mailto:christian.francq@univ-lille3.fr (C. Francq)

The class of the so-called weak ARMA models with uncorrelated but not necessarily independent errors is much more general and accommodates many nonlinear data-generating processes (see Francq, Roy and Zakoïan, 2005, and the references therein).

For standard ARMA models, it is well known that the asymptotic variance of the least squares estimator (LSE) is of the form $\sigma^2 J_{\theta_0}^{-1}$, where σ^2 is the variance of the errors and J_{θ_0} is an information matrix depending on the ARMA parameter θ_0 (see *e.g.* Brockwell and Davis, 1991). For weak ARMA models, the asymptotic variance of the LSE takes the sandwich form $J_{\theta_0}^{-1} I_{\theta_0} J_{\theta_0}^{-1}$ where I_{θ_0} is a second information matrix depending on θ_0 and on fourth-order moments of the errors. The estimation of the asymptotic information matrices J_{θ_0} and I_{θ_0} is thus necessary to evaluate the asymptotic accuracy of the LSE of weak ARMA models.

In the framework of (Gaussian) linear processes, the problem of computing the Fisher information matrices and of their inverses has been widely studied. Various expressions of these matrices have been given by Whittle (1953), Siddiqui (1958), Durbin (1959) and Box and Jenkins (1976). McLeod (1984), Klein and Mélard (1990, 2004) and Godolphin and Bane (2006) have given algorithms for their computation. For few particular cases of weak ARMA models, the matrices I_{θ_0} and J_{θ_0} have been computed by Francq and Zakoian (2000, 2007) and Francq, Roy and Zakoian (2005). In all the above-mentioned references, the information matrices are always computed at the true parameter value θ_0. For some applications, in particular to determine Bahadur's slopes under alternatives, it is necessary to compute the information matrices at $\theta \neq \theta_0$.

The aim of the present paper is to compute and estimate the information matrices J_θ and I_θ at a point θ which is not necessarily equal to θ_0.

The rest of the paper is organized as follows. In Section 2, we present the weak, strong and semi-strong ARMA representations and recall results concerning the estimation of the weak ARMA models. Section 3 displays the main results. We describe how to obtain numerical evaluations of I_θ and J_θ, up to some tolerance, and we propose consistent estimators for these information matrices. Section 4 studies the finite sample behavior of the estimators and compare the Bahadur slopes of two versions of the Lagrange multiplier test for testing linear restrictions on θ_0. For the latter application, it is necessary to compute J_θ at $\theta \neq \theta_0$. Concluding remarks are proposed in Section 5. The proofs of the main results are collected in the appendix.

2. Notations

We first introduce the notions of weak and strong ARMA representations, which differ by the assumptions on the error terms. We then recall results concerning the estimation of the weak ARMA models, and introduce extended information matrices.

2.1. Strong, semistrong and weak ARMA representations

For a linear model to be quite general, the error terms must be the linear innovations, which are uncorrelated by construction but are not independent, nor martingale differences, in general. Indeed, the Wold decomposition (see Brockwell and Davis (1991), Section 5.7) stipulates that any purely non deterministic stationary process can be expressed as

$$X_t = \sum_{\ell=0}^{\infty} \varphi_\ell \epsilon_{t-\ell}, \qquad (\epsilon_t) \sim \mathrm{WN}(0, \sigma^2), \tag{1}$$

where $\varphi_0 = 1$, $\sum_{\ell=0}^{\infty} \varphi_\ell^2 < \infty$, and the notation $(\epsilon_t) \sim \mathrm{WN}(0, \sigma^2)$ signifies that the linear innovation process (ϵ_t) is a *weak white noise*, that is a stationary sequence of centered and uncorrelated random variables with common variance σ^2. In practice the sequence φ_ℓ is often parameterized by assuming that X_t admits an ARMA(p, q) representation, *i.e.* that there exist integers p and q and constants $a_{01}, \ldots, a_{0p}, b_{01}, \ldots, b_{0q}$, such that

$$\forall t \in \mathbb{Z}, \quad X_t - \sum_{i=1}^{p} a_{0i} X_{t-i} = \epsilon_t + \sum_{j=1}^{q} b_{0j} \epsilon_{t-j}. \tag{2}$$

This representation is said to be a *weak* ARMA(p, q) representation under the assumption $(\epsilon_t) \sim \mathrm{WN}(0, \sigma^2)$. For the statistical inference of ARMA models, the weak white noise assumption is not sufficient and is often replaced by the *strong white noise* assumption $(\epsilon_t) \sim \mathrm{IID}(0, \sigma^2)$, *i.e.* the assumption that (ϵ_t) is an independent and identically distributed (iid) sequence of random variables with mean 0 and common variance σ^2. Sometimes an intermediate assumption is considered for the noise. The sequence (ϵ_t) is said to be a *semistrong white noise* or a *martingale-difference white noise*, and is denoted by $(\epsilon_t) \sim \mathrm{MD}(0, \sigma^2)$, if (ϵ_t) is a stationary sequence satisfying $E(\epsilon_t \mid \epsilon_u, u < t) = 0$ and $\mathrm{Var}(\epsilon_t) = \sigma^2$. An ARMA representation (2) will be called *strong* under the assumption $(\epsilon_t) \sim \mathrm{IID}(0, \sigma^2)$ and *semistrong* under the assumption $(\epsilon_t) \sim \mathrm{MD}(0, \sigma^2)$.

Obviously the strong white noise assumption is more restrictive than that of semistrong white noise, and the latter is more restrictive than the weak white noise assumption, because independence entails unpredictability and unpredictability entails uncorrelatedness, but the reverses are not true. Consequently the weak ARMA representation are more general that the semistrong and strong ones, what we schematize by

$$\{\text{Weak ARMA}\} \supset \{\text{Semistrong ARMA}\} \supset \{\text{Strong ARMA}\}. \tag{3}$$

Any process satisfying (1) is the limit, in L^2 as $n \to \infty$, of a sequence of processes satisfying weak ARMA(p_n, q_n) representations (see *e.g.* Francq and Zakoïan, 2005, page 244). In this sense, the subclass of the processes admitting weak ARMA(p_n, q_n) representations is dense in the set of the purely non deterministic stationary processes. Simple illustrations that the last inclusion of (3) is strict are given by the vast class of volatility models. Indeed GARCH-type models are generally martingale differences (because financial returns are generally assumed to be unpredictable) but they are not strong noises (in particular, because of the volatility clustering, the squared returns are predictable). Many nonlinear models, such as bilinear or Markov-switching models, illustrate the first inclusion in (3), since they admit weak ARMA representation (see Francq, Roy and Zakoïan, 2005, section 2.3) which are not semistrong, because the best predictor is generally not linear when the data generating process (DGP) is nonlinear. To fix ideas, we give below a simple illustrative example, which was not given by the above-mentioned references.

Example 2.1 (Integer-valued AR(1) and MA(1)). *McKenzie (2003) reviews the literature on models for integer-valued time series. Let $\circ$ be the thinning operator defined by*

$$a \circ X = \sum_{i=1}^{X} Y_j,$$

where (Y_j) is an iid counting series, independent of the integer-valued random variable X, with Bernoulli distribution of parameter $a \in [0, 1)$. The integer-valued autoregressive (INAR) model of order 1 is given by

$$\forall t \in \mathbb{Z}, \quad X_t = a \circ X_{t-1} + Z_t \tag{4}$$

where Z_t is an integer-valued iid sequence, independent of the counting series, with mean μ and variance σ^2. Clearly the best predictor of X_t is linear since $E(X_t \mid X_u, u < t) = aX_{t-1} + \mu$. Moreover we have $Var(X_t \mid X_u, u < t) = (1-a)aX_{t-1} + \sigma^2$. We thus have the semistrong AR(1) representation

$$X_t = aX_{t-1} + \mu + \epsilon_t, \qquad (\epsilon_t) \sim \text{IID}\left(0, \frac{a\mu}{1-a} + \sigma^2\right).$$

Similarly to (4) the integer-valued moving-average INMA(1) is defined by

$$\forall t \in \mathbb{Z}, \quad X_t = Z_t + a \circ Z_{t-1}.$$

Straightforward computations show that $EX_t = \mu(1+a)$, $Var(X_t) = \sigma^2 + a(1-a)\mu + a^2\sigma^2$ and $Cov(X_t, X_{t-1}) = a\sigma^2$, from which we deduce the weak MA(1) representation

$$X_t = \mu(1+a) + \epsilon_t + b\epsilon_{t-1}, \qquad (\epsilon_t) \sim \text{WN}\left(0, \sigma_\epsilon^2\right),$$

where $b \in [0,1)$ and $\sigma_\epsilon^2 > 0$ are solutions of $b/(1+b^2) = \rho_X(1)$ and $(1+b^2)\sigma_\epsilon^2 = Var(X_t)$. This MA(1) representation is not semistrong because $E(X_t \mid X_{t-1} = 0) = E(X_t \mid Z_{t-1} = 0) = \mu$ does not coincide with the linear prediction given by the MA(1) model when $a \neq 0$.

Finally we have shown that an INMA(1) is a weak MA(1) and that a INAR(1) is a semistrong AR(1).

2.2. Estimating weak ARMA representations

We now present the asymptotic behavior of the LSE in the case of weak ARMA models. The LSE is the standard estimation procedure for ARMA models and it coincides with the maximum-likelihood estimator in the Gaussian case. It will be convenient to write (2) as $\phi_0(B)X_t = \psi_0(B)\epsilon_t$, where B is the backshift operator, $\phi_0(z) = 1 - \sum_{i=1}^{p} a_{0i}z^i$ is the AR polynomial and $\psi_0(z) = 1 + \sum_{j=1}^{q} b_{0j}z^j$ is the MA polynomial. The unknown parameter $\theta_0 = (a_{01}, \ldots, a_{0p}, b_{01}, \ldots, b_{0q})$ is supposed to belong to the interior of a compact subspace Θ^* of the parameter space

$$\begin{aligned}\Theta \;:=\; \big\{&\theta = (\theta_1, \ldots, \theta_{p+q}) = (a_1, \ldots, a_p, b_1, \ldots, b_q) \in \mathbb{R}^{p+q} : \\ &\phi(z) = 1 - \sum_{i=1}^{p} a_i z^i \text{ and } \psi(z) = 1 + \sum_{i=1}^{q} b_i z^i \\ &\text{have all their zeros outside the unit disk}\big\}.\end{aligned}$$

Since $\theta \in \Theta$, the polynomials $\phi_0(z)$ and $\psi_0(z)$ have all their zeros outside the unit disk. We also assume that $\phi_0(z)$ and $\psi_0(z)$ have no zero in common, that $p+q > 0$ and $a_{0p}^2 + b_{0q}^2 \neq 0$ (by convention $a_{00} = b_{00} = 1$). These assumptions are standard and are also made for the usual strong ARMA models.

For all $\theta \in \Theta$, let

$$\epsilon_t(\theta) = \psi^{-1}(B)\phi(B)X_t = X_t + \sum_{i=1}^{\infty} c_i(\theta)X_{t-i}.$$

Given a realization of length n, $X_1, X_2, \ldots, X_n$, $\epsilon_t(\theta)$ can be approximated, for $0 < t \leq n$, by $e_t(\theta)$ defined recursively by

$$e_t(\theta) = X_t - \sum_{i=1}^{p} \theta_i X_{t-i} - \sum_{i=1}^{q} \theta_{p+i} e_{t-i}(\theta)$$

where the unknown starting values are set to zero: $e_0(\theta) = e_{-1}(\theta) = \ldots = e_{-q+1}(\theta) = X_0 = X_{-1} = \ldots = X_{-p+1} = 0$. The random variable $\hat{\theta}_n$ is called LSE if it satisfies, almost surely,

$$Q_n(\hat{\theta}_n) = \min_{\theta \in \Theta^*} Q_n(\theta), \qquad Q_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} e_t^2(\theta).$$

The asymptotic behavior of the LSE is well known in the strong ARMA case, *i.e.* under the assumption $(\epsilon_t) \sim \text{IID}(0, \sigma^2)$. This assumption being very restrictive, Francq and Zakoïan (1998) considered weak ARMA representations of stationary processes satisfying the following assumption.

A1 : $E|X_t|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$,

where $\alpha_X(k)$, $k = 0, 1, \ldots$, denote the strong mixing coefficients of the process (X_t) (see *e.g.* Bradley, 2005, for a review on strong mixing conditions). As noted by Francq and Zakoïan (2005), Assumption **A1** can be replaced by

A1' : $E|\epsilon_t|^{4+2\nu} < \infty$ and $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$ for some $\nu > 0$.

A straightforward extension of Francq and Zakoïan (1998) thus gives the following result.

