Identifying trend nature in time series using autocorrelation functions and R-routines based on stationarity tests

M Boutahar, M Royer-Carenzi

Identifying trend nature in time series using autocorrelation functions and R-routines based on stationarity tests

M. Boutahar[a] and M. Royer-Carenzi[a] and

[a]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, UMR 7373, Marseille, France

ARTICLE HISTORY
Compiled July 27, 2021

Abstract
Time series non-stationarity can be detected thanks to autocorrelation functions. But trend nature, either deterministic or either stochastic, is not identifiable.

Strategies based on Dickey-Fuller unit root-test are appropriate to choose between a linear deterministic trend or a stochastic trend. But all the observed deterministic trends are not linear, and such strategies fail in detecting a quadratic deterministic trend. Being a confounding factor, a quadratic deterministic trend makes appear a unit root spuriously.

We provide a new procedure, based on Ouliaris-Park-Phillips unit root test, convenient for time series containing polynomial trends with degree higher than one. Our approach is assessed on simulated data.

The strategy is finally applied on two real datasets : number of terminated pregnancies in Québec, Canada, and atmospheric CO2 concentration. Compared with Dickey-Fuller diagnosis, our strategy provides the model with the best performances.

Keywords: time series; stationarity; unit root test; Dickey-Fuller; quadratic trend.

1. Introduction

Time series non-stationarity can originate form various sources: either from a trend component or from a seasonal or even a cyclical component. In this paper, we will be interested in the non-stationarity caused by a trend. There are two kinds of trends: either a deterministic trend which can be modeled by some function of time (polynomial trend is generally considered), or a stochastic trend which presents unit roots. Deterministic and stochastic trends are two specific models suggested by Nelson and Plosser in [17].

$$\text{Deterministic trend (}\mathbf{DetS,d}\text{)} \quad Z_t = a_0 + a_1\, t + \cdots + a_d\, t^d + B_t \tag{1}$$

$$\text{Stochastic trend (}\mathbf{StoS,d}\text{)} \quad \Delta^d(Z_t) = B_t\,, \tag{2}$$

where we take $a_d \neq 0$, Δ is the differencing operator and $(B_t)_t$ is a moving average process

$$B_t = \sum_{j\in\mathbb{Z}} b_j \mathcal{E}_{t-j}, \tag{H1}$$

where $(\mathcal{E}_t)_t$ is a sequence of identically distributed and independent centered variables, such that

$$E(\mathcal{E}_t^{2k}) < \infty, \text{ for some } k \geq 2, \tag{H2}$$

CONTACT M. Boutahar. Email: mohamed.boutahar@univ-amu.fr

and where parameters b_j ($j \in \mathbb{Z}$) satisfy

$$\sum_{j\in\mathbb{Z}} |b_j| < \infty \qquad \sum_{j\in\mathbb{Z}} b_j \neq 0 \qquad \sum_{j\in\mathbb{Z}} b_j^2 |j| < \infty\,. \tag{H3}$$

(H3-a) (H3-b) (H3-c)

We denote $(B_t)_t$ as (**SN**), for *Stationary Noise*. It's well known that causal and invertible ARMA pocesses satisfy Hypotheses (H1) and (H3). When $B_t = \mathcal{E}_t$, it is called (**WN**) for *white noise*. In this case, the associated models defined in Equations 1 and 2, are referred as (**DetW,d**) and (**StoW,d**).

When modeling time series, specially for macroeconomic and financial data, it is very important to identify the nature of the trend: deterministic or stochastic. Indeed, every type of trend induces specific behaviors, that we can illustrate with moments properties. Let us consider a (**DetS,1**) model associated to a white noise, then $\mathbb{E}(Z_t) = a_0 + a_1 t$, and $\mathrm{var}(Z_t) = \sigma_B^2$, providing a stationary variance but a non-stationary mean. On the contrary, under a (**StoS,1**) model with $Z_0 = 0$, we have $\mathbb{E}(Z_t) = 0$, and $\mathrm{var}(Z_t) = t\sigma_B^2$, providing the opposite feature. Thus the source of non-stationarity differs with trend nature. Consequently identifying the correct trend is fundamental. Globally, time series with a deterministic trend always revert to the trend in the long run (the effects of shocks are eventually eliminated); and the forecast confident intervals have constant width. On the contrary, time series with a stochastic trend never recover from shocks to the mean (the effects of shocks are permanent); and the forecast confident intervals grow with the horizon. Several authors (see for instance [4] and [16]) studied the consequences of an inappropriate modeling choice, underlying the importance of developing procedures able to produce a reliable classification.

Autocorrelation function properties have been widely studied, and are helpful to specify accurate models for stationary time series. The theoretical autocorrelation function at lag h ($|h| < n$) is estimated from data $(Z_1, \cdots, Z_n)$ with the random variable

$$\Xi(h) \;=\; \frac{\sum_{t=1}^{n-h} (Z_{t+h} - \overline{Z})(Z_t - \overline{Z})}{\sum_{t=1}^{n} (Z_t - \overline{Z})^2}\,, \tag{3}$$

where

$$\overline{Z} \;=\; \frac{\sum_{t=1}^{n} Z_t}{n}\,,$$

is the random mean. We show that autocorrelation functions also have interesting properties in our framework, since they permit to identify the presence of a trend in a time series. Nevertheless it can not be used to distinguish between either a deterministic or a stochastic trend.

Dickey and Fuller developed a unit root test that is an essential tool in time series modeling [9]. The authors considered the following models

$$\begin{aligned}
\mathbf{M_1}: \quad Z_t &= \rho\, Z_{t-1} + \mathcal{E}_t \\
\mathbf{M_2}: \quad Z_t &= a_0 + \rho\, Z_{t-1} + \mathcal{E}_t \\
\mathbf{M_3}: \quad Z_t &= a_0 + a_1\, t + \rho\, Z_{t-1} + \mathcal{E}_t\,,
\end{aligned}$$

where $(\mathcal{E}_t)_t$ is a white noise. Dickey-Fuller procedure permits to test the null hypothesis of a unit

root ($\rho = 1$) against the alternative hypothesis of a stationary AR(1) model without drift (resp. with drift, resp. with trend), accordingly to model $\mathbf{M_1}$ (resp. $\mathbf{M_2}$, resp. $\mathbf{M_3}$). Note that a time series with a (linear) deterministic trend, as defined in (**DetW,1**), is included in $\mathbf{M_3}$ model, by taking $\rho = 0$. Consequently, under a (linear) deterministic trend, Dickey-Fuller unit-root test, led in $\mathbf{M_3}$ framework, usually rejects the null hypothesis. On the other hand, (**StoW,1**)-time series are included in $\mathbf{M_1}$ model, with $\rho = 1$, so that ($\underline{H_0}$), tested in $\mathbf{M_1}$ framework, is usually not rejected, accordingly with the rejection rate.

Several strategies based on Dickey-Fuller unit root-test have been developed ([8, 20, 10]), and they mostly appear to be appropriate to choose between a linear deterministic trend or a stochastic trend. But all the observed deterministic trends are not linear, trends with higher degrees can be involved. In this case, Ertur noticed in [11] that the usual strategies fail in detecting a quadratic deterministic trend. Indeed, under a quadratic trend for instance, Dickey-Fuller test generally concludes to the presence of a (spurious) unit root, even using model $\mathbf{M_3}$ that allows a linear trend.

In this paper, we aim to include polynomial trends with a degree higher than one. Let us introduce the general model :

$$\mathbf{M_{3,d}}: \quad Z_t \;=\; a_0 \,+\, a_1\, t \,+\, \cdots \,+\, a_d\, t^d \,+\, \rho\, Z_{t-1} \,+\, B_t \quad (a_d \neq 0)\,.$$

In [19], Ouliaris, Park and Phillips developed a test that corrects the bias caused by high degree trend when testing for a unit root. We included this test in a strategy, that correctly identifies either a deterministic trend or a stochastic one.

In Section 2, we detail autocorrelation functions convergence when time series are driven by a trend, either deterministic or stochastic. This result permits to detect the presence of a trend, but without precise identification. In Section 3, we study existing strategies based on Dickey-Fuller tests, and analyze their performance in classifying models (**DetW,1**), (**DetW,2**), (**StoW,1**) and (**StoW,2**). Note that we also consider the simplest model (**WN**) as a null-model. Next, we provide a new strategy, based on Trend Diagnosis Tests (TDT), able to identify trends type, even when (**WN**) is replaced by a (**SN**) underlying process, and when higher-degrees d are considered. Finally, we apply our strategy on real data such as abortion rate in Montreal, Québec, and also on the CO2 atmospheric concentration. All the functions are implemented in **R** language, and they are available at the website:
www.i2m.univ-amu.fr/perso/manuela.royer-carenzi/AnnexesR.TrendTS/TrendTS.html

2. Trend detection

It is essential to start by plotting the graph of the series in order to visualize its evolution, but the presence or absence of a trend is sometimes difficult to detect from the plot. Autocorrelation appears to be a powerful tool in the detection of a trend insofar as its behavior is specific in the presence of a trend. We recall that theoretical autocorrelations $\rho(h) = \mathrm{cor}(Z_t, Z_{t+h})$ are well-defined only if process $(Z_t)_t$ is square-integrable and stationary. However the associated random variables $\Xi(h)$, as defined in Equation 3, can be computed for any observed time series $(Z_1, \cdots, Z_n)$.

2.1. Autocorrelation functions behavior for (WN)-time series

Theorem 2.1.
Let $(Z_t)_t$ be a white noise. Then

$$\sqrt{n}\ {}^t(\Xi(1), \Xi(2), \cdots, \Xi(r)) \quad \xrightarrow[n\to+\infty]{\mathcal{L}} \quad \mathcal{N}_r({}^t(0, \cdots, 0),\, Id_r)\,,$$

where ${}^t v$ denotes the transpose of vector v, and Id_r is the identity $r \times r$-matrix.

Theorem 2.1 is a particular case of Theorems 7.2.1. or 7.2.2. in [3], that require Hypotheses (H1), (H3-a) and either Hypothesis (H2) or (H3-c). Thus, we get that the random variables $\Xi(1), \Xi(2), \cdots, \cdots, \Xi(r)$ are asymptotically independent and identically distributed as Gaussian random variables with zero mean and variance $1/n$. Consequently, for any fixed lag $h = 1, \cdots, r$, for large n, the sample autocorrelation function $\sqrt{n}\ \hat{\rho}(h)$ is expected to be a realization of a standard Gaussian, that is to be valued in the interval $[-1.96, 1.96]$, with 95% coverage. Thus, sample autocorrelation functions are used to assess for white noise. But even when the underlying process is a white noise, several autocorrelations among $\hat{\rho}(1), \cdots, \hat{\rho}(r)$ may lie out of the interval $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$. The asymptotic independence property for variables $\Xi(h)$ implies that, when sample size n is large, the number of observed autocorrelation functions out of this interval behaves as a Binomial $\mathcal{B}(r, 0.05)$ distribution. We developed an R function, based on the classical `acf()` function, that takes into account the multiple testing paradigm by incorporating the binomial exact test and Sidak correction for white noise diagnosis. Our function is called `acfG()` and is available at our website.

