Int. J. Computational Economics and Econometrics, Vol. 14, No. 1, 2024

SUPPLEMENTARY for Identifying trend nature in time series using autocorrelation functions and stationarity tests

M. Boutahar[a] and M. Royer-Carenzi[a]
[a]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, UMR 7373, Marseille, France

Contents

1. Simulations

1.1. Details on simulations

We recall that we defined two trend types :

$$\begin{aligned} \text{Deterministic trend } (\mathbf{Det,d}) \quad Z_t \; &= \; a_0 \,+\, a_1\, t \,+\, \cdots \,+\, a_d\, t^d \,+\, B_t && \text{(S1)} \\ \text{Stochastic trend } (\mathbf{Sto,d}) \quad \Delta^d(Z_t) \; &= \; B_t \,, && \text{(S2)} \end{aligned}$$

where we take $a_d \neq 0$, Δ is the 1-lag difference operator and $(B_t)_t$ is a $\mathcal{L}^2$-integrable, centered, stationary process, denoted as (**SN**), for *Stationary Noise*. When $(B_t)_t$ is merely a sequence of identically distributed and independent centered variables, it is called (**WN**) for *White Noise* and denoted by $(\mathcal{E}_t)_t$. In this case, the associated models defined in Equations (S1) and (S2), are referred as ($\mathbf{Det}_W$**,d**) and ($\mathbf{Sto}_W$**,d**).

In the main paper, we study processes constructed with an underlying white noise process, denoted as (**WN**). Actually, we simulate random independent centered gaussian variables $(\mathcal{E}_t)_t$ with a standard deviation $\sigma_{\mathcal{E}}$, taking value in $\{0.5, 1, 3,\ 5,\ 10,\ 20,\ 30,\ 50,\ 100,\ 200,\ 300,\ 500\}$. Next we construct the related processes ($\mathbf{Det}_W$**,1**), ($\mathbf{Det}_W$**,2**), ($\mathbf{Sto}_W$**,1**) and ($\mathbf{Sto}_W$**,2**). But in the Supplementary, we also analyze simulations with an underlying stationary noise, that can be a causal, invertible ARMA process. Thus we consider either simulations from a MA(2) process

$$B_t \;=\; \mathcal{E}_t \,+\, \frac{1}{2}\mathcal{E}_{t-1} \,-\, \frac{1}{5}\mathcal{E}_{t-2}\,, \tag{S3}$$

or an ARMA(1,1) process

$$B_t \,-\, \frac{1}{2}B_{t-1} \;=\; \mathcal{E}_t \,-\, \frac{1}{3}\mathcal{E}_{t-1}\,. \tag{S4}$$

And we deduce the associated (**Det,1**), (**Det,2**), (**Sto,1**) and (**Sto,2**) processes. In details, we perform the following steps :

1) fix a value for $\sigma_{\mathcal{E}}$ among $\{0.5, 1, 3,\ 5,\ 10,\ 20,\ 30,\ 50,\ 100,\ 200,\ 300,\ 500\}$.
2) simulate B_t as either
   a) $n = 300$ independent realizations of $\mathcal{N}(0, \sigma_{\mathcal{E}}^2)$
   b) or $n = 300$ realizations of MA(2) process, defined in Equation (S3)
   c) or $n = 300$ realizations of ARMA(1,1) process, defined in Equation (S4)
3) construct the related processes
   - (**SN**) $= B_t$
   - (**Det,1**) $= 5 + t + B_t$
   - (**Det,2**) $= 5 + t + t^2 + B_t$
   - (**Sto,1**) $= Z_t$ such that $\Delta(Z_t) = B_t$
   - (**Sto,2**) $= Z_t$ such that $\Delta^2(Z_t) = B_t$
4) run the stationarity test for every model process, generated in step 3) and compare the p-values with the nominal level $\alpha = 5\%$.

We repeat steps 2) to 4) 5 000 times. And we repeat the whole procedure, as $\sigma_{\mathcal{E}}$ successively takes values in $\{0.5, 1, 3,\ 5,\ 10,\ 20,\ 30,\ 50,\ 100,\ 200,\ 300,\ 500\}$.