Lemma 2.1 (Francq and Zakoïan, 1998). *Let (X_t) be a strictly stationary and ergodic process satisfying the weak ARMA model (2) with $(\epsilon_t) \sim \mathrm{WN}(0, \sigma^2)$. Under the previous assumptions and Assumption* **A1** *or* **A1'**,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \overset{d}{\rightsquigarrow} \mathcal{N}(0, \Omega = J^{-1} I J^{-1}) \text{ as } n \to \infty, \tag{5}$$

where $I = I_{\theta_0}$, $J = J_{\theta_0} = J^*_{\theta_0}$, with

$$\begin{aligned} I_\theta &= \sum_{h=-\infty}^{+\infty} Cov\left\{\epsilon_t(\theta)\frac{\partial \epsilon_t(\theta)}{\partial \theta}, \epsilon_{t-h}(\theta)\frac{\partial \epsilon_{t-h}(\theta)}{\partial \theta'}\right\}, \\ J_\theta &= E\frac{\partial \epsilon_t(\theta)}{\partial \theta}\frac{\partial \epsilon_t(\theta)}{\partial \theta'}, \quad J^*_\theta = E\epsilon_t(\theta)\frac{\partial^2 \epsilon_t(\theta)}{\partial \theta \partial \theta'} + E\frac{\partial \epsilon_t(\theta)}{\partial \theta}\frac{\partial \epsilon_t(\theta)}{\partial \theta'}. \end{aligned}$$

In the strong ARMA case, we have $I = I_s := \sigma^2 J$ and $\Omega = \Omega_s := \sigma^2 J^{-1}$. In the semistrong ARMA case, *i.e.* under the assumption $(\epsilon_t) \sim \mathrm{MD}(0, \sigma^2)$, we have

$$I = I_{ss} := E\epsilon_t^2 \frac{\partial \epsilon_t(\theta_0)}{\partial \theta}\frac{\partial \epsilon_t(\theta_0)}{\partial \theta'}.$$

Note that we introduce the two versions J_θ and J^*_θ because the following two estimators of J can be considered

$$\hat{J}_n = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial e_t(\hat{\theta}_n)}{\partial \theta}\frac{\partial e_t(\hat{\theta}_n)}{\partial \theta'}, \qquad \hat{J}^*_n = \frac{1}{n}\sum_{t=1}^{n} e_t(\hat{\theta}_n)\frac{\partial^2 e_t(\hat{\theta}_n)}{\partial \theta \partial \theta'} + \hat{J}_n. \tag{6}$$

The matrices J_θ, J^*_θ and I_θ can be called information matrices. As we will see in Section 4.2 they determine the asymptotic behavior of test procedures on θ_0. They are also involved in other inference steps, such as in portmanteau adequacy tests (see Francq, Roy and Zakoian, 2005).

3. Main results

McLeod (1978) gave a nice expression for J, as the variance of a VAR model involving only the ARMA parameter θ_0 (see (8.8.3) in Brockwell and Davis, 1991). Francq, Roy and Zakoïan (2005) obtained an expression of I involving the ARMA parameter θ_0 and the fourth-order moments of the weak noise (ϵ_t) (with their notations, $J = \mathbf{\Lambda}'_\infty \mathbf{\Lambda}_\infty$ and $I = \mathbf{\Lambda}'_\infty \mathbf{\Gamma}_{\infty,\infty} \mathbf{\Lambda}_\infty$ where $\mathbf{\Lambda}_\infty$ depends on θ_0 and $\mathbf{\Gamma}_{\infty,\infty}$ depends on moments of (ϵ_t)). For certain statistical applications, it is interesting to obtain similar expressions for I_θ, J_θ and J^*_θ when $\theta \neq \theta_0$. This is the subject of the next subsection.

3.1. Theoretical expressions for the information matrices

3.1.1. Matrix J_θ

Differentiating the two sides of the equation $\phi(B)X_t = \psi(B)\epsilon_t(\theta)$, for $i, k = 1, \ldots, p$ and $j, \ell = 1, \ldots, q$, we obtain

$$
\begin{aligned}
-X_{t-i} &= \psi(B)\frac{\partial}{\partial a_i}\epsilon_t(\theta), \qquad 0 = \epsilon_{t-j}(\theta) + \psi(B)\frac{\partial}{\partial b_j}\epsilon_t(\theta) \\
0 &= \psi(B)\frac{\partial^2}{\partial a_i \partial a_k}\epsilon_t(\theta), \qquad 0 = \frac{\partial}{\partial a_i}\epsilon_{t-j}(\theta) + \psi(B)\frac{\partial^2}{\partial b_j \partial a_i}\epsilon_t(\theta) \\
0 &= \frac{\partial}{\partial b_\ell}\epsilon_{t-j}(\theta) + \frac{\partial}{\partial b_j}\epsilon_{t-\ell}(\theta) + \psi(B)\frac{\partial^2}{\partial b_j \partial b_\ell}\epsilon_t(\theta).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
\frac{\partial}{\partial a_i}\epsilon_t(\theta) &= -\psi^{-1}(B)X_{t-i} = -\psi^{-1}\phi_0^{-1}\psi_0(B)\epsilon_{t-i} := -\sum_{h=0}^{\infty} c_h^a \epsilon_{t-i-h} \\
\frac{\partial}{\partial b_j}\epsilon_t(\theta) &= -\psi^{-2}\phi_0^{-1}\phi\psi_0(B)\epsilon_{t-j} := -\sum_{h=0}^{\infty} c_h^b \epsilon_{t-j-h}, \\
\frac{\partial^2}{\partial b_j \partial a_i}\epsilon_t(\theta) &= \psi^{-2}\phi_0^{-1}\psi_0(B)\epsilon_{t-i-j} := \sum_{h=0}^{\infty} c_h^{ab} \epsilon_{t-i-j-h}, \\
\frac{\partial^2}{\partial b_j \partial b_\ell}\epsilon_t(\theta) &= 2\psi^{-3}\phi_0^{-1}\phi\psi_0(B)\epsilon_{t-j-\ell} := \sum_{h=0}^{\infty} c_h^{bb} \epsilon_{t-j-\ell-h},
\end{aligned}
$$

and $\partial^2\epsilon_t(\theta)/\partial a_i \partial a_k = 0$. Moreover

$$
\epsilon_t(\theta) = \psi^{-1}\phi_0^{-1}\phi\psi_0(B)\epsilon_t := \sum_{h=0}^{\infty} c_h \epsilon_{t-h}.
$$

The following result immediately follows.

Proposition 3.1. *The elements of the matrix J_θ and J_θ^* are given by*

$$J_\theta(i,k) = J_\theta^*(i,k) = \sigma^2 \sum_{s=0}^{\infty} c_{s+k-i}^a c_s^a,$$

$$J_\theta(p+j,p+\ell) = \sigma^2 \sum_{s=0}^{\infty} c_{s+\ell-j}^b c_s^b,$$

$$J_\theta^*(p+j,p+\ell) = \sigma^2 \sum_{s=0}^{\infty} c_{s+j+\ell} c_s^{bb} + J_\theta(p+j,p+\ell),$$

$$J_\theta(i,p+\ell) = \sigma^2 \sum_{s=\max\{0,i-\ell\}}^{\infty} c_{s+\ell-i}^a c_s^b,$$

$$J_\theta^*(i,p+\ell) = \sigma^2 \sum_{s=0}^{\infty} c_{s+i+\ell} c_s^{ab} + J_\theta(i,p+\ell),$$

for $1 \le i \le k \le p$ *and* $1 \le j \le \ell \le q$.

On the web page of the authors, programs written in R are available for computing the information matrices defined in this paper, as well as their estimates. For example, the following function `infoJ()` computes J_θ when, in R language, θ_0`<-c(ar0,ma0)` and θ`<-c(ar1,ma1)`. The truncation parameter `M` is discussed in Section 3.2 below. This function uses the function `prod.poly()` which makes the product of the 2 polynomials, and the function `ARMAtoMA()` of the package `stats`.

```
# Product of 2 polynomials
prod.poly<- function(a,b) {
  p<-length(a); q<-length(b)
  if(p<=0|q<0)stop("a or b is invalid")
  c<-rep(0,(p+q))
  for(h in 2:(p+q)){
    imin<-max(1,h-q); imax<-min(p,h-1)
    for(i in (imin:imax))c[h]<-c[h]+a[i]*b[h-i]
  }
  c[2:(p+q)]
}
# Computation of the information matrix J at \theta=(ar1,ma1)
infoJ<- function(ar0,ma0,ar1,ma1,M=200){
  p<-length(ar1); q<-length(ma1); p0<-length(ar0); q0<-length(ma0)
  matJ.theta<-matrix(0,nrow=(p+q),ncol=(p+q))
  if(p>0){       # c_h^a + top-left corner of J
    p1<-p0+q
    if(p1==0) ar2 <- c()
    if(p1>0) ar2 <- -prod.poly(c(1,-1*ar0),c(1,ma1))[2:(p1+1)]
```

```
    ch.a<-c(1,ARMAtoMA(ar =ar2, ma=ma0, M))
    for(i in (1:p)){
       for(k in (1:p)){
         matJ.theta[i,k]<-sum(ch.a[(abs(k-i)+1):(M+1)]*ch.a[1:(M-abs(k-i)+1)])
       }
    }
  }
  if(q>0){  # c_h^b + bottom-right corner of  J
    p1<-p0+2*q
    if(p1==0) ar2 <- c()
    if(p1>0) ar2 <- -prod.poly(prod.poly(c(1,ma1),c(1,ma1)),c(1,-1*ar0))[2:(p1+1)]
    q1<-p+q0
    if(q1==0) ma2 <- c()
    if(q1>0) ma2 <- prod.poly(c(1,-1*ar1),c(1,ma0))[2:(q1+1)]
    ch.b<-c(1,ARMAtoMA(ar =ar2, ma=ma2, lag.max=M))
    for(j in (1:q)){
      for(l in (1:q)){
        matJ.theta[p+j,p+l]<-sum(ch.b[(abs(l-j)+1):(M+1)]*ch.b[1:(M-abs(l-j)+1)])
      }
    }
  }
  if(p>0&q>0){ # cross blocks
    for(i in (1:p)){
      for(l in (1:q)){
        indmin1<-max(0,i-l)+l-i+1
        indmin2<-max(0,i-l)+1
        indmax1<-M-max(0,i-l)
        indmax2<-indmax1-l+i
        matJ.theta[i,p+l]<-sum(ch.a[indmin1:indmax1]*ch.b[indmin2:indmax2])
        matJ.theta[p+l,i]<- matJ.theta[i,p+l]
      }
    }
  }
  matJ.theta
}
```

3.1.2. Matrix I_θ

We now search similar tractable expressions for I_θ. Let

$$\Gamma(m,m') = \sum_{h=-\infty}^{+\infty} \mathrm{Cov}\left(\epsilon_t\epsilon_{t-m}, \epsilon_h\epsilon_{h-m'}\right). \tag{7}$$

In the strong case, we have

$$\Gamma(0,0) = \mu_4 - \sigma^4, \quad \Gamma(m,m) = \Gamma(m,-m) = \sigma^4, \quad \Gamma(m',m'') = 0, \tag{8}$$

with $\mu_4 = E\epsilon_1^4$, $m \neq 0$ and $|m'| \neq |m''|$. Simplifications may also hold in semistrong cases. Indeed, consider the case $(\epsilon_t) \sim \mathrm{WN}\left(0, \sigma_\epsilon^2\right)$ under the following symmetry assumption

$$E\epsilon_{t_1}\epsilon_{t_2}\epsilon_{t_3}\epsilon_{t_4} = 0 \qquad \text{when} \qquad t_1 \neq t_2,\ t_1 \neq t_3 \text{ and } t_1 \neq t_4. \tag{9}$$

A similar assumption is made in Francq and Zakoian (2009b). In this paper, it is shown that, in particular, GARCH models with fourth-order moments and symmetric innovations satisfy (9). Many other martingale differences satisfy this assumption. In this semistrong case, we have

$$\Gamma(0,0)=\sum_{h=-\infty}^{\infty}\mathrm{Cov}(\epsilon_t^2,\epsilon_{t-h}^2),\quad \Gamma(m,m)=E\epsilon_t^2\epsilon_{t-m}^2,\quad \Gamma(m',m'')=0 \quad (10)$$

when $m\neq 0$ and $|m'|\neq|m''|$.

Example 3.1. *For a GARCH(1,1) model of the form*

$$\left\{\begin{array}{l}\epsilon_t=\sqrt{h_t}\eta_t,\quad t=1,2,\dots\\ h_t=\omega+\alpha\epsilon_{t-1}^2+\beta h_{t-1},\qquad (\eta_t)\sim \mathrm{IID}\,(0,1)\end{array}\right.$$

with $\omega>0$, $\alpha\geq 0$, $\beta\geq 0$ *and* $\alpha^2E\eta_1^4+\beta^2+2\alpha\beta<1$[1] *we obtain*

$$\begin{aligned}\Gamma(0,0) &= E\nu_t^2\frac{(1-\beta)^2}{(1-\alpha-\beta)^2},\\ \Gamma(1,1) &= E\nu_t^2\left(\alpha+\frac{\alpha^2(\alpha+\beta)}{1-(\alpha+\beta)^2}\right)+\left(E\sigma_t^2\right)^2\\ \Gamma(m,m) &= (\alpha+\beta)\Gamma(m-1,m-1)+\omega E\sigma_t^2,\qquad m>1,\end{aligned}$$

with $E\nu_t^2=E\eta_1^4\left(E\sigma_t^4+1-2E\sigma_t^2\right)$,

$$E\sigma_t^2 = \frac{\omega}{1-\alpha-\beta},\qquad E\sigma_t^4=\frac{\omega^2(1+\alpha+\beta)}{(1-\alpha^2E\eta_1^4-\beta^2-2\alpha\beta)(1-\alpha-\beta)}.$$

Proposition 3.2. *The elements of the matrix* I_θ *are given by*

$$\begin{aligned}I_\theta(i,k)&=\sum_{h_1,h_2,h_3,h_4=0}^{+\infty}c_{h_1}c_{h_2}^a c_{h_3}c_{h_4}^a\Gamma(h_2+i-h_1,h_4+k-h_3),\\ I_\theta(j,\ell)&=\sum_{h_1,h_2,h_3,h_4=0}^{+\infty}c_{h_1}c_{h_2}^b c_{h_3}c_{h_4}^b\Gamma(h_2+j-h_1,h_4+\ell-h_3),\\ I_\theta(i,\ell)&=\sum_{h_1,h_2,h_3,h_4=0}^{+\infty}c_{h_1}c_{h_2}^a c_{h_3}c_{h_4}^b\Gamma(h_2+i-h_1,h_4+\ell-h_3),\end{aligned}$$

[1]The latter conditions and necessary and sufficient for the existence of a nonanticipative stationary solution with fourth-order moments (see *e.g.* Example 2.3 in Francq and Zakoian, 2010).

for $1 \leq i \leq k \leq p$ *and* $1 \leq j \leq \ell \leq q$.

