INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
*Int. J. Adapt. Control Signal Process.* (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/acs.2424

# On the asymptotic behavior of the Durbin–Watson statistic for ARX processes in adaptive tracking

Bernard Bercu [1,*,†], Bruno Portier [2] and Victor Vazquez [3]

[1] *Institut de Mathématiques de Bordeaux, Université Bordeaux 1, UMR 5251, 351 cours de la libération, 33405 Talence Cedex, France*
[2] *Département de Génie Mathématiques, Laboratoire de Mathématiques, INSA de Rouen, LMI-EA 3226, BP 08, Avenue de l'Université, 76800 Saint-Etienne du Rouvray, France*
[3] *Facultad de Ciencias Fisico, Universidad Autónoma de Puebla, Matemáticas, Avenida San Claudio y Rio Verde, 72570 Puebla, Mexico*

SUMMARY

A wide literature is available on the asymptotic behavior of the Durbin–Watson statistic for autoregressive models. However, it is impossible to find results on the Durbin–Watson statistic for autoregressive models with adaptive control. Our purpose is to fill the gap by establishing the asymptotic behavior of the Durbin–Watson statistic for ARX models in adaptive tracking. On the one hand, we show the almost sure convergence as well as the asymptotic normality of the least squares estimators of the unknown parameters of the ARX models. On the other hand, we establish the almost sure convergence of the Durbin–Watson statistic and its asymptotic normality. Finally, we propose a bilateral statistical test for residual autocorrelation in adaptive tracking. Copyright © 2013 John Wiley & Sons, Ltd.

Received 4 December 2012; Revised 2 July 2013; Accepted 8 July 2013

KEY WORDS: Durbin–Watson statistic; estimation; adaptive control; almost sure convergence; central limit theorem; statistical test for serial autocorrelation

## 1. INTRODUCTION AND MOTIVATION

The Durbin–Watson statistic was introduced in the pioneer works of Durbin and Watson [1–3], in order to detect the presence of a first-order autocorrelated driven noise in linear regression models. A wide literature is available on the asymptotic behavior of the Durbin–Watson statistic for linear regression models, and it is well-known that the statistical test based on the Durbin–Watson statistic performs pretty well when the regressors are independent random variables. However, as soon as the regressors are lagged dependent variables, which is of course the most attractive case, its widespread use in inappropriate situations may lead to bad conclusions. More precisely, it was observed by Malinvaud [4] and Nerlove and Wallis [5] that the Durbin–Watson statistic may be asymptotically biased if the model itself and the driven noise are governed by first-order autoregressive processes. In order to prevent this misuse, Durbin [6] proposed a redesigned alternative test in the particular case of the first-order autoregressive process previously investigated in [4, 5]. More recently, Stocker [7] provided substantial improvements in the study of the asymptotic behavior of the Durbin–Watson statistic resulting from the presence of a first-order autocorrelated noise. We also refer the reader to Bercu and Proïa [8] for a recent sharp analysis on the asymptotic behavior of the Durbin–Watson statistic via a martingale approach, see also Proïa [9] for an extension to the multivariate case.

*Correspondence to: Bernard Bercu, Institut de Mathématiques de Bordeaux, Université Bordeaux 1, UMR 5251, 351 cours de la libération, 33405 Talence Cedex, France.
†E-mail: bernard.bercu@math.u-bordeaux1.fr

As far as the authors know, there are no results available on the Durbin–Watson statistic for autoregressive models with exogenous control such as ARX$(p, q)$ processes. One can observe that these models are widely used in many areas related to applied mathematics such as financial mathematics [10], robotics [11], engineering [12], medical physics [13], and neuroscience [14]. However, for these models, we do not yet have a tool for analyzing the non-correlation of the residuals, which is a crucial step in model validation. This is the reason why we have chosen to investigate the asymptotic behavior of the Durbin–Watson statistic for ARX processes in order to decide whether or not the residuals are autocorrelated. We shall focus our attention on the ARX$(p, 1)$ process, given for all $n \geqslant 0$, by

$$X_{n+1} = \sum_{k=1}^{p} \theta_k X_{n-k+1} + U_n + \varepsilon_{n+1} \tag{1.1}$$

where the driven noise $(\varepsilon_n)$ is given by the first-order autoregressive process

$$\varepsilon_{n+1} = \rho \varepsilon_n + V_{n+1}. \tag{1.2}$$

We assume that the serial autocorrelation parameter satisfies $|\rho| < 1$, and the initial values $X_0$, $\varepsilon_0$ and $U_0$ may be arbitrarily chosen. In all the sequel, we also assume that $(V_n)$ is a martingale difference sequence adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ stands for the $\sigma$-algebra of the events occurring up to time $n$. Moreover, we suppose that, for all $n \geqslant 0$, $\mathbb{E}\left[V_{n+1}^2 | \mathcal{F}_n\right] = \sigma^2$ a.s. with $\sigma^2 > 0$. Denote by $\theta$ the unknown parameter of (1.1)

$$\theta^t = (\theta_1, \theta_2, \ldots, \theta_p).$$

Our goal is to deal simultaneously with three objectives. The first one is to propose an efficient procedure in order to estimate the unknown parameters $\theta$ and $\rho$ of the ARX$(p, 1)$ process given by (1.1) and (1.2). The second one is to regulate the dynamic of the process $(X_n)$ by forcing $X_n$ to track step by step a predictable reference trajectory $(x_n)$. This second objective can be achieved by use of an appropriate version of the adaptive tracking control proposed by Aström and Wittenmark [15]. Finally, our last objective is to establish the asymptotic properties of the Durbin–Watson statistic in order to propose a bilateral test on the serial parameter $\rho$.

The paper is organized as follows. Section 2 is devoted to the parameter estimation procedure and the suitable choice of stochastic adaptive control. In Section 3, we establish the almost sure convergence of the least squares estimators of $\theta$ and $\rho$. The asymptotic normality of our estimates are given in Section 4. In Section 5, we shall be able to prove the almost sure convergence of the Durbin–Watson statistic as well as its asymptotic normality, which will lead us to propose a bilateral statistical test for residual autocorrelation. Some numerical simulations are provided in Section 6. Finally, all technical proofs are postponed in the Appendices.

## 2. ESTIMATION AND ADAPTIVE CONTROL

Relation (1.1) can be rewritten as

$$X_{n+1} = \theta^t \varphi_n + U_n + \varepsilon_{n+1} \tag{2.1}$$

where the regression vector

$$\varphi_n^t = (X_n, \ldots, X_{n-p+1}).$$

A naive strategy to regulate the dynamic of the process $(X_n)$ is to make use of the Aström–Wittenmark [15] adaptive tracking control

$$U_n = x_{n+1} - \hat{\theta}_n^t \varphi_n$$

where $\hat{\theta}_n$ stands for the least squares estimator of $\theta$. However, it is known that this strategy leads to biased estimation of $\theta$ and $\rho$. This is due to the fact that $(\varepsilon_n)$ is not a white noise but the first-order



DOI: 10.1002/acs

autoregressive process given by (1.2). Consequently, it is necessary to adopt a more appropriate strategy, which means a more suitable choice for the adaptive control $U_n$ in (2.1).

The construction of our control law is as follows. Starting from (1.1) together with (1.2), we can remark that the process $(X_n)$ satisfies the ARX$(p+1, 2)$ equation given, for all $n \geqslant 1$, by

$$\begin{aligned} X_{n+1} = (\theta_1 + \rho) X_n + (\theta_2 - \rho\theta_1) X_{n-1} + \cdots + (\theta_p - \rho\theta_{p-1}) X_{n-p+1} \\ - \rho\theta_p X_{n-p} + U_n - \rho U_{n-1} + V_{n+1} \end{aligned} \tag{2.2}$$

which can be rewritten as

$$X_{n+1} = \vartheta^t \Phi_n + U_n + V_{n+1} \tag{2.3}$$

where the new parameter $\vartheta \in \mathbb{R}^{p+2}$ is defined as

$$\vartheta = \begin{pmatrix} \theta \\ 0 \\ 0 \end{pmatrix} - \rho \begin{pmatrix} -1 \\ \theta \\ 1 \end{pmatrix} \tag{2.4}$$

and the new regression vector

$$\Phi_n^t = (X_n, \ldots, X_{n-p}, U_{n-1}).$$

The original idea of this paper is to control the model (1.1) using the adaptive control associated with the model (2.3) in order to a posteriori estimate the parameters $\theta$ and $\rho$ via the estimator of the parameter $\vartheta$. We shall now focus our attention on the estimation of the unknown parameter $\vartheta$. We propose to make use of the least squares estimator which satisfies, for all $n \geqslant 0$,

$$\hat{\vartheta}_{n+1} = \hat{\vartheta}_n + S_n^{-1} \Phi_n \left( X_{n+1} - U_n - \hat{\vartheta}_n^t \Phi_n \right) \tag{2.5}$$