We recall that 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
In this page, we called

- `Function acfG.R` the R-code for sample autocorrelation plots with Sidak correction and binomial exact test, as explained in the main paper, Section 2.1, and a script `Example of acfG use` to detail its use,
- `Function opp.test.R` the R-code for OPP test, and a script `Example of opp.test use` to detail its use,
- `Function trend.diag.tests.R` the code for TDT strategy, as explained in the main paper in Section 3.3, and its associated script `Example of trend.diag.tests use`,
- `Function trend.diag.high.R` the code for TDT strategy, generalized for higher degree trends, as introduced in Paragraph 2.2.5.3, and its associated script `Example of trend.diag.high use`,
- `Script_Tables2and3.R` the R-script to generate the simulations providing Table 2 , Table 3 and Figure F1 in the main paper, such as Figure S1 and Figure S2 in the Supplementary,
- `Script_Table4.R` the R-script to generate the simulations providing Table 4 in the main paper, such as Table S1 and Table S2 in the Supplementary.

1.2. Behavior of KPSS and OPP stationarity tests when the underlying noise is not (WN) – Complement to Table 3 and Figure F1

In the main paper, Table 3 and Figure F1 show that KPSS and OPP tests perform accurately on (**WN**), ($\mathbf{Det}_W$,**1**), ($\mathbf{Det}_W$,**2**), ($\mathbf{Sto}_W$,**1**) and ($\mathbf{Sto}_W$,**2**) simulations. Now we consider simulations with an underlying stationary noise, denoted as B_t, that is a causal, invertible ARMA(p,q) process. Figure S1 (respectively Figure S2) displays results when $(B_t)_t$ follows a MA(2) (resp. ARMA(1,1)), as defined in Equation (S3) (resp. Equation (S4)). We set that $\sigma_{\mathcal{E}}$ successively takes values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$. We still observe the same convenient behavior, whatever $(B_t)_t$.

1.3. Stability of diagnosis classification, when the underlying noise is not (WN) – Complement to Table 4

In order to identify the trend nature of a time series $(Z_t)_t$, 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 series $\Delta(Z_t)$.

Under a rejection risk $\alpha = 5\%$, we denote by **Null**, the case where the null hypothesis