Note that $c_0 = 1$ and that, at $\theta = \theta_0$, we have $c_h = 0$ for $h > 0$. The expression of $I = I_{\theta_0}$ thus simplifies to that given in Francq, Roy and Zakoian (2005). There is also a slight simplification in the strong and semistrong ARMA cases because, in view of (8) and (10), I_θ is then obtained by summing over 3 indices instead of 4.

3.1.3. Examples of analytic and numerical computations of J_θ *and* I_θ

Let us compute the information matrices of an ARMA(1,1) model at the point $\theta^* = (a, 0)'$ when $\theta_0 = (0, b_0)'$ (*i.e.* the DGP is a MA(1)). We have $\epsilon_t(\theta) = \sum_{h=0}^{\infty}(-b)^h(X_{t-h} - aX_{t-h-1})$. It follows that $\epsilon_t(\theta^*) = \epsilon_t + (b_0 - a)\epsilon_{t-1} - ab_0\epsilon_{t-2}$, $\partial\epsilon_t(\theta^*)/\partial a = -\epsilon_{t-1} - b_0\epsilon_{t-2}$, $\partial\epsilon_t(\theta^*)/\partial b = -\epsilon_{t-1} - (b_0 - a)\epsilon_{t-2} + ab_0\epsilon_{t-3}$, $\partial^2\epsilon_t(\theta^*)/\partial a^2 = 0$, $\partial^2\epsilon_t(\theta^*)/\partial a\partial b = \epsilon_{t-2} + b_0\epsilon_{t-3}$ and $\partial^2\epsilon_t(\theta^*)/\partial b^2 = 2\epsilon_{t-2} + 2(b_0 - a)\epsilon_{t-3} - 2ab_0\epsilon_{t-4}$.

Thus

$$\begin{aligned} J_{\theta^*} &= \sigma^2 \begin{pmatrix} 1 + b_0^2 & 1 + b_0(b_0 - a) \\ 1 + b_0(b_0 - a) & 1 + (b_0 - a)^2 + a^2 b_0^2 \end{pmatrix}, \\ J^*_{\theta^*} &= J_{\theta^*} + \sigma^2 \begin{pmatrix} 0 & -ab_0 \\ -ab_0 & -2ab_0 \end{pmatrix}. \end{aligned}$$

Now assume that ϵ_t is the weak white noise considered by Romano and Thombs (1996), defined by

$$\epsilon_t = \eta_t\eta_{t-1}\cdots\eta_{t-k}, \quad (\eta_t) \sim \text{IID } \mathcal{N}(0,1), \quad k \geq 0. \tag{11}$$

It seems impossible to obtain I_{θ^*} explicitly, but the information matrices can be obtained explicitly at θ_0:

$$\begin{aligned} J_{\theta_0} &= J^*_{\theta_0} = \begin{pmatrix} 1 & 1 \\ 1 & 1/(1 - b_0^2) \end{pmatrix}, \\ I_{\theta_0} &= 3^k \begin{pmatrix} 1 & 1 \\ 1 & \frac{1-(b^2/3)^{k+1}}{1-b^2/3} + \frac{b^{2(k+1)}}{3^k(1-b^2)} \end{pmatrix}. \end{aligned}$$

Note that $I_{\theta_0} = J_{\theta_0}$ in the strong case (*i.e.* when $k = 0$). For more complex models or at some point $\theta^* \neq \theta_0$ the evaluation of these information matrices is not feasible analytically but they can be easily obtained numerically. For

instance, on this example with $k = 3$, $\theta_0 = (0, 0.5)'$ and $\theta^* = (-0.4, -0.5)'$, we have

$$J_{\theta^*} = \begin{pmatrix} 2.33 & 4.33 \\ 4.33 & 11.25 \end{pmatrix}, \; J^*_{\theta^*} = \begin{pmatrix} 2.33 & 6.33 \\ 6.33 & 17.65 \end{pmatrix}, \; I_{\theta^*} = \begin{pmatrix} 1161.92 & 2177.66 \\ 2177.66 & 4187.63 \end{pmatrix}.$$

3.2. Approximation of the information matrices by finite sums

In practice the infinite sums involved in J_θ, J^*_θ and I_θ are truncated. This section concentrates on the choice of the truncation parameter for I_θ, the problem being similar, and somewhat simpler, for the other matrices. Matrix I_θ is truncated by the matrix I^M_θ of M^4 terms, defined by

$$I^M_\theta(i,k) = \sum_{h_1,h_2,h_3,h_4=0}^{M} c_{h_1} c^a_{h_2} c_{h_3} c^a_{h_4} \Gamma(h_2 + i - h_1, h_4 + k - h_3),$$

when $1 \le i \le k \le p$, and whose the other elements are defined similarly. The following proposition defines a value of M such that I^M_θ be equal to I_θ up to an arbitrarily small tolerance number ε. Let the matrix norm defined by $\|A\| = \sum_{i,j} |A(i,j)|$ with obvious notations.

Proposition 3.3. *Let $\overline{\rho}$ be the inverse of the largest modulus of the zeroes of the polynomials ϕ_0 and ψ, let*

$$\overline{\Gamma} = \max_{m,m' \ge 0} |\Gamma(m,m')|, \qquad \overline{\pi} = \left(1 + \sum_{j=1}^{q} |b_{0j}|\right) \max_{i=0,\dots,p} |a_i|$$

with $|a_0| = 1$, and let

$$K = \sqrt{2}(p+q+1)\overline{\pi}\overline{\Gamma}^{1/4} \left(\frac{-2(p+2q-1)}{\log \overline{\rho}}\right)^{(p+2q-1)} \overline{\rho}^{\,-0.5-(p+2q-1)/\log \overline{\rho}}.$$

For all $\varepsilon > 0$, if

$$M \ge M_\epsilon := \frac{\log \sqrt{\varepsilon}(1 - \sqrt{\log \overline{\rho}})^2/K^2}{\log \overline{\rho}}$$

then

$$\left\| I_\theta - I^M_\theta \right\| \le (p+q)^2 \varepsilon.$$

3.3. Estimation of the information matrices

Let $J_{\theta,n}$ and $J^*_{\theta,n}$ be defined as in (6), replacing $\hat{\theta}_n$ by θ in $\hat{J}_n$ and $\hat{J}^*_n$, so that $\hat{J}_n = J_{\hat{\theta}_n,n}$ and $\hat{J}^*_n = J^*_{\hat{\theta}_n,n}$. The following result shows that an estimator of J_{θ^*} is trivially deduced from one of θ^*.

Proposition 3.4. *Under the assumptions of Lemma 2.1, as* $n \to \infty$,

$$\text{if } \theta^*_n \to \theta^* \text{ a.s.} \quad \text{then} \quad J_{\theta^*_n,n} \to J_{\theta^*} \text{ and } J^*_{\theta^*_n,n} \to J^*_{\theta^*} \text{ a.s.}$$

The estimation of the long-run variance I_θ is more complicated. In the literature, two types of estimators are generally employed: Heteroskedasticity and Autocorrelation Consistent (HAC) estimators (see Newey and West (1987) and Andrews (1991) for general references, and Francq and Zakoian (2007) for an application to testing strong linearity in weak ARMA models) and spectral density estimators (see *e.g.* den Haan and Levin (1997) for a general reference and Francq, Roy and Zakoian (2005) for estimating I in the present context). We will extend the results of Francq, Roy and Zakoian (2005) for estimating I_θ when θ is not necessarily equal to θ_0.

3.3.1. An estimator based on a spectral density form for I_θ

Note $(2\pi)^{-1} I_{\theta^*}$ is the spectral density at frequency 0 (see Brockwell and Davis (1991) p. 459) of the process

$$\nabla_t = S_t - ES_t, \qquad S_t = \epsilon_t(\theta^*)\frac{\partial \epsilon_t(\theta^*)}{\partial \theta}. \tag{12}$$

For any given $\theta^* \in \Theta$, S_t is a measurable function of $\{X_u, u \leq t\}$. Let $\hat{S}_t$ be obtained by replacing the unknown initial values $\{X_u, u \leq 0\}$ by 0 and θ^* by θ^*_n in S_t. Let also

$$\hat{\nabla}_t = \hat{S}_t - \frac{1}{n}\sum_{t=1}^{n} \hat{S}_t.$$

The stationary process (∇_t) admits the Wold decomposition $\nabla_t = u_t + \sum_{i=1}^{\infty} B_i u_{t-i}$, where (u_t) is a $(p+q)$-variate weak white noise with covariance matrix Σ_u. Assume that Σ_u is non-singular, that $\sum_{i=1}^{\infty} \|B_i\| < \infty$, and that $\det\left(I_{p+q} + \sum_{i=1}^{\infty} B_i z^i\right) \neq 0$ when $|z| \leq 1$. Then (∇_t) admits an AR(∞) representation of the form

$$\mathcal{A}(B)\nabla_t := \nabla_t - \sum_{i=1}^{\infty} A_i \nabla_{t-i} = u_t, \tag{13}$$

such that $\sum_{i=1}^{\infty} \|A_i\| < \infty$ and $\det\{\mathcal{A}(z)\} \neq 0$ for all $|z| \leq 1$, and we obtain

$$I_\theta = \mathcal{A}^{-1}(1)\Sigma_u \mathcal{A}'^{-1}(1). \tag{14}$$

In the framework of univariate linear processes with independent innovations, Berk (1974) showed that the spectral density can be consistently estimated by fitting autoregressive models of order $r = r(n)$, whenever $r \to \infty$ and $r^3/n \to 0$ as $n \to \infty$. It can be shown that this result remains valid for the multivariate linear process (∇_t), though its innovation (u_t) is not an independent process. Another difference with Berk (1974), is that (∇_t) is not directly observed and is replaced by $(\hat{\nabla}_t)$.

Consider the regression of ∇_t on $\nabla_{t-1}, \ldots, \nabla_{t-r}$ defined by

$$\nabla_t = \sum_{i=1}^{r} A_{r,i}\nabla_{t-i} + u_{r,t}, \qquad u_{r,t} \perp \{\nabla_{t-1} \cdots \nabla_{t-r}\}. \tag{15}$$

The least squares estimators of $\underline{A}_r = (A_{r,1} \cdots A_{r,r})$ and $\Sigma_{u_r} = \mathrm{Var}(u_{r,t})$ are defined by

$$\hat{\underline{A}}_r = \hat{\Sigma}_{\hat{\nabla},\underline{\hat{\nabla}}_r} \hat{\Sigma}^{-1}_{\underline{\hat{\nabla}}_r} \qquad \text{and} \qquad \hat{\Sigma}_{u_r} = \frac{1}{n}\sum_{t=1}^{n} \left(\hat{\nabla}_t - \hat{\underline{A}}_r \underline{\hat{\nabla}}_{r,t}\right)\left(\hat{\nabla}_t - \hat{\underline{A}}_r \underline{\hat{\nabla}}_{r,t}\right)'$$

where $\underline{\hat{\nabla}}_{r,t} = (\hat{\nabla}'_{t-1} \cdots \hat{\nabla}'_{t-r})'$,

$$\hat{\Sigma}_{\hat{\nabla},\underline{\hat{\nabla}}_r} = \frac{1}{n}\sum_{t=1}^{n} \hat{\nabla}_t \underline{\hat{\nabla}}'_{r,t}, \qquad \hat{\Sigma}_{\underline{\hat{\nabla}}_r} = \frac{1}{n}\sum_{t=1}^{n} \underline{\hat{\nabla}}_{r,t} \underline{\hat{\nabla}}'_{r,t},$$

with by convention $\hat{\nabla}_t = 0$ when $t \leq 0$, and assuming $\hat{\Sigma}_{\underline{\hat{\nabla}}_r}$ is non singular (which holds true asymptotically).

Under mild regularity conditions (the precise statement of the result and its proof are available from the authors under request), it can be shown that if $\theta^*_n \to \theta^*$ almost surely,

$$I_{\theta^*_n,n} = \hat{\mathcal{A}}_r^{-1}(1)\hat{\Sigma}_{u_r}\hat{\mathcal{A}}_r'^{-1}(1) \to I_{\theta^*} \tag{16}$$

in probability when $r = r(n) \to \infty$ and $r^3/n \to 0$ as $n \to \infty$.

For the implementation of $I_{\theta^*_n,n}$, AR(r) models are obtained recursively for $r = 0, 1, \ldots, r_{max}$ (with $r_{max} = 15$ for the forthcoming applications), using the efficient Whittle's (1963) generalization of the Durbin-Levinson algorithm, described for instance in Brockwell and Davis (1991) Theorem 5.2.1. The order r is then selected using the AIC criterion.

4. Applications

A first set of experiments illustrates the finite sample behavior of our estimators of the information matrices I_{θ^*}, J_{θ^*} and $J^*_{\theta^*}$, for strong and weak ARMA models. We then study the impact of the estimator of J, *i.e.* the effect of choosing $\hat{J}_n$ or $\hat{J}^*_n$ defined by (6), on the asymptotic behavior of tests of linear restrictions on the ARMA parameters. For this study it will be necessary to evaluate J_{θ^*} at $\theta^* \neq \theta_0$. Finally, an application to financial data is presented.

