Matrix polynomials and statistical analysis of time-dependent models for vector time series

Guy Mélard*

*Université Libre de Bruxelles, Faculty SBS-EM, ECARES, Belgium
email: Guy.Melard@ulb.be

Version August 24, 2024
BSSM 2024
Brussels, Belgium
August 26, 2024

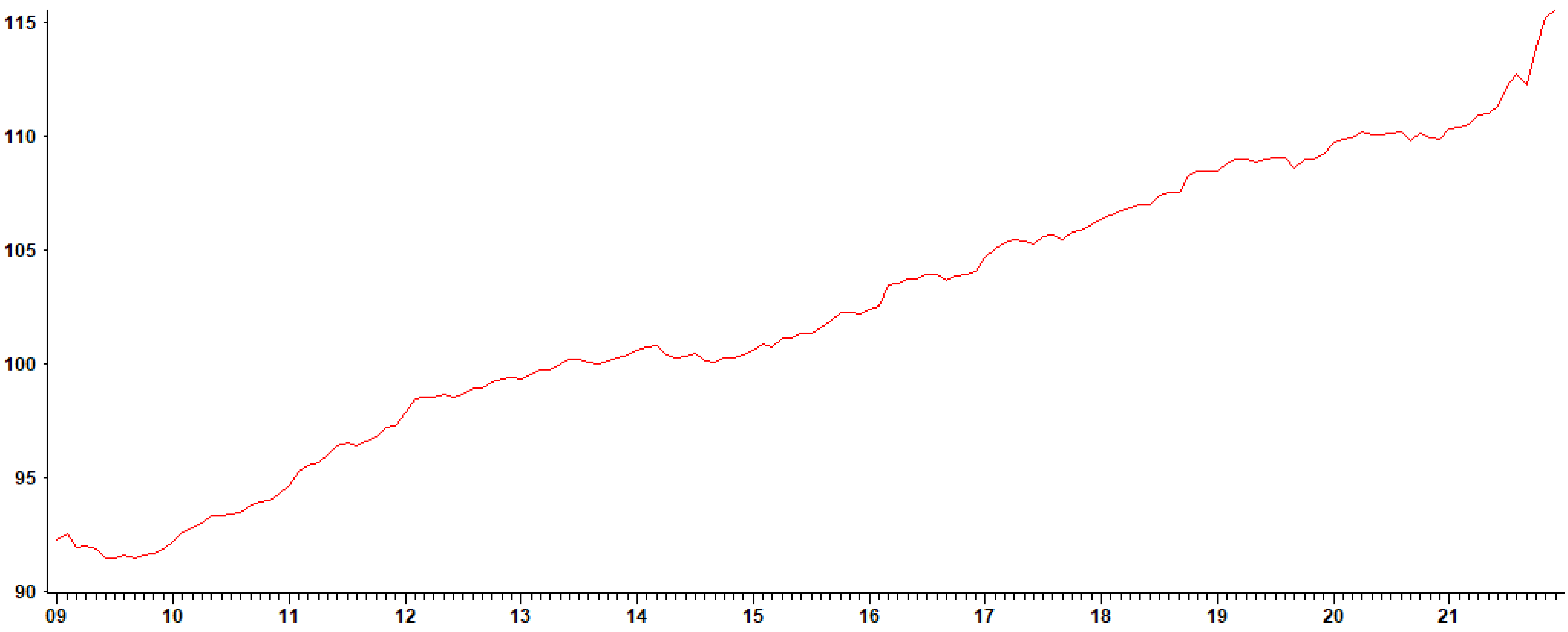
Outline

Introduction

- This presentation is a part of my work in the statistical analysis of time series, here with a focus on the mathematical aspects, particularly those related to polynomials
- Examples of time series are the Consumer Price Index (CPI) and industrial production but there are lots (from yearly data to observations every millisecond)
- I have worked mainly on time series modeling using the so-called **ARMA models** and their extensions
- This makes a heavy use of polynomials
- In particular, multivariate time series and, hence, **vector ARMA** or **VARMA models**, require polynomials with matrix coefficients or **matrix polynomials**
- My own research is more specifically on **time-dependent** or time-varying models, i.e., models where the coefficients vary with time, and I will mention that in most cases

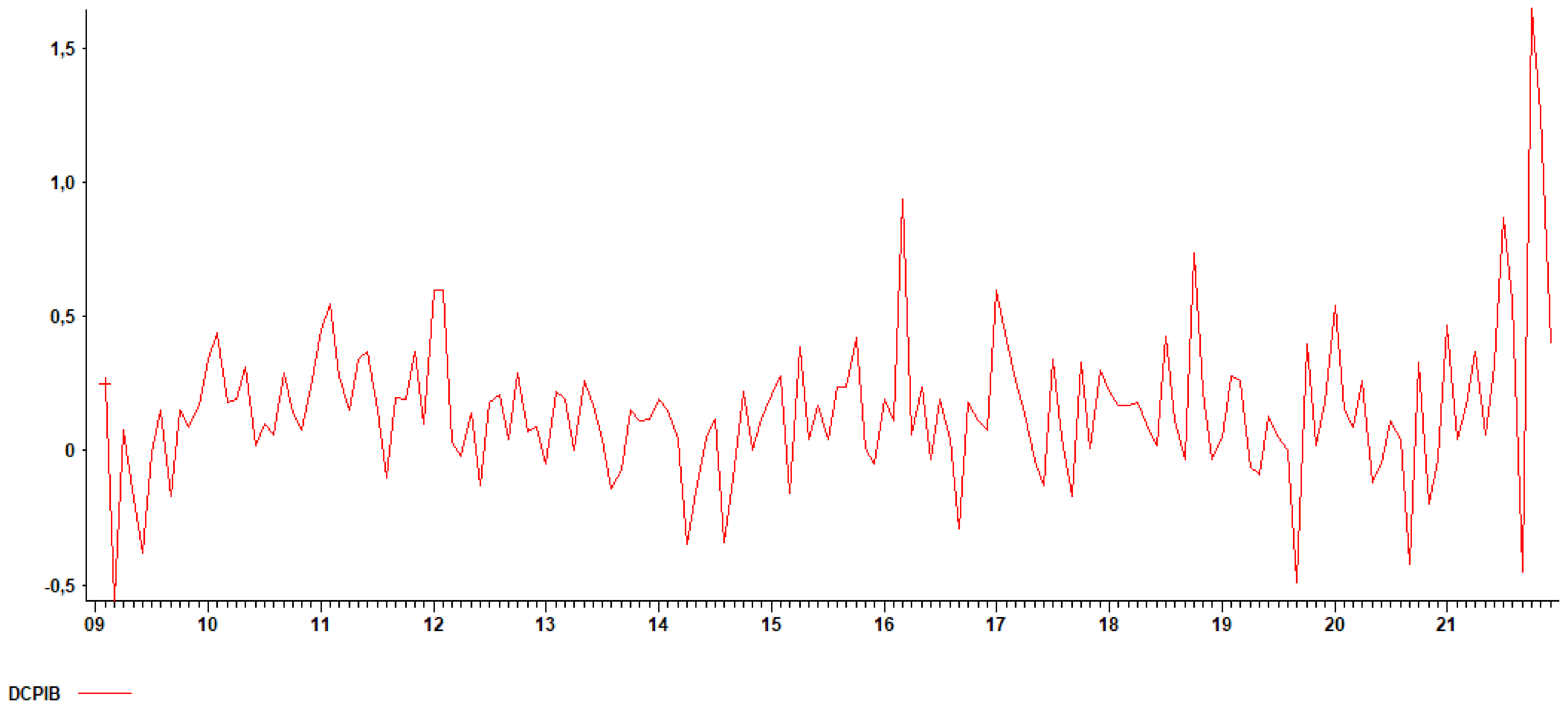
- I will start with two series and sketch how they can be modeled without explanations
- The purpose is just to show the models and the polynomials that define them
- If I have time at the end, I will shortly explain what are the AutoCorrelation Function (**ACF**) and Partial ACF (**PACF**) that I will use
- I will start with two monthly time series
  - the **consumer price index** or **CPI** for Belgium in a quiet period (2009.01 to 2021.12)
  - the Industrial Production of Belgium (**PRODIN**) from 2001.01 to 2019.12
- I will not show a complete example of a multivariate time series because it would be too difficult, only the final result
- I have put references at the end; the presentation is available upon request (I have also a less detailed presentation in French) (Guy.Melard@ulb.be)