As an example, we simulate a white noise, and compute sample autocorrelation functions with function `acfG()`. Figure 1 shows that 3 values lie out of the interval $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$, plotted with blue dashed lines. But binomial exact test (p-value = 0.1159) confirms that such a number remains consistent with white noise hypothesis. Moreover, a second set of interval, computed with Sidak correction, is provided, and plotted with red dotted lines. If at least one sample autocorrelation function lies out this global interval, then white noise hypothesis is rejected. In our simulation, white noise diagnosis is confirmed, both by graphics and by Binomial's test.

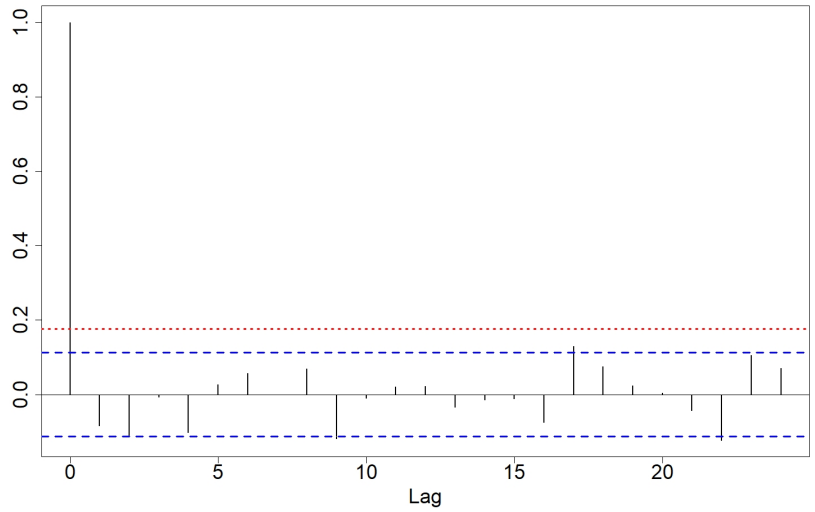

Figure 1.: Sample autocorrelation functions for a white noise simulation.

2.2. Autocorrelation functions for (DetS,d)-time series

Theorem 2.2.
Let $(Z_t)_t$ be a stochastic process such that $Z_t = \sum_{j=0}^{d} a_j t^j + B_t$, where $a_d \neq 0$ and $(B_t)_t$ is

*(**SN**) satisfying Hypotheses (H1) to (H3). Then*

$$\Xi(h) \xrightarrow[n\to+\infty]{\mathbb{P}} 1\,, \quad \forall\, h \neq 0\,.$$

Proof is given in Appendix A. Figure 2 (Left) illustrates the slow decreasing behavior of sample autocorrelation functions when time series are driven by a deterministic (**DetS,2**) trend.

2.3. Autocorrelation functions for (StoW,d)-time series

Theorem 2.3.
*Let $(Z_t)_t$ be a stochastic process such that $\Delta^d(Z_t) = \mathcal{E}_t$, with $Z_t = 0$, for any $t \leq 0$ and and $(B_t)_t$ is (**SN**) satisfying Hypotheses (H1) to (H3). Then*

$$\Xi(h) \xrightarrow[n\to+\infty]{\mathbb{P}} 1\,, \quad \forall\, h \neq 0\,.$$

Proof is given in Appendix B. Figure 2 (Right) illustrates the slow decreasing behavior of sample autocorrelation functions when time series are driven by a stochastic (**StoW,2**) trend.

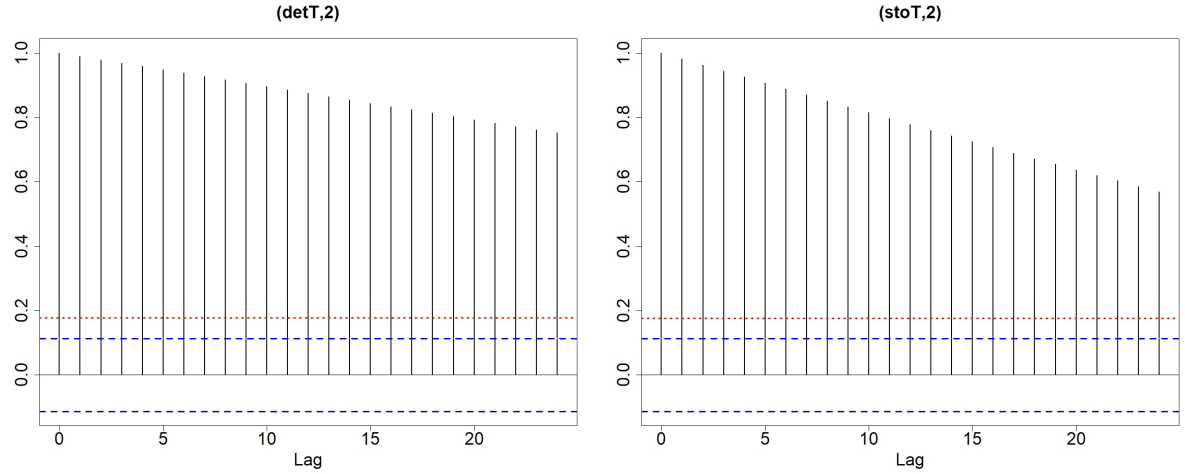

Figure 2.: Sample autocorrelation functions for simulations with either a deterministic (Left) or a stochastic trend (Right).

3. Trend-nature identification

In the previous section, we showed that the autocorrelation functions, computed from variables $Z_1, \cdots, Z_n$, have a particular asymptotic behavior in presence of a trend. But Figure 2 illustrates that the behavior is similar either for a deterministic or a stochastic trend. Consequently, a deeper study has to be led in order to specify the type of trend. We explore Dickey-Fuller unit-root testing diagnosis.

3.1. Dickey-Fuller tests failure

In [9], Dickey and Fuller introduced unit root tests adapted to models $\mathbf{M_1}, \mathbf{M_2}$ and $\mathbf{M_3}$. Every time the null and the alternative hypothesis are mathematically expressed in the same way:

$$(\underline{H_0}) : \rho = 1 \quad \text{against } (\underline{H_1}) : |\rho| < 1.$$

But the alternative hypothesis interpretation depends on the considered model.

$$\begin{array}{lll} (\mathbf{M_1}) & (\underline{H_1}) : & (Z_t)_t \textit{ is a stationary and centered AR(1) process;} \\ (\mathbf{M_2}) & (\underline{H_1}) : & (Z_t)_t \textit{ is a stationary not-centered AR(1) process;} \\ (\mathbf{M_3}) & (\underline{H_1}) : & (Z_t)_t \textit{ is (linear-)trend stationary (TS).} \end{array}$$

And test statistics do not have the same expression, possibly leading to opposing conclusions, even on the same data. To distinguish between the null or the alternative hypothesis, one has to use the suitable test statistics, adapted to every model.

In [9], Dickey and Fuller also developed joined tests :

$$\begin{array}{llll} \text{Test } \Phi_1 \text{ under } (\mathbf{M_2}) & (\underline{H_0}) : & (a_0, \rho) & = (0, 1)\,; \\ \text{Test } \Phi_2 \text{ under } (\mathbf{M_3}) & (\underline{H_0}) : & (a_0, a_1, \rho) & = (0, 0, 1)\,; \\ \text{Test } \Phi_3 \text{ under } (\mathbf{M_3}) & (\underline{H_0}) : & (a_0, a_1, \rho) & = (a_0, 0, 1)\,. \end{array}$$

3.1.1. Simulation study

We perform some simulations to explore the behavior of Dickey-Fuller based-tests, according to data generating process. We set $(\mathcal{E}_t)_t$ as a white noise with variance $\sigma^2_{\mathcal{E}}$. We consider processes $(Z_t)_t$ successively driven by one of the following models :

(**WN**)	$Z_t = a_0 + \mathcal{E}_t$	called "white noise with constant",
(**DetW,1**)	$Z_t = a_0 + a_1\, t + \mathcal{E}_t$	called "linear-trend stationary",
(**DetW,2**)	$Z_t = a_0 + a_1\, t + a_2\, t^2 + \mathcal{E}_t$	called "quadratic-trend stationary",
(**StoW,1**)	$\Delta(Z_t) = \mathcal{E}_t$	called "difference stationary",
(**StoW,2**)	$\Delta^2(Z_t) = \mathcal{E}_t$	called "second-order difference stationary".

Models (**StoW,1**) and (**StoW,2**) do contain a unit root and hence are not stationary, whereas models (**DetW,1**) and (**DetW,2**) are not stationary although they do not contain any unit root. Logically, Dickey-Fuller test should not reject the null hypothesis for almost realizations driven from models (**StoW,1**) and (**StoW,2**), precisely with a rate $(1-\alpha)\%$, where α stands for the significance level. Reciprocally, under (**DetW,1**) and (**DetW,2**) models, the null hypothesis should be rejected.

For simulations, we set $n = 300, a_0 = 5, a_1 = 1, a_2 = 1$. Random generations of $\mathcal{E}_t$ were taken from Gaussian centered variables with standard deviation $\sigma_{\mathcal{E}}$, a fixed value among $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. We ran 5 000 simulations from every model, and successively applied Dickey-Fuller tests, dedicated to models $\mathbf{M_1}, \mathbf{M_2}$ and $\mathbf{M_3}$. Table 1 presents the results when $\sigma_{\mathcal{E}} = 10$, and contains the rate of null-hypothesis rejecting, crossing every data generating process (DGP) with every stationarity test, under a standard level $\alpha = 5\%$.

It appears that Dickey-Fuller diagnosis is accurate for (**WN**) and (**StoW,1**) models. As expected, only the convenient model ($\mathbf{M_3}$), provides a correct answer for (**DetW,1**). And in Figure C1 from Appendix C, we illustrate that diagnosis is unclear for (**DetW,1**) realizations

Table 1.: Null-hypothesis rejection rate (%) in Dickey-Fuller tests, when $\sigma_{\mathcal{E}} = 10$.

	DGP[a]				
Test	**(WN)**	**(DetW,1)**	**(DetW,2)**	**(StoW,1)**	**(StoW,2)**
Test ρ under ($\mathbf{M_1}$)	100	0	0	5.06	0
Test ρ under ($\mathbf{M_2}$)	100	0	0	4.96	12.88
Test ρ under ($\mathbf{M_3}$)	100	100	0	5.04	32.4
Test Φ_1 under ($\mathbf{M_2}$)	100	0	100	4.86	95.46
Test Φ_2 under ($\mathbf{M_3}$)	100	100	100	4.92	99.08
Test Φ_3 under ($\mathbf{M_3}$)	100	100	100	5.1	90.4

[a]Data Generating Process.

when using any test under models ($\mathbf{M_1}$) or ($\mathbf{M_2}$). Indeed, results vary greatly according to parameter $\sigma_{\mathcal{E}}$, taking value in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. Surprisingly, unit root is far to be correctly detected in (**StoW,2**) realizations. Finally, diagnosis is mainly incorrect for model (**DetW,2**), since unit roots are systematically detected, whereas we have $\rho = 0$. Thus, Dickey-Fuller-based tests fail in diagnosing unit root for quadratic trends models.