4.1. Finite sample behavior of estimators of the information matrices

To investigate the finite sample performance of the estimators, we simulated $N = 1,000$ independent trajectories of size $n = 1,000$ and $n = 10,000$ of an ARMA(1,1) model with parameter $\theta_0 = (0.5, 0.7)$, in which the noise is defined by (11). Note that when $k = 0$ in (11), the ARMA model is strong, whereas the model is weak when $k > 0$.

4.1.1. Estimating the information matrices at a given point θ^*

The information matrices J_{θ^*}, $J^*_{\theta^*}$ and I_{θ^*} have been computed and estimated at 3 points θ^* chosen randomly in $(-1, 1)^2$. The estimators are $J_{\theta^*,n}$, $J^*_{\theta^*,n}$ and $I_{\theta^*,n}$ defined in Proposition 3.4 and (16). Table 1 displays the average, over the N replications, of the relative estimation errors

$$\frac{\|J_{\theta^*,n} - J_{\theta^*}\|}{\|J_{\theta^*}\|}, \quad \frac{\left\|J^*_{\theta^*,n} - J^*_{\theta^*}\right\|}{\|J^*_{\theta^*}\|} \quad \text{and} \quad \frac{\|I_{\theta^*,n} - I_{\theta^*}\|}{\|I_{\theta^*}\|}.$$

From Table 1, one can note that: 1) although the information matrices vary a lot with θ^*, the relative errors are not very sensitive to the value of θ^*; 2) as expected the relative errors decrease when n increases; 3) it is more difficult to estimate the information matrices when k is large; 4) it is easier to estimate J_{θ^*} than $J^*_{\theta^*}$, and easier to estimate $J^*_{\theta^*}$ than I_{θ^*}.

4.1.2. Estimating the asymptotic variance of the LSE

Several estimators of the asymptotic variance Ω involved in (5) can be considered. In view of Proposition 3.4 and (16), two estimators that are consistent under very general assumptions are

$$\hat{\Omega} = J^{-1}_{\hat{\theta}_n,n} I_{\hat{\theta}_n,n} J^{-1}_{\hat{\theta}_n,n} \qquad \text{and} \qquad \hat{\Omega}^* = J^{*-1}_{\hat{\theta}_n,n} I_{\hat{\theta}_n,n} J^{*-1}_{\hat{\theta}_n,n}.$$

Table 1: Average relative error for the estimators of the information matrices, over $N = 1000$ replications.

		$n = 1,000$			$n = 10,000$		
k	θ^*	J_{θ^*}	$J^*_{\theta^*}$	I_{θ^*}	J_{θ^*}	$J^*_{\theta^*}$	I_{θ^*}
0	(-0.9,-0.7)	0.10307	0.13466	0.34958	0.03472	0.04466	0.13070
	(0.5,-0.6)	0.07817	0.10840	0.28935	0.02635	0.03655	0.11602
	(0.7,0.9)	0.16050	0.23238	0.62677	0.05105	0.07239	0.35446
1	(-0.9,-0.7)	0.15183	0.17483	0.54247	0.04835	0.05592	0.23380
	(0.5,-0.6)	0.13454	0.15691	0.52612	0.04225	0.04979	0.22927
	(0.7,0.9)	0.19607	0.25395	0.81663	0.06119	0.08024	0.41813
2	(-0.9,-0.7)	0.24941	0.26726	1.01021	0.07945	0.08502	0.48073
	(0.5,-0.6)	0.23188	0.25324	1.02312	0.07372	0.08064	0.48426
	(0.7,0.9)	0.25566	0.29135	0.88026	0.08983	0.10374	0.52578
3	(-0.9,-0.7)	0.36007	0.37685	1.13567	0.13339	0.13941	0.80758
	(0.5,-0.6)	0.34200	0.36553	1.11668	0.12598	0.13487	0.80077
	(0.7,0.9)	0.38990	0.41912	1.35145	0.13595	0.14663	0.76092
4	(-0.9,-0.7)	0.49452	0.50613	1.19581	0.20527	0.21180	1.11480
	(0.5,-0.6)	0.47746	0.49784	1.19473	0.19569	0.20733	1.10979
	(0.7,0.9)	0.54117	0.55883	1.95453	0.21114	0.21816	1.07556
5	(-0.9,-0.7)	0.67610	0.68923	1.42507	0.31757	0.32562	1.66401
	(0.5,-0.6)	0.66101	0.68331	1.44809	0.30515	0.32038	1.67812
	(0.7,0.9)	0.69161	0.70621	1.77970	0.29799	0.30403	1.15614

In view of the consistency of the LSE stated in Lemma 2.1 and Proposition 3.1, the matrix J can be estimated by plugging. We then define the alternative estimator

$$\Omega(\hat{\theta}_n) = J^{-1}_{\hat{\theta}_n} I_{\hat{\theta}_n, n} J^{-1}_{\hat{\theta}_n}.$$

Other estimators of Ω that should be consistent in the strong ARMA case are defined by

$$\hat{\Omega}_s = \hat{\sigma}^2 J^{-1}_{\hat{\theta}_n, n} \qquad \text{and} \qquad \Omega_s(\hat{\theta}_n) = \hat{\sigma}^2 J^{-1}_{\hat{\theta}_n}, \quad \text{with} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} e_t^2(\hat{\theta}_n).$$

Finally, in the semi-strong case, an estimator is given by

$$\hat{\Omega}_{ss} = J_{\hat{\theta}_n,n}^{-1} \hat{I}_{ss} J_{\hat{\theta}_n,n}^{-1}, \quad \hat{I}_{ss} = \frac{1}{n} \sum_{t=1}^{n} e_t^2(\hat{\theta}_n) \frac{\partial e_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial e_t(\hat{\theta}_n)}{\partial \theta'}.$$

Table 2 indicates that, for all the consistent estimators (*i.e.* all estimators when $k = 0$, and $\hat{\Omega}$, $\hat{\Omega}^*$, $\Omega(\hat{\theta}_n)$ and $\hat{\Omega}_{ss}$ when $k > 0$) the relative errors decrease when n increases. As expected from Table 1, the estimation of the asymptotic matrices becomes more difficult when k increases. When $k > 0$ the estimator $\hat{\Omega}_{ss}$ is much more accurate than $\hat{\Omega}_s$ and $\Omega_s(\hat{\theta}_n)$ (which are actually not consistent in this semi-strong setting) and also slightly more accurate than the other ones. In the strong case (*i.e.* when $k = 0$ in (11)), the estimators $\hat{\Omega}_s$ and $\Omega_s(\hat{\theta}_n)$ are much more accurate than the other ones, but they are not consistent when $k > 0$ (the relative errors are almost the same for $n = 1,000$ and $n = 10,000$). This not surprising, because the noise defined in (11) is a semistrong one.

Table 2: Relative error of the asymptotic variance matrices of the LSE. The number of replications is $N = 1000$.

n	k	$\hat{\Omega}$	$\hat{\Omega}^*$	$\Omega(\hat{\theta}_n)$	$\hat{\Omega}_s$	$\hat{\Omega}_s(\hat{\theta}_n)$	Ω_{ss}
	0	0.13327	0.18264	0.15751	0.06353	0.07496	0.09014
	1	0.35768	0.36452	0.42471	0.47470	0.48099	0.24168
	2	0.52599	0.53392	0.65840	0.79349	0.79581	0.43806
1,000	3	0.69808	0.70883	0.84601	0.92835	0.92945	0.62591
	4	0.83908	0.85469	0.97604	0.97497	0.97583	0.76076
	5	0.92800	0.92828	0.95288	0.99173	0.99213	0.87238
	10	1.00805	0.99961	0.99892	1.00006	0.99983	1.02399
	0	0.04344	0.05567	0.05006	0.02002	0.02327	0.02906
	1	0.14336	0.14478	0.16351	0.47000	0.47020	0.08734
	2	0.31212	0.31412	0.35461	0.79448	0.79442	0.23533
10,000	3	0.50658	0.51049	0.58767	0.92943	0.92913	0.43183
	4	0.65129	0.66057	0.78561	0.97604	0.97626	0.62521
	5	0.76824	0.77414	0.88845	0.99198	0.99212	0.72576
	10	0.99694	0.99672	0.99761	0.99996	0.99996	0.99331

4.2. Bahadur's slopes of two versions of the Lagrange-Multiplier test

Let R be a given matrix of size $s_0 \times (p+q)$ and rank s_0, and let r_0 and r_1 be given vectors of size s_0 such that $r_1 \neq r_0$. Consider the testing problem

$$H_0 : R\theta_0 = r_0 \qquad \text{against} \qquad H_1 : R\theta_0 = r_1. \tag{17}$$

When the model is much simpler under the null than under the alternative, the Lagrange-Multiplier (LM) test is very attractive because, contrary to other tests, in particular the Wald and Likelihood-Ratio tests, the LM procedure only requires the estimation of the ARMA model under H_0 (see Engle (1984) for a general presentation of these tests). Let $\lambda \in \mathbb{R}^{s_0}$ be a Lagrange multiplier and let $\hat{\theta}_n^c$ be the LSE constrained by H_0:

$$(\hat{\theta}_n^c, \hat{\lambda}) = \arg \min_{\theta \in \Theta, \lambda \in \mathbb{R}^{s_0}} Q_n(\theta) - \lambda'(R\theta - r_0).$$

For simplicity, consider the strong ARMA case. The asymptotic variance of the LSE can then be estimated either by

$$\hat{\Omega}^c = \hat{\sigma}^{2c} \left(\hat{J}_n^c\right)^{-1} \qquad \text{or} \qquad \hat{\Omega}^{*c} = \hat{\sigma}^{2c} \left(\hat{J}_n^{*c}\right)^{-1}$$

where

$$\hat{J}_n^c = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \epsilon_t(\hat{\theta}_n^c)}{\partial \theta} \frac{\partial \epsilon_t(\hat{\theta}_n^c)}{\partial \theta'}, \qquad \hat{J}_n^{*c} = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t(\hat{\theta}_n^c) \frac{\partial^2 \epsilon_t(\hat{\theta}_n^c)}{\partial \theta \partial \theta'} + \hat{J}_n^c$$

and $\hat{\sigma}^{2c} = n^{-1} \sum_{t=1}^{n} e_t^2(\hat{\theta}_n^c)$. This leads to two versions of the LM statistic

$$\begin{aligned} \mathbf{LM} &:= \frac{n}{\hat{\sigma}^{2c}} \frac{\partial Q_n(\hat{\theta}_n^c)}{\partial \theta'} \left(\hat{J}_n^c\right)^{-1} \frac{\partial Q_n(\hat{\theta}_n^c)}{\partial \theta}, \\ \mathbf{LM}^* &:= \frac{n}{\hat{\sigma}^{2c}} \frac{\partial Q_n(\hat{\theta}_n^c)}{\partial \theta'} \left(\hat{J}_n^{*c}\right)^{-1} \frac{\partial Q_n(\hat{\theta}_n^c)}{\partial \theta}. \end{aligned}$$

The two versions have the same asymptotic distribution under the null:

$$\mathbf{LM} \xrightarrow{d} \chi^2_{s_0} \qquad \text{and} \qquad \mathbf{LM} \xrightarrow{d} \chi^2_{s_0} \qquad \text{under } H_0$$

but behaves differently under the alternative:

$$\frac{\mathbf{LM}}{n} \to c := D'_{\theta_0^c} \left(\sigma^{2c} J_{\theta_0^c}\right)^{-1} D_{\theta_0^c}, \quad \frac{\mathbf{LM}^*}{n} \to c^* := D'_{\theta_0^c} \left(\sigma^{2c} J^*_{\theta_0^c}\right)^{-1} D_{\theta_0^c}$$

under H_1 as $n \to \infty$, where

$$D_\theta \quad = \quad E\epsilon_t(\theta)\frac{\partial \epsilon_t(\theta)}{\partial \theta}, \qquad \sigma^{2c} = E\epsilon_1^2(\theta_0^c)$$

when $\hat{\theta}_n^c \to \theta_0^c$ a.s., where θ_0^c is such that $J_{\theta_0^c}$ and $J^*_{\theta_0^c}$ are positive-definite.

Note that $J_{\theta_0^c}$ is always positive-semidefinite, but this is not the case for $J^*_{\theta_0^c}$. When $J^*_{\theta_0^c}$ is not positive-definite, the $\mathbf{LM}^*$-test (*i.e.* the test of rejection region $\{\mathbf{LM}^* \geq \chi^2_1(1-\alpha)\}$, where α is the asymptotic level and $\chi^2_k(\alpha)$ denotes the α-quantile of the chi-square distribution with k degrees of freedom) may be inconsistent.

Thanks to the computation of Section 3, for any given alternative, we are able to determine which version is consistent and we are able to compute the Bahadur slopes. We now give a simple example in which hand-made computation of Bahadur's slopes is possible.