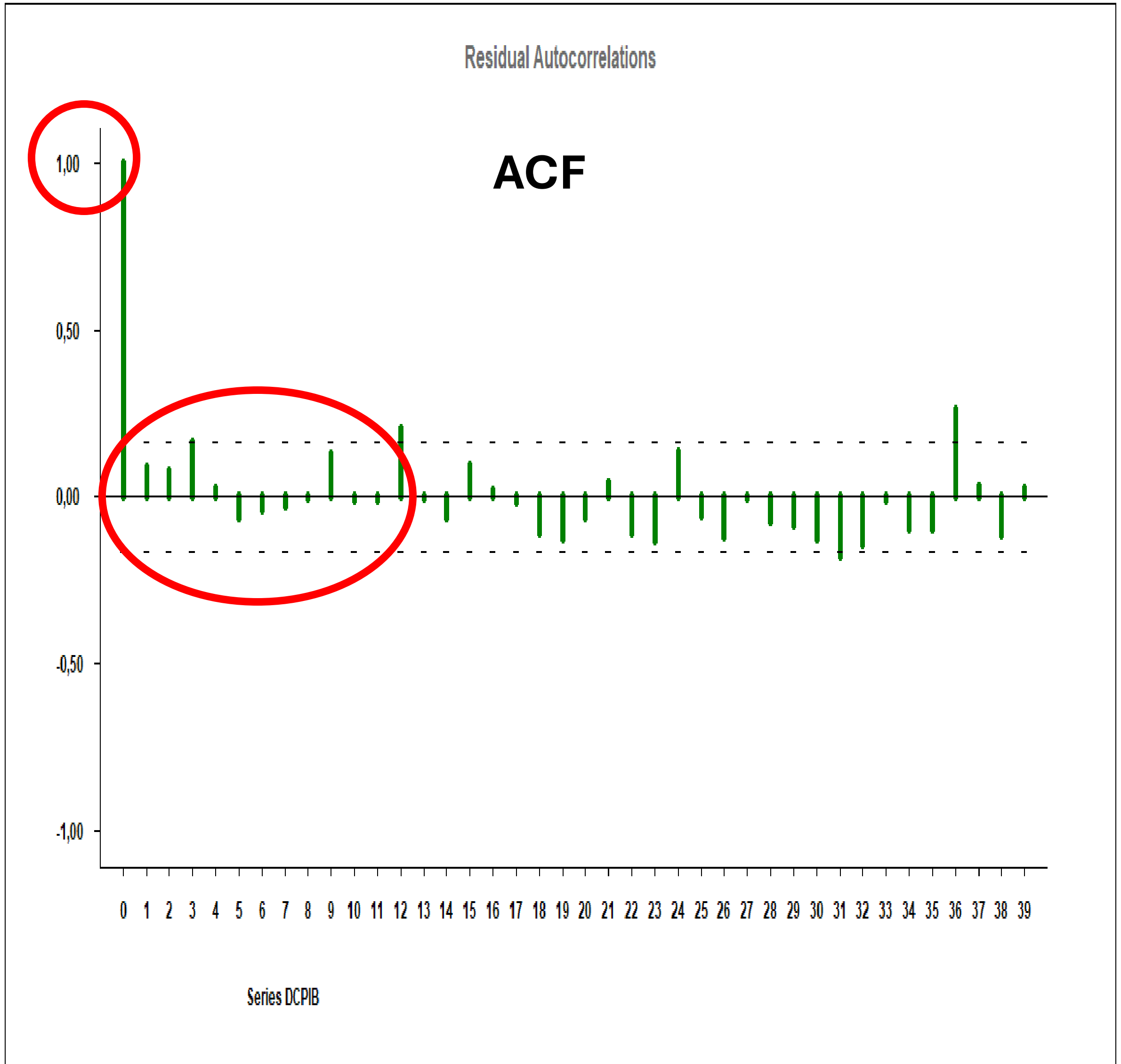
CPIB

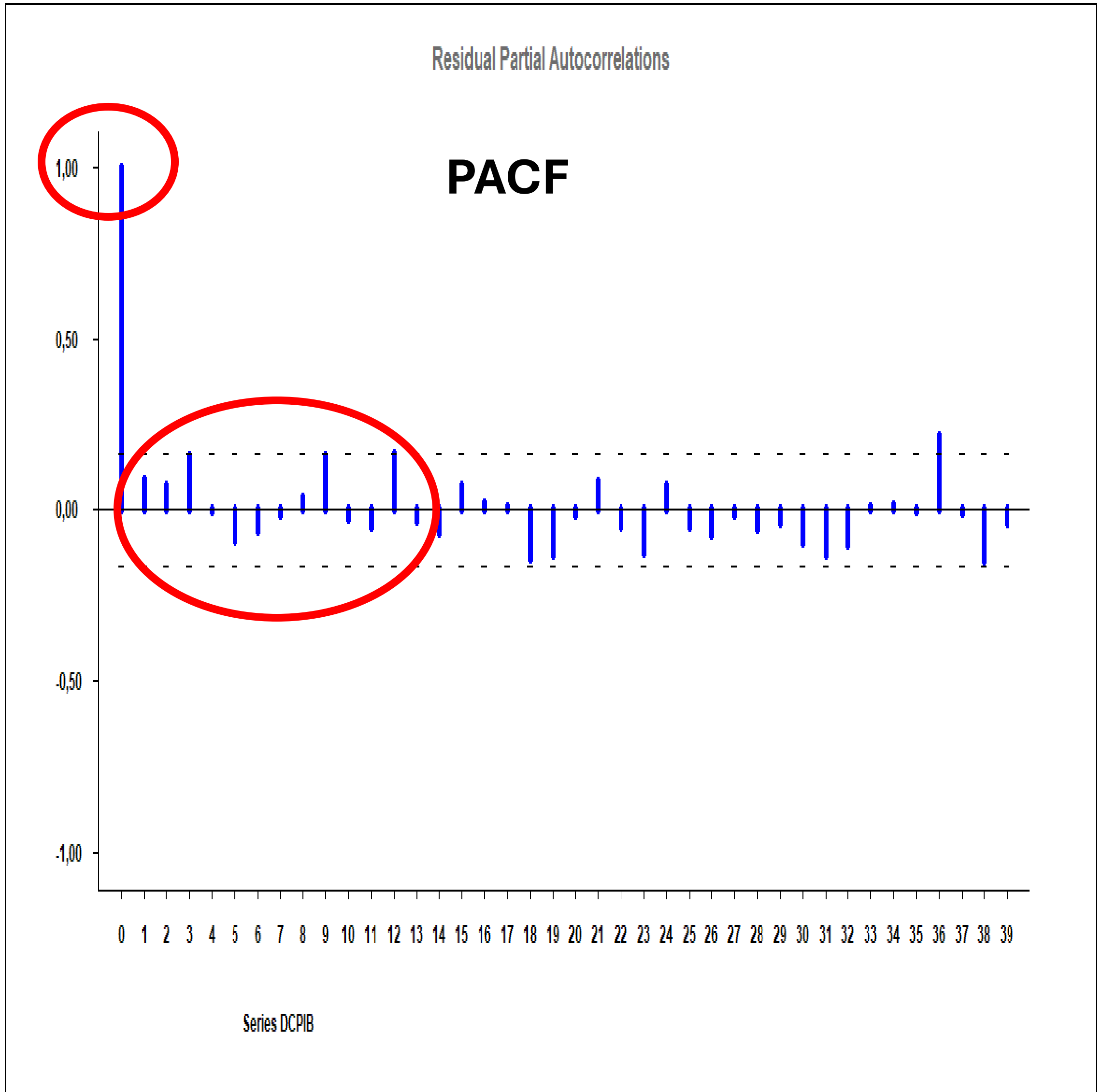


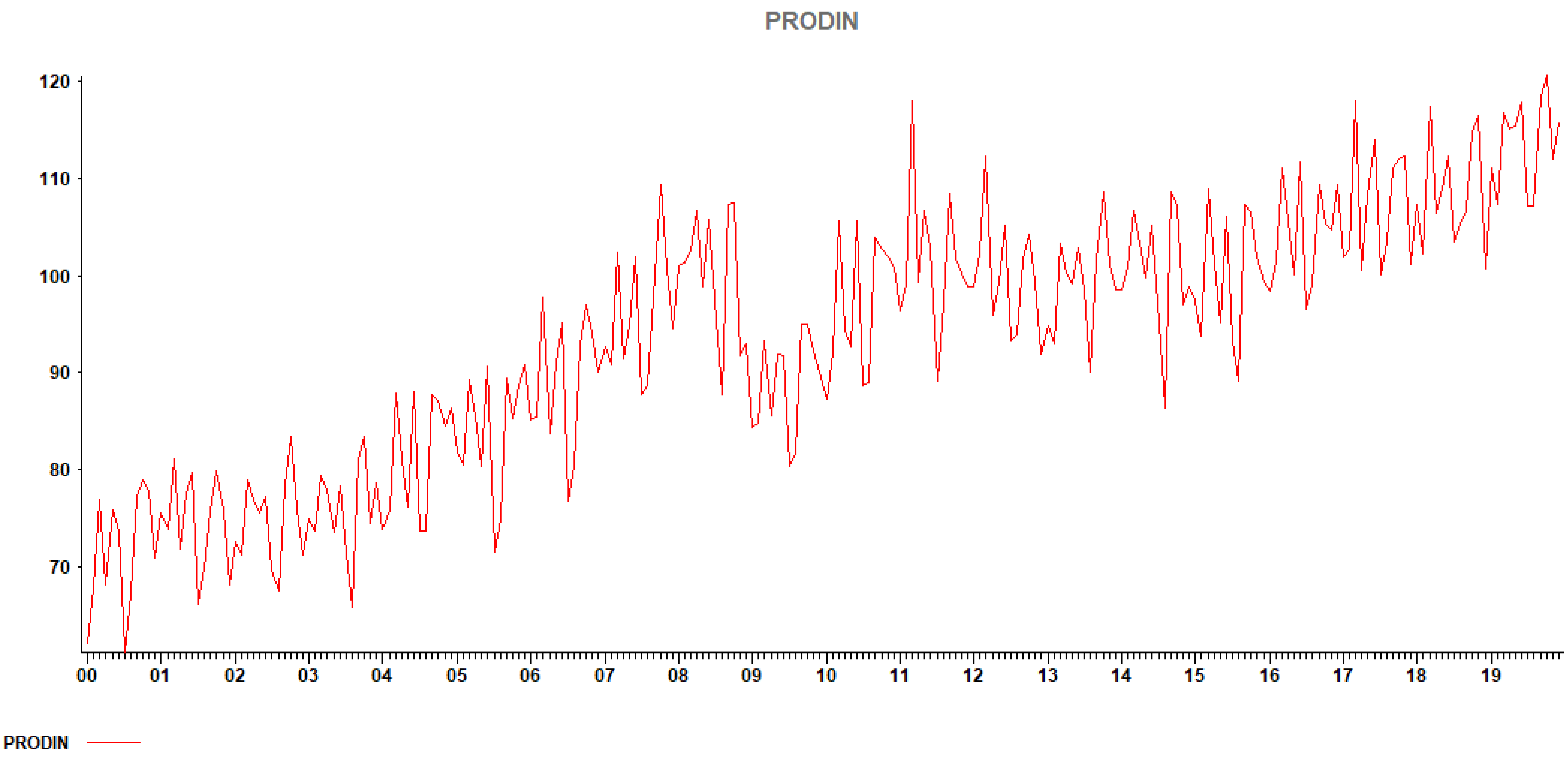
CPIB

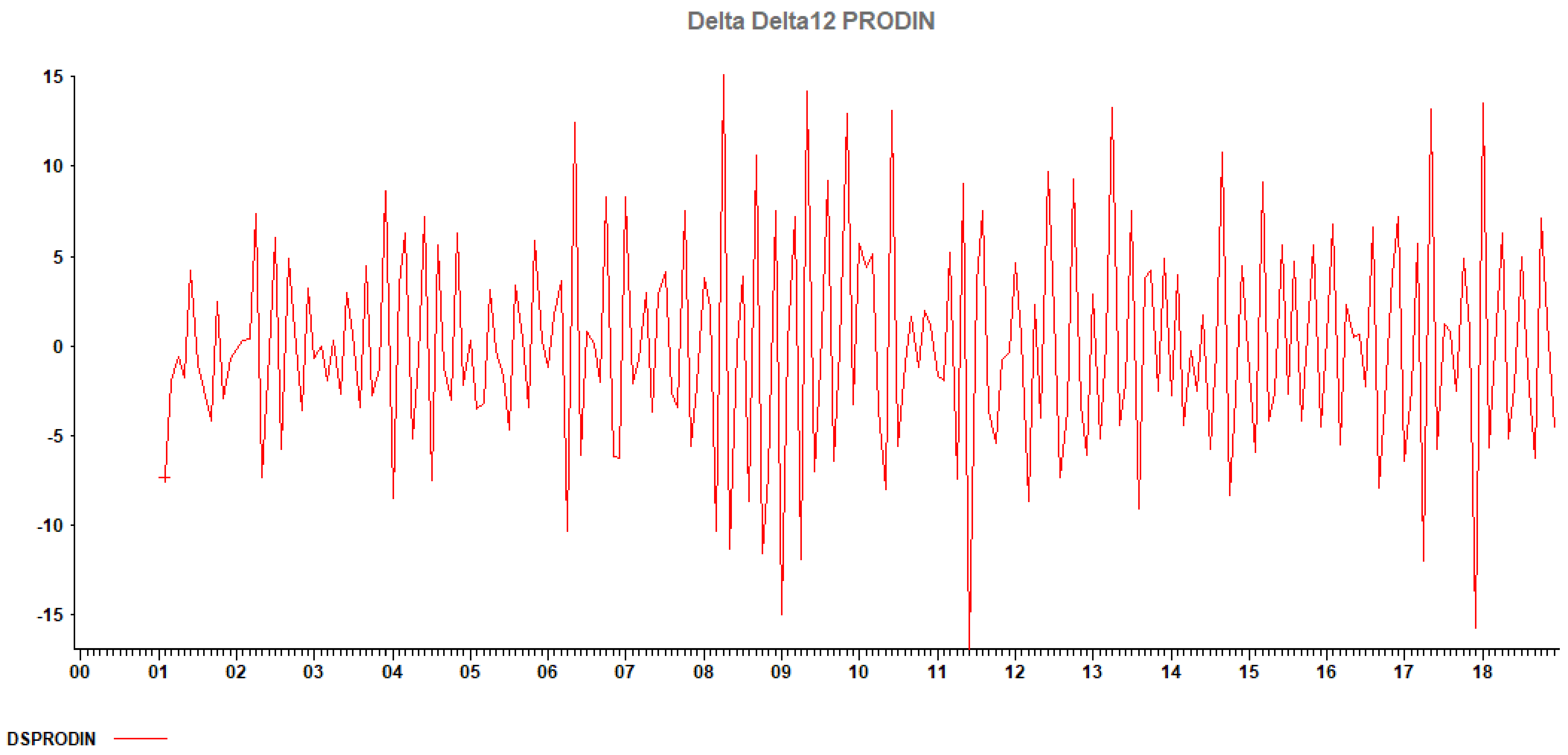
Delta CPIB

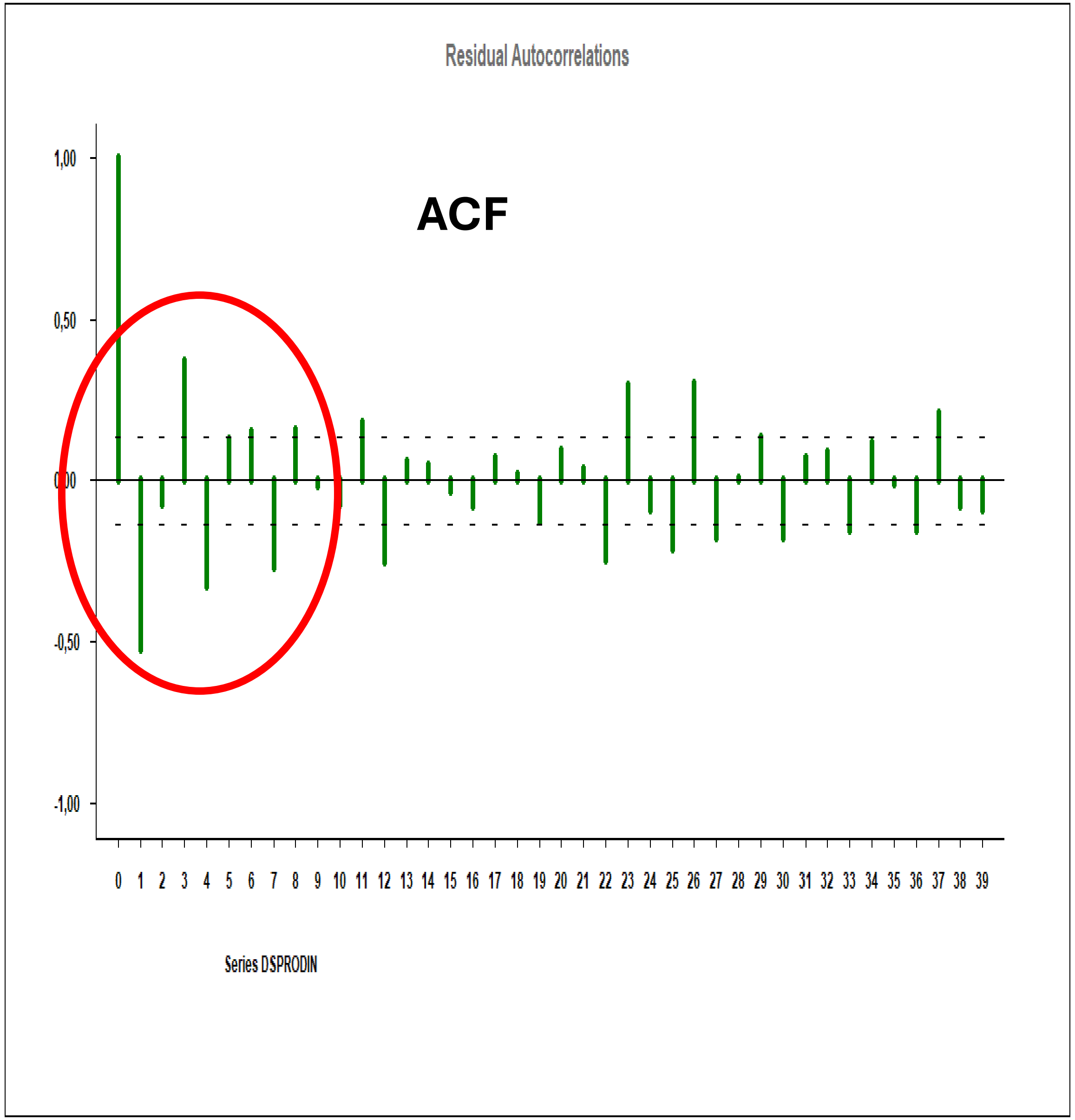


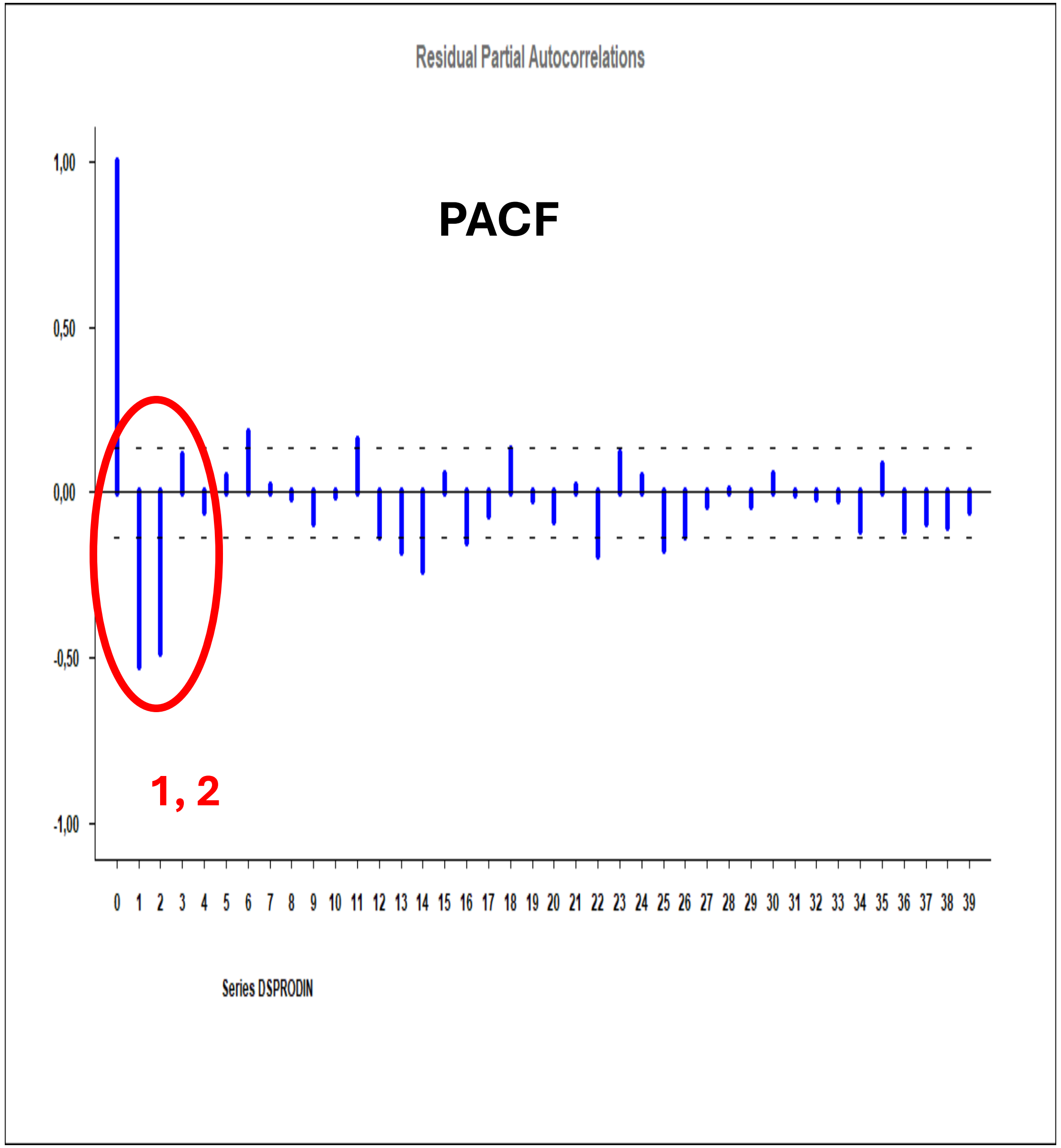

Residual Autocorrelations
ACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series DCPIB


Residual Partial Autocorrelations
PACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series DCPIB


PRODIN
120
110
100
90
80
70
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
PRODIN


Delta Delta12 PRODIN
15
10
5
0
-5
-10
-15
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
DSPRODIN


Residual Autocorrelations
ACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series DSPRODIN


Residual Partial Autocorrelations
PACF
1,00
0,50
0,00
-0,50
-1,00
1, 2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series DSPRODIN

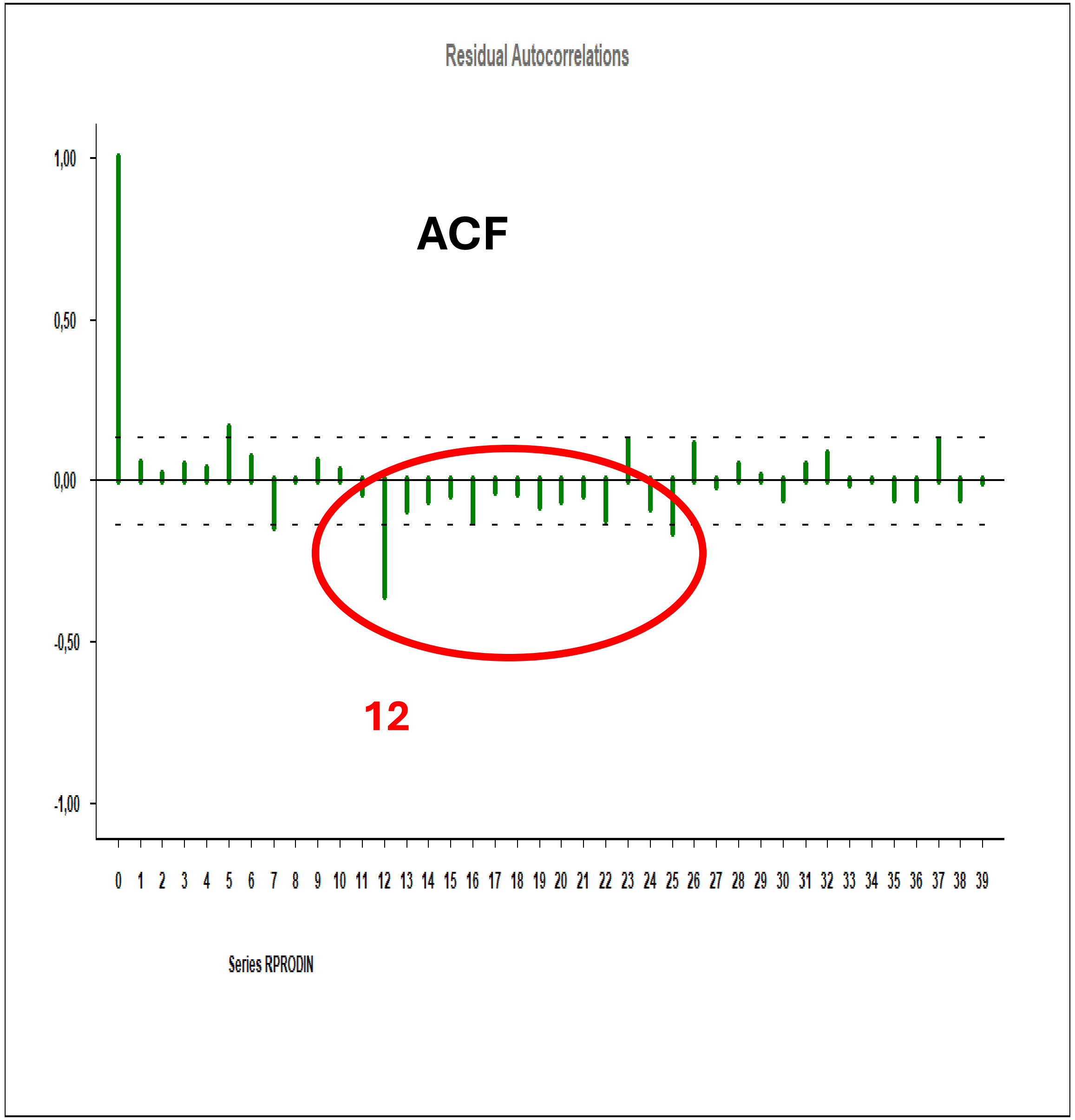
Fitted model (1)