3.1.2. Dickey-Fuller-based strategies

Several strategies based on Dickey-Fuller tests have been developed ([8, 20, 10]). From Table 1, we can deduce that all the strategies permit to discriminate between first-order deterministic or stochastic trend, but they do not plan to integrate second-order trends. Thus, the most simple strategy proposed in [8] falsely classifies (**DetW,2**) model as a random walk, whereas Perron predicts a linear-trend stationary process [20]. The most advanced strategy given in [10] nearly identifies (**DetW,2**) processes, by describing them as $\Delta(Z_t) = \beta_0 + \beta_1 t + \mathcal{E}_t$, instead of $\Delta(Z_t) = \beta_0 + \beta_1 t + \Delta(\mathcal{E}_t)$, where $\Delta(Z_t) = Z_t - Z_{t-1}$ is the differentiated series. But diagnosis is mainly incorrect for (**StoW,2**) processes. Surprisingly, only the most simple strategy suggested in [8] detects unit root in half realizations, and predicts either a linear-trend stationary process or a stationary AR(1) process otherwise. And other strategies predict linear-trend stationary process. Consequently, it appears useful to elaborate a new strategy to identify not only one-degree trends, but also higher-degree ones.

3.2. Accurate behavior for OPP and KPSS tests

3.2.1. Additional stationarity tests

3.2.1.1. OPP stationarity test.

In [19], Ouliaris, Park, and Phillips generalized Dickey-Fuller unit-root tests ρ to models $\mathbf{M_{3,d}}$ with polynomial trends, where $d = 2, 3, 4$ or 5. We denote this general test by OPP. Note that the invariance principle for partial sums, required in OPP test, applies to (**SN**) stationary linear processes satisfying Hypotheses (H1) to (H3). We implemented this test as a R function, denoted by `opp.test()`, estimating the long run variance with Newey West kernel (introduced in [18]).

3.2.1.2. KPSS stationarity test.

In [15], Kwiatkowski, Phillips, Schmidt and Shin developed another type of stationarity test, associated to an underlying (**SN**). Contrary to Dickey-Fuller tests, KPSS test takes the presence of unit root as the alternative hypothesis, and the stationarity as the null hypothesis. Actually, KPSS test can consider as null-hypothesis either level-stationarity or trend-stationarity

Table 2.: Null-hypothesis rejection rate (%) for KPSS and OPP stationarity tests, for DGP simulations when the underlying process is a white noise $(\mathcal{E}_t)_t$. We vary $\sigma_{\mathcal{E}}$ on the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. The final rejection rate is computed as the average of the rejection rates obtained for each $\sigma_{\mathcal{E}}$.

	DGP[a]				
Test	(**WN**)	(**DetW,1**)	(**DetW,2**)	(**StoW,1**)	(**StoW,2**)
OPP for Z_t	100	100	100	5.96	0
OPP for $\Delta(Z_t)$	100	100	100	100	5.86
KPSS for Z_t	4.8	98.6	100	98.9	100
KPSS for $\Delta(Z_t)$	0	0	100	4.8	98.9

[a]Data Generating Process.

(stationarity around a linear deterministic trend). Here, we refer to the level-stationarity test.

3.2.2. Simulation study for OPP and KPSS tests

Models (**StoS,1**) and (**StoS,2**) do contain a unit root and hence are not stationary, whereas models (**DetS,1**) and (**DetS,2**) are not stationary although they do not contain any unit root. Logically, KPSS's test should reject the null hypothesis for most realizations driven from all these models. Hence KPSS's test does not appear to be a good candidate for trend-nature identification. Nonetheless, KPSS's test reveals heterogeneous behaviors when applied to the differentiated series $\Delta(Z_t)$. Indeed, processes (Z_t) driven from models (**DetS,1**) and (**StoS,1**) do become stationary as soon as they are differentiated.

$$(\mathbf{SN}) \quad Z_t = B_t \implies \Delta(Z_t) = B_t - B_{t-1}, \tag{4}$$

$$(\mathbf{DetS,1}) \quad Z_t = a_0 + a_1 t + B_t \implies \Delta(Z_t) = a_1 + B_t - B_{t-1}, \tag{5}$$

$$(\mathbf{DetS,2}) \quad Z_t = a_0 + a_1 t + a_2 t^2 + B_t \implies \Delta(Z_t) = a_1 + 2a_2 t + B_t - B_{t-1}, \tag{6}$$

$$(\mathbf{StoS,1}) \quad Z_t = Z_{t-1} + B_t \implies \Delta(Z_t) = B_t, \tag{7}$$

$$(\mathbf{StoS,2}) \quad Z_t = 2Z_{t-1} - Z_{t-2} + B_t \implies \Delta(Z_t) = \Delta(Z_{t-1}) + B_t. \tag{8}$$

Moreover OPP's test should reject the null hypothesis only for realizations driven from models (**DetS,1**) and (**DetS,2**), vice versa for (**StoS,1**) and (**StoS,2**). Applied to the differentiated series, OPP's test should not reject the null hypothesis only for realizations initially driven from models (**StoS,2**).

We ran the suggested testing procedure on 5 000 simulations using every data generating process among (**SN**), (**DetS,1**), (**DetS,2**), (**StoS,1**) and (**StoS,2**), when the associated stationary noise is either a (**WN**), or (**SN**) such as a MA(2) or an ARMA(1,1) centered, causal and invertible process. We took again parameters values as $a_0 = 5, a_1 = 1, a_2 = 1, n = 300$. Table 2 presents the null-hypothesis rejection rate when $B_t = \mathcal{E}_t$ where $\sigma_{\mathcal{E}}$ varies in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$.

Table 2 shows that KPSS and OPP tests perform as expected, not only for simulations driven from one-order trends, but also for quadratic trends. Moreover, results remain identical for any value of parameter $\sigma_{\mathcal{E}}$, varying from 0.5 to 500, such as shown in Figure D1, Appendix D. We also present results for simulations with underlying stationary noises $(B_t)_t$, such as a MA(2) and an ARMA(1,1). In this case, both Supplementary Figures S1 and S2 show the same behavior for KPSS and OPP tests, whatever $\sigma_{\mathcal{E}}$ value.

4. Applications

4.1. A new strategy for trend identification

4.1.1. TDT strategy

We suggest to apply the following tests successively :

i) OPP test to series Z_t ;
ii) OPP test to series $\Delta(Z_t)$;
iii) KPSS test to series Z_t ;
iv) KPSS test to serie $\Delta(Z_t)$.

We call Trend Diagnosis Tests (TDT) the set of responses to tests **i)** to **iv)** computed on a time series. Let us denote by **Null**, the case where the null hypothesis can not be rejected, and by **Alt** otherwise. From Equations 4 to 8, we provide the expected diagnosis led by TDT strategy for every model. Hence,

(**SN**)	expected diagnosis :	**Alt/Alt/Null/Null** ,
(**DetS,1**)	expected diagnosis :	**Alt/Alt/Alt/Null** ,
(**DetS,2**)	expected diagnosis :	**Alt/Alt/Alt/Alt** ,
(**StoS,1**)	expected diagnosis :	**Null/Alt/Alt/Null** ,
(**StoS,2**)	expected diagnosis :	**Null/Null/Alt/Alt** .

We ran again 5 000 simulations under every data generating process among (**WN**), (**DetW,1**), (**DetW,2**), (**StoW,1**) and (**StoW,2**), with $a_0 = 5, a_1 = 1, a_2 = 1, n = 300$ and $\sigma_\mathcal{E}$, taking value in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. Then we applied successively tests **i)** to **iv)** to any simulation, using a risk $\alpha = 5\%$. In Table 3, every column sums the diagnosis percentage associated to the corresponding DGP. The percentage in bold refers to the expected diagnosis. For example, if the DGP is (**WN**), only two sets of responses were obtained to TDT : **Alt/Alt/Null/Null** for 95.247% of (**WN**) simulations (percentage written in bold), and **Alt/Alt/Alt/Null** for the other 4.753% simulations. Note that the two TDT diagnosis associated to (**WN**) only differ due to KPSS response on the initial series. The two obtained percentages are consistent with the Type I error for this test, when α is fixed to 5%. In the same way, if the DGP is (**StoW,1**), several responses are possible. Note that among tests **i)** to **iv)**, only the first and the last one are led under the Null hypothesis. Consequently, when $\alpha = 5\%$, each of tests **i)** to **iv)** may approximately produce the **Null** response in 95% simulations, and the **Alt** response in the other 5%. This theoretically yields to percentages

$95\% \times 95\%$	=	90.25%	(**StoW,1**)-simulations with response	**Null/ - / - /Null**
$95\% \times 5\%$	=	4.75%	(**StoW,1**)-simulations with response	**Null/ - / - /Alt**
$95\% \times 5\%$	=	4.75%	(**StoW,1**)-simulations with response	**Alt/ - / - /Null**
$5\% \times 5\%$	=	0.25%	(**StoW,1**)-simulations with response	**Alt/ - / - /Alt**

We do observe almost these expected percentages by summing the corresponding TDT percentages in Table 3, column (**StoW,1**). For instance, in relation to the expected diagnosis **Null/ - / - /Null**, we observed either **Null/Alt/Alt/Null** or **Null/Alt/Null/Null** diagnosis for (**StoW,1**)-simulations, with respective percentages 86.162% and 3.216%, summing to 89.378%, that is very close to the expected 90% percentage.

Table 3.: Percentage of Trend Diagnosis Tests (TDT) associated to every Data Generating Process (DGP)when $\sigma_{\mathcal{E}}$ takes values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$.

	DGP[a]				
TDT[b]	(**WN**)	(**DetW,1**)	(**DetW,2**)	(**StoW,1**)	(**StoW,2**)
Alt/Alt/Alt/Alt	0	0	**100**[c]	0.438	0
Alt/Alt/Null/Alt	0	0	0	0	0
Alt/Null/Alt/Alt	0	0	0	0	0
Alt/Null/Null/Alt	0	0	0	0	0
Null/Alt/Alt/Alt	0	0	0	4.273	6.213
Null/Alt/Null/Alt	0	0	0	0.005	0
Null/Null/Alt/Alt	0	0	0	0	**90.427**
Null/Null/Null/Alt	0	0	0	0	0.007
Alt/Alt/Alt/Null	4.753	**99.993**	0	5.822	0
Alt/Alt/Null/Null	**95.247**	0.007	0	0.084	0
Alt/Null/Alt/Null	0	0	0	0	0
Alt/Null/Null/Null	0	0	0	0	0
Null/Alt/Alt/Null	0	0	0	**86.162**	0.089
Null/Alt/Null/Null	0	0	0	3.216	0
Null/Null/Alt/Null	0	0	0	0	3.220
Null/Null/Null/Null	0	0	0	0	0.015
Total percentage	100	100	100	100	100

[a]Data Generating Process
[b]Trend Diagnosis Tests
[c]Bold font highlights the expected TDT diagnosis associated to every DGP.