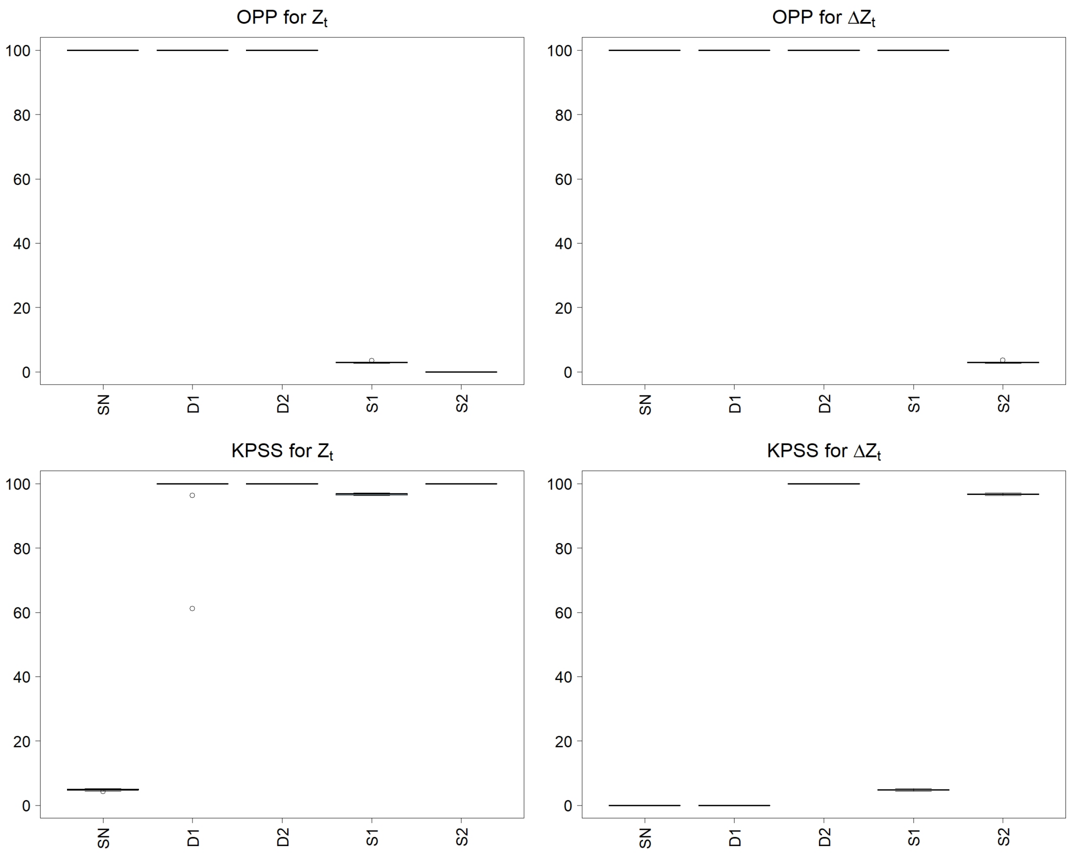

Figure S1.: Null hypothesis rejection rate for either KPSS or OPP stationarity tests applied upon either the initial or the differentiated series, with respect to the underlying generating process used for simulations. All the simulations, driven with a MA(2), where $\sigma_{\mathcal{E}}$ takes successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$, are gathered.

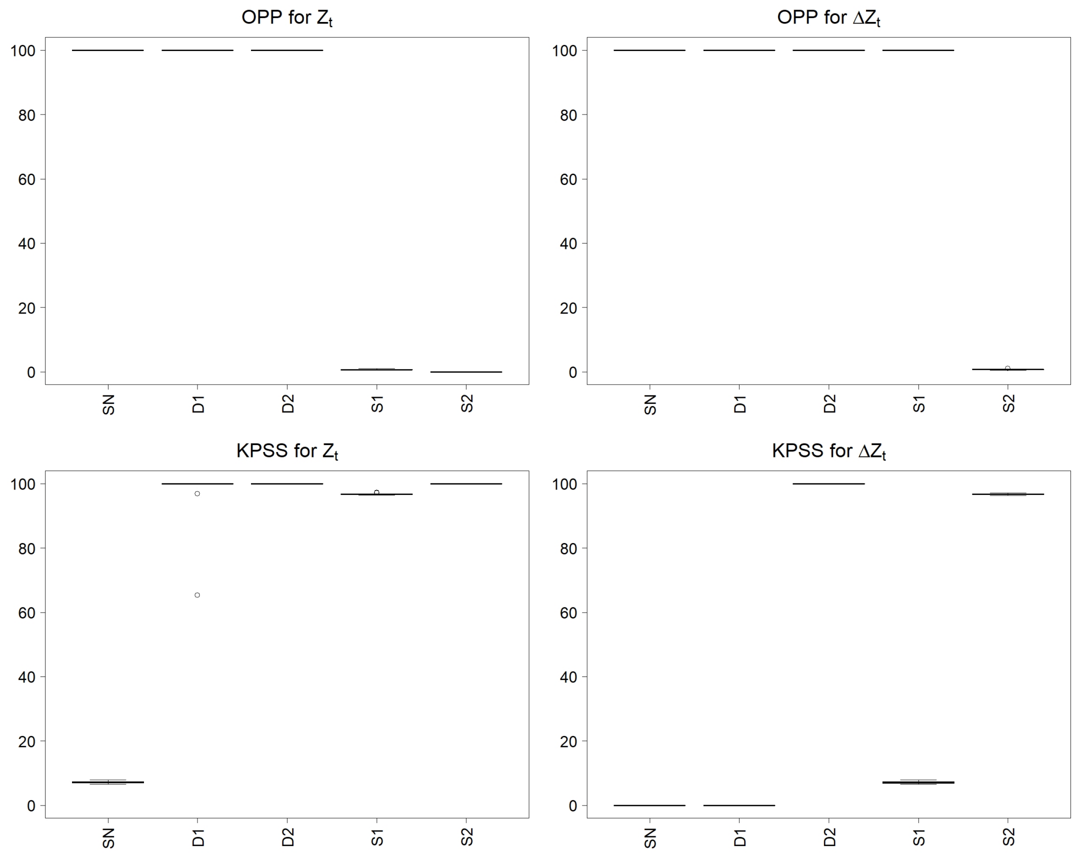

Figure S2.: Null hypothesis rejection rate for either KPSS or OPP stationarity tests applied upon either the initial or the differentiated series, with respect to the underlying generating process used for simulations. All the simulations, driven with a ARMA(1,1), where $\sigma_{\mathcal{E}}$ takes successive values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300, 500\}$, are gathered.