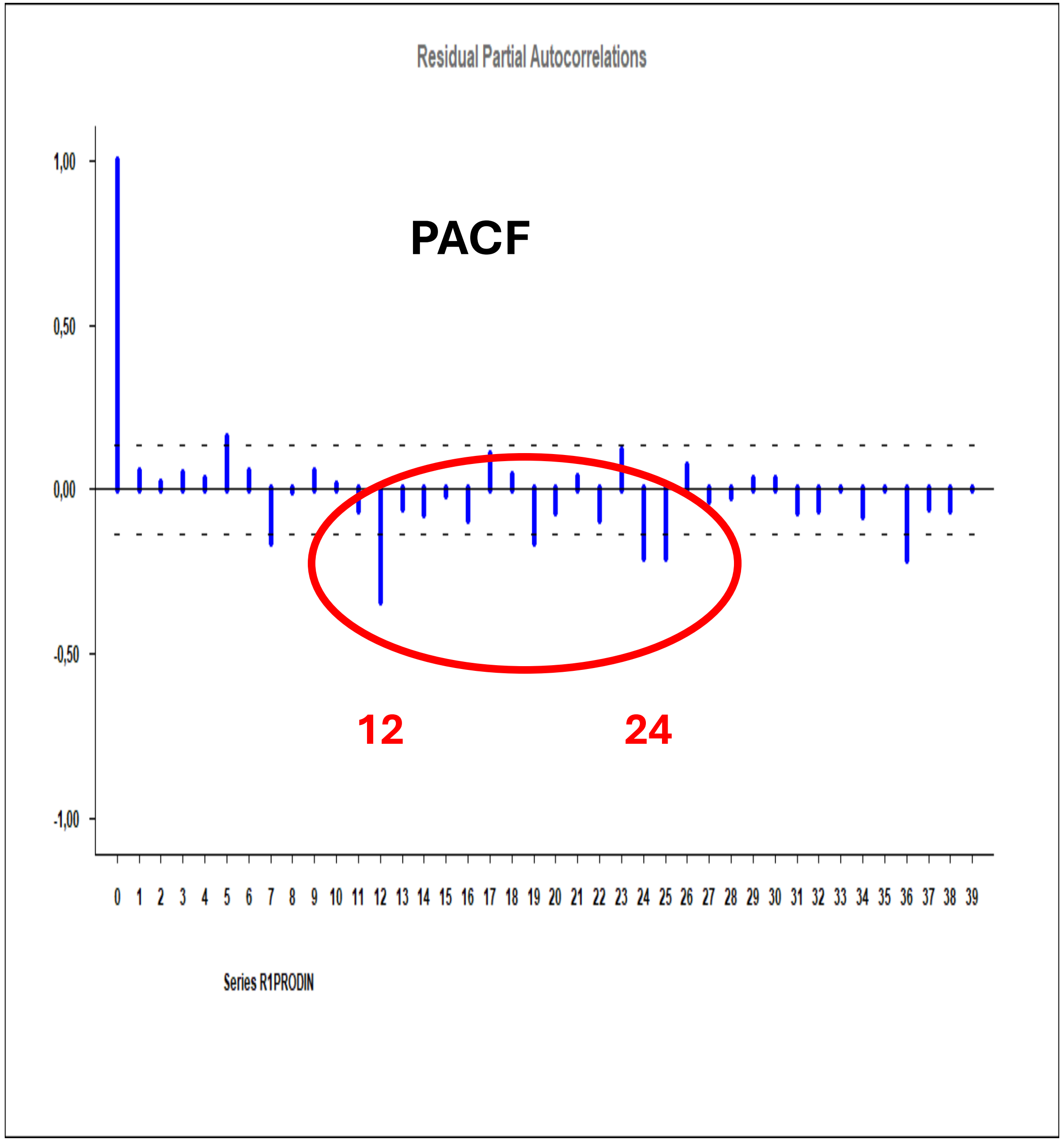
```
=== MODEL DESCRIPTION                          FORM      DEGREE/ORD PARAMETERS
NUMBER
 - SEASONAL PERIOD                                           12
 - DIFFERENCE                       REGULAR                   1
 - DIFFERENCE                       SEASONAL                  1
 - ARMA MODEL
   AUTOREGRESSIVE POLYNOMIAL REGULAR   2     AR    2
NON LINEAR ESTIMATION:    …
FINAL VALUES OF THE PARAMETERS
   NAME        VALUE           STD ERROR    T-VALUE
1  AR    1    -.78932        5.98608E-02    -13.2
2  AR    2    -.49156        5.96132E-02     -8.2
*** WARNING-A MEAN LEVEL IS NOT INCLUDED IN THE MODEL
=== SUMMARY MEASURES  <V>
 SUM OF SQUARES: COMPUTED =   4018.60 ADJUSTED 4002.15
 VARIANCE ESTIMATES: BIASED =   18.61 UNBIASED    18.79
 NUMBER OF PARAMETERS: 2 STANDARD DEVIATION =    4.33468
 INFORMATION CRITERIA: AIC =   1320.99   SBIC =   1331.90
 === RESIDUAL ANALYSIS WITH  215 RESIDUALS, BEGINNING AT TIME FEB2001===
 MEAN = -.101266, T-STATISTIC = -.34 (FOR TESTING ZERO MEAN)
```

- Seasonal period: 12
- Delta or Δ
- Delta12 or Δ_{12}

- $1 - (-0.779)\,L - (-0.492)\,L^2$

- $\sigma^2 = 18{,}79$


Residual Autocorrelations
ACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
12
Series RPRODIN


Residual Partial Autocorrelations
PACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
12
24
Series R1PRODIN

