A Durbin-Watson serial correlation test for ARX processes via excited adaptive tracking

Bernard Bercu, Bruno Portier, Victor Vazquez

A DURBIN-WATSON SERIAL CORRELATION TEST FOR ARX PROCESSES VIA EXCITED ADAPTIVE TRACKING

BERNARD BERCU, BRUNO PORTIER, AND VICTOR VAZQUEZ

Abstract. We propose a new statistical test for the residual autocorrelation in ARX adaptive tracking. The introduction of a persistent excitation in the adaptive tracking control allows us to build a bilateral statistical test based on the well-known Durbin-Watson statistic. We establish the almost sure convergence and the asymptotic normality for the Durbin-Watson statistic leading to a powerful serial correlation test. Numerical experiments illustrate the good performances of our statistical test procedure.

1. Introduction

Model validation is an important and essential final step in the identification of stochastic dynamical systems. This validation step is often done through the analysis of residuals of the model considered. In particular, testing the non-correlation of the residuals is a crucial task since many theoretical results require independence of the driven noise of the systems. Moreover, non-compliance with this hypothesis can lead to misinterpretation of the theoretical results. For example, it is well known that for linear autoregressive models with autocorrelated residuals, the least squares estimator is asymptotically biased, see e.g. [5], [11], [14], [15], and therefore the estimated model is not the correct one. Consequently, to ensure a good interpretation of the results, it is necessary to have a powerful tool allowing to detect the possible autocorrelation of the residuals. The well-known statistical test of Durbin-Watson was introduced to deal with this question, and more specifically, for detecting the presence of a first-order autocorrelated noise in linear regression models [8], [9], [10], [11], firstly and for linear autoregressive models [5], [14], [16], [17], [15], secondly.

To the best of our knowledge, no such serial correlation statistical test is available for controlled autoregressive processes. The aim of this paper is to carry out a serial correlation test, based on the Durbin-Watson statistic, for the ARX$(p,1)$ process given, for all $n \geq 0$, 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 (ε_n) is given by the first-order autoregressive process

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

and the control objective is the tracking of a given reference trajectory. More precisely, we shall propose a bilateral statistical test allowing to decide between the null hypothesis

2000 *Mathematics Subject Classification.* Primary: 62G05 Secondary: 93C40, 62F05, 60F15, 60F05.

Key words and phrases. Durbin-Watson statistic, Estimation, Adaptive tracking control, Persistent excitation, Almost sure convergence, Asymptotic normality, Statistical test for serial autocorrelation.

$\mathcal{H}_0$: «$\rho = 0$» which ensures that the driven noise is not correlated, and the alternative one $\mathcal{H}_1$: «$\rho \neq 0$» which means that the residual process is effectively first-order autocorrelated. The choice of the Durbin-Watson statistic, instead of any other statistical tests, is governed by its efficienty for autoregressive processes without control, see [5], [11], [14], [15].

In contrast with the recent work [6], we propose to make use of a different strategy via a modification of the adaptive control law. This modification relies on the introduction of an additional persistent excitation. Since the pioneering works of Anderson [1] and Moore [13], the concept of persistent excitation has been successfully developped in many fields of applied mathematics such as identification of complex systems, feedback adaptive control, etc. While it was not possible in [6] to test the non correlation of the driven noise (ε_n), that is to test whether or not $\rho = 0$, the introduction of an additional persistent excitation term in the control law will be the key point to build our serial correlation test. Moreover, we wish to mention that all previous works devoted to non correlation test based on the Durbin-Watson statistic were only related to uncontrolled processes. Therefore, thanks to the persistent excitation, our statistical test is at our knowledge the first one in the context of linear processes with adaptive control.

The paper is organized as follows. Section 2 is devoted to the ARX process and to the persistently excited adaptive control law. In Section 3, we establish the asymptotic properties of the Durbin-Watson statistic as well as a bilateral statistical test for residual autocorrelation. Some numerical experiments are provided in Section 4. Finally, all technical proofs are postponed in the Appendices.

2. Model and excited adaptive tracking

We focus our attention on the ARX$(p, 1)$ process given, for all $n \geq 0$, by

$$X_{n+1} = \sum_{k=1}^{p} \theta_k X_{n-k+1} + U_n + \varepsilon_{n+1} \tag{2.1}$$

where the driven noise (ε_n) is given by the first-order autoregressive process

$$\varepsilon_{n+1} = \rho\, \varepsilon_n + V_{n+1}. \tag{2.2}$$

We assume that the autocorrelation parameter satisfies $|\rho| < 1$ and the initial values X_0, ε_0 and U_0 may be arbitrarily chosen. We also assume that (V_n) is a martingale difference sequence adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$ where $\mathcal{F}_n$ is the σ-algebra of the events occurring up to time n, such that, for all $n \geq 0$, $\mathbb{E}\left[V_{n+1}^2 | \mathcal{F}_n\right] = \sigma^2$ a.s. with $\sigma^2 > 0$. We denote by θ the unknown parameter of the ARX$(p, 1)$ process,