Table S1.: Percentage of Trend Diagnosis Tests (TDT) associated to every Data Generating Process (DGP). Simulations are driven with a MA(2), when $\sigma_{\mathcal{E}}$ takes values in $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$.

	DGP[a]				
TDT[b]	**(SN)**	**(Det,1)**	**(Det,2)**	**(Sto,1)**	**(Sto,2)**
Alt/Alt/Alt/Alt	0	0	**100**[c]	0.178	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	4.562	3.131
Null/Alt/Null/Alt	0	0	0	0.005	0.002
Null/Null/Alt/Alt	0	0	0	0	**93.463**
Null/Null/Null/Alt	0	0	0	0	0.033
Alt/Alt/Alt/Null	4.744	**99.638**	0	2.940	0
Alt/Alt/Null/Null	**95.256**	0.362	0	0.040	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	**88.986**	0.044
Null/Alt/Null/Null	0	0	0	3.289	0
Null/Null/Alt/Null	0	0	0	0	3.313
Null/Null/Null/Null	0	0	0	0	0.014
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.

can not be rejected, and by **Alt** otherwise. So that any time series can be associated to a single classification among the 2^4 possibilities. We call Trend Diagnosis Tests (TDT) the set of responses to tests **i)** to **iv)** computed on a time series.

In the main paper, Table 4 shows results for simulations driven by a white noise (**WN**), denoted as $(\mathcal{E}_t)_t$. 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 (**Det**$_W$**,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 (**Det**$_W$**,1**) and (**WN**) naturally increases with $\sigma_{\mathcal{E}}$. In this case, the true model (**Det**$_W$**,1**) might no longer be the most suitable for the series.

Here, we consider simulations associated to a more general noise (**SN**), denoted as B_t, that is a causal, invertible ARMA(p,q) process. Table S1 (respectively Table S2) displays results when $(B_t)_t$ follows a MA(2) (resp. ARMA(1,1)), as defined in Equation (S3) (resp. Equation (S4)), when $\sigma_{\mathcal{E}}$ takes values in the set $\{0.5, 1, 3, 5, 10, 20, 30, 50, 100, 200, 300\}$. Furthermore Figures S3 and S4 illustrate the stability of the classification associated to every model as $\sigma_{\mathcal{E}}$ varies.

Table S2.: Percentage of Trend Diagnosis Tests (TDT) associated to every Data Generating Process (DGP). Simulations are driven with a ARMA(1,1), when $\sigma_{\mathcal{E}}$ takes values in $\{0.5, 1, 3,\ 5,\ 10,\ 20,\ 30,\ 50,\ 100,\ 200,\ 300\}$.

	DGP[a]				
TDT[b]	**(SN)**	**(Det,1)**	**(Det,2)**	**(Sto,1)**	**(Sto,2)**
Alt/Alt/Alt/Alt	0	0	**100**[c]	0.085	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	6.747	0.698
Null/Alt/Null/Alt	0	0	0	0.007	0
Null/Null/Alt/Alt	0	0	0	0	**96.118**
Null/Null/Null/Alt	0	0	0	0	0.029
Alt/Alt/Alt/Null	6.809	**99.651**	0	0.611	0
Alt/Alt/Null/Null	**93.191**	0.349	0	0.014	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	**89.465**	0.011
Null/Alt/Null/Null	0	0	0	3.071	0
Null/Null/Alt/Null	0	0	0	0	3.124
Null/Null/Null/Null	0	0	0	0	0.020
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.

2. Nelson-Plosser analysis

In the main paper, we studied the *money stock* series from the macroeconomic Nelson-Plosser data. Actually, we applied our strategy on whole the 14 American macroeconomic indexes, contained in `tseries` R-package. Let us first study the *unemployment rate* series. Sample autocorrelation functions, plotted in Figure S3 do not show the typical behavior associated to series with a trend. Consequently, it should be modeled with a (**SN**) model.