But the interest of Table 3 lies is in its reverse reading. Let us associate a DGP to a TDT diagnosis. As an example, since none simulation under (**WN**), (**DetW,1**), (**DetW,2**) or (**StoW,1**) DGP led to responses **Null/Null/Alt/Alt**, then if one obtains such a response on its time series, this means that a (**StoW,2**) model is suitable. Alternately **Alt/Alt/Alt/Null** diagnosis could lead either to a (**WN**), a (**DetW,1**) or a (**StoW,1**) model. But referring to occurrence percentages, $99.993/(4.753 + 99.993 + 5.822) = 90.436\%$ of simulations with TDT **Alt/Alt/Alt/Null** are produced by a (**DetW,1**) DGP, only 4.299% by a (**WN**) and 5.266% by a (**StoW,1**). Then a (**DetW,1**) model appears as the best candidate, but (**WN**) and (**StoW,1**) models can not be totally excluded. As shown in Figure E1, Appendix E, the relative occurrence of these three models for diagnosis **Alt/Alt/Alt/Null** remains the same, whatever $\sigma_{\mathcal{E}}$ value in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. In this precise case, even if a linear trend is present, the relevance of either a (**WN**) or a (**DetW,1**) model depends on the intensity of the trend in relation to the variance of the associated noise. The convenient choice can be ruled out by previously computing autocorrelation functions, with `acfG()` function. If several models remain acceptable, then we suggest to construct and compare them.

In order to evaluate the effect of autocorrelation on TDT strategy, we also run simulations with autocorrelated noises. In other words, we replaced $\mathcal{E}_t$ either by a MA(2) process or by an ARMA(1,1) process, denoted as B_t, see Supplementary. Percentage of TDT diagnosis associated to every Data Generating Process, driven by (**SN**), are given in Supplementary Tables S1 and S2. Percentage diagnosis remains similar regardless of the underlying model for B_t. From Table 3 and Supplementary Tables S1 and S2, we suggest to associate each model with some sets of responses to TDT. More specifically, with diagnoses ranked by risk of occurrence :

- **Alt/Alt/Null/Null** is associated with (**SN**),
- **Alt/Alt/Alt/Null** is associated with (**DetS,1**),
- **Alt/Alt/Alt/Alt** is associated with (**DetS,2**),
- **Null/Alt/Alt/Null**
Null/Alt/Null/Null
Null/Alt/Null/Alt are associated with (**StoS,1**),
- **Null/Null/Alt/Alt**
Null/Null/Alt/Null
Null/Null/Null/Alt
Null/Null/Null/Null are associated with (**StoS,2**),
- **Null/Alt/Alt/Alt** is associated either with (**StoS,1**) or (**StoS,2**).

Figure E1 and Supplementary Figures S3 and S4 illustrate that this classification remains stable as $\sigma_{\mathcal{E}}$ varies, whenever the underlying noise is (**WN**) or (**SN**). We implemented this diagnosis strategy as the R function `trend.diag.tests()`, with an argument `nb.mod` taking values among `c("single", "multiple")`, specifying either if only the main model is returned or if all possibilities are returned, even the most occasional one. In this case, all the suggested models can be constructed and compared.

4.1.2. Higher-degree trends

Actually, it is possible to detect higher-degree trends, either deterministic (**DetS,d**) or stochastic (**StoS,d**), with $d = 3, 4, 5$, by iterating OPP and KPPS tests on the successive differentiated series. More precisely,

⋆ **Step 0 :**
Compute sample autocorrelation functions with `acfG()` in order to distinguish between a white noise and a time series with a trend. If the series is driven by a trend, then run the following steps.

⋆ **Step 1 :**
Run OPP test on the given time series.
If the null is rejected, we identify a (**DetS,d**) model, otherwise (**StoS,d**), with $d \geq 1$. It remains to precise d.

⋆ **Step 2 :**
Case 2a :
If a (**StoS,d**) model is detected in Step 1, differentiate the current time series, and apply OPP test. Iterate this step until the null is rejected. Then d corresponds to the number of necessary differentiations.
Case 2b :
If a (**DetS,d**) model is detected in Step 1, differentiate the current time series, and apply KPSS test. Iterate this step until the null is rejected. Then d corresponds to the number of necessary differentiations.

We implemented this diagnosis strategy as the R function `trend.diag.high()`. When applying this strategy, a model is suggested, leading either to Equation 1 or 2, with parameters to be determined. In particular, process $(B_t)_t$ is rarely a white noise, and should rather be modeled by an ARMA(p,q) process. The validity of the global model has to be confirmed with residuals diagnosis.

Table 4.: p-values provided by several tests on the initial and the differentiated abortion series.

	Test				
Series	**OPP**	**KPSS**	**ρ under (M_1)**	**ρ under (M_2)**	**ρ under (M_3)**
Z_t	0.2	0.01	0.9784	0.011	0.99
$\Delta(Z_t)$	0.0852	0.02	0.01	0.01	0.01

4.2. Application on data

4.2.1. Terminated pregnancies in Québec, Canada

Since 1969, abortion is legal in Canada at all stages of pregnancy. Since 1971, *la Régie de l'assurance maladie du Québec* provides the number of voluntarily terminated pregnancies in Québec, Canada, plotted in Figure 3 (Left). Autocorrelation functions, given in Figure 3 (Right), suggest that data are driven by a trend.

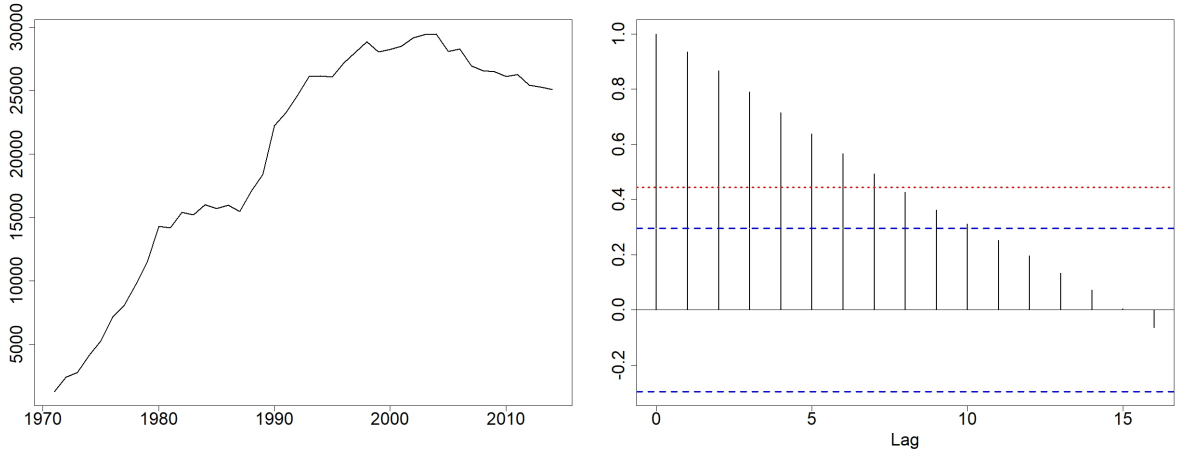

Figure 3.: Abortion data evolution (Left) and autocorrelation functions (Right).

Table 4 shows that our strategy TDT applied to abortion data suggests a (**StoS,2**) model. Both functions `trend.diag.tests()` and `trend.diag.high()` confirm this diagnosis. Dickey-Fuller tests rather suggest a (**StoS,1**) model. Indeed, most Dickey-Fuller tests do not reject the null for the initial series, but they are unable to detect a unit root in the differentiated series. We construct both a (**StoS,1**) and a (**StoS,2**) model, as in Equation 2, by modeling the stationary process $(B_t)_t$ by the first clearly valid model among all ARMA(p,q) with $p, q \leq 2$, sorted by minimizing Schwarz's Bayesian Criterion [22]:

- $(B_t)_t$ is AR(1) for (**StoS,2**) model, suggested by our strategy TDT.
- $(B_t)_t$ is AR(2) for (**StoS,1**) model, suggested by Dickey-Fuller tests.

In Table 5, we compare both models relevance in terms of information criteria such as AIC, BIC, AICc [1, 22, 13] and prediction criterion computed between the observed series and predictions for the last 4 values(almost 10%), such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Table 5 shows that the (**StoS,2**) model provided by our strategy is the most suitable.

Figure 4 plots forecasts for abortion data. Residuals being normally distributed, we plot both

Table 5.: Models comparison for abortion series.

	Criterion				
Model	**AIC**	**AICc**	**BIC**	**RMSE**	**MAPE**
TDT[a]: (StoS,2) with AR(1)	695.4	695.7	698.9	206.2	0.731
DF[b]: (StoS,1) with AR(2)	710.1	710.8	715.4	382	1.469

[a]Trend Diagnosis Tests.
[b]Dickey-Fuller tests.

forecasts and 80% and 95% prediction intervals, computed from gaussian quantiles.

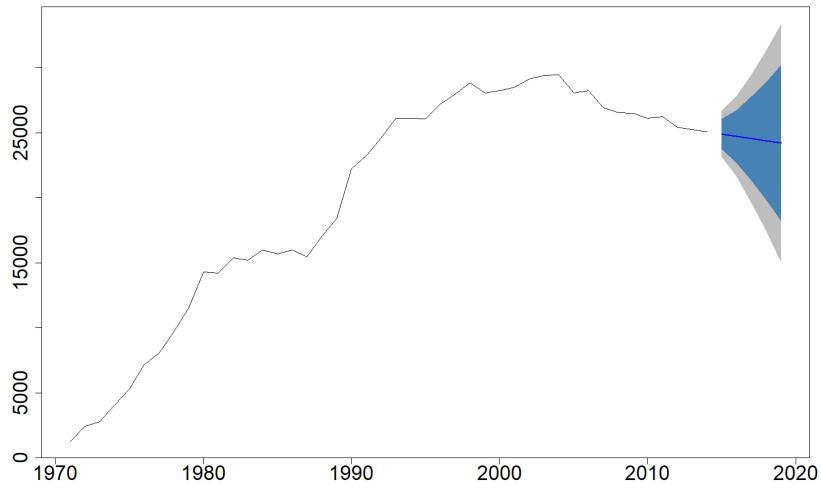

Figure 4.: Predictions for abortion series with TDT model, where prediction intervals colored in steel blue (resp. light grey) represent 80% (resp. 95%) confidence level.

4.2.2. Evolution of atmospheric CO2 concentration

Since 1959, atmospheric CO2 concentrations (ppm) has been measured monthly in situ air measurements, at Mauna Loa, Observatory, Hawaï see [14]. Figure 5 reveals that the global average concentration of atmospheric carbon dioxide has a clear increasing trend, and also a seasonal monthly component.