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

Our control strategy is to regulate the dynamic of the process (X_n) by forcing X_n to track a bounded reference trajectory (x_n). We assume that (x_n) is predictable which means that for all $n \geq 1$, x_n is $\mathcal{F}_{n-1}$-measurable. For the sake of simplicity, we also assume that

$$\sum_{k=1}^{n} x_k^2 = o(n) \qquad \text{a.s.} \tag{2.3}$$

In order to regulate the dynamic of the process (X_n) given by (2.1), we propose to make use of the adaptive control law introduced in [6] together with additional persistent excitation. The strategy consists of using a control associated with a higher order model than the initial ARX$(p,1)$, and more precisely an ARX$(p+1,2)$ model. The introduction of an additional excitation in the control law will be the key point to build our serial correlation test for the driven noise (ε_n), that is to test wether or not $\rho=0$. Denote by (ξ_n) a centered exogenous noise with known variance $\nu^2>0$, which will play the role of the additional excitation. We assume that (ξ_n) is independent of (V_n), of (x_n) and of the initial state of the system. One can observe that these assumptions are not at all restrictive as we have in our own hands the additional excitation (ξ_n)

The excited adaptive control law is given, for all $n \geq 0$, by

$$U_n = x_{n+1} - \widehat{\vartheta}_n^{\,t}\,\Phi_n + \xi_{n+1} \tag{2.4}$$

where $\widehat{\vartheta}_n$ stands for the least squares estimator of the unknown parameter of the ARX$(p+1,2)$ model with uncorrelated driven noise

$$X_{n+1} = \vartheta^t\Phi_n + U_n + V_{n+1} \tag{2.5}$$

where the new parameter $\vartheta \in \mathbb{R}^{p+2}$ is related to θ and ρ by the identity

$$\vartheta = \begin{pmatrix} \theta \\ 0 \\ 0 \end{pmatrix} - \rho \begin{pmatrix} -1 \\ \theta \\ 1 \end{pmatrix} \tag{2.6}$$

and the new regression vector is given by

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

It is well-known that $\widehat{\vartheta}_n$ satisfies the recursive relation

$$\widehat{\vartheta}_{n+1} = \widehat{\vartheta}_n + S_n^{-1}\Phi_n\Big(X_{n+1} - U_n - \widehat{\vartheta}_n^{\,t}\Phi_n\Big) \tag{2.7}$$

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

$$S_n = \sum_{k=0}^{n} \Phi_k\Phi_k^t + \mathrm{I}_{p+2}. \tag{2.8}$$

As usual, the identity matrix I_{p+2} is added in order to avoid useless invertibility assumption. One can immediately see from (2.6) that the last component of the vector ϑ is $-\rho$. Consequently, we obtain an estimator of ρ by simply taking the opposite of the last coordinate of $\widehat{\vartheta}_n$ which will be denoted by $\widehat{\rho}_n$. In addition, one can also deduce from (2.6) that

$$\begin{pmatrix} \theta \\ \rho \end{pmatrix} = \Delta\,\vartheta \tag{2.9}$$

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

$$\Delta = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ \rho & 1 & 0 & \cdots & \cdots & 0 & \rho \\ \rho^2 & \rho & 1 & 0 & \cdots & 0 & \rho^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & \rho & 1 & 0 & \rho^{p-1} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}. \tag{2.10}$$

Then, starting from (2.9) and replacing ρ by $\widehat{\rho}_n$ in (2.10), we can estimate θ by

$$\widehat{\theta}_n = \begin{pmatrix} \mathrm{I}_p & 0 \end{pmatrix} \widehat{\Delta}_n \widehat{\vartheta}_n \tag{2.11}$$

where $\widehat{\vartheta}_n$ is given by (2.7) and

$$\widehat{\Delta}_n = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ \widehat{\rho}_n & 1 & 0 & \cdots & \cdots & 0 & \widehat{\rho}_n \\ \widehat{\rho}_n^2 & \widehat{\rho}_n & 1 & 0 & \cdots & 0 & \widehat{\rho}_n^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \widehat{\rho}_n^{p-1} & \widehat{\rho}_n^{p-2} & \cdots & \widehat{\rho}_n & 1 & 0 & \widehat{\rho}_n^{p-1} \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & -1 \end{pmatrix}. \tag{2.12}$$

From the almost sure convergence of $\widehat{\vartheta}_n$ to ϑ, we easily deduce the almost sure convergences of $\widehat{\theta}_n$ and $\widehat{\rho}_n$ to θ and ρ, respectively.