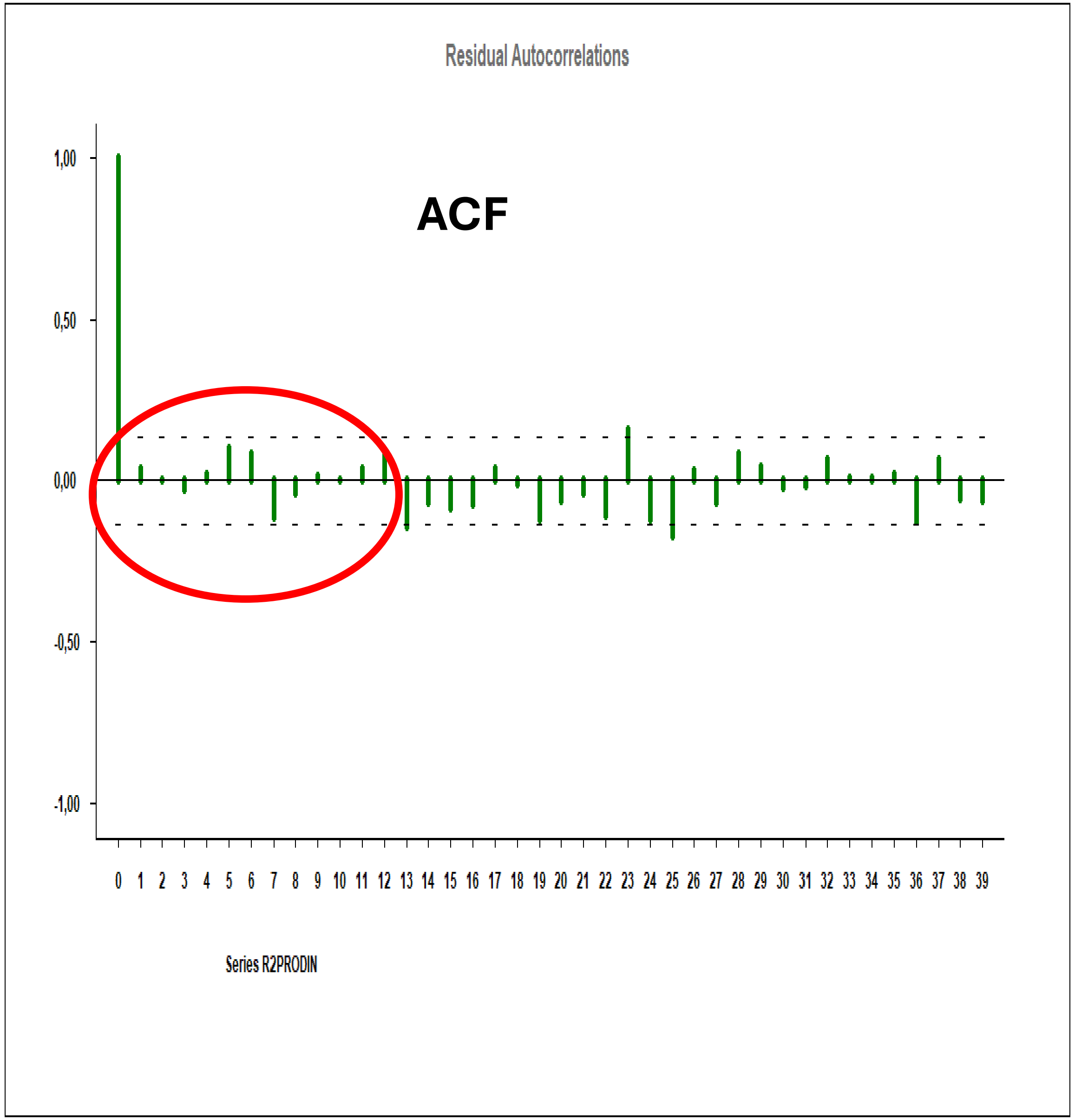
Fitted model (2)

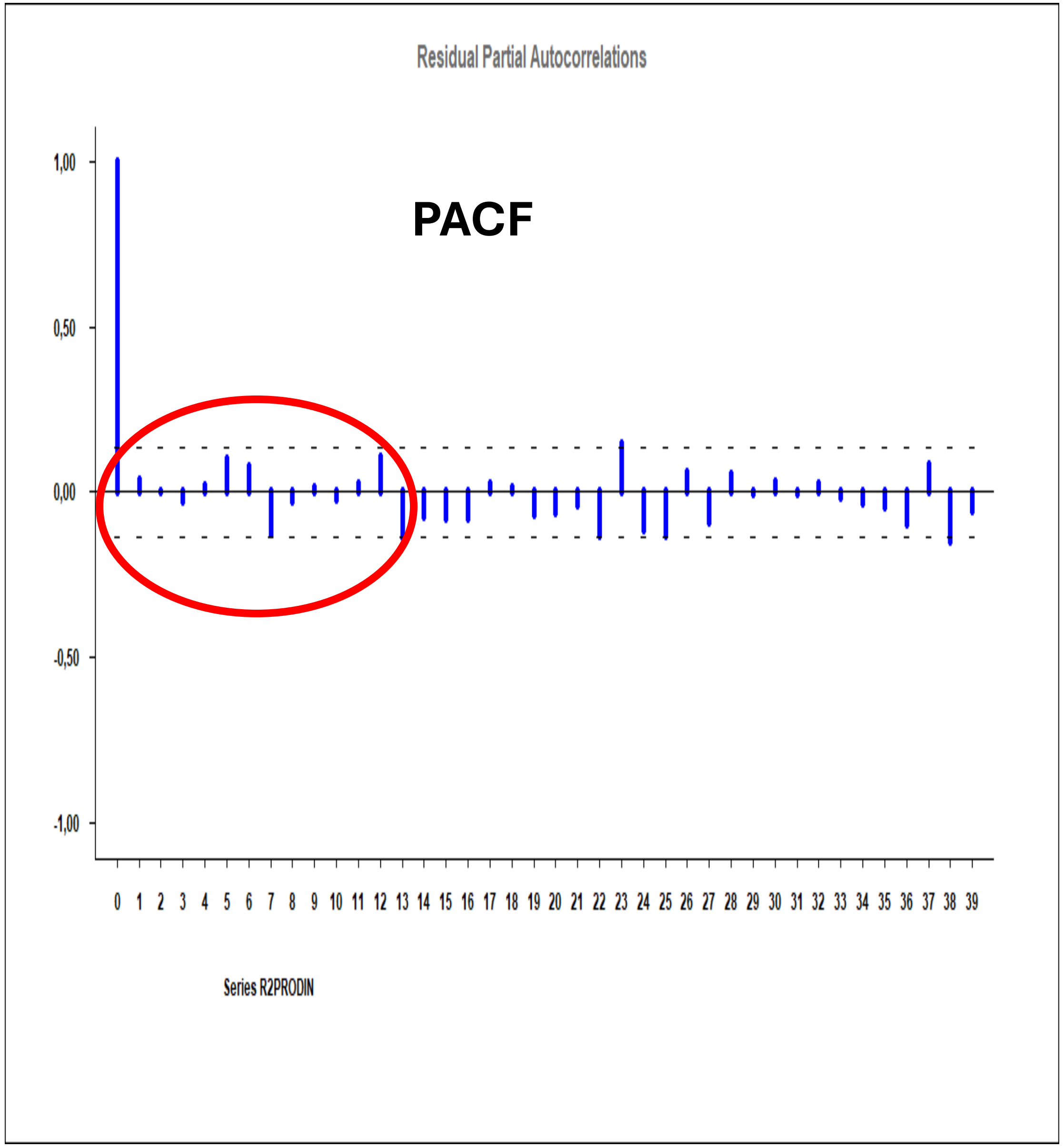
```
=== MODEL DESCRIPTION                        FORM      DEGREE/ORD PARAMETERS
 - SEASONAL PERIOD                                      12
 - DIFFERENCE                       REGULAR              1
 - DIFFERENCE                       SEASONAL             1
 - ARMA MODEL
   AUTOREGRESSIVE POLYNOMIAL REGULAR    2      AR   nn
   MOVING AVERAGE POLYNOMIAL SEASONAL   1      SMA  nn
 NON LINEAR ESTIMATION: …
FINAL VALUES OF THE PARAMETERS
   NAME      VALUE           STD ERROR    T-VALUE
1  AR    1   -.87746       5.64437E-02    -15.5
2  AR    2   -.58596       5.59854E-02    -10.5
3  SMA   1    .82206       4.27206E-02     19.2
 === ROOTS OF AR AND MA POLYNOMIALS <Z>
 AR ROOTS            MODULUS  PERIOD
    COMPLEX PAIR      1.306     2.88
 MA ROOTS            MODULUS  PERIOD
    REAL              1.016
 === SUMMARY MEASURES
 SUM OF SQUARES: COMPUTED = 2694.14 ADJUSTED 2694.14
 VARIANCE ESTIMATES: BIASED = 12.65 UNBIASED = 12.83
 NUMBER OF PARAMETERS: 3 STANDARD DEVIATION = 3.58179
=== RESIDUAL ANALYSIS WITH  213 RESIDUALS, BEGINNING AT TIME APR2001===
 MEAN = -.336177, T-STATISTIC = -1.37 (FOR TESTING ZERO MEAN)
```