4.2.1. Testing an AR(1) against an ARMA(1,1)

We now consider an ARMA(1,1) model and we test for an AR(1). We thus have $R = (0, 1)$, $r_0 = 0$ and $\theta_0^c = (a_0^c, 0)'$ where

$$a_0^c = \arg\min_a E(X_t - aX_{t-1})^2 = \rho(1),$$

where $\rho(h) = \gamma(h)/\gamma(0)$ and $\gamma(h) = \mathrm{Cov}(X_t, X_{t-h})$ denote respectively the autocorrelation and autocovariance of (X_t) at lag h. Standard computations show that the constrained estimator satisfies

$$\hat{\theta}_n^c = \begin{pmatrix} \hat{a}^c \\ 0 \end{pmatrix}, \qquad \frac{\partial Q_n(\hat{\theta}_n^c)}{\partial \theta'} = \begin{pmatrix} 0 \\ \hat{\lambda} \end{pmatrix}$$

with

$$\hat{a}^c = \frac{\sum_{t=2}^n X_t X_{t-1}}{\sum_{t=2}^n X_{t-1}^2}, \qquad \hat{\lambda} = \hat{a}^c \frac{1}{n}\sum_{t=3}^n X_t X_{t-2} - (\hat{a}^c)^2 \frac{1}{n}\sum_{t=3}^n X_{t-1}X_{t-2}.$$

Other tedious computations show that

$$\begin{aligned}
\sigma^{2c} &= \left\{1+(a_0^c)^2\right\}\gamma(0) - 2a_0^c\gamma(1), \\
D_{\theta_0^c} &= -E\left(X_t - a_0^c X_{t-1}\right)\begin{pmatrix} X_{t-1} \\ X_{t-1} - a_0^c X_{t-2} \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ -(a_0^c)^2\gamma(1) + (a_0^c)\gamma(2) \end{pmatrix}, \\
J_{\theta_0^c} &= \begin{pmatrix} \gamma(0) & \gamma(0) - a_0^c\gamma(1) \\ \gamma(0) - a_0^c\gamma(1) & \gamma(0) + (a_0^c)^2\gamma(0) - 2a_0^c\gamma(1) \end{pmatrix}, \\
J^*_{\theta_0^c} &= J_{\theta_0^c} + E(X_t - a_0^c X_{t-1})\begin{pmatrix} 0 & X_{t-2} \\ X_{t-2} & 2\left(X_{t-2} - a_0^c X_{t-3}\right) \end{pmatrix} \\
&= J_{\theta_0^c} + \begin{pmatrix} 0 & \gamma(2) - a_0^c\gamma(1) \\ \gamma(2) - a_0^c\gamma(1) & 2\left\{1+(a_0^c)^2\right\}\gamma(2) - 2a_0^c\left\{\gamma(1)+\gamma(3)\right\} \end{pmatrix}.
\end{aligned}$$

The Bahadur slopes of the two versions of the LM tests are thus

$$c = \frac{\left(a_0^c\gamma(1) - \gamma(2)\right)^2\gamma(0)}{\left\{\gamma^2(0) - \gamma^2(1)\right\}\sigma^{2c}}$$

and, under the assumption that the denominator is strictly positive,

$$c^* = \frac{\left(a_0^c\gamma(1) - \gamma(2)\right)^2\gamma(0)}{\left\{\gamma^2(0) + 2\gamma(0)\gamma(2) - 2\frac{\gamma(0)\gamma(3)}{a_0^c} - \frac{\gamma^2(2)}{a_0^{c2}} - 4\gamma^2(1) + 4\frac{\gamma(1)\gamma(2)}{a_0^c}\right\}\sigma^{2c}}.$$

In particular, it follows that, in the Bahadur sense, the $\mathbf{LM}^*$ version is more efficient than the **LM** one for MA(1) alternatives of the form $X_t = \epsilon_t + b_0\epsilon_{t-1}$. Moreover, the asymptotic relative efficiency c^*/c tends to infinity as $|b_0|$ approches 1. For strong ARMA(1,1) alternatives of the form

$$X_t - 0.5X_{t-1} = \epsilon_t + b_0\epsilon_{t-1}, \quad \epsilon_t \text{ iid } \mathcal{N}(0,1), \tag{18}$$

tedious computations show that the $\mathbf{LM}^*$ version is inconsistent for $b_0 \leq -0.5807...$ but is more efficient than the **LM** version when $b_0 > -0.5807....$ Figure 1 shows the percentage of Bahadar asymptotic relative efficiency (ARE) gain of $\mathbf{LM}^*$ with respect to **LM**, as measured by $100(c^* - c)/c$, when b_0 varies from -0.45 to 0.55.

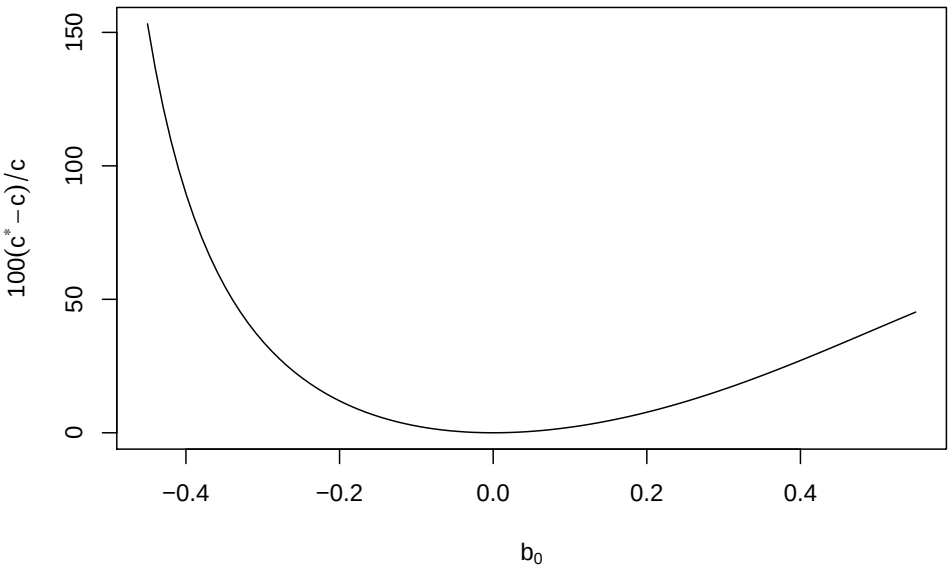

Figure 1: Relative improvement (in percentage) of the Bahadur slope of the $\mathbf{LM}^*$ test with respect to that of the **LM**-test, when the null is AR(1) and the alternative is the ARMA(1,1) model (18).

4.2.2. Finite sample comparison of the two versions

In order to determine whether the ARE computed in the previous section provide valuable insights on the actual behavior of the two tests, we simulated $N = 1\ 000$ independent trajectories of size $n = 100$, $n = 1\ 000$ and $n = 10\ 000$ of the ARMA(1,1) models (18). Table 3 displays the averaged p-values of the **LM** and $\mathbf{LM}^*$ tests of the null hypothesis of an AR(1), *i.e.* $H_0 : b_0 = 0$. The two lines in bold correspond to the null hypothesis H_0. For the line $b_0 = 0$, the DGP is an AR(1) and the test statistics **LM** and $\mathbf{LM}^*$ are asymptotically χ^2_1-distributed because

$$J_{\theta_0^c} = J^*_{\theta_0^c} = \gamma(0) \begin{pmatrix} 1 & 1-a_0^2 \\ 1-a_0^2 & 1-a_0^2 \end{pmatrix}, \quad a_0 = 1/2,$$

is invertible and the arguments of Section 4.2.1 apply. For the line $b_0 = -0.5$, the DGP is a white noise, which can also be written as an ARMA(1,1) with $a_0 = b_0 = 0$, but the arguments of Section 4.2.1 do not apply because

$$J_{\theta_0^c} = J^*_{\theta_0^c} = \gamma(0) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

is singular. As expected, the average p-value of the **LM** and $\mathbf{LM}^*$ tests are close to 0.5 when $b_0 = 0$. We also note the average p-value of the **LM**-test is

close to 0.5 when $b_0 = -0.5$, and is close to zero under the alternative when n is large. In accordance with the theoretical results of the previous section, we also note that the $\mathbf{LM}^*$-test is inconsistent for $b_0 < -0.5$ because the p-values do not tend to zero as n increases. When n is small and $b_0 > 0.5$, the p-values of the $\mathbf{LM}^*$-test are slightly smaller than those of the $\mathbf{LM}$-test, but these p-values tend rapidly to zero as n increases. For an easier comparison of the empirical behavior of the two tests, Table 4 reports the averages of the estimated Bahadur slopes $\mathbf{LM}/n$ and $\mathbf{LM}^*/n$. As expected from the asymptotic theory illustrated by Figure 1, the $\mathbf{LM}^*$ statistic is in average larger than the $\mathbf{LM}$ statistic for alternatives such that $b_0 > -0.5$. Note also that the $\mathbf{LM}$ statistic is always positive, whereas negative values of $\mathbf{LM}^*$ are observed, because $\hat{J}_n^c$ is semi-definite positive, whereas it is not the case for $\hat{J}_n^{*c}$.

To conclude this section, although the $\mathbf{LM}^*$ version may be asymptotically more efficient in Bahadur's sense than the $\mathbf{LM}$ version for particular alternatives, the $\mathbf{LM}$ version seems globally preferable because it is unbiased and consistent for a larger set of alternatives.

4.3. Testing weak ARMA models for stock returns

We now consider an application to the daily returns of 10 stock market indices (CAC, DAX, FTSE, HSI, Nikkei, NSE, SMI, SP500, SPTSX and SSE). The observations cover the period from the starting date of each index to July 26, 2010. In Financial Econometrics, the returns are often assumed to be martingale increments, and the squares of the returns have often second-order moments close to those of an ARMA(1,1) (which is compatible with a GARCH(1,1) model for the returns).

We will test these hypotheses by fitting weak ARMA models on the returns and on their squares. In view of Section 4.1.2, let $\hat{\Omega} = J_{\hat{\theta}_n,n}^{-1} I_{\hat{\theta}_n,n} J_{\hat{\theta}_n,n}^{-1}$, $\hat{\Omega}_s = \hat{\sigma}^2 J_{\hat{\theta}_n,n}^{-1}$ and $\hat{\Omega}_{ss} = J_{\hat{\theta}_n,n}^{-1} \hat{I}_{ss} J_{\hat{\theta}_n,n}^{-1}$. We will consider three versions of the Wald test of the null hypothesis defined in (17). Under the assumptions of Proposition 3.4 and the assumption that I and I_{ss} are invertible, the Wald statistics

$$\mathbf{W}_S = n(R_0\hat{\theta}_n - r_0)'(R_0\hat{\Omega}_s R_0')^{-1}(R_0\hat{\theta}_n - r_0),$$
$$\mathbf{W}_{SS} = n(R_0\hat{\theta}_n - r_0)'(R_0\hat{\Omega}_{ss} R_0')^{-1}(R_0\hat{\theta}_n - r_0),$$
$$\mathbf{W}_W = n(R_0\hat{\theta}_n - r_0)'(R_0\hat{\Omega} R_0')^{-1}(R_0\hat{\theta}_n - r_0)$$

Table 3: Averages of the p-values of **LM** and **LM*** for testing the null hypothesis of an AR(1), *i.e.* $H_0 : b_0 = 0$, in the ARMA(1,1) model (18). The number of replications is $N = 1000$.

	$n = 100$		$n = 1,000$		$n = 10,000$	
b_0	**LM**	**LM***	**LM**	**LM***	**LM**	**LM***
-0.9	0.14185	0.82045	0.00000	0.98700	0.00000	1.00000
-0.8	0.19998	0.78249	0.00028	0.94300	0.00000	1.00000
-0.7	0.32798	0.77807	0.00917	0.81042	0.00000	0.98900
-0.6	0.46288	0.81641	0.18710	0.67573	0.00004	0.61900
-0.5	**0.49549**	**0.85000**	**0.50789**	**0.84556**	**0.51899**	**0.85018**
-0.4	0.48500	0.77410	0.29030	0.48105	0.00252	0.01066
-0.3	0.46271	0.65144	0.14221	0.13635	0.00000	0.00000
-0.2	0.44848	0.50338	0.15298	0.14171	0.00000	0.00000
-0.1	0.49491	0.51222	0.29239	0.28933	0.00368	0.00345
0.0	**0.50143**	**0.49780**	**0.51095**	**0.50941**	**0.51854**	**0.51830**
0.1	0.44112	0.42938	0.17563	0.17287	0.00003	0.00003
0.2	0.29159	0.27676	0.00483	0.00388	0.00000	0.00000
0.3	0.14343	0.12986	0.00005	0.00000	0.00000	0.00000
0.4	0.05982	0.04610	0.00000	0.00000	0.00000	0.00000
0.5	0.01696	0.01194	0.00000	0.00000	0.00000	0.00000
0.6	0.00752	0.00758	0.00000	0.00000	0.00000	0.00000
0.7	0.00227	0.01223	0.00000	0.00000	0.00000	0.00000
0.8	0.00156	0.01510	0.00000	0.00000	0.00000	0.00000
0.9	0.00112	0.01109	0.00000	0.00000	0.00000	0.00000

asymptotically follow a $\chi^2_{s_0}$ distribution under H_0. At the asymptotic level α, each Wald test consists in rejecting H_0 when its statistic is greater than $\chi^2_{s_0}(1-\alpha)$.

4.3.1. Testing a white noise against an AR(1)

In this section, we fit AR(1) models on each series of daily returns, and we apply the above-mentioned Wald tests for testing the hypothesis that the returns constitute a white noise. This testing problem can be trivially written under the form (17). Table 5 displays the p-values of the standard and

Table 4: Estimated Bahadur slopes of the **LM** and **LM*** tests, when the null hypothesis is an AR(1) and the alternative is the ARMA(1,1) model (18). The number of replications is $N = 1000$.