3. A Durbin-Watson serial correlation test

We are in the position to introduce our serial correlation test based on the Durbin-Watson statistic which is certainly the most commonly used statistics for testing the presence of serial autocorrelation. Our goal is to test

$$\mathcal{H}_0 \,:\, \langle\!\langle \rho = 0 \rangle\!\rangle \qquad \text{vs} \qquad \mathcal{H}_1 \,:\, \langle\!\langle \rho \neq 0 \rangle\!\rangle.$$

For that purpose, we consider the Durbin-Watson statistic [5], [8], [9], [10], [11] given, for all $n \geq 1$, by

$$\widehat{D}_n = \frac{\sum_{k=1}^{n} \left(\widehat{\varepsilon}_k - \widehat{\varepsilon}_{k-1}\right)^2}{\sum_{k=0}^{n} \widehat{\varepsilon}_k^2} \tag{3.1}$$

where the residuals $\widehat{\varepsilon}_k$ are defined, for all $0 \leq k \leq n$, by

$$\widehat{\varepsilon}_k = X_k - U_{k-1} - \widehat{\theta}_n^t \varphi_{k-1} \tag{3.2}$$

with $\widehat{\theta}_n$ given by (2.11) and $\varphi_n^t = (X_n, \ldots, X_{n-p+1})$. The initial value $\widehat{\varepsilon}_0$ may be arbitrarily chosen and we take $\widehat{\varepsilon}_0 = X_0$.

On the one hand, we would like to emphasize that it is not possible to perform this statistical test if the control law is not persistently excited [6]. On the other hand, one can notice that it is also possible to estimate the serial correlation parameter ρ by the least squares estimator

$$\overline{\rho}_n = \frac{\sum_{k=1}^{n} \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1}}{\sum_{k=1}^{n} \widehat{\varepsilon}_{k-1}^2} \tag{3.3}$$

which is certainly the more natural estimator of ρ. The Durbin-Watson statistic $\widehat{D}_n$ is related to $\overline{\rho}_n$ by the linear relation

$$\widehat{D}_n = 2(1 - \overline{\rho}_n) + \zeta_n \tag{3.4}$$

where the remainder ζ_n plays a negligeable role. The almost sure properties of $\widehat{D}_n$ and ρ_n are as follows.

Theorem 3.1. *Assume (V_n) has a finite conditional moment of order > 2. Then, $\overline{\rho}_n$ converges almost surely to ρ*

$$(\overline{\rho}_n - \rho)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{3.5}$$

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

$$\left(\widehat{D}_n - D\right)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{3.6}$$

Proof. The proofs are given in Appendix A. □

Let us now give the asymptotic normality of the Durbin-Watson statistic which will be useful to build our serial correlation test.

Theorem 3.2. *Assume that (V_n) has finite conditional moments of order > 2. Then, we have*

$$\sqrt{n}(\overline{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2\right) \tag{3.7}$$

where the asymptotic variance τ^2 is given by

$$\begin{aligned} \tau^2 &= \frac{(1-\rho^2)}{(\sigma^2+\nu^2)(\nu^2+\sigma^2\rho^{2(p+1)})}\Big[\Big((\sigma^2-\nu^2)-(p+1)\sigma^2\rho^{2p}+(p-1)\sigma^2\rho^{2(p+1)}\Big)^2 \\ &\quad + \sigma^2(\nu^2+\sigma^2\rho^{2(p+1)})\Big(4-(4p+3)\rho^{2p}+4p\rho^{2(p+1)}-\rho^{2(2p+1)}\Big)\Big]. \end{aligned} \tag{3.8}$$

Moreover, if (V_n) has finite conditional moments of order > 4, we also have

$$\sqrt{n}(\widehat{D}_n - D) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\tau^2\right). \tag{3.9}$$

Proof. The proofs are given in Appendix B. □

Remark 3.1. We now point out the crucial role played by the additional excitation in the control law given by (2.4). It follows from (3.8) that if $\rho = 0$, then τ^2 reduces to

$$\tau^2 = \frac{\sigma^2+\nu^2}{\nu^2}.$$

Consequently, if $\nu^2 = 0$ i.e. there is no persistent excitation, then this variance explodes. Therefore, the persistent excitation allows to investigate the important case $\rho = 0$ and more generally to stabilize the asymptotic variance of the Durbin-Watson statistic.

We are now in the position to test whether or not the serial correlation parameter $\rho = 0$. According to Theorem 3.1, we have under the null hypothesis $\mathcal{H}_0$,

$$\lim_{n\to\infty} \widehat{D}_n = 2 \qquad \text{a.s.}$$

In addition, we clearly have from (3.9) that under $\mathcal{H}_0$,

$$\frac{n}{4\tau^2}\left(\widehat{D}_n - 2\right)^2 \xrightarrow{\mathcal{L}} \chi^2 \tag{3.10}$$

where χ^2 stands for a Chi-square distribution with one degree of freedom. It remains to accurately estimate the asymptotic variance τ^2. It is not hard to see that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} X_k^2 = \sigma^2 + \nu^2 \qquad \text{a.s.} \tag{3.11}$$