- Seasonal period: 12
- Delta or Δ
- Delta12 or Δ_{12}

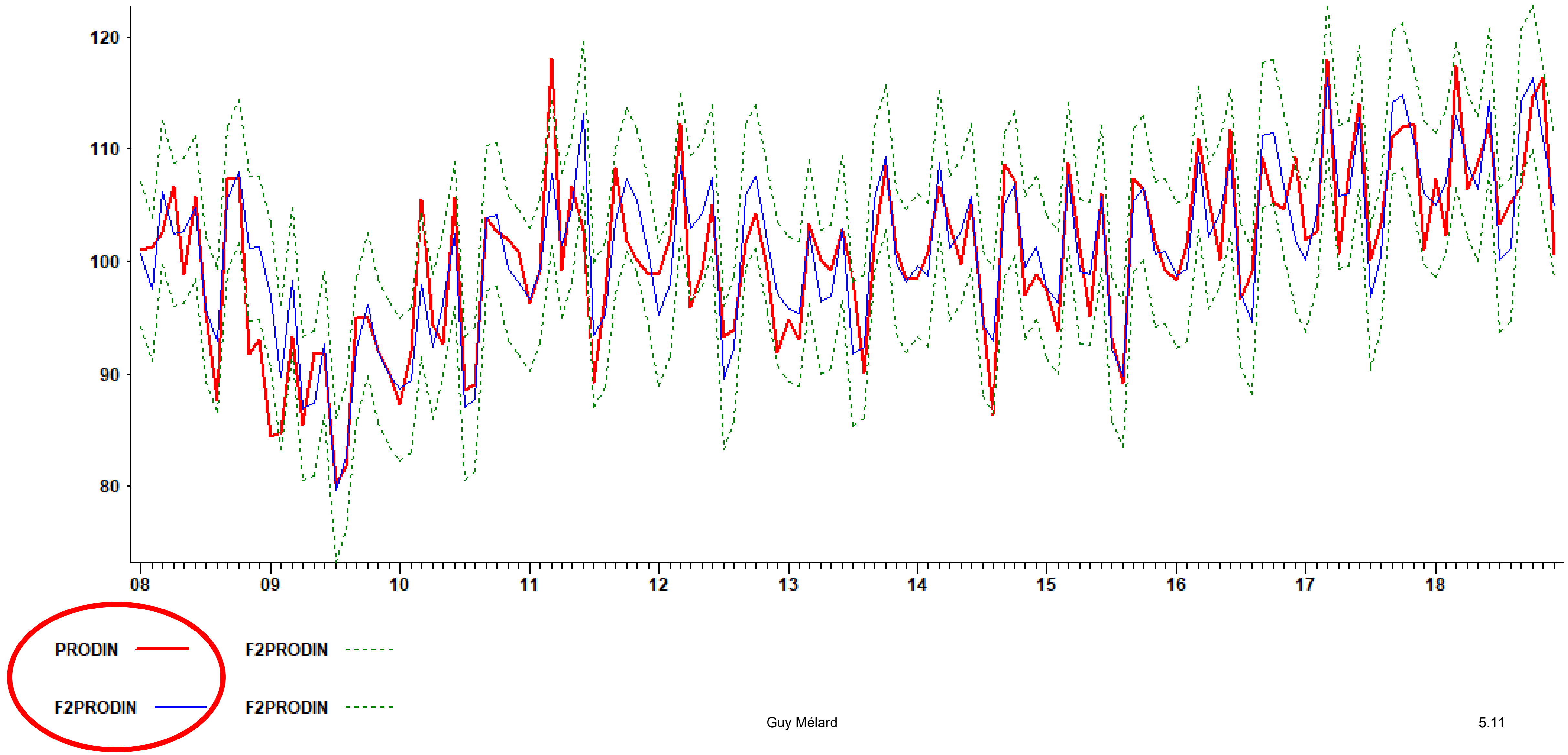
- $1 - (-0.877)\,L - (-0.586)\,L^2$
- $1 - (0.822)\,L^{12}$

- $\sigma^2 = 18{,}79$

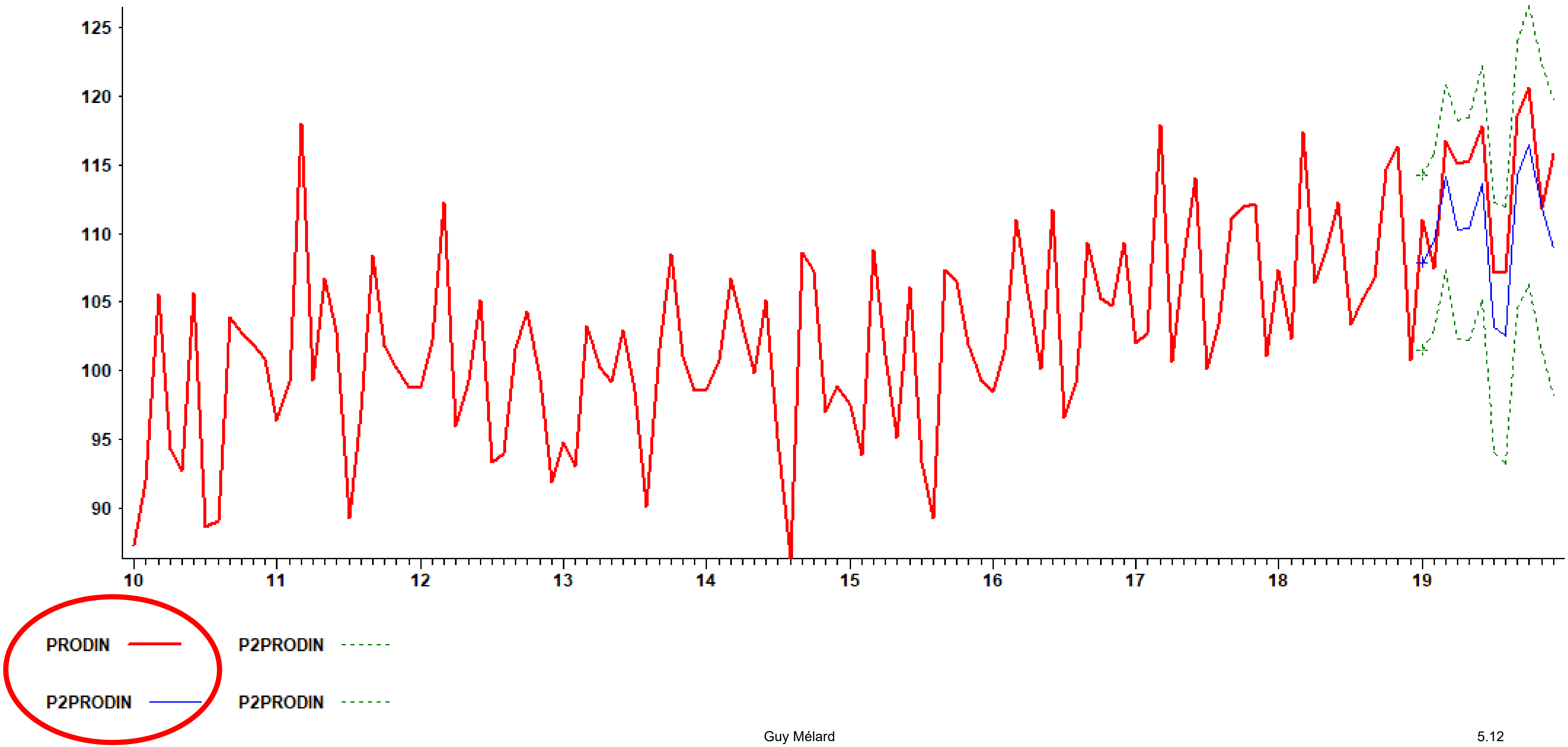

Residual Autocorrelations
ACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series R2PRODIN


Residual Partial Autocorrelations
PACF
1,00
0,50
0,00
-0,50
-1,00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Series R2PRODIN

Fitted value plot



Forecast plot



Conclusion of the statistical analysis

- For the Belgian CPI, we have found a very simple model

$$\Delta \mathrm{CPI}_t = \mathrm{CPI}_t - \mathrm{CPI}_{t-1} = \mu + \epsilon_t, \qquad (1)$$

where the estimate of μ is 0.123 and the errors ϵ_t have mean 0 and standard deviation 0.221. **Δ is the difference operator** (often denoted ∇)

- For the Belgian industrial production, we found a more complex model for $\widetilde{\mathrm{PRODIN}}_t = \Delta\Delta_{12}\mathrm{PRODIN}_t$, where $\Delta_{12}\mathrm{PRODIN}_t = \mathrm{PRODIN}_t - \mathrm{PRODIN}_{t-12}$ is the **seasonal difference operator**, hence

$$\widetilde{\mathrm{PRODIN}}_t = \Delta\Delta_{12}\mathrm{PRODIN}_t$$
$$= \mathrm{PRODIN}_t - \mathrm{PRODIN}_{t-1} - \mathrm{PRODIN}_{t-12} + \mathrm{PRODIN}_{t-13} \qquad (2)$$

- Using that notation, the series is represented by

$$\widetilde{\mathrm{PRODIN}}_t + 0.877\,\widetilde{\mathrm{PRODIN}}_{t-1} + 0.586\,\widetilde{\mathrm{PRODIN}}_{t-2} = \epsilon_t - 0.822\epsilon_{t-12} \qquad (3)$$