For all of the other indexes, we ran all the stationarity tests presented in the main paper; results are given in Table S3. For every series, OPP and DF tests both detect a unit root in the series, but not in the differenced one, this would naturally lead to construct a (**Sto,1**) model. Our strategy, and simulation results summarized in Equations (20), suggests that a (**Sto,1**) model is convenient, but it additionally suggests that a (**Sto,2**) model should be explored, the best model being used for predictions.

Table S3.: p-values provided by several tests on the initial and the differentiated Nelson-Plosser series.

Series	**Test**				
cpi	**OPP**	**KPSS**	**ρ under (M$_1$)**	**ρ under (M$_2$)**	**ρ under (M$_3$)**
Z_t	0.2	0.01	0.99	0.99	0.99
$\Delta(Z_t)$	0.01	0.038	0.01	0.01	0.01
ip					
Z_t	0.17	0.01	0.99	0.727	0.084
$\Delta(Z_t)$	0.01	0.1	0.01	0.01	0.01
gnp.nom					
Z_t	0.2	0.01	0.99	0.99	0.912
$\Delta(Z_t)$	0.037	0.1	0.01	0.01	0.01
vel					
Z_t	0.2	0.01	0.012	0.084	0.741
$\Delta(Z_t)$	0.01	0.042	0.01	0.01	0.01
emp					
Z_t	0.2	0.01	0.99	0.894	0.436
$\Delta(Z_t)$	0.01	0.1	0.01	0.01	0.01
int.rate					
Z_t	0.2	0.01	0.84	0.861	0.833
$\Delta(Z_t)$	0.01	0.1	0.01	0.01	0.01
nom.wages					
Z_t	0.2	0.01	0.99	0.99	0.853
$\Delta(Z_t)$	0.03	0.1	0.01	0.01	0.01
gnp.def					
Z_t	0.2	0.01	0.99	0.99	0.952
$\Delta(Z_t)$	0.01	0.055	0.01	0.01	0.01
money.stock					
Z_t	0.2	0.01	0.99	0.99	0.943
$\Delta(Z_t)$	0.09	0.1	0.01	0.01	0.01
gnp.real					
Z_t	0.2	0.01	0.99	0.964	0.412
$\Delta(Z_t)$	0.012	0.1	0.01	0.01	0.01
stock.prices					
Z_t	0.2	0.01	0.99	0.99	0.653
$\Delta(Z_t)$	0.01	0.1	0.01	0.01	0.01
gnp.capita					
Z_t	0.2	0.01	0.99	0.953	0.371
$\Delta(Z_t)$	0.011	0.1	0.01	0.01	0.01
real.wages					
Z_t	0.2	0.01	0.99	0.679	0.938
$\Delta(Z_t)$	0.01	0.1	0.01	0.01	0.01

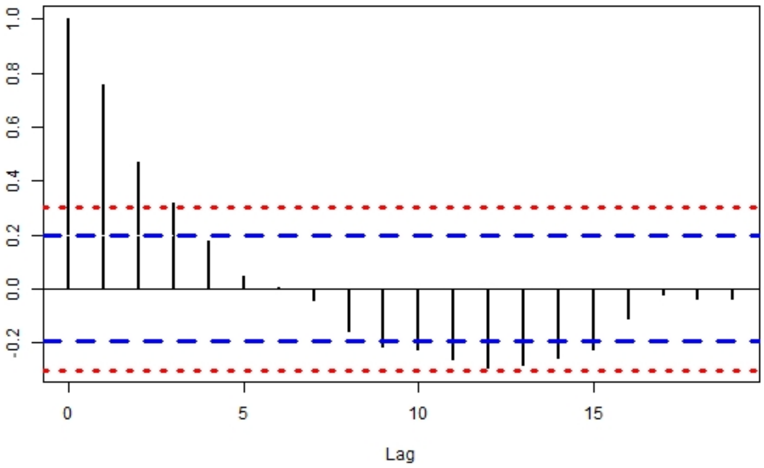

Figure S3.: Sample autocorrelation functions for unemployment series.