Our strategy does not take into account seasonal component, then we deseasonalize, by regressing the CO2 series on the seasonal dummy variables and by retaining the residuals from this regression. Autocorrelation functions plotted in Figure 6 (Right) show that the remaining series is driven by a trend. From Table 6, we see that Dickey-Fuller tests applied to the deseasonalized series clearly suggests a (**StoS,1**) model, whereas our TDT strategy produces responses **Alt/Alt/Alt/Alt** to tests **i)** to **iv)**, suggesting rather a (**DetS,2**) model. We construct both a (**StoS,1**) and a (**DetS,2**) model, as in Equations 1 and 2, by modeling the stationary process

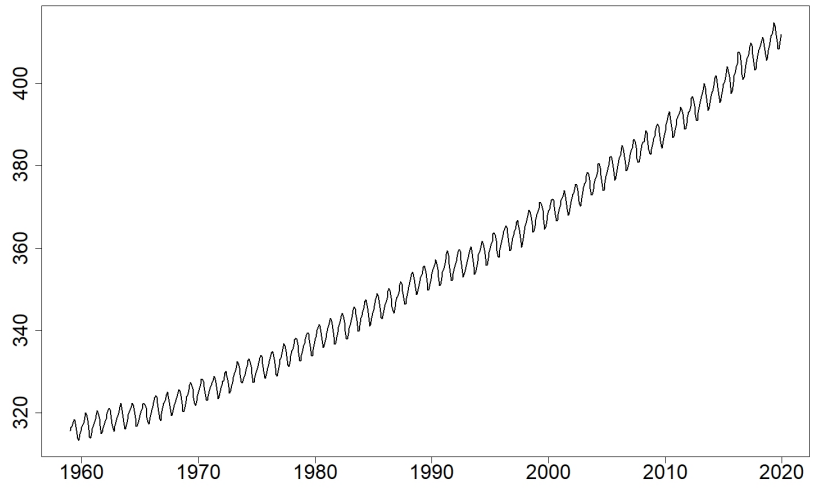

Figure 5.: CO2 atmospheric concentration evolution.

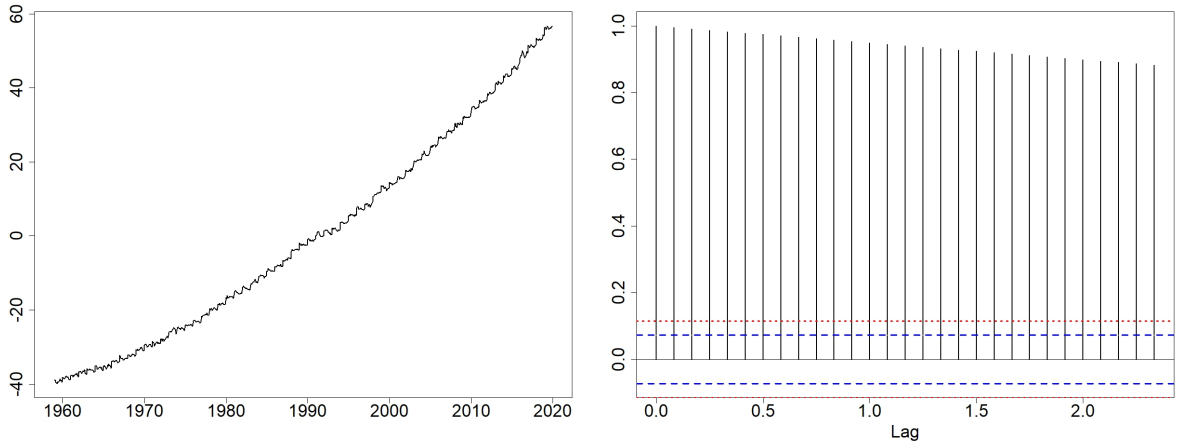

Figure 6.: CO2 deseasonalized series evolution (Left) and associated autocorrelation functions (Right).

Table 6.: p-values provided by several tests on CO2 deseasonalized series.

	Test				
Series	**OPP**	**KPSS**	ρ **under** (M_1)	ρ **under** (M_2)	ρ **under** (M_3)
Z_t	0.01	0.01	0.9791	0.99	0.5465
$\Delta(Z_t)$	0.01	0.01	0.01	0.01	0.01

Table 7.: Models comparison for CO2 deseasonalized series.

Model	Criterion				
	AIC	**AICc**	**BIC**	**RMSE**	**MAPE**
TDT[a]: (DetS,2) with SARMA(1,2)(1,1)[12]	376.8	377	413.6	0.78	1.291
DF[b]: (StoS,1) with SARMA(1,0)(1,1)[12]	400	400.1	418.4	1.757	2.816

[a]Trend Diagnosis Tests.
[b]Dickey-Fuller tests.

$(B_t)_t$ by the first valid model among all SARMA(p,q)(P,Q)[12] with $p, q, P, Q \leq 2$, sorted by minimizing Schwarz's Bayesian Criterion [22]:

- $(B_t)_t$ is SARMA(1,2)(1,1)[12] for (**DetS,2**) model, suggested by our strategy TDT.
- $(B_t)_t$ is SARMA(1,0)(1,1)[12] for (**StoS,1**) model, suggested by Dickey-Fuller tests.

In Table 7, we compare both models relevance in terms of information criteria such as AIC, BIC, AICc [1, 22, 13] and prediction criterion computed between the observed series and predictions for the last six years (almost 10%), such as RMSE and MAPE. Table 7 shows that the (**DetS,2**) model provided by our strategy is the most suitable.

Figure 7 shows that forecasts for CO2 atmospheric concentration maintain the same trajectory, with great accuracy. Indeed, prediction intervals are so thin that they are hardly visible.

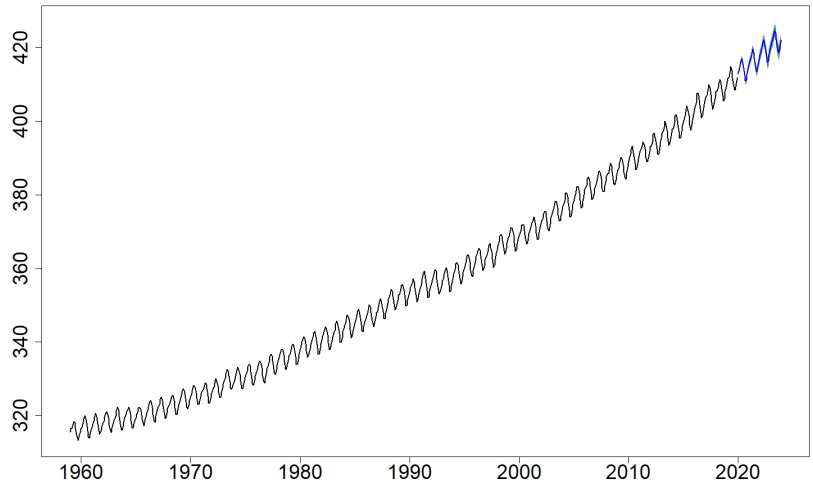

Figure 7.: Predictions for CO2 atmospheric concentration with TDT model, where prediction intervals colored in steel blue (resp. light grey) represent 80% (resp. 95%) confidence level.

5. Conclusion

We present a strategy to detect and identify trend component in time series. We recall that, as a first analysis, visualizing time series plot is indispensable, since it may already suggest the presence of a trend. Then trend can be confirmed by analyzing autocorrelation functions with `acfG()`. Next, whence a trend is detected, it remains to identify its nature. Indeed deterministic

or stochastic trends do not produce the same forecasts.

From Dickey-Fuller tests behavior, it appears useful to elaborate a new strategy to identify not only one-degree trends, but also higher-degree ones. For instance diagnosis for data generated from a (**StoS,2**) model results misleading. Indeed, the tests being ran on a single realization, the conclusion provided by each test will not necessary correspond to the majority response, that was obtained over a large number of simulations. Thus realizations of a (**StoS,2**) process can be falsely inferred as generated from either a (**DetS,1**) or a (**DetS,2**) or a (**StoS,1**) model. Then we propose a new strategy, involving other stationarity tests.

In this paper, we suggest a strategy based on OPP and KPSS tests, called TDT, and implemented in `diag.trend.tests()` and `diag.trend.high()` functions, in order to select between either a (**DetS,d**) or a (**StoS,d**) generating process. Our TDT strategy was assessed on simulations, and compared with Dickey-Fuller tests diagnosis. Of course, both procedures, TDT and Dickey-Fuller tests, may result in the same model suggestion. But diagnosis can be different, especially for time series with a quadratic trend, that reduces Dickey-Fuller tests reliability. Hence when both procedures suggest different models, both model candidates should be computed, validated and compared. Applied on two real data sets, our automatic TDT strategy provides a better model than the one provided using the classical Dickey-Fuller tests. Thus we recommend TDT strategy to improve and facilitate modeling procedure.

References

[1] H. Akaike, *Information theory as an extension of the maximum likelihood principle*, Second International Symposium on Information Theory, B.N. Petrov and F. Csaki, eds., vol. 1557, Springer, Akademiai Kiado, Budapest, Hungary, 1973, pp. 30–32.

[2] M. Boutahar, *Comparison of non-parametric and semi-parametric tests in detecting long memory*, Journal of Applied Statistics **36** (2009), no. 9, 945–972.

[3] PJ Brockwell and RA. Davis, *Time series : Theory and methods*, Springer-Verlag, New York, 1991, Second Edition.

[4] K.H. Chan, J.C. Hayya, and J.K. Ord, *A Note on Trend Removal Methods : The Case of Polynomial Regression versus Variate Differencing*, Econometrica **45** (1977), no. 3, 737–744.

[5] N.H. Chan and C.Z. Wei, *Limiting distributions of least squares estimates of unstable autoregressive processes*, Ann. Statist. **16** (1988), no. 1, 367–401.

[6] J.H. Conway and R.K. Guy, *The book of numbers*, Springer-Verlag, 1996.

[7] Y.A. Davydov, *The invariance principle of stationary processes*, Theory of probability and its application **XV** (1970), no. 3, 487–498.

[8] D.A. Dickey, W.R. Bell, and R.B. Miller, *Unit Roots in Time Series Models: Tests and Implications*, J. Am. Stat. Assoc. **40** (1986), no. 1, 12–26.

[9] D.A. Dickey and W.A. Fuller, *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, J. Am. Stat. Assoc. **74** (1979), no. 366, 427–431.

[10] J.J. Dolado, T. Jenkinson, and S. Sosvilla-Rivero, *Cointegration and unit roots*, J. Econ. Surv. **4** (1990), 249–273.

[11] C. Ertur, *Méthodologies de test de la racine unitaire*, Rapport de recherche, Laboratoire d'Analyses et de techniques économiques (LATEC), Université de Bourgogne, hal-01527262, 1998.

[12] G.H. Hardy and E.M. Wright, *An introduction to the theory of numbers*, 4^{th} with corrections ed., Springer-Verlag, Oxford, Clarendon Press, 1975.

[13] C. M. Hurvich and C.-L. Tsai, *Bias of the corrected AIC criterion for underfitted regression and time series models*, Biometrika **78** (1991), no. 3, 499–509.

[14] C.D. Keeling, S.C. Piper, R.B. Bacastow, M. Wahlen, T.P. Whorf, M. Heimann, and H.A.