where the initial value $\hat{\vartheta}_0$ may be arbitrarily chosen and

$$S_n = \sum_{k=0}^{n} \Phi_k \Phi_k^t + \mathrm{I}_{p+2}$$

where the identity matrix $\mathrm{I}_{p+2}$ is added in order to avoid useless invertibility assumption. On the other hand, we are concern with the crucial choice of the adaptive control $U_n$. The role played by $U_n$ is to regulate the dynamic of the process $(X_n)$ by forcing $X_n$ to track step by step a bounded reference trajectory $(x_n)$. We assume that $(x_n)$ is predictable which means that for all $n \geqslant 1$, $x_n$ is $\mathcal{F}_{n-1}$-measurable. In order to control the dynamic of $(X_n)$ given by (1.1), we propose to make use of the Aström–Wittenmark adaptive tracking control associated with (2.3) and given, for all $n \geqslant 0$, by

$$U_n = x_{n+1} - \hat{\vartheta}_n^t \, \Phi_n. \tag{2.6}$$

This suitable choice of $U_n$ will allow us to control the dynamic of the process (1.1) while maintaining the optimality of the tracking and then estimate without bias the parameters $\theta$ and $\rho$. In all the sequel, we assume that the reference trajectory $(x_n)$ satisfies

$$\sum_{k=1}^{n} x_k^2 = o(n) \quad \text{a.s.} \tag{2.7}$$

## 3. ALMOST SURE CONVERGENCE

All our asymptotic analysis relies on the following keystone lemma. First of all, let $L$ be the identity matrix of order $p+1$ and denote by $H$ the positive real number

$$H = \sum_{k=1}^{p} \left( \theta_k + \rho^k \right)^2 + \frac{\rho^{2(p+1)}}{1-\rho^2}. \tag{3.1}$$



DOI: 10.1002/acs

In addition, for $1 \leqslant k \leqslant p$, let $K_k = -\left(\theta_k + \rho^k\right)$ and denote by $K$ the line vector

$$K = \left(0, K_1, K_2, \ldots, K_p\right). \tag{3.2}$$

Moreover, let $\Lambda$ be the symmetric square matrix of order $p+2$,

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix}. \tag{3.3}$$

*Lemma 3.1*
Assume that $(V_n)$ has a finite conditional moment of order $> 2$. Then, we have

$$\lim_{n\to\infty} \frac{1}{n} S_n = \sigma^2 \Lambda \quad \text{a.s.} \tag{3.4}$$

where the limiting matrix $\Lambda$ is given by (3.3). In addition, as soon as the correlation parameter $\rho \neq 0$, the matrix $\Lambda$ is invertible and

$$\Lambda^{-1} = \frac{1-\rho^2}{\rho^{2(p+1)}} \begin{pmatrix} SL + K^t K & -K^t \\ -K & 1 \end{pmatrix} \tag{3.5}$$

where $S = H - ||K||^2$ is the Schur complement of $L$ in $\Lambda$,

$$S = \frac{\rho^{2(p+1)}}{1-\rho^2}. \tag{3.6}$$

*Proof*
The proof is given in Appendix A. □

*Remark 3.1*
As $L$ is the identity matrix of order $p+1$, we clearly have

$$\det(\Lambda) = \frac{\rho^{2(p+1)}}{1-\rho^2}.$$

Consequently, as long as $\rho \neq 0$, $\det(\Lambda) \neq 0$ which of course implies that the matrix $\Lambda$ is invertible. The identity (3.5) comes from the block matrix inversion formula given, for example, by Horn and Johnson [16], page 18.

We start with the almost sure properties of the least squares estimator $\hat{\vartheta}_n$ of $\vartheta$ which are well-known as the process $(X_n)$ is controllable.

*Theorem 3.1*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, $\hat{\vartheta}_n$ converges almost surely to $\vartheta$,

$$\| \hat{\vartheta}_n - \vartheta \|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{3.7}$$

*Proof*
The proof is given in Appendix A. □

We shall now explicit the estimators of $\theta$ and $\rho$ and their convergence results. It follows from (2.4) that

$$\begin{pmatrix} \theta \\ \rho \end{pmatrix} = \Delta\vartheta \tag{3.8}$$



DOI: 10.1002/acs

where $\Delta$ is the rectangular matrix of size $(p+1)\times(p+2)$ given by

$$\Delta = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ \rho & 1 & 0 & \cdots & \cdots & 0 & \rho \\ \rho^2 & \rho & 1 & 0 & \cdots & 0 & \rho^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 & 0 & \rho^{p-1} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}. \tag{3.9}$$

Consequently, a natural choice to estimate the initial parameters $\theta$ and $\rho$ is to make use of

$$\begin{pmatrix} \hat{\theta}_n \\ \hat{\rho}_n \end{pmatrix} = \hat{\Delta}_n \hat{\vartheta}_n \tag{3.10}$$

where $\hat{\rho}_n$ is simply the opposite of the last coordinate of $\hat{\vartheta}_n$ and

$$\hat{\Delta}_n = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ \hat{\rho}_n & 1 & 0 & \cdots & \cdots & 0 & \hat{\rho}_n \\ \hat{\rho}_n^2 & \hat{\rho}_n & 1 & 0 & \cdots & 0 & \hat{\rho}_n^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_n^{p-1} & \hat{\rho}_n^{p-2} & \cdots & \hat{\rho}_n & 1 & 0 & \hat{\rho}_n^{p-1} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}. \tag{3.11}$$

*Corollary 3.1*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, $\hat{\theta}_n$ and $\hat{\rho}_n$ both converge almost surely to $\theta$ and $\rho$,

$$\| \hat{\theta}_n - \theta \|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{3.12}$$

$$(\hat{\rho}_n - \rho)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{3.13}$$

*Proof*
One can immediately see from (3.8) that the last component of the vector $\vartheta$ is $-\rho$. The same is true for the estimator $\hat{\rho}_n$ of $\rho$. Consequently, we deduce from (3.7) that $\hat{\rho}_n$ converges a.s. to $\rho$ with the almost sure rate of convergence given by (3.13). Therefore, we obtain from (3.9) and (3.11) that

$$\| \hat{\Delta}_n - \Delta \|^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.}$$

which ensures via (3.7) and (3.10) that $\hat{\theta}_n$ converges a.s. to $\theta$ with the almost sure rate of convergence given by (3.12). □

## 4. ASYMPTOTIC NORMALITY

This section is devoted to the asymptotic normality of the couple $(\hat{\theta}_n, \hat{\rho}_n)$, which is obtained from the one of the estimator $\hat{\vartheta}_n$ of $\vartheta$.

*Theorem 4.1*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, we have

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Lambda^{-1}\right) \tag{4.1}$$

where the matrix $\Lambda^{-1}$ is given by (3.5).

In order to provide the joint asymptotic normality of the estimators of $\theta$ and $\rho$, denote, for all $1 \leqslant k \leqslant p-1$,

$$\xi_k = \sum_{i=1}^{k} \rho^{k-i} \theta_i$$

and let $\nabla$ be the rectangular matrix of size $(p+1)\times(p+2)$ given by

$$\nabla = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ \rho & 1 & 0 & \cdots & \cdots & 0 & \rho - \xi_1 \\ \rho^2 & \rho & 1 & 0 & \cdots & 0 & \rho^2 - \xi_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 & 0 & \rho^{p-1} - \xi_{p-1} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}. \tag{4.2}$$

*Corollary 4.1*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, we have

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta \\ \hat{\rho}_n - \rho \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma) \tag{4.3}$$

where $\Sigma = \nabla \Lambda^{-1} \nabla^t$. In particular,

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1-\rho^2}{\rho^{2(p+1)}}\right). \tag{4.4}$$

*Proof*
The proof is given in Appendix B. □

## 5. ON THE DURBIN–WATSON STATISTIC

We now investigate the asymptotic behavior of the Durbin–Watson statistic [1–3] given, for all $n \geqslant 1$, by

$$\hat{D}_n = \frac{\sum_{k=1}^{n} (\hat{\varepsilon}_k - \hat{\varepsilon}_{k-1})^2}{\sum_{k=0}^{n} \hat{\varepsilon}_k^2} \tag{5.1}$$

where the residuals $\hat{\varepsilon}_k$ are defined, for all $1 \leqslant k \leqslant n$, by

$$\hat{\varepsilon}_k = X_k - U_{k-1} - \hat{\theta}_n^t \varphi_{k-1} \tag{5.2}$$

with $\hat{\theta}_n$ given by (3.10). The initial value $\hat{\varepsilon}_0$ may be arbitrarily chosen, and we take $\hat{\varepsilon}_0 = X_0$. One can observe that it is also possible to estimate the serial correlation parameter $\rho$ by the least squares estimator

$$\overline{\rho}_n = \frac{\sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_{k-1}}{\sum_{k=1}^{n} \hat{\varepsilon}_{k-1}^2} \tag{5.3}$$

which is the natural estimator of $\rho$ in the autoregressive framework without control. The Durbin–Watson statistic $\hat{D}_n$ is related to $\overline{\rho}_n$ by the linear relation

$$\hat{D}_n = 2(1 - \overline{\rho}_n) + \zeta_n \tag{5.4}$$

where the remainder term $\zeta_n$ plays a negligible role. The almost sure properties of $\hat{D}_n$ and $\bar{\rho}_n$ are as follows.