	$n = 100$		$n = 1,000$		$n = 10,000$	
b_0	**LM**	**LM***	**LM**	**LM***	**LM**	**LM***
-0.9	0.04712	4.17203	0.03717	-0.05981	0.03631	-0.04923
-0.8	0.03654	-0.13111	0.02654	5.93785	0.02589	-0.04407
-0.7	0.02320	0.98988	0.01376	-0.03605	0.01269	-0.04327
-0.6	0.01280	-0.00635	0.00404	0.01628	0.00313	0.00297
-0.5	**0.01035**	**-0.00211**	**0.00102**	**0.00020**	**0.00010**	**0.00001**
-0.4	0.01050	-0.01204	0.00269	0.00062	0.00186	0.00449
-0.3	0.01282	0.00389	0.00482	0.00794	0.00425	0.00591
-0.2	0.01240	0.01819	0.00487	0.00600	0.00417	0.00471
-0.1	0.01060	0.00009	0.00271	0.00287	0.00180	0.00185
0.0	**0.01036**	**0.00721**	**0.00092**	**0.00095**	**0.00008**	**0.00008**
0.1	0.01393	0.01720	0.00422	0.00438	0.00335	0.00343
0.2	0.02607	0.03886	0.01600	0.01750	0.01510	0.01627
0.3	0.04626	0.05910	0.03781	0.04417	0.03660	0.04254
0.4	0.07466	0.22160	0.06637	0.08452	0.06550	0.08327
0.5	0.10383	0.17138	0.09796	0.13724	0.09764	0.13601
0.6	0.13199	0.28353	0.12860	0.19518	0.12850	0.19402
0.7	0.15988	0.28823	0.15401	0.25127	0.15339	0.24780
0.8	0.17431	0.34168	0.16972	0.29062	0.17080	0.28910
0.9	0.18064	0.33781	0.17949	0.31841	0.18062	0.31362

modified Wald tests. For the NSE, SMI, SP500 and SPTSX series, the white noise hypothesis is rejected by the $\mathbf{W}_S$ test at the nominal level $\alpha = 5\%$. This is not surprising because the $\mathbf{W}_S$ test required the iid assumption and, in particular in view of the so-called volatility clustering, it is well known that the strong white noise model is not adequate for these series. By contrast, the white noise hypothesis is not rejected by the modified tests based on $\mathbf{W}_{SS}$ and $\mathbf{W}_W$. To summary, the outputs of Table 5 are in accordance with the common belief that these series are not strong white noises, but could be weak white noises (or even martingale increments).

We now turn to the dynamics of the squared returns.

Table 5: For standard and modified versions of Wald tests, p-values of the null hypothesis that the returns are white noises. The p-values which are less than $\alpha = 5\%$ are displayed in bold.

Returns	Length n	$\mathbf{W}_S$	$\mathbf{W}_{SS}$	$\mathbf{W}_W$
CAC	5154	0.386	0.570	0.486
DAX	4966	0.343	0.521	0.349
FTSE	6647	0.705	0.857	0.760
HSI	5849	0.144	0.631	0.356
Nikkei	6530	0.057	0.328	0.159
NSE	1990	**0.038**	0.250	0.082
SMI	4963	**0.035**	0.264	0.060
SP500	15237	**0.000**	0.073	0.019
SPTSX	2665	**0.042**	0.321	0.105
SSE	2716	0.707	0.781	0.758

4.3.2. Testing the ARMA$(1,1)$ model for the squared returns

We fitted ARMA(p,q) models with $p = 1$ and $q > 1$, or $q = 1$ and $p > 1$, on the squares of the previous daily returns, and we applied Wald tests for testing the null hypothesis of an ARMA$(1,1)$ model. The p-values of the standard and modified Wald tests are displayed in Table 6. The standard Wald test frequently rejects the ARMA(1,1) model. The validity of this test is however questionable, because the assumption of iid linear innovations is not very plausible, as well for the squared returns than for the returns themselves (as was discussed in the previous section). If the returns are assumed to follow a GARCH(1,1), which is one of the most widely used model for such series, then the squared returns follow a semi-strong ARMA(1,1), and higher-order powers follow ARMA models which are only weak (see). The tests based on the statistics $\mathbf{W}_{SS}$ and $\mathbf{W}_W$ thus appear as more reliable, a priori. These tests also frequently reject the ARMA(1,1) model in favor of more complex models. This leads us to reconsider the common belief that the GARCH(1,1) model is sufficient to capture the dynamics of most financial returns, and that higher-order models would be unnecessarily complicated. Francq and Zakoïan (2009a) drew the same conclusion from parametric tests

on GARCH models. The advantage of the present study is that it leads to reconsider not only the GARCH(1,1) model, but also any parametric model leading to a weak ARMA(1,1) for the squares.

5. Conclusion

The asymptotic variance of the LSE of ARMA models depend on information matrices I and J computed at the true value of the parameter θ_0. It is sometime necessary to evaluate these matrices at some point $\theta \neq \theta_0$. In the case of strong ARMA models, I_θ and J_θ depend only on θ_0 and on the moments $\sigma^2 = E\epsilon_t^2$ and $\mu_4 = E\epsilon_t^4$ of the iid noise ϵ_t. In the much more general case of weak ARMA models, I_θ also depends on the autocovariances of the weak white noise ϵ_t^2.

We proposed here algorithms for the exact computation of I_θ and J_θ from the model, and for the estimation of these matrices from the data. It is possible to define estimators of the information matrices which are consistent in the general weak case, or in the more restrictive semi-strong case, or only in the strong case. Simulations experiments confirmed the domain of validity of the different estimators, and also that an efficiency loss is the price to pay for having more robust estimators. As an illustration of the interest of considering I_θ and J_θ at $\theta \neq \theta_0$, we computed and compared Bahadur's slopes of two versions of the Lagrange-Multiplier test for testing general linear restriction on θ_0 in the strong ARMA case. The two versions are based on two estimators $\hat{J}_n^c$ and $\hat{J}_n^{*c}$ of J under the null. The standard estimator of J is $\hat{J}_n^c$, whereas $\hat{J}_n^{*c}$ contains an extra term which is asymptotically negligible under the null but may have importance under the alternative. We showed, analytically and also by means of simulations, that the version based on $\hat{J}_n^{*c}$ may be asymptotically much more efficient than the standard version, but is consistent for a narrower set of alternatives. Applying different Wald tests based on different estimators of the information matrices, and applying them for testing weak ARMA specifications on daily stock returns and on their squares, we reconsidered models such as the popular GARCH(1,1) for which the squares follow a weak ARMA(1,1).

Table 6: As Table 5, but for the null hypothesis that the squared returns follow an ARMA(1, 1) model.

Alternative	Returns	$\mathbf{W}_S$	$\mathbf{W}_{SS}$	$\mathbf{W}_W$
	CAC	**0.000**	0.228	0.167
	DAX	**0.000**	**0.013**	**0.000**
	FTSE	0.767	0.949	0.962
	HSI	**0.000**	0.351	**0.000**
ARMA(2, 1)	Nikkei	0.483	0.922	0.940
	NSE	**0.027**	0.485	0.570
	SMI	0.589	0.886	0.803
	SP500	**0.014**	0.641	0.503
	SPTSX	**0.009**	0.472	0.443
	SSE	**0.042**	0.281	0.139
	CAC	**0.000**	0.288	0.064
	DAX	**0.000**	0.060	**0.013**
	FTSE	0.828	0.957	0.941
	HSI	**0.000**	**0.005**	**0.000**
ARMA(1, 2)	Nikkei	0.594	0.936	0.938
	NSE	0.053	0.305	0.191
	SMI	0.668	0.897	0.863
	SP500	0.069	0.737	0.793
	SPTSX	**0.001**	0.312	0.571
	SSE	**0.040**	0.067	**0.025**
	CAC	**0.003**	0.617	0.228
	DAX	**0.000**	**0.012**	**0.000**
	FTSE	**0.000**	0.396	0.181
	HSI	**0.000**	**0.000**	**0.000**
ARMA(1, 3)	Nikkei	**0.000**	0.419	**0.000**
	NSE	0.286	0.742	0.621
	SMI	**0.000**	0.328	0.041
	SP500	**0.000**	**0.000**	**0.000**
	SPTSX	**0.000**	**0.001**	**0.000**
	SSE	**0.006**	**0.022**	**0.001**

A. Proofs

Proof of Proposition 3.3. The first result follows from (7) and

$$I_\theta(i,k) = \sum_{h=-\infty}^{+\infty} \sum_{h_1,h_2,h_3,h_4=0}^{+\infty} c_{h_1} c^a_{h_2} c_{h_3} c^a_{h_4} \mathrm{Cov}\left(\epsilon_{t-h_1}\epsilon_{t-h_2-i}, \epsilon_{t-h-h_3}\epsilon_{t-h-h_4-k}\right).$$

The other results follow similarly. □

Proof of Proposition 3.3. Let $\tilde{c}_h = c_h$ for $0 \le h \le M$ and $\tilde{c}_h = 0$ for $h > M$. Similarly, we defined $\tilde{c}^a_h$ and $\tilde{c}^b_h$. For $1 \le i \le k \le p$, we have

$$\begin{aligned} I_\theta(i,k) - I^M_\theta(i,k) = \sum_{h_1,h_2,h_3,h_4=0}^{\infty} \big\{ &(c_{h_1} - \tilde{c}_{h_1}) c^a_{h_2} c_{h_3} c^a_{h_4} + \tilde{c}_{h_1}(c^a_{h_2} - \tilde{c}^a_{h_2}) c_{h_3} c^a_{h_4} \\ &+ \tilde{c}_{h_1} \tilde{c}^a_{h_2} (c_{h_3} - \tilde{c}_{h_3}) c^a_{h_4} + \tilde{c}_{h_1} \tilde{c}^a_{h_2} \tilde{c}_{h_3} (c^a_{h_4} - \tilde{c}^a_{h_4}) \big\} \Gamma(h_2 + i - h_1, h_4 + k - h_3). \end{aligned}$$

Note that if

$$\max_{i=1,\ldots,k} |\rho_i| \le \overline{\rho} < 1$$

then for all $|z| \le 1$,

$$\frac{1}{\prod_{i=1}^k (1 - \rho_i z)} = \sum_{h=0}^{\infty} d_h z^h, \qquad \text{with} \quad |d_h| \le (h+1)^{k-1} \overline{\rho}^h.$$

Note also that

$$\phi(z)\psi_0(z) = \left(1 - \sum_{i=1}^p a_i z^i\right)\left(1 + \sum_{j=1}^q b_{0j} z^j\right) = \sum_{\ell=0}^{p+q} \pi_\ell z^\ell,$$

where, with the convention $a_0 = -1$, $b_{00} = 1$ and $a_i = 0$ for $i < 0$,

$$|\pi_\ell| = \left| \sum_{j=0}^q b_{0j} a_{\ell - j} \right| \le \overline{\pi}.$$

We thus have

$$\max\left\{|c_h|, |c^a_h|, \left|c^b_h\right|\right\} \le (p+q+1)\overline{\pi}(h+1)^{k_0}\overline{\rho}^h \le K_0 \overline{\rho}^{h/2}$$

with

$$k_0 = p + 2q - 1, \qquad K_0 = (p+q+1)\overline{\pi}\left(\frac{-2k_0}{\log \overline{\rho}}\right)^{k_0} \overline{\rho}^{\,-0.5 - k_0/\log\overline{\rho}}.$$

We then obtain

$$\left|I_\theta(i,k) - I_\theta^M(i,k)\right| \leq 4\overline{\Gamma}\left(\frac{K}{1-\overline{\rho}^{1/2}}\right)^4 \overline{\rho}^{(M+1)/2}$$

and the result follows. □

In the following proofs, K and ρ denote generic constant such as $K > 0$ and $\rho \in (0,1)$, whose exact values are unimportant.

Proof of Proposition 3.4. Note that because the roots ψ and ϕ_0 are outside the unit circle,

$$\max\left\{\left|c_h^a\right|, \left|c_h^b\right|, \left|c_h^{ab}\right|, \left|c_h^{bb}\right|\right\} \leq K\rho^h. \tag{19}$$

Because Θ^* is compact, this inequality holds uniformly in $\theta \in \Theta^*$. Note that this entails the existence of J_{θ^*} and $J^*_{\theta^*}$. The ergodic theorem then shows that

$$J_{\theta^*,n} \to J_{\theta^*} \text{ and } J^*_{\theta^*,n} \to J^*_{\theta^*} \text{ a.s.} \tag{20}$$

By the previous arguments, for all $i, i = 1, \ldots, p+q$, we have

$$E \sup_{\theta\in\Theta^*} \left\| \frac{\partial}{\partial\theta} \frac{\partial\epsilon_t(\theta)}{\partial\theta_i} \frac{\partial\epsilon_t(\theta)}{\partial\theta_j} \right\| < \infty. \tag{21}$$

A Taylor expansion yields

$$J_{\theta_n^*,n}(i,j) = J_{\theta^*,n}(i,j) + (\theta_n^* - \theta^*)' \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta}\left\{\frac{\partial\epsilon_t(\theta)}{\partial\theta_i}\frac{\partial\epsilon_t(\theta)}{\partial\theta_j}\right\}(\theta^{**})$$

for some θ^{**} between θ_n^* and θ^*. The consistency of $J_{\theta_n^*,n}$ follows from (20), (21) and the convergence of θ_n^* to θ^*. The consistency of $J^*_{\theta_n^*,n}$ is shown similarly. □

References

Andrews, D.W.K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817–858, 1991.

Berk, K.N. Consistent Autoregressive Spectral Estimates. *Ann. Stat.* 2:489–502, 1974.

Box, G.E.P. and Jenkins, G.M. *Time Series Analysis: Forecasting and Control, 2nd edition.* Holden-Day, San Francisco, 1976.