Consequently, as ν^2 is known, it immediately follows from (3.11) that

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \nu^2$$

converges almost surely to σ^2. Hence, we can propose to make use of

$$\begin{aligned}\widehat{\tau}_n^2 \;&=\; \frac{(1-\overline{\rho}_n^2)}{(\widehat{\sigma}_n^2+\nu^2)(\nu^2+\widehat{\sigma}_n^2\overline{\rho}_n^{2(p+1)})}\Bigg[\Big((\widehat{\sigma}_n^2-\nu^2)-(p+1)\widehat{\sigma}_n^2\overline{\rho}_n^{2p}+(p-1)\widehat{\sigma}_n^2\overline{\rho}_n^{2(p+1)}\Big)^2 \\ &+\; \widehat{\sigma}_n^2(\nu^2+\widehat{\sigma}_n^2\overline{\rho}_n^{2(p+1)})\Big(4-(4p+3)\overline{\rho}_n^{2p}+4p\overline{\rho}_n^{2(p+1)}-\overline{\rho}_n^{2(2p+1)}\Big)\Bigg].\end{aligned} \tag{3.12}$$

Therefore, our bilateral statistical test relies on the following results.

Corollary 3.1. *Assume that (x_n) and (V_n) have finite conditional moments of order > 4. Then, under the null hypothesis $\mathcal{H}_0$: "$\rho = 0$",*

$$T_n = \frac{n}{4\widehat{\tau}_n^2}\left(\widehat{D}_n - 2\right)^2 \xrightarrow{\mathcal{L}} \chi^2 \tag{3.13}$$

In addition, under the alternative hypothesis $\mathcal{H}_1$: "$\rho \neq 0$",

$$\lim_{n\to\infty} T_n = +\infty \qquad \text{a.s.} \tag{3.14}$$

From a practical point of view, for a significance level α 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_α 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 $T_n \leq a_\alpha$, and will be rejected otherwise.

Let us now make a few comments. First of all, under $\mathcal{H}_0$, we already saw that τ^2 reduces to $(\sigma^2+\nu^2)/\nu^2$. It can be estimated by $(\widehat{\sigma}_n^2+\nu^2)/\nu^2$. Therefore, it is also possible to consider the test statistic associated with

$$\mathcal{T}_n = \frac{n^2\nu^2}{4(\widehat{\sigma}_n^2+\nu^2)}\left(\widehat{D}_n - 2\right)^2. \tag{3.15}$$

Intuitively, one may think that the statistical test based on $\mathcal{T}_n$ is more efficient under $\mathcal{H}_0$ since we do not estimate the parameter ρ, but less powerful under $\mathcal{H}_1$. This point will be examined in Section 4. Next, the acceptance of $\mathcal{H}_0$ after our statistical test procedure should lead to a change of control law. As a matter of fact, if we accept $\rho = 0$, the driven noise (ε_n) is not correlated. It means that we can implement the usual control law [2] associated with model (2.1) given, for all $n \geq 0$, by

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

where $\widehat{\theta}_n$ stands for the standard least-squares estimator associated with (2.1). Finally, the test provided by Corollary 3.1 may be of course extended if we replace zero by any $\rho_0 \in \mathbb{R}$ with $|\rho_0| < 1$ in the null hypothesis. To be more precise, we are able as in [6] to test $\mathcal{H}_0$: « $\rho = \rho_0$ » versus $\mathcal{H}_1$: « $\rho \neq \rho_0$ ». We wish to mention that the asymptotic variance τ^2 is smaller than the one obtained in [6].

4. Numerical Experiments

This section is devoted to the application of our Durbin-Watson serial correlation test. Although this test has several potential of being applied in concrete situations, a large search in the literature did not offer any one. We then consider artificial models for illustrative purposes and for studying the empirical level and power of our test for sample sizes from small to moderate, that is $n = 50$, 100, 200, 500, 1000 and 2000.

In order to keep this section brief, we restrict ourself to the three explosive models in open-loop

$$X_{n+1} = \frac{3}{2}X_n + U_n + \varepsilon_{n+1} \tag{4.1}$$

$$X_{n+1} = -X_n + 2X_{n-1} + U_n + \varepsilon_{n+1} \tag{4.2}$$

$$X_{n+1} = X_n + \frac{1}{2}X_{n-1} + \frac{1}{4}X_{n-2} + U_n + \varepsilon_{n+1} \tag{4.3}$$

where the driven noise (ε_n) is given by (2.2) and (V_n) is a sequence of independent and identically distributed random variables with $\mathcal{N}(0,1)$ distribution. The control law U_n is given by (2.4) where, for the sake of simplicity, the reference trajectory $x_n = 0$ and the persistent excitation (ξ_n) s a sequence of independent and identically distributed random variables with $\mathcal{N}(0,\nu^2)$ distribution.

For each model, we based our numerical simulations on $N = 1000$ realizations of sample size n. We use a short learning period of 100 time steps. This learning period allows us to forget the transitory phase. The level of significance is set to $\alpha = 5\%$. For the statistical tests based on T_n and $\mathcal{T}_n$, we are interested in the empirical level under $\mathcal{H}_0$ to be compared to the theoretical level 5%, and the empirical power under $\mathcal{H}_1$, to be compared with 1.