DOI: 10.1002/acs

*Theorem 5.1*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, $\overline{\rho}_n$ converges almost surely to $\rho$,

$$(\overline{\rho}_n - \rho)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{5.5}$$

In addition, $\hat{D}_n$ converges almost surely to $D = 2(1-\rho)$. Moreover, if $(V_n)$ has a finite conditional moment of order $> 4$, we also have

$$\left(\hat{D}_n - D\right)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{5.6}$$

Our next result deals with the asymptotic normality of the Durbin–Watson statistic. For that purpose, it is necessary to introduce some notations. Denote

$$\alpha = \begin{pmatrix} 1 \\ -\theta_1 \\ \vdots \\ -\theta_p \\ -1 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 1 \\ \rho \\ \vdots \\ \rho^{p-1} \\ 0 \end{pmatrix}. \tag{5.7}$$

In addition, let

$$\gamma = \Lambda\alpha + (1-\rho^2)\nabla^t \beta. \tag{5.8}$$

*Theorem 5.2*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 2$. Then, we have

$$\sqrt{n}(\overline{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2\right) \tag{5.9}$$

where the asymptotic variance $\tau^2 = (1-\rho^2)^2 \gamma^t \Lambda^{-1} \gamma$. Moreover, if $(V_n)$ has a finite conditional moment of order $> 4$, we also have

$$\sqrt{n}(\hat{D}_n - D) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\tau^2\right) \tag{5.10}$$

*Proof*
The proofs are given in Appendix C. □

*Remark 5.1*
It follows from (3.5) together with tedious but straightforward calculations that for all $p \geqslant 1$,

$$\tau^2 = \frac{(1-\rho^2)}{\rho^{2(p+1)}} \left[ \rho^{2(p+1)} \left(4 - (4p+3)\rho^{2p} + 4p\rho^{2(p+1)} - \rho^{2(2p+1)}\right) + \left(1 - (p+1)\rho^{2p} + (p-1)\rho^{2(p+1)}\right)^2 \right]. \tag{5.11}$$

For example, in the particular case $p = 1$, we obtain that

$$\tau^2 = \frac{(1-\rho^2)}{\rho^4} \left(1 - 4\rho^2 + 8\rho^4 - 7\rho^6 + 4\rho^8 - \rho^{10}\right). \tag{5.12}$$

Moreover, it is not hard to see by a convexity argument that we always have for all $p \geqslant 1$,

$$\tau^2 \leqslant \frac{1-\rho^2}{\rho^{2(p+1)}}.$$



DOI: 10.1002/acs

In other words, the least squares estimator $\overline{\rho}_n$ performs better than $\hat{\rho}_n$ for the estimation of $\rho$. It means that a statistical test procedure built on the Durbin–Watson statistic should be really powerful.

We are now in the position to propose our new bilateral statistical test built on the Durbin–Watson statistic $\hat{D}_n$. First of all, we shall not investigate the case $\rho = 0$ because our approach is only of interest for ARX processes where the driven noise is given by a first-order autoregressive process. For a given value $\rho_0$ such that $|\rho_0| < 1$ and $\rho_0 \neq 0$, we wish to test whether or not the serial correlation parameter is equal to $\rho_0$. It means that we wish to test

$$\mathcal{H}_0 \,:\, \text{“}\rho = \rho_0\text{”} \quad \text{against} \quad \mathcal{H}_1 \,:\, \text{“}\rho \neq \rho_0\text{”}.$$

According to Theorem 5.1, we have under the null hypothesis $\mathcal{H}_0$

$$\lim_{n\to\infty} \hat{D}_n = D_0 \quad \text{a.s.}$$

where $D_0 = 2(1-\rho_0)$. In addition, we clearly have from (5.10) that under $\mathcal{H}_0$

$$\frac{n}{4\tau^2}\left(\hat{D}_n - D_0\right)^2 \xrightarrow{\mathcal{L}} \chi^2 \tag{5.13}$$

where $\chi^2$ stands for a Chi-square distribution with one degree of freedom. Via (5.11), an efficient strategy to estimate the asymptotic variance $\tau^2$ is to make use of

$$\hat{\tau}_n^2 = \frac{(1-\overline{\rho}_n^2)}{\overline{\rho}_n^{2(p+1)}}\left[\overline{\rho}_n^{2(p+1)}\left(4-(4p+3)\overline{\rho}_n^{2p} + 4p\overline{\rho}_n^{2(p+1)} - \overline{\rho}_n^{2(2p+1)}\right)\right.$$
$$\left. + \left(1-(p+1)\overline{\rho}_n^{2p} + (p-1)\overline{\rho}_n^{2(p+1)}\right)^2\right]. \tag{5.14}$$

Therefore, our new bilateral statistical test relies on the following result.

*Theorem 5.3*
Assume that the serial correlation parameter $\rho \neq 0$ and that $(V_n)$ has a finite conditional moment of order $> 4$. Then, under the null hypothesis $\mathcal{H}_0 \,:\, \rho = \rho_0$,

$$\frac{n}{4\hat{\tau}_n^2}\left(\hat{D}_n - D_0\right)^2 \xrightarrow{\mathcal{L}} \chi^2 \tag{5.15}$$

where $\chi^2$ stands for a Chi-square distribution with one degree of freedom. In addition, under the alternative hypothesis $\mathcal{H}_1 \,:\, \rho \neq \rho_0$,

$$\lim_{n\to\infty} \frac{n}{4\hat{\tau}_n^2}\left(\hat{D}_n - D_0\right)^2 = +\infty \quad \text{a.s.} \tag{5.16}$$

*Proof*
The proof is given in Appendix C. □

From a practical point of view, for a significance level $\alpha$ where $0 < \alpha < 1$, the acceptance and rejection regions are given by $\mathcal{A} = [0, a_\alpha]$ and $\mathcal{R} = ]a_\alpha, +\infty[$ where $a_\alpha$ stands for the $(1-\alpha)$-quantile of the Chi-square distribution with one degree of freedom. The null hypothesis $\mathcal{H}_0$ will be accepted if

$$\frac{n}{4\hat{\tau}_n^2}\left(\hat{D}_n - D_0\right)^2 \leqslant a_\alpha,$$

and will be rejected otherwise.



DOI: 10.1002/acs

## 6. NUMERICAL EXPERIMENTS

The purpose of this section is to provide some numerical experiments in order to illustrate our main theoretical results. In order to keep this section brief, we shall only consider the ARX$(p, 1)$ process $(X_n)$ given by (1.1) in the particular cases $p = 1$ and $p = 2$, where the driven noise $(\varepsilon_n)$ satisfies (1.2). Moreover, for the sake of simplicity, the reference trajectory $(x_n)$ is chosen to be identically zero, and $(V_n)$ is a Gaussian white noise with $\mathcal{N}(0, 1)$ distribution. Finally, our numerical simulations are based on 500 realizations of sample size $N = 1000$. First of all, consider the ARX$(1, 1)$ process given, for all $n \geqslant 1$, by

$$X_{n+1} = \theta X_n + U_n + \varepsilon_{n+1} \quad \text{and} \quad \varepsilon_{n+1} = \rho \varepsilon_n + V_{n+1} \tag{6.1}$$

where we have chosen $\theta = 8/5$ and $\rho = -4/5$ which implies that $D = 18/5$ and the Schur complement $S = 16^2/15^2$. This choice has been made in order to obtain simple expressions for the matrices $\Lambda$ and $\Sigma$. One can easily see from (3.2) to (3.5) that

$$\Lambda = \frac{1}{45}\begin{pmatrix} 45 & 0 & 0 \\ 0 & 45 & -36 \\ 0 & -36 & 80 \end{pmatrix}$$