Meijer, *Exchanges of atmospheric co2 and ^{13}co2 with the terrestrial biosphere and oceans from 1978 to 2000. i. global aspects*, SIO Reference Series **6** (2001), no. 1, 1–28.
[15] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin, *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*, Journal of Econometrics **54** (1992), 159–178.
[16] C.R. Nelson and H. Kang, *Spurious Periodicity in Inappropriately Detrended Time Series*, Econometrica **49** (1981), no. 3, 741–751.
[17] C.R. Nelson and C. Plosser, *Trends and random walks in macroeconomics time series : some evidences and applications*, Journal of Monetary Economics **10** (1982), no. 2, 139–162.
[18] W.K. Newey and K.D. West, *A Simple, positive, semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix*, Econometrica **55** (1987), no. 3, 703–708.
[19] S. Ouliaris, J.Y. Park, and P.C.B. Phillips, *Testing for a Unit Root in the Presence of a Maintained Trend*, Advances in Econometrics and Modeling, Kluwer Academic Publishers, Needham, MA, 1989.
[20] P. Perron, *Trends and Random Walk in Macroeconomics Time Series : Further Evidence from a New Approach*, JEDC **12** (1988), 297–332.
[21] Y.V. Prokhorov, *Convergence of random processes and limit theorems in probability theory*, Theory Probab. Appl **1** (1956), no. 2, 157–214.
[22] G.E. Schwartz, *Estimating the dimension of a model*, Ann. Stat. **6** (1978), no. 2, 461–464.

Appendix A. Sample autocorrelation functions behavior for (DetS,d) models – Proof of Theorem 2.2

We consider the polynomial case with degree $d \geq 1$. Let us define as $S_j(n)$ the sum of the j-th power of the first n integers:

$$S_j(n) \;=\; \sum_{k=1}^{n} k^j \,.$$

From Faulhaber's formula, detailed in [6], we know that

$$S_j(n) \;=\; \frac{n^{j+1}}{j+1} \;+\; \frac{1}{2}n^j \;+\; \frac{1}{j+1}\sum_{p=2}^{j} \mathcal{B}_p \binom{j+1}{p} n^{j-p+1}\,, \tag{A1}$$

where $\mathcal{B}_p$ are rational numbers called Bernoulli numbers.

Lemma A.1. *For any $j \in \mathbb{N}^*$, the sum of the j-th power of the first n integers, $S_j(n)$, is a $(j+1)$-degree polynomial with leading coefficient $\frac{1}{j+1}$. More precisely,*

$$S_j(n) \;=\; \frac{n^{j+1}}{j+1} \;+\; o(n^{j+1}), \tag{A2}$$

where $o(f)$ is one of the Landau symbols, as defined in [12], p. 7-8. $o(f)$ is called "little-O of f", and expresses the convergence to 0 of a given function, when it is divided by f.

Then,

$$\overline{Z} \;=\; \frac{a_d}{d+1}n^d \;+\; o(n^d) \;+\; \overline{B}\,. \tag{A3}$$

In order to study the asymptotic behavior of variables $\Xi(h)$, estimators of the theoretical autocorrelation function for stationary square-integrable processes, we first compute variables $\Gamma(h)$, estimators of the theoretical autocovariance function $\mathrm{cov}(Z_t, Z_{t+h})$.

Definition A.2. *From random variables $(Z_1, \cdots, Z_n)$, we define the autocovariance function as*

$$\Gamma(h) \;=\; \frac{\sum_{k=1}^{n-h} (Z_{k+h} - \overline{Z})(Z_k - \overline{Z})}{n}\,, \quad |h| \leq n.$$

Using Equation A3, we can express $\Gamma(h)$, by summing from $k = 1$ to $k = n - h$ the products of

$$Z_k - \overline{Z} \;=\; \sum_{j=0}^{d} a_j\, k^j \;-\; \frac{a_d}{d+1} n^d \;+\; o(n^d) \;+\; B_k - \overline{B}$$

with

$$Z_{k+h} - \overline{Z} \;=\; \sum_{j=0}^{d} a_j\, (k+h)^j \;-\; \frac{a_d}{d+1} n^d \;+\; o(n^d) \;+\; B_{k+h} - \overline{B}\,.$$

We have to study every term, and clarify its asymptotic behavior.

a) First, we consider all product terms involving either $\overline{B}$ or $\sum_{k=1}^{n-h} \frac{B_k}{n}$, or $\sum_{k=1}^{n-h} \frac{B_{k+h}}{n}$. Let us denote by T_a the sum of all these terms. Since $(B_t)_t$ is (**SN**) satisfying Hypotheses (H1) to (H3), then we can apply the weak law of large numbers for moving averages (see [3], Prop 6.3.10) and obtain

$$\overline{B} \;\xrightarrow[n\to+\infty]{\mathbb{P}}\; \mathbb{E}(B_1) \;=\; 0\,,$$

so do converge $\sum_{k=1}^{n-h} \frac{B_k}{n}$ and $\sum_{k=1}^{n-h} \frac{B_{k+h}}{n}$. Since $\overline{B}$, $\sum_{k=1}^{n-h} \frac{B_k}{n}$, or $\sum_{k=1}^{n-h} \frac{B_{k+h}}{n}$ do multiply either themselves or polynomials with degree $\leq d$, then these product terms converge $\mathbb{P}$ to 0, as soon as they are divided by n^d.

$$T_a \;=\; o_{\mathbb{P}}(n^d) \tag{A4}$$

b) Next, we study the behavior of the following terms :

$$\begin{aligned}
T_{b,1} &= \sum_{k=1}^{n-h} \frac{B_k\, B_{k+h}}{n}\,,\\
T_{b,2} &= \sum_{k=1}^{n-h} \sum_{j=0}^{d} a_j \frac{k^j\, B_{k+h}}{n}\,,\\
T_{b,3} &= \sum_{k=1}^{n-h} \sum_{j=0}^{d} a_j \frac{(k+h)^j\, B_k}{n}\,.
\end{aligned}$$

Applying again the weak law of large numbers for moving averages (see [3], Prop 7.3.5), we

obtain the $\mathbb{P}$-convergence of term $T_{b,1}$ to $\gamma_B(h)$, and then

$$T_{b,1} \;=\; o_{\mathbb{P}}(n^d)$$

On the other hand, we need Cauchy-Schwarz's inequality to study terms $T_{b,2}$ and $T_{b,3}$ in the same way. We get

$$\begin{aligned} T_{b,2} &= \sum_{j=0}^{d} a_j \sum_{k=1}^{n-h} \left(\frac{k^j}{\sqrt{n}} \times \frac{B_{k+h}}{\sqrt{n}} \right) \\ &\leq \sum_{j=0}^{d} a_j \left[\left(\sum_{k=1}^{n-h} \frac{k^{2j}}{n} \right)^{1/2} \times \left(\sum_{k=1}^{n-h} \frac{B_{k+h}^2}{n} \right)^{1/2} \right] \\ &\leq \left(\frac{n-h}{n} \sum_{k=1}^{n-h} \frac{B_{k+h}^2}{n-h} \right)^{1/2} \sum_{j=0}^{d} a_j \left(\frac{n^{2j}}{2j+1} + o(n^{2j}) \right)^{1/2} . \end{aligned}$$

The weak law of large numbers for moving averages and Prop 7.3.5 in [3] imply the $\mathbb{P}$-convergence of the left hand term to $\gamma_B(0)^{1/2} = \sigma_B$. In addition, the right hand term is $o(n^{d+1})$. Consequently,

$$T_b \;=\; T_{b,1} + T_{b,2} + T_{b,3} \;=\; o_{\mathbb{P}}(n^{d+1}) \tag{A5}$$

c) Let us denote by T_c all the product terms involving $o(n^d)$, not studied yet. From Equation A1, we get that $o(n^d)$ multiplies polynomials with degree $\leq d$. Then all product terms converge to 0, as soon as they are divided by n^d. Consequently, we have

$$T_c \;=\; o(n^{2d})\,. \tag{A6}$$

d) It remains to specify terms with polynomial products, in order to explicit the leading coefficient. Let us consider

$$\begin{aligned} T_{d,1} &= \frac{a_d^2}{(d+1)^2} \frac{n-h}{n} \, n^{2d}\,, \\ T_{d,2} &= -\frac{a_d}{d+1} n^d \times \sum_{j=0}^{d} a_j \sum_{k=1}^{n-h} \frac{k^j}{n}\,, \\ T_{d,3} &= -\frac{a_d}{d+1} n^d \times \sum_{j=0}^{d} a_j \sum_{k=1}^{n-h} \frac{(k+h)^j}{n}\,, \\ T_{d,4} &= \sum_{k=1}^{n-h} \left(\sum_{i=0}^{d} a_i \frac{k^i}{\sqrt{n}} \times \sum_{j=0}^{d} a_j \frac{(k+h)^j}{\sqrt{n}} \right) . \end{aligned}$$

All the terms $T_{d,1}$ to $T_{d,4}$ contain a leading term, associated to degree $2d$. There is nothing

to do for $T_{d,1}$. Equation A2 provides

$$\begin{aligned} T_{d,2} &= -\frac{a_d}{d+1} n^d \times \sum_{j=0}^{d} a_j \left(\frac{n^j}{j+1} + o(n^j) \right) \\ &= -\frac{a_d^2}{(d+1)^2} n^{2d} + o(n^{2d}) \,. \end{aligned}$$

We get the same formula for T_{d3}. In addition, Equation A2 also provides

$$\begin{aligned} T_{d,4} &= \sum_{i,j=0}^{d} \frac{a_i \, a_j}{n} \sum_{k=1}^{n-h} k^i \, (k+h)^j \,, \\ &= \frac{a_d^2}{2d+1} n^{2d} + o(n^{2d}) \,. \end{aligned}$$

Finally,

$$\begin{aligned} T_d &= T_{d,1} + T_{d,2} + T_{d,3} + T_{d,4} \\ &= \left(\frac{a_d^2}{(d+1)^2} \frac{n-h}{n} - 2\frac{a_d^2}{(d+1)^2} + \frac{a_d^2}{2d+1} \right) n^{2d} + o(n^{2d}) \end{aligned} \tag{A7}$$

Adding Equations A4 to A7, we obtain

$$\Gamma(h) = \left(\frac{a_d^2}{(d+1)^2} \frac{n-h}{n} - 2\frac{a_d^2}{(d+1)^2} + \frac{a_d^2}{2d+1} \right) n^{2d} + o_{\mathbb{P}}(n^{2d})$$

Finally since $a_d \neq 0$,

$$\begin{aligned} \Xi(h) &= \frac{\Gamma(h)}{\Gamma(0)} \\ &= \frac{\left(\frac{a_d^2}{(d+1)^2} \frac{n-h}{n} - 2\frac{a_d^2}{(d+1)^2} + \frac{a_d^2}{2d+1} \right) n^{2d} + o_{\mathbb{P}}(n^{2d})}{\left(\frac{-a_d^2}{(d+1)^2} + \frac{a_d^2}{2d+1} \right) n^{2d} + o_{\mathbb{P}}(n^{2d})} \\ &\xrightarrow[n \to +\infty]{\mathbb{P}} 1 \,. \end{aligned}$$

Appendix B. Sample autocorrelation functions behavior for (StoS,d) models – Proof of Theorem 2.3

We just give the proof for $d = 1$, since the general case can be deduced using the decomposition technique suggested in [5].