Bradley, R.C. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2:107–144, 2005.

Brockwell, P.J. and Davis, R.A. *Time series: theory and methods.* Springer Verlag, New York, 1991.

Durbin, J. Efficient Estimation of Parameters in Moving Average Models. *Biometrika* 46:306–316, 1959.

den Haan, W.J. and Levin, A. A Practitioner's Guide to Robust Covariance Matrix Estimation. *In Handbook of Statistics* 15, Rao, C.R. and G.S. Maddala (eds), 291–341, 1997.

Engle, R.F. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of Econometrics* 2, Z. Griliches and M.D. Intriligator (eds), Amsterdam: North-Holland, 775–826, 1984.

Francq, C., Roy, R. and Zakoïan, J-M. Diagnostic checking in ARMA Models with Uncorrelated Errors. *Journal of the American Statistical Association* 100:532–544, 2005.

Francq, C. and Zakoïan, J-M. Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68:145–165, 1998.

Francq, C. and Zakoïan, J-M. Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* 83:369–394, 2000.

Francq, C. and Zakoïan, J-M. Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems,* Chap. 12 (eds P. Duchesne and B. Rémillard). New York: Springer Verlag, 241–265, 2005.

Francq, C. and Zakoïan, J-M. HAC estimation and strong linearity testing in weak ARMA models. *Journal of Multivariate Analysis* 98:114–144, 2007.

Francq, C. and Zakoïan, J-M. Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104:313–324, 2009a.

Francq, C. and Zakoïan, J-M. Bartlett's formula for a general class of non linear processes. *Journal of Time Series Analysis* 30:449–465, 2009b.

Francq, C. and Zakoïan, J-M. *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley, 2010.

Godolphin, E.J. and Bane, S.R. On the evaluation of the information matrix for multiplicative seasonal time series. *Journal of Time Series Analysis* 27: 167–190, 2006.

Klein, A. and Mélard, G. Fisher's information matrix for seasonal moving average models. *Journal of Time Series Analysis* 11: 231–237, 1990.

Klein, A. and Mélard, G. An algorithm for computing the asymptotic Fisher information matrix for seasonal SISO models. *Journal of Time Series Analysis* 25: 627–648, 2004.

McKenzie, E. Discrete variate time series. *In Handbook of Statistics*, Rao, C.R. and D. Shanbhag (eds), Elsevier Science, Amsterdam, 573–606, 2003.

McLeod, A.I. On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society* B 40:296–302, 1978.

McLeod, A.I. Duality and other properties of multiplicative seasonal autoregressive-moving average models. *Biometrika* 71:207–211, 1984.

Newey, W.K. and West, K.D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708, 1987.

Romano, J.L. and Thombs, L.A. Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91:590–600, 1996.

Siddiqui, M.M. On the inversion of the sample covariance matrix in a stationary autogressive process. *Annals of Mathematical Statistics* 29:585–588, 1958.

Whittle, P. Estimation and information in stationary time series. *Arkiv för Matematik* 2: 423–434, 1953.

Whittle, P. On the fitting of multivariate autoregressions and the approximate canonical factorization of a special density matrix. *Biometrika* 40:129–134, 1963.

Computing and estimating information matrices of weak ARMA models: a complementary result which is not submitted for publication

A. Asymptotic properties of the spectral density estimator of the long-run variance I_{θ^*}

Theorem A.1. *Let the assumptions of Lemma 2.1 be satisfied. Assume that the process* (∇_t) *defined by (12) admits the* $AR(\infty)$ *representation (13), where* $\|A_i\| = o(i^{-2})$ *as* $i \to \infty$*, the roots of* $\det(\mathcal{A}(z)) = 0$ *are outside the unit disk, and* Σ_u *is non-singular. Assume moreover that* $E|\epsilon_t|^{8+4\nu} < \infty$ *and* $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$*. Then, if* $\theta_n^* \to \theta^*$ *almost surely,*

$$I_{\theta_n^*,n} = \hat{\mathcal{A}}_r^{-1}(1)\hat{\Sigma}_{u_r}\hat{\mathcal{A}}_r'^{-1}(1) \to I_{\theta^*}$$

in probability when $r = r(n) \to \infty$ *and* $r^3/n \to 0$ *as* $n \to \infty$*.*

The proof of Theorem A.1 is based on a series of lemmas. We use the multiplicative matrix norm defined by: $\|A\| = \sup_{\|x\|\leq 1}\|Ax\| = \varrho^{1/2}(A'A)$, where A is a $d_1 \times d_2$ matrix, $\|x\|$ is the Euclidean norm of the vector $x \in \mathbb{R}^{d_2}$, and $\varrho(\cdot)$ denotes the spectral radius. This norm satisfies

$$\|A\|^2 \leq \sum_{i,j} a_{i,j}^2 \tag{22}$$

with obvious notations. This choice of the norm is crucial for the following lemma to hold (with e.g. the Euclidean norm, this result is not valid). Let

$$\Sigma_{\nabla,\underline{\nabla}_r} = E\nabla_t\underline{\nabla}_{r,t}', \quad \Sigma_\nabla = E\nabla_t\nabla_t', \quad \Sigma_{\underline{\nabla}_r} = E\underline{\nabla}_{r,t}\underline{\nabla}_{r,t}'.$$

In the sequel, K and ρ denote generic constant such as $K > 0$ and $\rho \in (0,1)$, whose exact values are unimportant.

Lemma 1. *Under the assumptions of Theorem A.1,*

$$\sup_{r\geq 1}\max\left\{\left\|\Sigma_{\nabla,\underline{\nabla}_r}\right\|, \left\|\Sigma_{\underline{\nabla}_r}\right\|, \left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|\right\} \leq \infty.$$

Proof. We readily have

$$\|\Sigma_{\underline{\nabla}_r} x\| \leq \|\Sigma_{\underline{\nabla}_{r+1}}(x', 0'_{p+q})'\| \quad \text{and} \quad \|\Sigma_{\nabla,\underline{\nabla}_r} x\| \leq \|\Sigma_{\underline{\nabla}_{r+1}}(0'_{p+q}, x')'\|$$

for any $x \in \mathbb{R}^{(p+q)r}$. Therefore

$$0 < \|\mathrm{Var}\,(\nabla_t)\| = \left\|\Sigma_{\underline{\nabla}_1}\right\| \leq \left\|\Sigma_{\underline{\nabla}_2}\right\| \leq \cdots$$

and

$$\left\|\Sigma_{\nabla,\underline{\nabla}_r}\right\| \leq \left\|\Sigma_{\underline{\nabla}_{r+1}}\right\|.$$

Let $f(\lambda)$ be the spectral density of ∇_t. Because the autocovariance function of ∇_t is absolutely summable, $\|f(\lambda)\|$ is bounded by a finite constant K, say. Denoting by $\delta = (\delta'_1, \ldots, \delta'_r)'$ an eigenvector of $\Sigma_{\underline{\nabla}_r}$ associated with its largest eigenvalue, such that $\|\delta\| = 1$ and $\delta_i \in \mathbb{R}^{p+q}$ for $i = 1, \ldots, r$, we have

$$\begin{aligned} \left\|\Sigma_{\underline{\nabla}_r}\right\| &= \varrho^{1/2}(\Sigma^2_{\underline{\nabla}_r}) = \varrho(\Sigma_{\underline{\nabla}_r}) = \delta'\Sigma_{\underline{\nabla}_r}\delta \\ &= \sum_{j,k=1}^{r} \delta'_j \int_{-\pi}^{\pi} e^{i(k-j)\lambda} f(\lambda) d(\lambda) \delta_k \leq 2\pi K. \end{aligned}$$

By similar arguments, the smallest eigenvalue of $\Sigma_{\underline{\nabla}_r}$ is greater than a positive constant independent of r. Using the fact that $\|\Sigma^{-1}_{\underline{\nabla}_r}\|$ is equal to the inverse of the smallest eigenvalue of $\Sigma_{\underline{\nabla}_r}$, the proof is completed. □

Denote by $\nabla_t(i)$ the i-th element of ∇_t.

Lemma 2. *Under the assumptions of Theorem A.1, there exits a finite constant K_1 such that for $m_1, m_2 = 1, \ldots, p+q$*

$$\sup_{s\in\mathbb{Z}} \sum_{h=-\infty}^{\infty} |Cov\{\nabla_1(m_1)\nabla_{1+s}(m_2), \nabla_{1+h}(m_1)\nabla_{1+s+h}(m_2)\}| < K_1.$$

Proof. Without loss of generality, we can take the supremum over the integers $s > 0$, and write the proof in the case $m_1 = m_2 = m$. In view of (19), we have

$$\begin{aligned} &\sum_{h=-\infty}^{\infty} |\mathrm{Cov}\,\{\nabla_1(m)\nabla_{1+s}(m), \nabla_{1+h}(m)\nabla_{1+s+h}(m)\}| \\ \leq\ & K^8 \sum_{h_1,\ldots,h_8=0}^{\infty} \rho^{\sum_{i=1}^{8} h_i} \sum_{h=-\infty}^{\infty} |\mathrm{Cov}\,(Y_{1,h_1,h_2} Y_{1+s,h_3,h_4}, Y_{1+h,h_5,h_6} Y_{1+s+h,,h_7,h_8})| \end{aligned}$$

where

$$Y_{t,h_1,h_2} = \epsilon_{t-h_1}\epsilon_{t-h_2-m} - E\epsilon_{t-h_1}\epsilon_{t-h_2-m}.$$

A slight extension of Corollary A.3 in Francq and Zakoian (2010) concludes. □

Let $\hat{\Sigma}_{\underline{\nabla}_r}$, $\hat{\Sigma}_{\nabla}$ and $\hat{\Sigma}_{\nabla,\underline{\nabla}_r}$ be the matrices obtained by replacing $\hat{\nabla}_t$ by ∇_t in $\hat{\Sigma}_{\underline{\hat{\nabla}}_r}$, $\hat{\Sigma}_{\hat{\nabla}}$ and $\hat{\Sigma}_{\hat{\nabla},\underline{\hat{\nabla}}_r}$.

Lemma 3. *Under the assumptions of Theorem A.1,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\nabla}_r} - \Sigma_{\underline{\nabla}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\nabla} - \Sigma_{\nabla}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\nabla,\underline{\nabla}_r} - \Sigma_{\nabla,\underline{\nabla}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = o(n^{1/3})$.

Proof. For $1 \leq m_1, m_2 \leq p+q$ and $1 \leq r_1, r_2 \leq r$, the element of the $\{(r_1-1)(p+q)+m_1\}$-th row and $\{(r_2-1)(p+q)+m_2\}$-th column of $\hat{\Sigma}_{\underline{\nabla}_r}$ is of the form $n^{-1}\sum_{t=1}^{n} Z_t$ where $Z_t = \nabla_{t-r_1}(m_1)\nabla_{t-r_2}(m_2)$. By stationarity of (Z_t), we have

$$\mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n} Z_t\right) = \frac{1}{n^2}\sum_{h=-n+1}^{n-1}(n-|h|)\,\mathrm{Cov}\,(Z_t, Z_{t-h}) \leq \frac{K_1}{n}, \tag{23}$$

where, by Lemma 2, K_1 is a constant independent of r_1, r_2, m_1, m_2 and r, n.

In view of (22) and (23) we have

$$\begin{aligned} E\left\{r\|\hat{\Sigma}_{\nabla} - \Sigma_{\nabla}\|^2\right\} &\leq E\left\{r\|\hat{\Sigma}_{\nabla,\underline{\nabla}_r} - \Sigma_{\nabla,\underline{\nabla}_r}\|^2\right\} \\ &\leq E\left\{r\|\hat{\Sigma}_{\underline{\nabla}_r} - \Sigma_{\underline{\nabla}_r}\|^2\right\} \leq \frac{K_1(p+q)^2 r^3}{n} = o(1) \end{aligned}$$

as $n \to \infty$ when $r = o(n^{1/3})$. The result follows. □

We now show that the previous lemma applies when ∇_t is replaced by $\hat{\nabla}_t$.

Lemma 4. *Under the assumptions of Theorem A.1,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r} - \Sigma_{\underline{\nabla}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{\nabla}} - \Sigma_{\nabla}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\hat{\nabla},\underline{\hat{\nabla}}_r} - \Sigma_{\nabla,\underline{\nabla}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = o(n^{1/3})$.