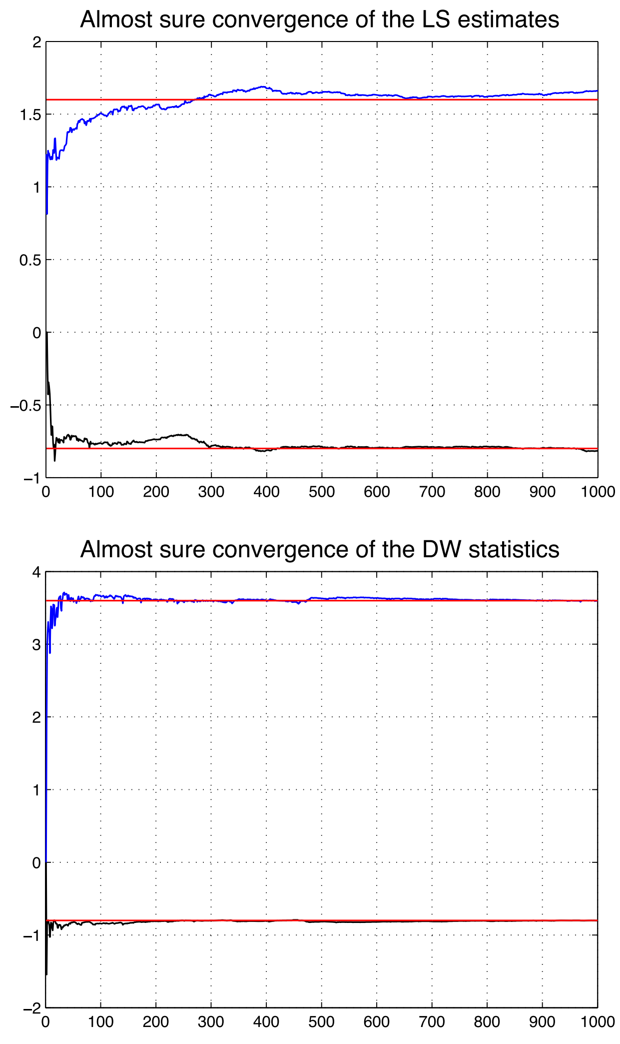


Figure 1. Almost sure convergence in the particular case $p = 1$.



DOI: 10.1002/acs

as well as

$$\Sigma = \nabla \Lambda^{-1} \nabla^t = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \left(\frac{15}{16}\right)^2 \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Figure 1 illustrates the almost sure convergence of $\hat{\theta}_n$, $\hat{\rho}_n$, $\overline{\rho}_n$, and $\widehat{D}_n$. One can see that the almost sure convergence is very satisfactory.

We shall now focus our attention to the asymptotic normality. We compare the empirical distributions of the least squares estimates

$$\frac{\sqrt{nS}}{\sqrt{1+S}}\left(\hat{\theta}_n - \theta\right) \quad \text{and} \quad \sqrt{nS}\,(\hat{\rho}_n - \rho)$$

with the standard $\mathcal{N}(0,1)$ distribution. We proceed in the same way for the Durbin–Watson statistics

$$\frac{\sqrt{n}}{\tau}\,(\overline{\rho}_n - \rho) \quad \text{and} \quad \frac{\sqrt{n}}{2\tau}\left(\widehat{D}_n - D\right)$$

where $\tau^2$ is given by (5.12). We use the natural estimates of $S$ and $\tau^2$ by replacing $\rho$ by $\hat{\rho}_n$ and $\overline{\rho}_n$, respectively. One can see in Figure 2 that the approximation by a standard $\mathcal{N}(0,1)$ distribution performs pretty well. These results are very promising in order to built a statistical test based on these statistics.

Next, we are interested in the ARX(2, 1) process given, for all $n \geqslant 1$, by

$$X_{n+1} = \theta_1 X_n + \theta_2 X_{n-1} + U_n + \varepsilon_{n+1} \quad \text{and} \quad \varepsilon_{n+1} = \rho \varepsilon_n + V_{n+1} \tag{6.2}$$

where we have chosen $\theta_1 = 1$, $\theta_2 = 4/5$ and $\rho = -9/10$ which leads to $D = 19/5$ and $S = 9^6/(19 \times 10^6)$. It follows from (3.2) to (3.5) that

$$\Lambda = \frac{1}{9500}\begin{pmatrix} 9500 & 0 & 0 & 0 \\ 0 & 9500 & 0 & -950 \\ 0 & 0 & 9500 & -15295 \\ 0 & -950 & -1529 & 51292 \end{pmatrix}.$$

In addition, the diagonal entries of the covariance matrix $\Sigma = \nabla \Lambda^{-1} \nabla^t$ are respectively given by

$$1 + \frac{1}{S} = \frac{721441}{531441}, \quad 1 + \rho^2 + \frac{4\rho^2}{S} = \frac{1947541}{656100}, \quad \frac{1}{S} = \frac{190000}{531441}.$$

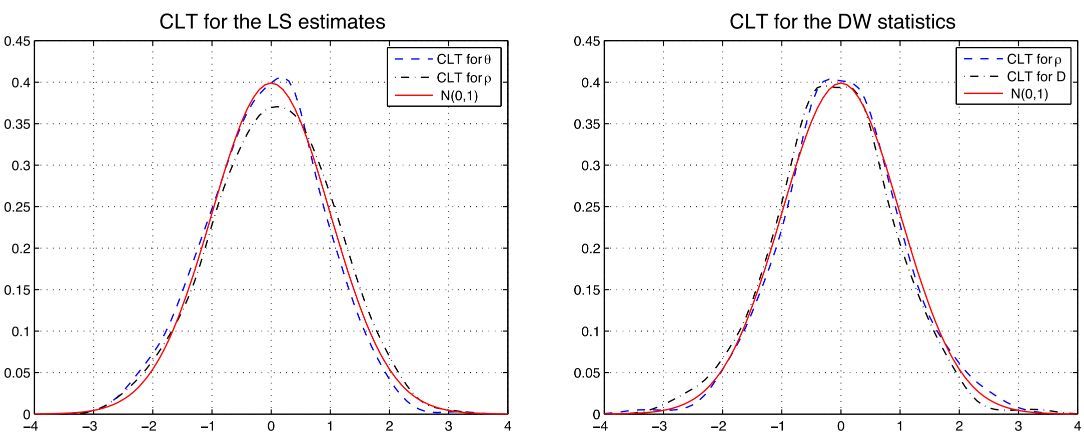


Figure 2. Asymptotic normality in the particular case $p = 1$.



DOI: 10.1002/acs

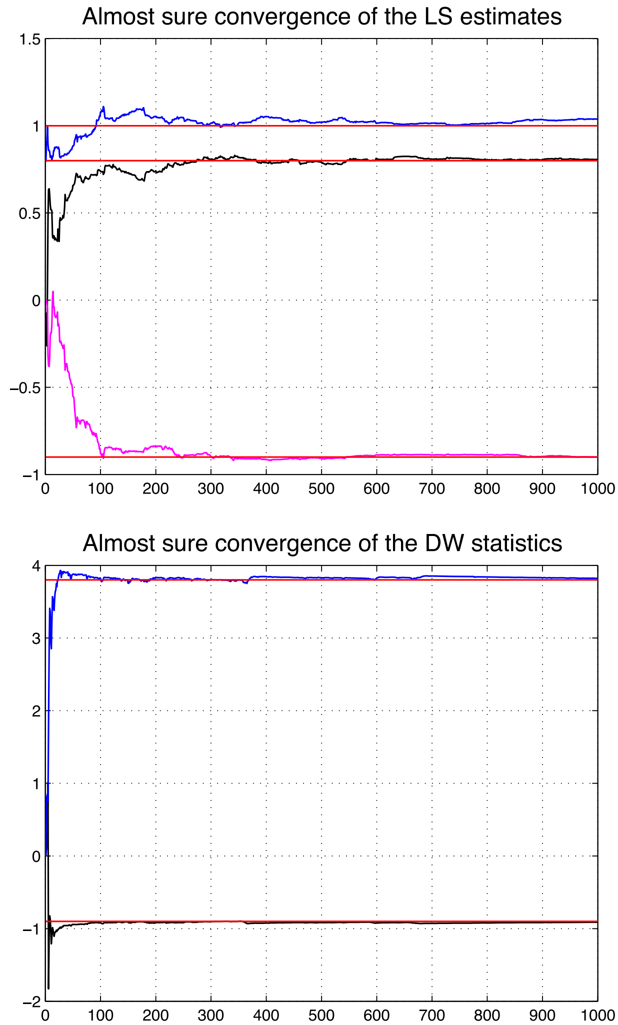


Figure 3. Almost sure convergence in the particular case $p = 2$.

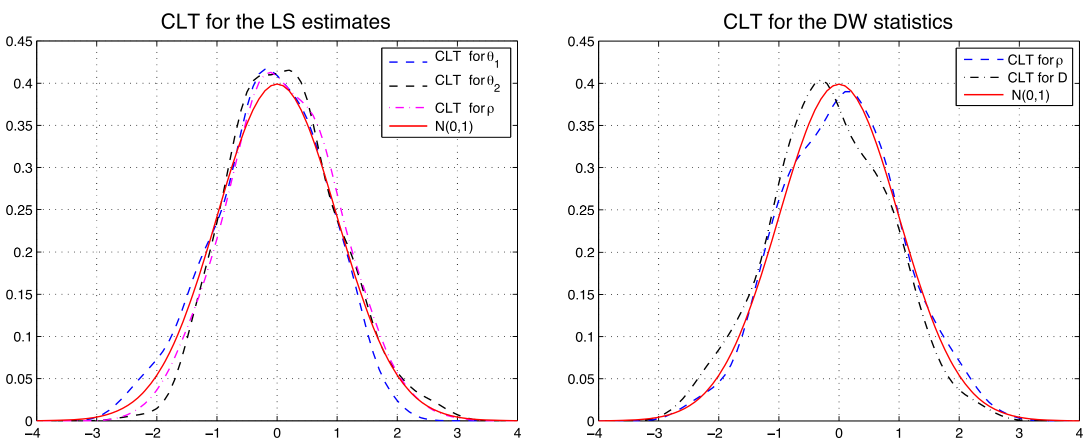


Figure 4. Asymptotic normality in the particular case $p = 2$.