First of all, let us study the effect of the variance ν^2 of the exogenous noise (ξ_n) on the behavior of the statistical test under $\mathcal{H}_0$.

	n	50	100	200	500	1000	2000
$\nu = 0.5$	T_n	0.9%	1.6%	2%	3%	2.9%	4.9%
	$\mathcal{T}_n$	0%	0.1%	0.5%	2.1%	2.4%	4.4%
$\nu = 1$	T_n	2.5%	2.5%	3.3%	4.4%	4.4%	4.9%
	$\mathcal{T}_n$	1.3%	1.3%	2.5%	4.1%	4%	4.8%
$\nu = 2$	T_n	5.2%	4.1%	4.9%	5.3%	4.7%	4.6%
	$\mathcal{T}_n$	3.7%	3.7%	4.1%	5.1%	4.7%	4.6%
$\nu = 3$	T_n	5.8%	5.1%	5.6%	4.3%	4.3%	4.8%
	$\mathcal{T}_n$	4.5%	4.6%	5.1%	4.2%	4.3%	4.7%

TABLE 1. Model (4.1). Percentage of rejections of our test under $\mathcal{H}_0$ (to be compared to the 5% theoretical level).

It is clear from Table 1, where one can find the results obtained for different values of ν, that the variance of the persistent excitation (ξ_n) in the control law plays a crucial role. Indeed, one can observe that if it is too small, then the empirical level of the test is bad for sample sizes from small to moderate $n \leq 1000$. Of course, a high value of ν^2 improves the performance of the test under $\mathcal{H}_0$, but degrades the performance of the tracking. The value $\nu = 2$ realizes a good compromise and allows a good calibration of the test under $\mathcal{H}_0$.

	n	50	100	200	500	1000	2000
Model (4.1)	T_n	5.2%	4.1%	4.9%	5.3%	4.7%	4.6%
	$\mathcal{T}_n$	3.7%	3.7%	4.1%	5.1%	4.7%	4.6%
Model (4.2)	T_n	5.9%	3%	3.9%	4.6%	4.8%	5.2%
	$\mathcal{T}_n$	4.7%	2.5%	3.7%	4.5%	4.8%	5.1%
Model (4.3)	T_n	4.8%	4.7%	4.1%	5.2%	4.9%	6%
	$\mathcal{T}_n$	3.8%	3.5%	3.9%	5%	4.9%	5.9%

TABLE 2. Percentage of rejections of our test under $\mathcal{H}_0$ (to be compared to the 5% theoretical level). $\nu = 2$.

One can find in Table 2 the percentage of rejections of our test under $\mathcal{H}_0$ for the three different models (4.1) to (4.3). The empirical levels of the test are close to the 5% theoretical level even for small sample sizes. Both statistical tests based on T_n and $\mathcal{T}_n$ are comparable even if the test statistic $\mathcal{T}_n$ systematically tends to less reject $\mathcal{H}_0$ than the test statistic T_n.

Let us now study the empirical power of our statistical test. One can find in Tables 3 to 5 the results obtained for each of the three models (4.1) to (4.3). As expected, it is difficult to reject $\mathcal{H}_0$ when $\rho = 0.05$ for small sample sizes or $\rho = 0.1$ to a lesser extent. However, the test performs pretty well as the percentage of correct decisions increases with the sample size.

	n	50	100	200	500	1000	2000
$\rho = 0.05$	T_n	6.4%	7.1%	9.6%	18.8%	30.6%	50.3%
	$\mathcal{T}_n$	4.8%	6.2%	9%	18.4%	30.5%	50.2%
$\rho = 0.1$	T_n	11.1%	17.4%	25.9%	56.8%	81.9%	98.2%
	$\mathcal{T}_n$	8.9%	15.2%	24.7%	56.2%	81.6%	98.1%
$\rho = 0.2$	T_n	32%	47.5%	77.6%	98.9%	100%	100%
	$\mathcal{T}_n$	10.7%	24.1%	49.5%	91.2%	99.5%	100%
$\rho = 0.3$	T_n	57%	83%	97.7%	100%	100%	100%
	$\mathcal{T}_n$	51.7%	81.7%	97.5%	100%	100%	100%
$\rho = 0.4$	T_n	78.8%	97.2%	99.7%	100%	100%	100%
	$\mathcal{T}_n$	73.4%	96.5%	99.7%	100%	100%	100%

TABLE 3. Model (4.1). Percentage of correct decisions of our test under $\mathcal{H}_1$.