We first differentiate the initial series at a given lag h :

$$V_{k,h} = Z_k - Z_{k-h} = \sum_{j=0}^{h-1} B_{k-j} \,.$$

Let L denotes de lag operator i.e. $LX_t = X_{t-1}$, hence $V_{k,h}$ can be written as

$$V_{t,h} = A(L)B_t,$$

where A is the following polynomial

$$A(z) = 1 + z + ... + z^{h-1}.$$

Since $B_t = \Psi(L)\mathcal{E}_t$, where $\Psi(z) = \sum_{j\in\mathbb{Z}} b_j z^j$, it follows that

$$V_{t,h} = A(L)\Psi(L)\mathcal{E}_t = V(L)\mathcal{E}_t,$$

where

$$V(z) = A(z)\Psi(z) = \sum_{j\in\mathbb{Z}} v_j z^j, \quad \text{with } v_j = \sum_{i=0}^{h-1} b_{j-i}.$$

Consequently the process $(V_{t,h})$ is also a moving average, by straightforward calculations we can show that $(V_{t,h})$ satisfies Hypotheses (H1) to (H3).
Let us set

$$\begin{aligned} B_n(t) &= \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} B_k \,, \\ V_{n,h}(t) &= \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} V_{k,h}, \text{ for all t } \in [0,1] \,, \end{aligned}$$

where $[x]$ is the integer part of x. Then using [2] and Theorem 2 in [7], we get the weak convergence :

$$\begin{aligned} B_n(.) &\xrightarrow[n\to+\infty]{D[0,1]} \sqrt{2\pi f_B(0)} W. \\ V_{n,h}(.) &\xrightarrow[n\to+\infty]{D[0,1]} \sqrt{2\pi f_V(0)} W. \,, \end{aligned}$$

where $D[0,1]$ is the set of càdlàg functions with Skorokhod topology, and where $(W_t)_t$ is a standard Brownian motion. Moreover f_B and f_V are the spectral densities associated to processes (B_t) and $(V_{t,h})$:

$$f_B(\lambda) = \frac{\sigma_{\mathcal{E}}^2}{2\pi} \left| \sum_{j\in\mathbb{Z}} b_j e^{ij\lambda} \right|^2 \tag{B1}$$

$$f_V(\lambda) = \frac{\sigma_{\mathcal{E}}^2}{2\pi} \left| A(e^{ij\lambda}) \sum_{j\in\mathbb{Z}} b_j e^{ij\lambda} \right|^2 \tag{B2}$$

We also define

$$Z_n(t) = Z_{[nt]} - \overline{Z}\,, \text{ such that } Z_n\left(\frac{k}{n}\right) = Z_k - \overline{Z}\,.$$

We recall that

$$\frac{1}{\sqrt{n}} Z_n(t) \;=\; B_n(t) - \frac{1}{n}\sum_{j=1}^{n} B_n\left(\frac{j}{n}\right)\,,$$

Then by weak convergence continuity, we obtain that

$$\frac{1}{\sqrt{n}} Z_n(.) \xrightarrow[n\to+\infty]{D[0,1]} \sqrt{2\pi f_B(0)}\; W_{1,.}\,,$$

with

$$W_{1,t} = \left(W_t - \int_0^1 W_s ds\right)\,.$$

Autocorrelation function definition was given in Equation 3. We deduce that

$$\Xi(h) \;=\; 1 + \frac{\sum_{k=1}^{n-h} (Z_k - \overline{Z})\; V_{k+h,h}}{\sum_{k=1}^{n} (Z_k - \overline{Z})^2} + O_{\mathbb{P}}\left(\frac{1}{n}\right)\,.$$

Then,

$$\begin{aligned}
n\;(\hat{\Xi}(h) - 1) \qquad &= \qquad n\;\frac{\sum_{k=1}^{n-h} (Z_k - \overline{Z})\; V_{k+h,h}}{\sum_{k=1}^{n} (Z_k - \overline{Z})^2} + O_{\mathbb{P}}(1) \\
&= \qquad \frac{\sum_{k=1}^{n-h} \frac{Z_n\left(\frac{k}{n}\right)}{\sqrt{n}}\;\left(V_n\left(\frac{k+h}{n}\right) - V_n\left(\frac{k+h-1}{n}\right)\right)}{\frac{1}{n}\sum_{k=1}^{n}\left(\frac{Z_n\left(\frac{k}{n}\right)}{\sqrt{n}}\right)^2} + O_{\mathbb{P}}(1) \\
&\xrightarrow[n\to+\infty]{\mathcal{L}} \quad \frac{\sqrt{2\pi f_V(0)}}{\sqrt{2\pi f_B(0)}}\;\frac{\int_0^1 W_{1,s}\, dW_s}{\int_0^1 W_{1,s}^2\, ds} \;=\; |h|\;\frac{\int_0^1 W_{1,s}\, dW_s}{\int_0^1 W_{1,s}^2\, ds}\,.
\end{aligned}$$

Prokhorov theorem ([21]) permits to deduce that $n\;(\hat{\Xi}(h) - 1) \;=\; O_{\mathbb{P}}(1)$.

Appendix C. Boxplot of null-hypothesis rejection rate when $\sigma_{\mathcal{E}}$ varies – Complement to Table 1

In the main paper, Table 1 shows results for $\sigma_{\mathcal{E}} = 10$. Here, Figure C1 displays results when all the simulations with $\sigma_{\mathcal{E}}$ taking successive values in

$\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$ are gathered. This illustrates the stability of Dickey-Fuller-testing response, as $\sigma_{\mathcal{E}}$ varies, for most data generating process, except for (**DetW,1**) simulations, showing high variability.

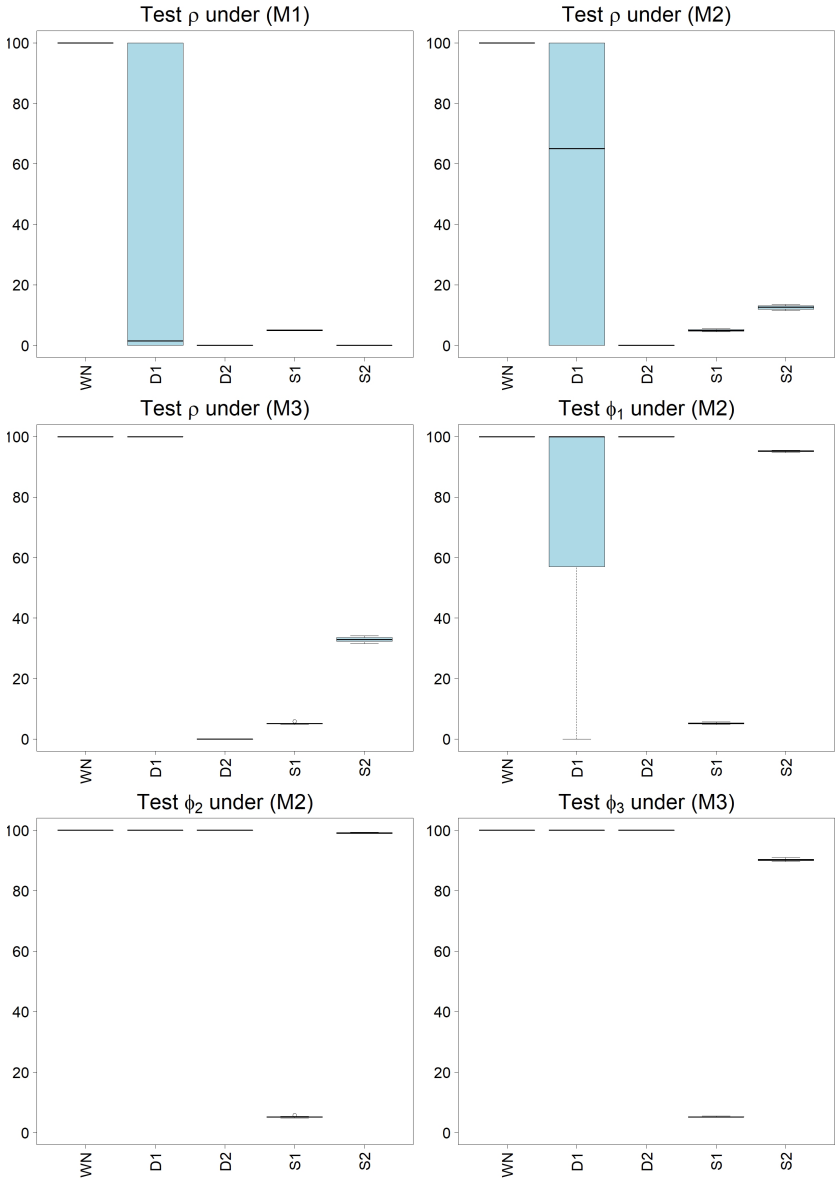

Figure C1.: Null hypothesis rejection rate for Dickey-Fuller tests, with respect to the underlying generating process used for simulations. All the simulations with $\sigma_{\mathcal{E}}$ taking successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$ are gathered.

Appendix D. Boxplot of null-hypothesis rejection rate when $\sigma_{\mathcal{E}}$ varies, and when noise is (WN) – Complement to Table 2

In the main paper, Table 2 shows that KPSS and OPP tests perform accurately on (**WN**), (**DetW,1**), (**DetW,2**), (**StoW,1**) and (**StoW,2**) simulations. In Table, 2, $\sigma_{\mathcal{E}}$ successively takes values in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$ and the final rejection rate is computed by gathering all the simulations obtained for each $\sigma_{\mathcal{E}}$. Here, Figure D1 illustrates the stability of testing procedure for every data generating process as $\sigma_{\mathcal{E}}$ varies. Note that an outlier is observed when applying KPSS test to (**DetS,1**) simulations. This means that KPSS test generally rejects the null hypothesis, as expected, in almost all cases. Actually, KPSS test sometimes fails to reject the null for several (**DetS,1**) simulations with $\sigma_{\mathcal{E}} = 500$, that is to say when noise intensity is too high in relation to the linear coefficient a_1, so that the trend becomes imperceptible. Thus $\sigma_{\mathcal{E}} = 500$ is above the high-limit for noise intensity.

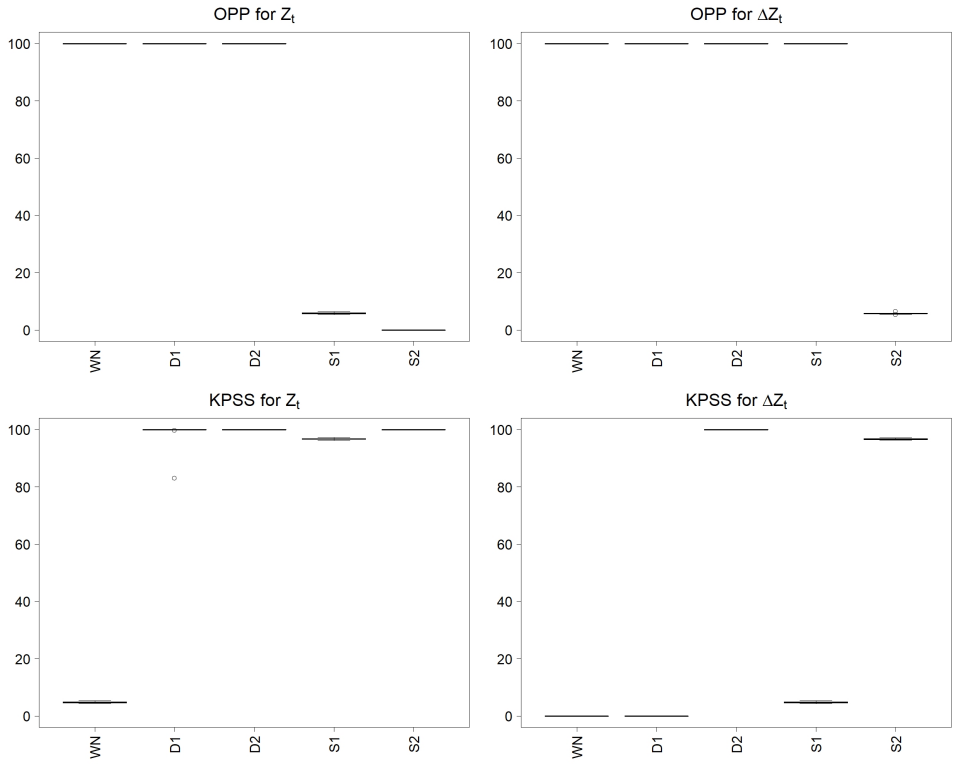

Figure D1.: Null hypothesis rejection rate for either KPSS or OPP stationarity tests applied upon either the initial or the differentiated series, with respect to the underlying generating process used for simulations. All the simulations, driven with a (**WN**), where $\sigma_{\mathcal{E}}$ takes successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$, are gathered.