Table I. Durbin–Watson test in the particular case $p = 1$ and $\rho = -0.8$.

| Durbin–Watson | Values of $\rho_0$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | −0.9 | −0.8 | −0.7 | −0.6 | −0.4 | −0.2 | 0.2 | 0.4 | 0.6 | 0.7 | 0.8 | 0.9 |
| $N = 50$ | 0.20 | 0.02 | 0.12 | 0.38 | 0.79 | 0.95 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 |
| | (0.80) | (0.98) | (0.88) | (0.62) | (0.21) | (0.05) | (0.01) | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) |
| $N = 100$ | 0.51 | 0.03 | 0.25 | 0.66 | 0.97 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.49) | (0.97) | (0.75) | (0.34) | (0.03) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $N = 1000$ | 1.00 | 0.05 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.00) | (0.95) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |

Table II. Durbin–Watson test in the particular case $p = 2$ and $\rho = -0.9$.

| Durbin–Watson | Values of $\rho_0$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | −0.9 | −0.8 | −0.7 | −0.6 | −0.4 | −0.2 | 0.2 | 0.4 | 0.6 | 0.7 | 0.8 | 0.9 |
| $N = 50$ | 0.06 | 0.17 | 0.52 | 0.76 | 0.92 | 0.96 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.94) | (0.83) | (0.48) | (0.24) | (0.08) | (0.04) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) |
| $N = 100$ | 0.05 | 0.38 | 0.82 | 0.95 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.95) | (0.62) | (0.18) | (0.05) | (0.01) | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| $N = 1000$ | 0.05 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (0.95) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |

Figure 3 shows the almost sure convergence of $\hat{\theta}_{n,1}$, $\hat{\theta}_{n,2}$, $\hat{\rho}_n$, $\overline{\rho}_n$, and $\hat{D}_n$ while Figure 4 illustrates their asymptotic normality. As in the case $p = 1$, one can observe that the approximation by a standard $\mathcal{N}(0, 1)$ distribution works pretty well.

We shall achieve this section by illustrating the behavior of the Durbin–Watson statistical test. We wish to test $\mathcal{H}_0 : \rho = \rho_0$ against $\mathcal{H}_1 : \rho \neq \rho_0$ at 5% level of significance for the ARX processes given by (6.1) and (6.2). More precisely, we compute the frequency for which $\mathcal{H}_0$ is rejected for different values of $\rho_0$,

$$\mathbb{P}\left(\text{rejecting } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true}\right)$$

via 500 realizations of different sample sizes $N = 50, 100$, and 1000. In Tables I and II, one can appreciate the empirical power of the statistical test which means that the Durbin–Watson statistic performs very well.

## APPENDIX A: PROOFS OF THE ALMOST SURE CONVERGENCE RESULTS

Denote by $A$ and $B$ the polynomials given, for all $z \in \mathbb{C}$, by

$$A(z) = 1 - \sum_{k=1}^{p+1} a_k z^k \quad \text{and} \quad B(z) = 1 - \rho z \tag{A.1}$$

where $a_1 = \theta_1 + \rho$, $a_{p+1} = -\rho\theta_p$ and, for $2 \leqslant k \leqslant p$, $a_k = \theta_k - \rho\theta_{k-1}$. The ARX$(p + 1, 2)$ equation given by (2.2) may be rewritten as

$$A(R)X_n = B(R)U_{n-1} + V_n \tag{A.2}$$

where $R$ stands for the shift-back operator $RX_n = X_{n-1}$. On the one hand, $B(z) = 0$ if and only if $z = 1/\rho$ with $\rho \neq 0$. Consequently, as $|\rho| < 1$, $B$ is clearly causal and for all $z \in \mathbb{C}$ such that $|\rho z| < 1$,

$$B^{-1}(z) = \frac{1}{1 - \rho z} = \sum_{k=0}^{\infty} \rho^k z^k.$$



DOI: 10.1002/acs

On the other hand, let $P$ be the polynomial given, for all $z \in \mathbb{C}$, by

$$P(z) = B^{-1}(z)(A(z) - 1) = \sum_{k=1}^{\infty} p_k z^k. \tag{A.3}$$

It is not hard to see from (A.3) that, for $1 \leqslant k \leqslant p$, $p_k = -(\theta_k + \rho^k)$ while, for all $k \geqslant p+1$, $p_k = -\rho^k$. Consequently, as soon as $\rho \neq 0$, we deduce from [17] that the process $(X_n)$ given by (A.2) is strongly controllable. One can observe that in our situation, the usual notion of controllability is the same as the concept of strong controllability. To be more precise, the assumption that $\rho \neq 0$ implies that the polynomials $A - 1$ and $B$, given by (A.1), are coprime. It is exactly the so-called controllability condition. We refer the reader to [17] for more details on the links between the notions of controllability and strong controllability. Finally, we clearly obtain Lemma 3.1 and Theorem 3.1 from (2.3) together with Theorem 5 of [17].

## APPENDIX B: PROOFS OF THE ASYMPTOTIC NORMALITY RESULTS

Theorem 4.1 immediately follows from Theorem 8 of [17]. We shall now proceed to the proof of Corollary 4.1. First of all, denote for $0 \leqslant k \leqslant p-1$,

$$s_k(\vartheta) = \sum_{i=1}^{k+1} \rho^{k-i+1} \vartheta_i + \rho^k \vartheta_{p+2}$$

where $\rho = -\vartheta_{p+2}$ and $s_p(\vartheta) = \rho$. In addition, let

$$g(\vartheta) = \Delta\vartheta = \begin{pmatrix} s_0(\vartheta) \\ s_1(\vartheta) \\ \vdots \\ s_p(\vartheta) \end{pmatrix}. \tag{B.1}$$

One can easily check that the gradient of the function $g$ is given by

$$\nabla g(\vartheta) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & \xi_0(\theta) \\ \rho & 1 & 0 & \cdots & \cdots & 0 & \rho - \xi_1(\theta) \\ \rho^2 & \rho & 1 & 0 & \cdots & 0 & \rho^2 - \xi_2(\theta) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 & 0 & \rho^{p-1} - \xi_{p-1}(\theta) \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & \xi_p(\theta) \end{pmatrix} \tag{B.2}$$

where $\xi_0(\theta) = 1$, $\xi_p(\theta) = -1$ and, for all $1 \leqslant k \leqslant p-1$,

$$\xi_k(\theta) = \sum_{i=1}^{k} \rho^{k-i} \theta_i.$$

The gradient of $g$ coincides with the matrix $\nabla$ given by (4.2). On the one hand, it follows from (3.8) and (B.1) that

$$g(\vartheta) = \begin{pmatrix} \theta \\ \rho \end{pmatrix}. \tag{B.3}$$

On the other hand, we already saw from (4.1) that

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda^{-1}). \tag{B.4}$$

Consequently, we deduce from (B.3) and (B.4) together with the well-known delta method that

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta \\ \hat{\rho}_n - \rho \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

where $\Sigma = \nabla \Lambda^{-1} \nabla^t$, which completes the proof of Corollary 4.1.