We further observe, as expected, that for a fixed value of the sample size n, the higher the value of ρ is, the more the percentage of correct decisions increases. We also notice that for a fixed value of ρ, the empirical power increases with the sample size. In conclusion, the test performs very well under $\mathcal{H}_1$. Moreover, higher values of the order p does not degrade the performances of our statistical test.

	n	50	100	200	500	1000	2000
$\rho = 0.05$	T_n	5.2%	6.2%	11%	17.6%	28.5%	54.6%
	$\mathcal{T}_n$	4.5%	5.6%	10.7%	17.4%	28.3%	54.6%
$\rho = 0.1$	T_n	11.5%	15%	25.3%	53%	78.1%	97.3%
	$\mathcal{T}_n$	9.7%	14.1%	24.6%	52.5%	78.1%	97.3%
$\rho = 0.2$	T_n	28.5%	45.8%	70.9%	98.3%	99.9%	100%
	$\mathcal{T}_n$	24.4%	44.1%	69.6%	98.3%	99.9%	100%
$\rho = 0.3$	T_n	53.1%	80.6%	97%	100%	100%	100%
	$\mathcal{T}_n$	48.4%	79.1%	96.9%	100%	100%	100%
$\rho = 0.4$	T_n	77.4%	96.4%	100%	100%	100%	100%
	$\mathcal{T}_n$	74%	95.9%	100%	100%	100%	100%

TABLE 4. Model (4.2). Percentage of correct decisions of our test under $\mathcal{H}_1$.

Finally, one can realize that for small sample sizes, the statistical test based on $\mathcal{T}_n$ is less powerful than the one associated with T_n. We also wish to mention that, by symmetry, the performance of our statistical tests are the same for negative values of ρ.

	n	50	100	200	500	1000	2000
$\rho = 0.05$	T_n	6.9%	7%	9.6%	16%	28.5%	50.6%
	$\mathcal{T}_n$	6.1%	6.6%	9.4%	15.9%	28.5%	50.5%
$\rho = 0.1$	T_n	11%	14.2%	23.8%	53.6%	83.5%	98%
	$\mathcal{T}_n$	9.6%	12.6%	23.6%	53%	83.1%	98%
$\rho = 0.2$	T_n	29.7%	46.2%	77.1%	97.7%	100%	100%
	$\mathcal{T}_n$	25.9%	44.6%	76.3%	97.7%	100%	100%
$\rho = 0.3$	T_n	52.5%	81.1%	97.4%	100%	100%	100%
	$\mathcal{T}_n$	48.8%	79.2%	97.3%	100%	100%	100%
$\rho = 0.4$	T_n	75.9%	96.9%	100%	100%	100%	100%
	$\mathcal{T}_n$	73.3%	96.6%	100%	100%	100%	100%

Table 5. Model (4.3). Percentage of correct decisions of our test under $\mathcal{H}_1$.

5. Conclusion

Thanks to the introduction of a persistent excitation in the control law used to regulate an ARX(p,1) process, we were able to propose a non correlation test for the driven noise based on the well-known Durbin-Watson statistic. In addition, we have shown through a simulation study on artificial models the efficiency of our statistical test procedure.

Of course, many questions remain open. In particular, the extension of our results to ARX(p, q) processes where $q > 1$, would be a very attractive challenge for the control community. Even though our test is a potentially useful tool, we have seen in the literature that ARX$(p, 1)$ models are often too simple for being applied to real physical models. However, the study of such models is much more difficult to handle. On the one hand, it will be necessary to make an additional assumption of strong controllability, see [3], [4]. On the other hand, the asymptotic variance given by (3.8) will be much more complicated as well as its estimate given by (3.12).

Finally, this work may be seen as a first step towards a serial correlation test for ARX(p, q) processes. The implementation of our statistical test within a real physical model will allow to fully validate its efficiency.

Appendix A

PROOFS OF THE ALMOST SURE CONVERGENCE RESULTS

The almost sure convergence results rely on the following keystone lemma.

Lemma A.1. *Assume that (V_n) has a finite conditional moment of order > 2. Then, we have*

$$\lim_{n\to\infty} \frac{S_n}{n} = \Lambda \qquad \textit{a.s.} \tag{A.1}$$

where Λ is the symmetric square matrix of order $p + 2$ given by

$$\Lambda = \begin{pmatrix} L & K^t \\ K & H \end{pmatrix} \tag{A.2}$$

with $L = (\sigma^2 + \nu^2)\mathrm{I}_{p+1}$ where I_{p+1} stands for the identity matrix of order $p+1$, K is the line vector of $\mathbb{R}^{p+1}$

$$K = \Big(K_0, K_1, K_2, \ldots, K_p\Big) \tag{A.3}$$

with $K_0 = \nu^2$ and for all $1 \le k \le p$, $K_k = -(\sigma^2+\nu^2)\theta_k - \sigma^2\rho^k$ and H is the positive real number given by

$$H = \nu^2 + \sigma^2 \sum_{k=1}^{p} (\theta_k + \rho^k)^2 + \nu^2 \sum_{k=1}^{p} \theta_k^2 + \frac{\sigma^2 \rho^{2(p+1)}}{1-\rho^2}. \tag{A.4}$$