Appendix E. Classification stability of Trend Diagnosis Tests (TDT) when noise is (WN) – Complement to Table 3

We ran 5 000 simulations under every data generating process among (**WN**), (**DetW,1**), (**DetW,2**), (**StoW,1**) and (**StoW,2**), with $a_0 = 5, a_1 = 1, a_2 = 1, n = 300$ and $\sigma_{\mathcal{E}}$, taking value in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$. Then we applied successively tests

i) to **iv)** to any simulation, using a risk $\alpha = 5\%$.

Given a data generating process, several diagnosis were observed, but not all of them. In the paper, Table 3 shows TDT diagnosis, when gathering simulations associated to a fixed DGP, whatever $\sigma_{\mathcal{E}}$ value in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. Here, Figure E1 illustrates the stability of the classification associated to every model as $\sigma_{\mathcal{E}}$ varies.

Note that the classification remains stable when $\sigma_{\mathcal{E}}$ keeps growing. But when noise intensity is too high in relation to the linear coefficient a_1, the trend becomes imperceptible, and KPSS test sometimes fails to reject the null for several **(detT,1)** simulations with $\sigma_{\mathcal{E}} > 300$. Whereas **Alt/Alt/Alt/Null** diagnosis is accurately associated to almost 99.9% of **(detT,1)** simulations while $\sigma_{\mathcal{E}} \leq 300$, 83.6% of **(detT,1)** simulations with $\sigma_{\mathcal{E}} = 500$ have the convenient diagnosis **Alt/Alt/Alt/Null**, but the 16.4% other simulations are associated to diagnosis **Alt/Alt/Null/Null**, that is accurate for (**WN**). And the confusion between **(detT,1)** and (**WN**) naturally increases with $\sigma_{\mathcal{E}}$. In this case, the true model **(detT,1)** might no longer be the most suitable for the series.

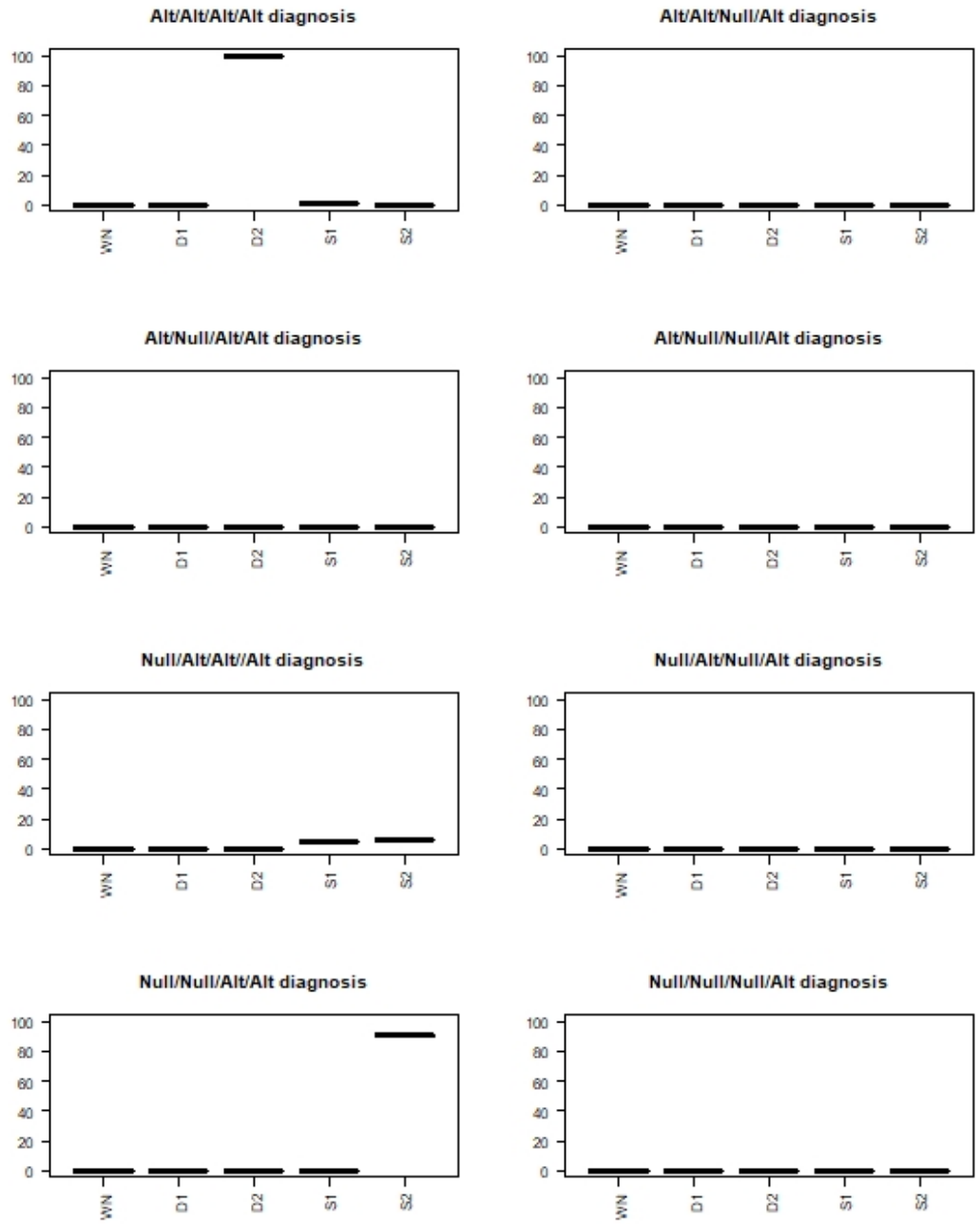

Figure E1.: (First part) Percentage of simulations with the associated diagnosis for TDT tests, versus every data generating process. All the simulations, driven with a (**WN**), where $\sigma_{\mathcal{E}}$ takes successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$, are gathered.

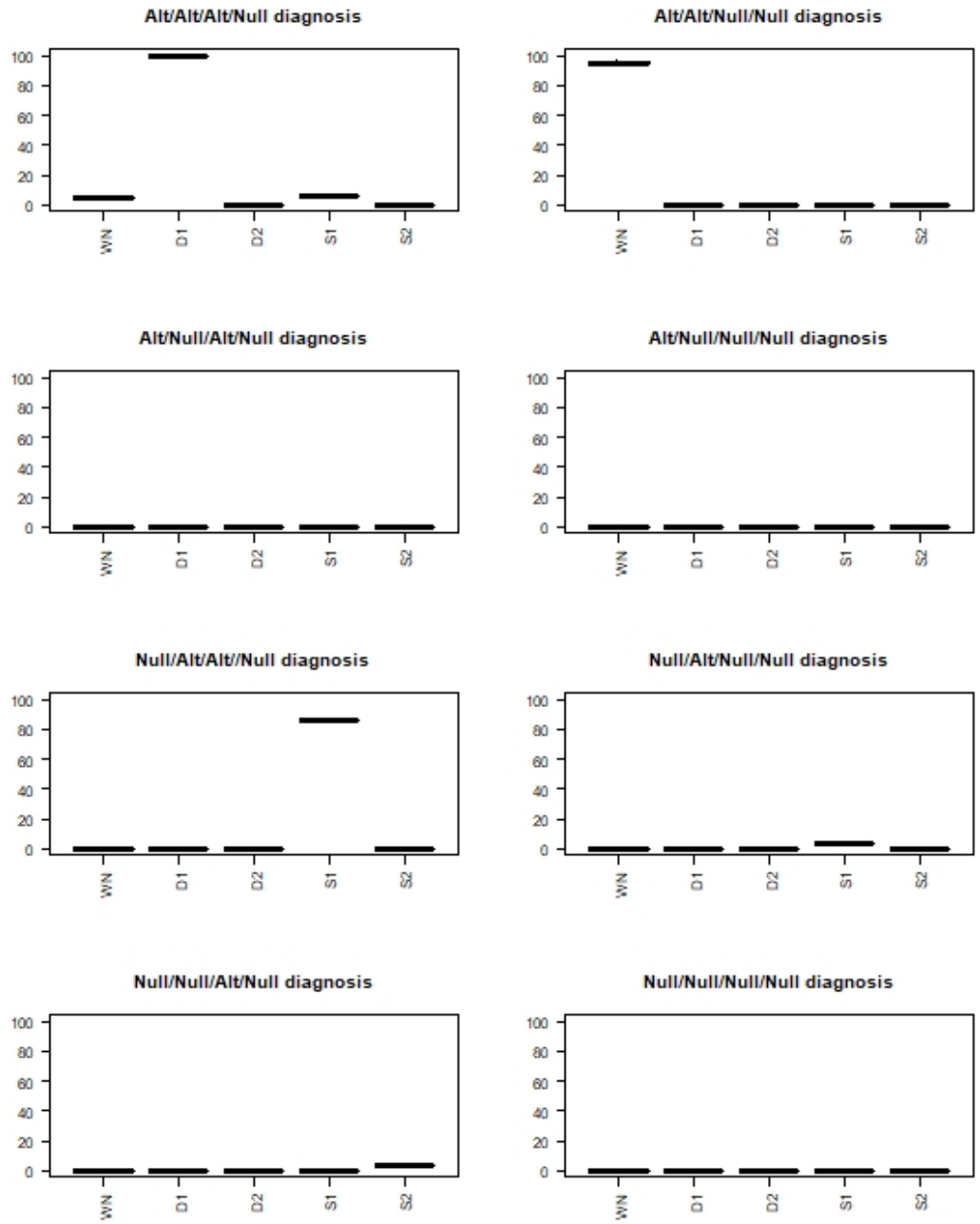

Figure E1.: (Last part) Percentage of simulations with the associated diagnosis for TDT tests, versus every data generating process. All the simulations, driven with a (**WN**), where $\sigma_{\mathcal{E}}$ takes successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$, are gathered.