DOI: 10.1002/acs

## APPENDIX C: PROOFS OF THE DURBIN–WATSON STATISTIC RESULTS

*Proof of Theorem 5.1*

We are now in position to investigate the asymptotic behavior of the Durbin–Watson statistic. First of all, we start with the proof of Theorem 5.1. Recall from (2.1) together with (5.2) that the residuals are given, for all $1 \leqslant k \leqslant n$, by

$$\hat{\varepsilon}_k = X_k - U_{k-1} - \hat{\theta}_n^t \varphi_{k-1} = \varepsilon_k - \tilde{\theta}_n^t \varphi_{k-1} \tag{C.1}$$

where $\tilde{\theta}_n = \hat{\theta}_n - \theta$. For all $n \geqslant 1$, denote

$$I_n = \sum_{k=1}^{n} \hat{\varepsilon}_k \hat{\varepsilon}_{k-1} \quad \text{and} \quad J_n = \sum_{k=0}^{n} \hat{\varepsilon}_k^2.$$

It is not hard to see that

$$I_n = \hat{\varepsilon}_0 \hat{\varepsilon}_1 + P_n^I - \tilde{\theta}_n^t Q_n^I + \tilde{\theta}_n^t S_{n-1}^I \tilde{\theta}_n, \tag{C.2}$$

$$J_n = \hat{\varepsilon}_0^2 + P_n^J - 2\tilde{\theta}_n^t Q_n^J + \tilde{\theta}_n^t S_{n-1}^J \tilde{\theta}_n \tag{C.3}$$

where

$$P_n^I = \sum_{k=2}^{n} \varepsilon_k \varepsilon_{k-1}, \quad Q_n^I = \sum_{k=2}^{n} (\varphi_{k-2}\varepsilon_k + \varphi_{k-1}\varepsilon_{k-1}), \quad S_n^I = \sum_{k=1}^{n} \varphi_k \varphi_{k-1}^t,$$

and

$$P_n^J = \sum_{k=1}^{n} \varepsilon_k^2, \quad Q_n^J = \sum_{k=1}^{n} \varphi_{k-1}\varepsilon_k, \quad S_n^J = \sum_{k=0}^{n} \varphi_k \varphi_k^t.$$

We deduce from (1.2) that

$$(1-\rho^2) P_n^J = \rho^2(\varepsilon_0^2 - \varepsilon_n^2) + 2\rho N_n + L_n \tag{C.4}$$

where

$$N_n = \sum_{k=1}^{n} \varepsilon_{k-1} V_k \quad \text{and} \quad L_n = \sum_{k=1}^{n} V_k^2.$$

Moreover, we assume that $(V_n)$ has a finite conditional moment of order $a > 2$. Then, it follows from Proposition 1.3.23 page 25 of [18] that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} V_k^2 = \sigma^2 \quad \text{a.s.} \tag{C.5}$$

In addition, we also have from Corollary 1.3.21 page 23 of [18] that for all $2 \leqslant b < a$,

$$\sum_{k=1}^{n} |V_k|^b = \mathcal{O}(n) \quad \text{a.s.} \tag{C.6}$$

and

$$\sup_{1\leqslant k\leqslant n} |V_k| = o(n^{1/b}) \quad \text{a.s.} \tag{C.7}$$

However, we clearly obtain from (1.2) that

$$\sup_{1\leqslant k\leqslant n} |\varepsilon_k| \leqslant \frac{1}{1-|\rho|} \left( |\varepsilon_0| + \sup_{1\leqslant k\leqslant n} |V_k| \right) \tag{C.8}$$



DOI: 10.1002/acs

and

$$\sum_{k=1}^{n} |\varepsilon_k|^b \leqslant (1-|\rho|)^{-b} \left( |\varepsilon_0|^b + \sum_{k=1}^{n} |V_k|^b \right) \tag{C.9}$$

which of course implies that

$$\sup_{1 \leqslant k \leqslant n} |\varepsilon_k| = o(n^{1/b}) \quad \text{a.s.} \tag{C.10}$$

and

$$\sum_{k=1}^{n} |\varepsilon_k|^b = \mathcal{O}(n) \quad \text{a.s.} \tag{C.11}$$

In the particular case $b = 2$, we find that

$$\sup_{1 \leqslant k \leqslant n} \varepsilon_k^2 = o(n) \quad \text{and} \quad \sum_{k=1}^{n} \varepsilon_k^2 = \mathcal{O}(n) \quad \text{a.s.} \tag{C.12}$$

Hereafter, $(N_n)$ is a locally square-integrable real martingale with predictable quadratic variation given, for all $n \geqslant 1$, by

$$\langle N \rangle_n = \sigma^2 \sum_{k=0}^{n-1} \varepsilon_k^2.$$

Therefore, we deduce from (C.12) and the strong law of large numbers for martingales given, for example, by Theorem 1.3.15 page 20 of [18] that

$$\lim_{n \to \infty} \frac{N_n}{n} = 0 \quad \text{a.s.} \tag{C.13}$$

Hence, we obtain from (C.4) together with (C.5), (C.12), and (C.13) that

$$\lim_{n \to \infty} \frac{P_n^J}{n} = \frac{\sigma^2}{1-\rho^2} \quad \text{a.s.} \tag{C.14}$$

Furthermore, convergence (3.4) immediately implies that

$$\lim_{n \to \infty} \frac{1}{n} S_n^J = \sigma^2 \mathrm{I}_p \quad \text{a.s.} \tag{C.15}$$

We also obtain from the Cauchy–Schwarz inequality, (C.12), and (C.15), that

$$\| Q_n^J \| = \mathcal{O}(n) \quad \text{a.s.}$$

Consequently, we find from the conjunction of (3.12), (C.3), (C.13), and (C.15) that

$$\lim_{n \to \infty} \frac{J_n}{n} = \frac{\sigma^2}{1-\rho^2} \quad \text{a.s.} \tag{C.16}$$

By the same token, as

$$P_n^I = \rho P_{n-1}^J + N_n + \rho \varepsilon_0^2 - \varepsilon_0 \varepsilon_1, \tag{C.17}$$

it follows from (C.13) and (C.14) that

$$\lim_{n \to \infty} \frac{P_n^I}{n} = \frac{\sigma^2 \rho}{1-\rho^2} \quad \text{a.s.} \tag{C.18}$$



DOI: 10.1002/acs

which leads via (C.2) to

$$\lim_{n\to\infty} \frac{I_n}{n} = \frac{\sigma^2 \rho}{1-\rho^2} \quad \text{a.s.} \tag{C.19}$$

Therefore, we obtain from definition (5.3) together with (C.16) and (C.19) that

$$\lim_{n\to\infty} \overline{\rho}_n = \lim_{n\to\infty} \frac{I_n}{J_{n-1}} = \rho \quad \text{a.s.} \tag{C.20}$$

In order to establish the almost sure rate of convergence given by (5.5), it is necessary to make some sharp calculations. We infer from (C.2), (C.3), and (C.17) that

$$I_n - \rho J_{n-1} = N_n - Q_n + R_n \tag{C.21}$$

where $Q_n = \left(Q_n^I - 2\rho Q_{n-1}^J\right)^t \tilde{\theta}_n$ and

$$R_n = \hat{\varepsilon}_0 \hat{\varepsilon}_1 - \varepsilon_0 \varepsilon_1 + \rho \varepsilon_0^2 - \rho \hat{\varepsilon}_0^2 + \tilde{\theta}_n^t \left(S_{n-1}^I - \rho S_{n-2}^J\right) \tilde{\theta}_n.$$

On the one hand, it follows from convergence (3.4) together with (3.12) and the Cauchy–Schwarz inequality, that

$$|Q_n| = \mathcal{O}\left(\sqrt{n \log n}\right) \quad \text{and} \quad |R_n| = \mathcal{O}(\log n) \quad \text{a.s.}$$

On the other hand, as $\langle N \rangle_n = \mathcal{O}(n)$ a.s., we deduce from Theorem 1.3.24 page 26 of [18] related to the almost sure rate of convergence in the strong law of large numbers for martingales that $|N_n| = \mathcal{O}\left(\sqrt{n \log n}\right)$ a.s. Therefore, we can conclude from (C.16) and (C.21) that

$$(\overline{\rho}_n - \rho)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{C.22}$$

The proof of the almost sure convergence of $\hat{D}_n$ to $D = 2(1-\rho)$ immediately follows from (C.20). As a matter of fact, it follows from (5.1) that

$$\left(J_{n-1} + \hat{\varepsilon}_n^2\right) \hat{D}_n = 2\left(J_{n-1} - I_n\right) + \hat{\varepsilon}_n^2 - \hat{\varepsilon}_0^2. \tag{C.23}$$

Dividing both sides of (C.23) by $J_{n-1}$, we obtain that

$$\hat{D}_n = 2(1 - f_n)\,(1 - \overline{\rho}_n) + g_n \tag{C.24}$$

where

$$f_n = \frac{\hat{\varepsilon}_n^2}{J_n} \quad \text{and} \quad g_n = \frac{\hat{\varepsilon}_n^2 - \hat{\varepsilon}_0^2}{J_n}.$$

However, convergence (C.16) ensures that $f_n$ and $g_n$ both tend to zero a.s. Consequently, (C.20) immediately implies that

$$\lim_{n\to\infty} \hat{D}_n = 2(1-\rho) \quad \text{a.s.} \tag{C.25}$$

The almost sure rate of convergence given by (5.6) requires some additional assumption on $(V_n)$. Hereafter, assume that the noise $(V_n)$ has a finite conditional moment of order $> 4$. We clearly obtain from (3.4), (3.12) together with (C.1) and (C.10) with $b = 4$ that

$$\sup_{1\leqslant k\leqslant n} \hat{\varepsilon}_k^2 = o(\sqrt{n}) + o(\log n) = o(\sqrt{n}) \quad \text{a.s.} \tag{C.26}$$

which leads by (C.16) to

$$f_n = o\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad g_n = o\left(\frac{1}{\sqrt{n}}\right) \quad \text{a.s.} \tag{C.27}$$



DOI: 10.1002/acs

In addition, it follows from (C.24) that

$$\hat{D}_n - D = -2(1 - f_n)(\overline{\rho}_n - \rho) + 2(\rho - 1) f_n + g_n \tag{C.28}$$

where $D = 2(1-\rho)$. Consequently, we obtain by (C.22) and (C.27) that

$$\left(\hat{D}_n - D\right)^2 = \mathcal{O}\left((\overline{\rho}_n - \rho)^2\right) + \mathcal{O}\left(f_n^2\right) = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{C.29}$$

which achieves the proof of Theorem 5.1.