The proof of Lemma A.1 is left to the reader as it follows exactly the same lines as the one of Theorem 4.1 in [4]. Denote by S the Schur complement of L in Λ,

$$S = H - \frac{1}{\sigma^2+\nu^2} \parallel K \parallel^2 = \frac{\sigma^2(\nu^2 + \sigma^2\rho^{2(p+1)})}{(1-\rho^2)(\sigma^2+\nu^2)}.$$

We deduce from (A.2) that

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

Consequently, whatever the value of the parameter ρ with $|\rho| < 1$, $\det(\Lambda) \neq 0$ which means that the matrix Λ is always invertible. The almost sure convergence of the least squares estimator $\widehat{\vartheta}_n$ of the parameter ϑ associated with the ARX$(p+1, 2)$ process given by (2.5) is as follows.

Theorem A.1. *Assume that (V_n) has a finite conditional moment of order > 2. Then, $\widehat{\vartheta}_n$ converges almost surely to ϑ,*

$$\parallel \widehat{\vartheta}_n - \vartheta \parallel^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad \textit{a.s.} \tag{A.5}$$

Proof. We deduce from (2.7) and (2.8) that

$$\widehat{\vartheta}_n - \vartheta = S_{n-1}^{-1}\Big(M_n + \widehat{\vartheta}_0 - \vartheta\Big) \tag{A.6}$$

where

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

The sequence (M_n) is a locally square-integrable $(p+2)$-dimensional martingale with increasing process

$$<M>_n = \sigma^2 \sum_{k=0}^{n-1} \Phi_k \Phi_k^t.$$

Then, it follows from the strong law of large numbers for martingales given e.g. in Theorem 4.3.16 of [7] that

$$\parallel \widehat{\vartheta}_{n+1} - \vartheta \parallel^2 = \mathcal{O}\left(\frac{\log \lambda_{max}(S_n)}{\lambda_{min}(S_n)}\right) \qquad \text{a.s.} \tag{A.7}$$

Therefore, we clearly obtain (A.5) from (A.1) and (A.7). □

We immediately deduce from Theorem A.1 the almost sure convergence of the least squares estimators $\widehat{\theta}_n$ and $\widehat{\rho}_n$ to θ and ρ.

Corollary A.1. *Assume that (V_n) has a finite conditional moment of order > 2. Then, $\widehat{\theta}_n$ and $\widehat{\rho}_n$ both converge almost surely to θ and ρ,*

$$\parallel \widehat{\theta}_n - \theta \parallel^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad a.s. \tag{A.8}$$

$$(\widehat{\rho}_n - \rho)^2 = \mathcal{O}\left(\frac{\log n}{n}\right) \qquad a.s. \tag{A.9}$$

Proof of Theorem 3.1. The proof of Theorem 3.1 relies on Corollary A.1. It is left to the reader inasmuch as it follows essentially the same lines as those in Appendix C of [6]. □

Appendix B

PROOFS OF THE ASYMPTOTIC NORMALITY RESULTS

We shall now prove Theorem 3.2. First of all, we obtain from (3.3) that

$$\overline{\rho}_n = \frac{I_n}{J_{n-1}} \tag{B.1}$$

where

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

As in [6], we deduce from (A.6) and (B.1) the martingale decomposition

$$\sqrt{n}\begin{pmatrix} \widehat{\vartheta}_n - \vartheta \\ \overline{\rho}_n - \rho \end{pmatrix} = \frac{1}{\sqrt{n}}\mathcal{A}_n Z_n + \mathcal{B}_n \tag{B.2}$$

where (Z_n) is the locally square-integrable $(p+3)$-dimensional martingale given by

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

with

$$M_n = \sum_{k=1}^{n} \Phi_{k-1} V_k \qquad \text{and} \qquad N_n = \sum_{k=1}^{n} \varepsilon_{k-1} V_k.$$

In addition, it follows from Lemma A.1 that the sequences $(\mathcal{A}_n)$ and $(\mathcal{B}_n)$ converge almost surely to $\mathcal{A}$ and $\mathcal{B}$ given by

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

where 0_{p+2} stands for the null vector of $\mathbb{R}^{p+2}$ and Λ is the matrix given by (A.2). Moreover, the vector $\mathcal{C}$ belongs to $\mathbb{R}^{p+2}$ with

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

where $\mathrm{J}_p = (\mathrm{I}_p \;\; 0_p)$, T is the vector of $\mathbb{R}^p$ given by $T^t = (1, \rho, \ldots, \rho^{p-1})$ and ∇ is the rectangular matrix of size $(p+1) \times (p+2)$ given by

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

where, for all $1 \leq k \leq p-1$, ξ_k is the weighted sum

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

We already saw that (Z_n) is a martingale with predictable quadratic variation given, for all $n \geq 1$, by

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

Hence, we deduce once again from Lemma A.1 that

$$\lim_{n\to\infty} \frac{1}{n} \langle Z \rangle_n = \mathcal{Z} \qquad \text{a.s.}$$