Proof. We first show that the replacement of the unknown initial values $\{X_u,\ u \leq 0\}$ by zero is asymptotically unimportant. Let $\hat{\Sigma}_{\underline{\nabla}_{r,n}}$ be the matrix obtained by replacing $e_t(\theta_n^*)$ by $\epsilon_t(\theta_n^*)$ in $\hat{\Sigma}_{\underline{\hat{\nabla}}_r}$. Because $\epsilon_t(\theta)$ and their

derivatives have ARMA representations (see Section 3), it is easy to show that

$$\sup_{\theta\in\Theta^*}|e_t(\theta)-\epsilon_t(\theta)|\leq K\rho^t,\quad \sup_{\theta\in\Theta^*}\left\|\frac{\partial}{\partial\theta}e_t(\theta)-\frac{\partial}{\partial\theta}\epsilon_t(\theta)\right\|\leq K\rho^t.$$

It can be deduced that $\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\hat{\Sigma}_{\underline{\nabla}_{r,n}}\|=O_P(rn^{-1})$. We thus have

$$\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\hat{\Sigma}_{\underline{\nabla}_{r,n}}\|=o_P(1). \tag{24}$$

Taylor expansions around θ^* yield

$$|\epsilon_t(\theta_n^*)-\epsilon_t(\theta^*)|\leq r_t\|\theta_n^*-\theta^*\|,\quad \left|\frac{\partial\epsilon_t(\theta_n^*)}{\partial\theta_m}-\frac{\partial\epsilon_t(\theta^*)}{\partial\theta_m}\right|\leq s_t\|\theta_n^*-\theta^*\|$$

with $r_t=\left\|\frac{\partial}{\partial\theta'}\epsilon_t(\overline{\theta})\right\|$, $s_t=\left\|\frac{\partial^2}{\partial\theta'\partial\theta_m}\epsilon_t(\underline{\theta})\right\|$ where $\overline{\theta}$ and $\underline{\theta}$ are between θ_n^* and θ^*. Define Z_t as in the proof of Lemma 3, and let $Z_{t,n}$ be obtained by replacing $\nabla_t(m)$ by $\nabla_{t,n}(m)=\epsilon_t(\theta_n^*)\partial\epsilon_t(\theta_n^*)/\partial\theta_m$ in Z_t. We have

$$\begin{aligned}|\nabla_t(m)-\nabla_{t,n}(m)| &\leq r_t\|\theta_n^*-\theta^*\|\left|\frac{\partial}{\partial\theta_m}\epsilon_t(\theta^*)\right|+s_t\|\theta_n^*-\theta^*\||\epsilon_t(\theta_n^*)|\\ &:= \|\theta_n^*-\theta^*\|d_{t,n,m},\end{aligned}$$

and thus

$$|Z_t-Z_{t,n}| \leq \|\theta_n^*-\theta^*\|D_{t,n,m_1,m_2,r_1,r_2},$$

where

$$D_{t,n,m_1,m_2,r_1,r_2}=|d_{t-r_1,n,m_1}\nabla_{t-r_2}(m_2)|+|\nabla_{t-r_1,n}(m_1)d_{t-r_2,n,m_2}|.$$

Note that $E|D_{t,n,m_1,m_2,r_1,r_2}|\leq K$ for some constant K independent of n,r_1,r_2,m_1 and m_2. Thus

$$\|\hat{\Sigma}_{\underline{\nabla}_{r,n}}-\hat{\Sigma}_{\underline{\nabla}_r}\|^2\leq r^2\|\theta_n^*-\theta^*\|^2O_P(1).$$

Since $\|\theta_n^*-\theta^*\|=O_P\left(n^{-1/2}\right)$, we obtain for $r=o(n^{1/3})$

$$\sqrt{r}\|\hat{\Sigma}_{\underline{\nabla}_{r,n}}-\hat{\Sigma}_{\underline{\nabla}_r}\|=o_P(1). \tag{25}$$

By Lemma 3 , (24) and (25) show that $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\|=o_P(1)$. The other results are obtained similarly. □

Write $\underline{A}_r^*=(A_1\cdots A_r)$ where the A_i's are defined by (13).

Lemma 5. *Under the assumptions of Theorem A.1,*

$$\sqrt{r}\,\|\underline{A}_r^* - \underline{A}_r\| \to 0,$$

as $r \to \infty$.

Proof. Recall that by (13) and (15)

$$\nabla_t \;=\; \underline{A}_r\underline{\nabla}_{r,t} + u_{r,t} = \underline{A}_r^*\underline{\nabla}_{r,t} + \sum_{i=r+1}^{\infty} A_i\nabla_{t-i} + u_t := \underline{A}_r^*\underline{\nabla}_{r,t} + u_{r,t}^*.$$

Hence, using the orthogonality conditions in (13) and (15)

$$\underline{A}_r^* - \underline{A}_r \;=\; -\Sigma_{u_r^*,\underline{\nabla}_r}\Sigma_{\underline{\nabla}_r}^{-1} \tag{26}$$

where $\Sigma_{u_r^*,\underline{\nabla}_r} = Eu_{r,t}^*\underline{\nabla}_{r,t}'$. Using arguments and notations of the proof of Lemma 2, there exists a constant K_2 independent of s and m_1, m_2 such that

$$E\,|\nabla_1(m_1)\nabla_{1+s}(m_2)| \le K^4 \sum_{h_1,\ldots,h_4=0}^{\infty} \rho^{h_1+\cdots+h_4}\|\epsilon_1\|_4^4 \le K_2.$$

By the Cauchy-Schwarz inequality and (22), we then have

$$\left\|\mathrm{Cov}\left(\nabla_{t-r-h}, \underline{\nabla}_{r,t}\right)\right\| \le K_2 r^{1/2}(p+q).$$

Thus,

$$\begin{aligned}\|\Sigma_{u_r^*,\underline{\nabla}_r}\| \;&=\; \|\sum_{i=r+1}^{\infty} A_i E\nabla_{t-i}\underline{\nabla}_{r,t}'\| \le \sum_{h=1}^{\infty}\|A_{r+h}\|\left\|\mathrm{Cov}\left(\nabla_{t-r-h}, \underline{\nabla}_{r,t}\right)\right\| \\ &=\; O(1)r^{1/2}\sum_{h=1}^{\infty}\|A_{r+h}\|. \end{aligned} \tag{27}$$

Note that the assumption $\|A_i\| = o\left(i^{-2}\right)$ entails $r\sum_{h=1}^{\infty}\|A_{r+h}\| = o(1)$ as $r \to \infty$. The lemma therefore follows from (26), (27) and Lemma 1. □

The following lemma is similar to Lemma 3 in Berk (1974).

Lemma 6. *Under the assumptions of Theorem A.1,*

$$\sqrt{r}\|\hat{\Sigma}_{\hat{\underline{\nabla}}_r}^{-1} - \Sigma_{\underline{\nabla}_r}^{-1}\| \;=\; o_P(1)$$

as $n \to \infty$ *when* $r = o(n^{1/3})$ *and* $r \to \infty$.

Proof. We have

$$\begin{aligned}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}\right\| &= \left\|\left\{\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}+\Sigma_{\underline{\nabla}_r}^{-1}\right\}\left\{\Sigma_{\underline{\nabla}_r}-\hat{\Sigma}_{\underline{\hat{\nabla}}_r}\right\}\Sigma_{\underline{\nabla}_r}^{-1}\right\| \\ &\leq \left(\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}\right\|+\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|\right)\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|.\end{aligned}$$

Iterating this inequality, we obtain

$$\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}\right\| \leq \left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|\sum_{i=1}^{\infty}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|^i\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|^i.$$

Thus, for every $\varepsilon>0$,

$$\begin{aligned}& P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}\right\|>\varepsilon\right) \\ \leq\ & P\left(\sqrt{r}\frac{\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|^2\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|}{1-\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|}>\varepsilon \text{ and } \left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|<1\right) \\ & +P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|\geq 1\right) \\ \leq\ & P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|>\frac{\varepsilon}{\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|^2+\varepsilon r^{-1/2}\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|}\right) \\ & +P\left(\sqrt{r}\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}-\Sigma_{\underline{\nabla}_r}\right\|\geq\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|^{-1}\right)=o(1)\end{aligned}$$

by Lemmas 3 and 1. This establishes Lemma 6. □

Lemma 7. *Under the assumptions of Theorem A.1,*

$$\sqrt{r}\left\|\underline{\hat{A}}_r-\underline{A}_r\right\|=o_P(1)$$

as $r\to\infty$ *and* $r=o(n^{1/3})$.

Proof. By the triangle inequality and Lemmas 1 and 6, we have

$$\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}\right\|\leq\left\|\hat{\Sigma}_{\underline{\hat{\nabla}}_r}^{-1}-\Sigma_{\underline{\nabla}_r}^{-1}\right\|+\left\|\Sigma_{\underline{\nabla}_r}^{-1}\right\|=O_P(1). \tag{28}$$

Note that the orthogonality conditions in (15) entail that $\underline{A}_r = \Sigma_{\nabla,\underline{\nabla}_r}\Sigma_{\underline{\nabla}_r}^{-1}$. By Lemmas 1, 3, 6, and (28), we then have

$$\begin{aligned}
& \sqrt{r}\left\|\hat{\underline{A}}_r - \underline{A}_r\right\| = \sqrt{r}\left\|\hat{\Sigma}_{\hat{\nabla},\hat{\underline{\nabla}}_r}\hat{\Sigma}_{\hat{\underline{\nabla}}_r}^{-1} - \Sigma_{\nabla,\underline{\nabla}_r}\Sigma_{\underline{\nabla}_r}^{-1}\right\| \\
= \; & \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{\nabla},\hat{\underline{\nabla}}_r} - \Sigma_{\nabla,\underline{\nabla}_r}\right)\hat{\Sigma}_{\hat{\underline{\nabla}}_r}^{-1} + \Sigma_{\nabla,\underline{\nabla}_r}\left(\hat{\Sigma}_{\hat{\underline{\nabla}}_r}^{-1} - \Sigma_{\underline{\nabla}_r}^{-1}\right)\right\| = o_P(1).
\end{aligned}$$

□

Proof of Theorem A.1. In view of (14), it suffices to show that $\hat{\mathcal{A}}_r(1) \to \mathcal{A}(1)$ and $\hat{\Sigma}_{u_r} \to \Sigma_u$ in probability. Let the $r \times 1$ vector $\mathbf{1}_r = (1, \ldots, 1)'$ and the $r(p+q) \times (p+q)$ matrix $\mathbf{E}_r = \mathbb{I}_{p+q} \otimes \mathbf{1}_r$, where $\otimes$ denotes the matrix Kronecker product and $\mathbb{I}_d$ the $d \times d$ identity matrix. Using (22), and Lemmas 5, 7, we obtain

$$\begin{aligned}
\left\|\hat{\mathcal{A}}_r(1) - \mathcal{A}(1)\right\| &\leq \left\|\sum_{i=1}^{r}\hat{A}_{r,i} - A_{r,i}\right\| + \left\|\sum_{i=1}^{r}A_{r,i} - A_i\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&= \left\|\left(\hat{\underline{A}}_r - \underline{A}_r\right)\mathbf{E}_r\right\| + \left\|\left(\underline{A}_r^* - \underline{A}_r\right)\mathbf{E}_r\right\| + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&\leq \sqrt{p+q}\sqrt{r}\left\{\left\|\hat{\underline{A}}_r - \underline{A}_r\right\| + \|\underline{A}_r^* - \underline{A}_r\|\right\} + \left\|\sum_{i=r+1}^{\infty}A_i\right\| \\
&= o_P(1).
\end{aligned}$$

Now note that

$$\hat{\Sigma}_{u_r} = \hat{\Sigma}_{\hat{\nabla}} - \hat{\underline{A}}_r\hat{\Sigma}'_{\hat{\nabla},\hat{\underline{\nabla}}_r}$$

and, by (13)

$$\begin{aligned}
\Sigma_u &= Eu_tu_t' = Eu_t\nabla_t' = E\left\{\left(\nabla_t - \sum_{i=1}^{\infty}A_i\nabla_{t-i}\right)\nabla_t'\right\} \\
&= \Sigma_\nabla - \sum_{i=1}^{\infty}A_iE\nabla_{t-i}\nabla_t' = \Sigma_\nabla - \underline{A}_r^*\Sigma'_{\nabla,\underline{\nabla}_r} - \sum_{i=r+1}^{\infty}A_iE\nabla_{t-i}\nabla_t'.
\end{aligned}$$

Thus,

$$\begin{aligned}\left\|\hat{\Sigma}_{u_r}-\Sigma_u\right\| &= \Big\|\hat{\Sigma}_{\hat{\nabla}}-\Sigma_{\nabla}-\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r} \\ &\quad -\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r}-\Sigma'_{\nabla,\underline{\nabla}_r}\right)+\sum_{i=r+1}^{\infty}A_iE\nabla_{t-i}\nabla_t'\Big\| \\ &\leq \left\|\hat{\Sigma}_{\hat{\nabla}}-\Sigma_{\nabla}\right\|+\left\|\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\left(\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r}-\Sigma'_{\nabla,\underline{\nabla}_r}\right)\right\| \\ &\quad +\left\|\left(\underline{\hat{A}}_r-\underline{A}_r^*\right)\Sigma'_{\nabla,\underline{\nabla}_r}\right\|+\left\|\underline{A}_r^*\left(\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r}-\Sigma'_{\nabla,\underline{\nabla}_r}\right)\right\| \\ &\quad +\left\|\sum_{i=r+1}^{\infty}A_iE\nabla_{t-i}\nabla_t'\right\|. \qquad (29)\end{aligned}$$

In the right-hand side of this inequality, the first norm is $o_P(1)$ by Lemma 3. By Lemmas 5 and 7, we have $\|\underline{\hat{A}}_r-\underline{A}_r^*\|=o_p(r^{-1/2})=o_p(1)$, and by Lemma 3, $\|\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r}-\Sigma'_{\nabla,\underline{\nabla}_r}\|=o_p(r^{-1/2})=o_p(1)$. Therefore the second norm in the right-hand side of (29) tends to zero in probability. The third norm tends to zero in probability because $\|\underline{\hat{A}}_r-\underline{A}_r^*\|=o_p(1)$ and, by Lemma 1, $\|\Sigma'_{\nabla,\underline{\nabla}_r}\|=O(1)$. The fourth norm tends to zero in probability because, in view of Lemma 3, $\|\hat{\Sigma}'_{\hat{\nabla},\underline{\hat{\nabla}}_r}-\Sigma'_{\nabla,\underline{\nabla}_r}\|=o_p(1)$, and, in view of (22), $\|\underline{A}_r^*\|^2\leq\sum_{i=1}^{\infty}\mathrm{Tr}(A_iA_i')<\infty$. Clearly, the last norm tends to zero, which completes the proof. □