*Proof of Theorem 5.2*

The proof of Theorem 5.2 is much more difficult to handle. We already saw from (C.21) that

$$J_{n-1}(\overline{\rho}_n - \rho) = N_n - Q_n + R_n \tag{C.30}$$

where the remainder $R_n$ plays a negligible role. This is of course not the case for $Q_n = \left(Q_n^I - 2\rho Q_{n-1}^J\right)^t \tilde{\theta}_n$. We know from (3.8) and (3.10) that

$$\begin{pmatrix} \hat{\theta}_n - \theta \\ \hat{\rho}_n - \rho \end{pmatrix} = \hat{\Delta}_n \hat{\vartheta}_n - \Delta\vartheta = \hat{\Delta}_n \left(\hat{\vartheta}_n - \vartheta\right) + \left(\hat{\Delta}_n - \Delta\right)\vartheta. \tag{C.31}$$

One can observe that in the particular case $p = 1$, the right-hand side of (C.31) reduces to the vector

$$\Delta\left(\hat{\vartheta}_n - \vartheta\right)$$

because

$$\hat{\Delta}_n = \Delta = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

For all $1 \leqslant k \leqslant p-1$, denote

$$s_n(k) = \sum_{i=0}^{k} \hat{\rho}_n^i \rho^{k-i}.$$

It is easily check that $\hat{\Delta}_n - \Delta$ can be rewritten as $\hat{\Delta}_n - \Delta = (\hat{\rho}_n - \rho)A_n$ where $A_n$ is the rectangular matrix of size $(p+1)\times(p+2)$ given by

$$A_n = \begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & \cdots & 0 & 0 & 1 \\ s_n(1) & 1 & 0 & 0 & \cdots & 0 & 0 & s_n(1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ s_n(p-2) & s_n(p-3) & \cdots & s_n(1) & 1 & 0 & 0 & s_n(p-2) \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

It was already proven that $\hat{\rho}_n$ converges almost surely to $\rho$ which implies that for all $1 \leqslant k \leqslant p-1$,

$$\lim_{n\to\infty} s_n(k) = (k+1)\rho^k \quad \text{a.s.}$$

It immediately leads to the almost sure convergence of $A_n$ to the matrix $A$ given by

$$A = \begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & \cdots & 0 & 0 & 1 \\ 2\rho & 1 & 0 & 0 & \cdots & 0 & 0 & 2\rho \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (p-1)\rho^{p-2} & (p-2)\rho^{p-3} & \cdots & 2\rho & 1 & 0 & 0 & (p-1)\rho^{p-2} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \end{pmatrix}. \tag{C.32}$$

Denote by $e_{p+2}$ the last vector of the canonical basis of $\mathbb{R}^{p+2}$. We clearly have from (C.31) that

$$\hat{\rho}_n - \rho = -e_{p+2}^t \left(\hat{\vartheta}_n - \vartheta\right)$$

which implies that

$$\hat{\Delta}_n \hat{\vartheta}_n - \Delta\vartheta = B_n \left(\hat{\vartheta}_n - \vartheta\right) \tag{C.33}$$

where $B_n = \hat{\Delta}_n - A_n \vartheta e_{p+2}^t$. By the same token, let $0_p$ be the null vector of $\mathbb{R}^p$ and denote by $\mathrm{J}_p$ the rectangular matrix of size $p\times(p+1)$ given by

$$\mathrm{J}_p = \left(\mathrm{I}_p \ 0_p\right).$$

We deduce from (C.31) and (C.33) that

$$\tilde{\theta}_n = \hat{\theta}_n - \theta = \mathrm{J}_p \begin{pmatrix} \hat{\theta}_n - \theta \\ \hat{\rho}_n - \rho \end{pmatrix} = \mathrm{J}_p \left(\hat{\Delta}_n \hat{\vartheta}_n - \Delta\vartheta\right) = \mathrm{J}_p B_n \left(\hat{\vartheta}_n - \vartheta\right). \tag{C.34}$$

We also have from (2.5) that

$$\hat{\vartheta}_n - \vartheta = S_{n-1}^{-1} M_n \tag{C.35}$$

where

$$M_n = \sum_{k=1}^{n} \Phi_{k-1} V_k.$$

Consequently, it follows from (C.30), (C.34), and (C.35) that

$$J_{n-1}\left(\overline{\rho}_n - \rho\right) = N_n - C_n^t M_n + R_n$$

where $C_n = S_{n-1}^{-1} B_n^t \mathrm{J}_p^t T_n$ with $T_n = Q_n^I - 2\rho Q_{n-1}^J$, which leads to the main decomposition

$$\sqrt{n}\begin{pmatrix} \hat{\vartheta}_n - \vartheta \\ \overline{\rho}_n - \rho \end{pmatrix} = \frac{1}{\sqrt{n}} \mathcal{A}_n Z_n + \mathcal{B}_n \tag{C.36}$$

where

$$Z_n = \begin{pmatrix} M_n \\ N_n \end{pmatrix},$$

$$\mathcal{A}_n = n \begin{pmatrix} S_{n-1}^{-1} & 0_{p+2} \\ J_{n-1}^{-1} C_n^t & J_{n-1}^{-1} \end{pmatrix} \quad \text{and} \quad \mathcal{B}_n = \sqrt{n} \begin{pmatrix} 0_{p+2} \\ J_{n-1}^{-1} R_n \end{pmatrix}$$

where $0_{p+2}$ stands for the null vector of $\mathbb{R}^{p+2}$. The random sequence $(Z_n)$ is a locally square-integrable $(p+3)$-dimensional martingale with predictable quadratic variation given, for all $n \geqslant 1$, by

$$\langle Z \rangle_n = \sigma^2 \sum_{k=0}^{n-1} \begin{pmatrix} \Phi_k \Phi_k^t & \Phi_k \varepsilon_k \\ \Phi_k^t \varepsilon_k & \varepsilon_k^2 \end{pmatrix}.$$

We already saw from (3.4) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \Phi_k \Phi_k^t = \sigma^2 \Lambda \quad \text{a.s.} \tag{C.37}$$

In addition, it follows from (C.14) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\varepsilon_k^2=\frac{\sigma^2}{1-\rho^2}\quad \text{a.s.} \tag{C.38}$$

Furthermore, it is not hard to see that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_kV_k=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k\varepsilon_k=\sigma^2\quad \text{a.s.}$$

Moreover, we obtain from (1.2) that for all $n\geqslant p$ and for all $1\leqslant \ell\leqslant p$,

$$\varepsilon_n=\rho^\ell\varepsilon_{n-\ell}+\sum_{i=0}^{\ell-1}\rho^iV_{n-i}.$$

Consequently,

$$\begin{aligned}\sum_{k=1}^{n}X_{k-\ell}\varepsilon_k&=\sum_{k=1}^{n}X_{k-\ell}\left(\rho^\ell\varepsilon_{k-\ell}+\sum_{i=0}^{\ell-1}\rho^iV_{k-i}\right),\\&=\rho^\ell\sum_{k=1}^{n}X_{k-\ell}\varepsilon_{k-\ell}+\sum_{i=0}^{\ell-1}\rho^i\sum_{k=1}^{n}X_{k-\ell}V_{k-i},\end{aligned}$$

which implies that for all $1\leqslant \ell\leqslant p$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_{k-\ell}\varepsilon_k=\sigma^2\rho^\ell\quad \text{a.s.}$$

On the other hand, we infer from (1.1) that

$$\sum_{k=1}^{n}U_{k-1}\varepsilon_k=\sum_{k=1}^{n}X_k\varepsilon_k-\sum_{k=1}^{n}\varepsilon_k^2-\sum_{i=1}^{p}\theta_i\sum_{k=1}^{n}X_{k-i}\varepsilon_k.$$

Hence, we find that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}U_{k-1}\varepsilon_k=-\sigma^2\left(\frac{\rho^2}{1-\rho^2}+\sum_{i=1}^{p}\theta_i\rho^i\right)\quad \text{a.s.}$$

Consequently, we obtain that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\Phi_k\varepsilon_k=\sigma^2\zeta\quad \text{a.s.} \tag{C.39}$$

where $\zeta$ is the vector of $\mathbb{R}^{p+2}$ such that $\zeta^t=(1,\rho,\ldots,\rho^p,\varrho_p)$ with

$$\varrho_p=-\eta\rho^2-\sum_{i=1}^{p}\theta_i\rho^i\quad\text{and}\quad \eta=\frac{1}{1-\rho^2}.$$