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

$$\mathcal{Z} = \sigma^4 \begin{pmatrix} \sigma^{-2}\Lambda & \zeta \\ \zeta^t & \eta \end{pmatrix}$$

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

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

As (Z_n) satisfies the Lindeberg condition, we deduce from the central limit theorem for multidimensional martingales given e.g. by Corollary 2.1.10 in [7] that

$$\frac{1}{\sqrt{n}} Z_n \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \mathcal{Z}\Big)$$

which, via the martingale decomposition (B.2) and Slutsky's lemma, leads to

$$\sqrt{n} \begin{pmatrix} \widehat{\vartheta}_n - \vartheta \\ \overline{\rho}_n - \rho \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \mathcal{A}\mathcal{Z}\mathcal{A}'\Big). \tag{B.3}$$

Therefore, we immediately obtain from (B.3) that

$$\sqrt{n}(\overline{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \tau^2\right) \tag{B.4}$$

where the asymptotic variance τ^2 is given by $\tau^2 = (1-\rho^2)^2(\sigma^{-2}\mathcal{C}^t \Lambda \mathcal{C} + 2\mathcal{C}^t\zeta + \eta)$. It follows from tedious but straighforward calculations that τ^2 coincides with the expansion given by (3.8). Finally, as

$$\widehat{D}_n - D = -2(\overline{\rho}_n - \rho) + R_n \tag{B.5}$$

where the remainder R_n is negligeable which means that

$$R_n = o\left(\frac{1}{\sqrt{n}}\right) \qquad \text{a.s.}$$

we obtain (3.9) from (B.4) and (B.5), which achieves the proof of Theorem 3.2. □

Acknowledgements. The authors would like to thanks the anonymous reviewers for their constructive comments which helped to improve the paper substantially.

References

[1] B. D. O. Anderson, Exponential convergence and persistent excitation, 21th IEEE Conference on Decision and Control, 1982.
[2] K. J. Aström and B. Wittenmark, Adaptive Control, 2nd edition, Addison-Wesley, New York, 1995.
[3] B. Bercu and V. Vázquez, A new concept of strong controllability via the Schur complement for ARX models in adaptive tracking, Automatica, Vol. 46, pp. 1799-1805, 2010.
[4] B. Bercu and V. Vázquez, On the usefulness of persistent excitation in ARX adaptive tracking, International Journal of Control, Vol. 83, pp. 1145-1154, 2010.
[5] B. Bercu and F. Proia, A sharp analysis on the asymptotic behavior of the Durbin-Watson for the first-order autoregressive process, ESAIM PS, Vol. 17, pp. 500-530, 2013.
[6] B. Bercu, B. Portier and V. Vázquez, On the asymptotic behavior of the Durbin-Watson statistic for ARX processes in adaptive tracking, To appear in International Journal of Adaptive Control and Signal Processing, DOI: 10.1002/acs.2424, Vol. 28, 2014.
[7] M. Duflo, Random Iterative Models, Springer Verlag, Berlin, 1997.
[8] J. Durbin and G. S. Watson, Testing for serial correlation in Least Squares regression I, Biometrika, Vol. 37, pp. 409-428, 1950.
[9] J. Durbin and G. S. Watson, Testing for serial correlation in Least Squares regression II, Biometrika, Vol. 38, pp. 159-178, 1951.
[10] J. Durbin and G. S. Watson, Testing for serial correlation in Least Squares regression III, Biometrika, Vol. 58, pp. 1-19, 1971.
[11] J. Durbin, Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, Econometrica, Vol. 38, pp. 410-421, 1970.
[12] L. Guo and H. F. Chen, The Aström Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, IEEE Trans. Automat. Control, Vol. 36, pp. 802-812, 1991.
[13] J. B. Moore, Persistency of excitation in extended least squares, IEEE Trans. Automat. Control, Vol. 28, pp. 60-68, 1983.
[14] M. Nerlove and K. F. Wallis, Use of the Durbin Watson statistic in inappropriate situations, Econometrica, Vol. 34, pp. 235-238. 1966.
[15] F. Proïa, Further results on the h-test of Durbin for stable autoregressive processes, Journal of Multivariate Analysis, Vol. 118, pp. 77-101, 2013.
[16] M.S. Srivastava. Asymptotic distribution of Durbin Watson statistic, Economics Letters, Vol. 24, pp. 157-160, 1987.
[17] T. Stocker, On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors, Statist. Papers, Vol. 48, pp. 81-93, 2007.

Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR 5251, 351 cours de la libération, 33405 Talence cedex, France.

Normandie Université, Département de Génie Mathématiques, Laboratoire de Mathématiques, INSA de Rouen, LMI-EA 3226, place Emile Blondel, BP 08, 76131 Mont-Saint-Aignan cedex, France

Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, Avenida San Claudio y Rio Verde, 72570 Puebla, Mexico.