We deduce from (C.37), (C.38), and (C.39) that

$$\lim_{n\to\infty}\frac{1}{n}\langle Z\rangle_n=\mathcal{Z}\quad \text{a.s.} \tag{C.40}$$

where $\mathcal{Z}$ is the positive-semidefinite symmetric matrix given by

$$\mathcal{Z}=\sigma^4\begin{pmatrix}\Lambda & \zeta\\ \zeta^t & \eta\end{pmatrix}. \tag{C.41}$$



DOI: 10.1002/acs

One can observe that $\mathcal{Z}$ is not positive-definite as $\det(\mathcal{Z}) = 0$. Nevertheless, it is not hard to see that $(Z_n)$ satisfies the Lindeberg condition. Therefore, we can conclude from the central limit theorem for multidimensional martingales given, for example, by Corollary 2.1.10 of [18] that

$$\frac{1}{\sqrt{n}} Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{Z}). \tag{C.42}$$

Furthermore, we already saw from (C.32) that

$$\lim_{n\to\infty} A_n = A \quad \text{a.s.}$$

which implies that

$$\lim_{n\to\infty} B_n = \Delta - A\vartheta e_{p+2}^t \quad \text{a.s.}$$

One can easily check from (3.9) and (C.32) that

$$\Delta - A\vartheta e_{p+2}^t = \nabla$$

where the matrix $\nabla$ is given by (4.2). Moreover, it follows from the previous calculation that

$$\lim_{n\to\infty} \frac{1}{n} T_n = \sigma^2 \left(1-\rho^2\right) T \quad \text{a.s.}$$

where $T$ is the vector of $\mathbb{R}^p$ given by $T^t = (1, \rho, \ldots, \rho^{p-1})$. Consequently, as the vector $C_n = S_{n-1}^{-1} B_n^t \mathrm{J}_p^t T_n$, we obtain from (3.4) that

$$\lim_{n\to\infty} C_n = C \quad \text{a.s.}$$

where

$$C = \left(1-\rho^2\right) \Lambda^{-1} \nabla^t \mathrm{J}_p^t T.$$

Hence, we obtain from (3.4) and (C.16) that

$$\lim_{n\to\infty} \mathcal{A}_n = \mathcal{A} \quad \text{a.s.} \tag{C.43}$$

where

$$\mathcal{A} = \sigma^{-2} \begin{pmatrix} \Lambda^{-1} & 0_{p+2} \\ \left(1-\rho^2\right) C^t & \left(1-\rho^2\right) \end{pmatrix}.$$

In addition, we clearly have from (C.16) that

$$\lim_{n\to\infty} \mathcal{B}_n = \begin{pmatrix} 0_{p+2} \\ 0 \end{pmatrix} \quad \text{a.s.} \tag{C.44}$$

Finally, we deduce from the conjunction of (C.36), (C.42), (C.43), (C.44), together with Slutsky's lemma that

$$\sqrt{n} \begin{pmatrix} \hat{\vartheta}_n - \vartheta \\ \overline{\rho}_n - \rho \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathcal{A}\mathcal{Z}\mathcal{A}'\right)$$

which leads to

$$\sqrt{n}(\overline{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2\right)$$

where the asymptotic variance $\tau^2$ is given by

$$\tau^2 = (1-\rho^2)^2 \left(C^t \Lambda C + 2C^t \zeta + \eta\right).$$

However, one can easily see from (5.7) and (5.8) that

$$\begin{aligned}\tau^2 &= \left(1-\rho^2\right)^2 \parallel \Lambda^{1/2}\alpha + (1-\rho^2)\Lambda^{-1/2}\nabla^t\beta \parallel^2,\\ &= \left(1-\rho^2\right)^2 \parallel \Lambda^{-1/2}(\Lambda\alpha + (1-\rho^2)\nabla^t\beta) \parallel^2,\\ &= \left(1-\rho^2\right)^2 \parallel \Lambda^{-1/2}\gamma \parallel^2,\\ &= \left(1-\rho^2\right)^2 \gamma^t\Lambda^{-1}\gamma,\end{aligned}$$

which completes the proof of (5.9). Finally, (5.10) immediately follows from (5.9) together with (C.27) and (C.28), which achieves the proof of Theorem 5.2.

*Proof of Theorem 5.3*

The proof of Theorem 5.3 is straightforward. As a matter of fact, we already know from (5.10) that under the null hypothesis $\mathcal{H}_0$,

$$\sqrt{n}(\hat{D}_n - D_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\tau^2\right) \tag{C.45}$$

where the asymptotic variance $\tau^2$ is given by (5.11). In addition, it follows from (5.14) that

$$\lim_{n\to\infty} \hat{\tau}_n^2 = \tau^2 \quad \text{a.s.} \tag{C.46}$$

Hence, we deduce from (C.45), (C.46), and Slutsky's lemma that under the null hypothesis $\mathcal{H}_0$,

$$\frac{\sqrt{n}}{2\hat{\tau}_n}\left(\hat{D}_n - D_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

which obviously implies (5.15). It remains to show that under the alternative hypothesis $\mathcal{H}_1$, our test statistic goes almost surely to infinity. Under $\mathcal{H}_1$, we already saw from Theorem 5.1 that

$$\lim_{n\to\infty} \overline{\rho}_n - \rho_0 = \rho - \rho_0 \quad \text{a.s.}$$

and this limit is different from zero. Consequently,

$$\lim_{n\to\infty} n\left(\overline{\rho}_n - \rho_0\right)^2 = +\infty \quad \text{a.s.} \tag{C.47}$$

However, we clearly find from (C.28) that

$$\hat{D}_n - D_0 = -2\left(\overline{\rho}_n - \rho_0\right) + e_n \tag{C.48}$$

where $e_n = -2f_n(1-\overline{\rho}_n) + g_n$. Finally, (C.47) and (C.48) clearly lead to (5.16), completing the proof of Theorem 5.3.

## REFERENCES


1. Durbin J, Watson GS. Testing for serial correlation in least squares regression I. *Biometrika* 1950; **37**:409–428.
2. Durbin J, Watson GS. Testing for serial correlation in least squares regression II. *Biometrika* 1951; **38**:159–178.
3. Durbin J, Watson GS. Testing for serial correlation in least squares regression III. *Biometrika* 1971; **58**:1–19.
4. Malinvaud E. Estimation et prévision dans les modéles économiques autorégressifs. *Review of the International Institute of Statistics* 1961; **29**:1–32.
5. Nerlove M, Wallis KF. Use of the Durbin Watson statistic in inappropriate situations. *Econometrica* 1966; **34**:235–238.
6. Durbin J. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* 1970; **38**:410–421.
7. Stocker T. On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors. *Statistic Papers* 2006; **48**:81–93.
8. Bercu B, Proia F. A sharp analysis on the asymptotic behavior of the Durbin-Watson for the first-order autoregressive process. *ESAIM PS* 2013; **17**:500–530.
9. Proïa F. Further results on the h-test of Durbin for stable autoregressive processes. *Journal of Multivariate Analysis* 2013; **118**:77–101.




DOI: 10.1002/acs

10. Huang KY, Jane C-J. A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. *Expert Systems with Applications* 2009; **36**:5387–5392.
11. Kamaldar M, Mahjoob MJ, Vahid Alizadeh H. Robust speed control of a spherical robot using ARX uncertain modeling. *IEEE International Symposium on Robotic and Sensors Environments*, Montreal, QC, 17-18 September 2011; 196–201.
12. Kon J, Yamashita Y, Tanaka T, Tashiro A, Daiguji M. Practical application of model identification based on ARX models with transfer functions. *Control Engineering Practice* 2013; **21**:195–203.
13. Jeon YJ, Eom Ys, Hwang JO, Yang HK, Im JJ, Park SH, Jung NS. Estimation of an aortic pressure waveform from radial pulse using ARX model. *Proceeding of the World Congress on Medical Physics and Biomedical Engineering* 2006; **14**:183–186.
14. Liu Y, Birch AA, Allen R. Dynamic cerebral auto regulation assessment using an ARX model: comparative study using step response and phase shift analysis. *Medical Engineering and Physics* 2003; **25**:647–653.
15. Aström KJ, Wittenmark B. *Adaptive Control*, 2nd ed. Addison-Wesley: New York, 1995.
16. Horn RA, Johnson CR. *Matrix Analysis*. Cambridge University Press: New York, 1990.
17. Bercu B, Vázquez V. A new concept of strong controllability via the Schur complement in adapative tracking. *Automatica* 2010; **46**:1799–1805.
18. Duflo M. *Random Iterative Models*. Springer Verlag: Berlin, 1997.



DOI: 10.1002/acs