- To write it algebraically, we need to introduce polynomials in the **lag operator** L (often B instead) such that $Ly_t = y_{t-1}$, $L^2 y_t = y_{t-2}$, $L^{12}\epsilon_t = \epsilon_{t-12}$, etc
- In particular $\Delta = 1 - L$ and $\Delta_{12} = 1 - L^{12}$, and this explains (2) since $(1-L)(1-L^{12}) = 1 - L - L^{12} + L^{13}$

AR and MA polynomials

- With $\widetilde{\mathrm{PRODIN}}_t = \Delta\Delta_{12}\mathrm{PRODIN}_t$, we have seen the model equation (3)

$$\widetilde{\mathrm{PRODIN}}_t + 0.877\,\widetilde{\mathrm{PRODIN}}_{t-1} + 0.586\,\widetilde{\mathrm{PRODIN}}_{t-2} = \epsilon_t - 0.822\,\epsilon_{t-12}$$

- Let us define the **autoregressive** or **AR operator** $A(L) = 1 - a_1 L - a_2 L^2$, with $a_1 = -0.877$ and $a_2 = -0.586$, a polynomial of degree $p = 2$ acting on $\widetilde{\mathrm{PRODIN}}_t$ and the so-called **moving average** or **MA operator** $B(L) = 1 - b_{12} L^{12}$, $b_{12} = 0.822$, a polynomial of degree $q = 12$ (called a seasonal MA polynomial)
- We have therefore what is called an ARMA(2,12) model for $\Delta\Delta_{12}\mathrm{PRODIN}_t$ and it can be written in a one-line equation

$$[1 - (-0.877)L - (-0.586)L^2]\Delta\Delta_{12}\mathrm{PRODIN}_t = (1 - 0.822L^{12})\epsilon_t$$

- For an **ARMA**(p, q) **model**, the general equation is

$$(1 - a_1 L - \ldots - a_p L^p) y_t = (1 - b_1 L - \ldots - b_q L^q)\epsilon_t$$

we will denote the AR coefficients $-a_1, -a_2, \ldots, -a_p$ and the MA coefficients $-b_1, -b_2, \ldots, -b_q$ [The minus sign is purely conventional, mainly for the MA polynomial, Box & Jenkins (1970)]

Seasonal ARMA and ARIMA models

- More generally, we can have the regular and the seasonal AR polynomials and the regular and the seasonal MA polynomials, Box et al. (2015)
- If their respective degrees are denoted p, P, q, and Q, we can speak of a seasonal ARMA model of orders (p, q) and (P, Q) with period 12 here
- In our example $p = 2$, $P = 0$, $q = 0$, and $Q = 1$, hence a seasonal $\text{ARMA}(2, 0)(0, 1)_{12}$, where the subscript 12 reminds the seasonal period
- Also, noting that here we have $s = 1$ and $S = 1$ as degrees of the (regular) difference and seasonal difference, respectively, I can mention the now traditional **(seasonal) ARIMA** notation: $\text{ARIMA}(2, 1, 0)(0, 1, 1)_{12}$, where the middle integers refer to the differences
- The letter "I" refers to integration, the inverse operator of the difference. We will say that our series CPI and PRODIN are integrated. The latter is even seasonally integrated.
- In this introductory talk, we will only treat the products of the regular and seasonal polynomials and speak of the AR and MA polynomials
- So we have an ARMA(2, 12) model, with AR coefficients $a_1 = -0.877$ and $a_2 = -0.586$, and MA coefficients $b_1 = 0$, $b_2 = 0$, ..., $b_{11} = 0$, and $b_{12} = 0.822$

Digression on statistics

- Most statistical theories and applications are about **random samples**, meaning that the observations are **independent** of each other
- This is not true in time series: there is no reason why the Belgian industrial production of this month is independent of that in the previous months
- We can perhaps consider the realizations of the ϵ_t as being independent like the differences of the Belgian CPI_t
- This is the basis of the theory of time series which treats them as realizations of a **stochastic process**, a sequence of possibly dependent random variables
- Usually, the theory is about **stationary stochastic processes**
- A stationary stochastic process is a sequence of possibly dependent random variables that have the **same mean** and the **same variance** (+ a condition on lag-dependency, see later)
- The differences of CPI_t, $t = 1, \ldots, n$, and the residuals of the model for $\Delta\Delta_{12}\text{PRODIN}_t$, $t = 1, \ldots, n$, can perhaps be considered as realizations of a stationary stochastic process, but neither CPI_t, $t = 1, \ldots, n$, nor PRODIN_t, $t = 1, \ldots, n$, because of the **trend**, the **variations in level**, and/or the **seasonality**

Differences versus AR and MA polynomials

- We come back to the model for PRODIN
- The difference operators have the particularity that they have a **unit root**: $\Delta = 1 - L$ has root 1 while $\Delta_{12} = 1 - L^{12}$ has 12 roots including two real ones 1 and -1 and 10 complex roots, the 12-th complex roots of 1
- The AR and MA operators have the particularity that they have **roots larger than 1** (in modulus)
- In the example, for the AR polynomial $[1 - (-0,877)L - (-0,586)L^2]$ with complex roots since $\delta = 0.877^2 - 4 * 0.586 = -1.574 < 0$: $-0.439 \pm 0.627i$ with product $1/0.586 = 1.707 = 1.306^2$, and for the MA polynomial $(1 - 0,822L^{12})$, the 12-th complex roots of $1/0.822 = 1.016^{12} > 1$
- Indeed, an ARMA model like the one defined by the model for $z_t = \Delta\Delta_{12}\widetilde{\mathrm{PRODIN}}_t =$

$$(1 - (-0.877)L - (-0.586)L^2)z_t = (1 - 0.822L^{12})\epsilon_t \qquad (4)$$

  is considered as a stationary stochastic process, if we suppose that the ϵ_t, $t = 1, ..., n$, are independent random variables with mean 0 and a constant variance $\sigma^2 = 12.83$
- Note that no AR root is an MA root and vice-versa
- If we have $(1 - 0.5L)y_t = (1 - 0.5L)\epsilon_t$, a simpler model is $y_t = \epsilon_t$ and it is not possible to estimate a_1 and b_1. This is **non-identifiability**

Parameter estimation

- We are in a statistical context
- We have a **sample of n observations** of a time series $y_t, t = 1, ..., n$
- We want to infer the **true unknown population**
- For instance, find a model for PRODIN and estimate its parameters
- Let us denote the **parameters** by θ and their unknown **true value** by θ^0 (as a matter of fact, we even do not know the true model, if it exists)
- In simpler contexts (estimation of the mean, of a correlation or regression coefficient), we can invoke principles like **least-squares** (minimizing the sum of squared errors), **maximum likelihood**, etc.
- For instance, if we suppose $y_t = a_1 y_{t-1} + \epsilon_t$, we write $y_t - \theta y_{t-1} = e_t(\theta)$; an estimate of θ is obtained by minimizing the sum of $e_t^2(\theta)$ where the residual $e_t(\theta) = y_t - \theta y_{t-1}$, and this yields an estimate $\widehat{\theta} = (\sum_{t=2}^{n} y_t y_{t-1})/(\sum_{t=2}^{n} y_{t-1}^2)$
- More generally, obtaining the **residuals** $e_t(\theta)$ and being able to obtain **their derivatives** with respect to θ is crucial
- In time series analysis, we use least squares or, better, the Gaussian likelihood method that has the least starting effects
- In general, except for AR(p) models which are linear, we need to use **numerical optimization**

Implications

- It should be clear that most (not all) of the statistical results in time series suppose a **stationary stochastic process** behind
- This is the case, in particular, for the ACF and PACF presented quickly at the beginning but also for the estimation method that led to our estimates (based on a maximum likelihood principle) that was largely skipped
- Moreover, statistical estimation results are often (if not always) accepted if and only if they are supported by
  - a **law of large numbers**: convergence in some sense of the estimator $\widehat{\theta}_n$ to the true value θ when $n \to \infty$, and
  - a **central limit theorem**: the difference $\widehat{\theta}_n - \theta$ times $\sqrt{n}$ converges in law to a normal distribution when $n \to \infty$
- These results are standard for most of statistics but were harder to obtain for time series and practically only under the **stationarity assumption**
- There is well a theory for testing the presence of a unit root in an AR polynomial but it rests on more advanced results in probability theory (Brownian and Wiener processes)
- As said above, the models for CPI and for PRODIN are not stationary, only those for ΔCPI and $\Delta\Delta_{12}$PRODIN can be seen as stationary

MA form

- Let $(1 - a_1 L - \ldots - a_p L^p) y_t = (1 - b_1 L - \ldots - b_q L^q)\epsilon_t$
- An **MA form** consists of writing y_t as a function of present and past ϵ_t's
- Let θ be the set of parameters, the a_j's and the b_j's, θ^0 their true value
- We imagine that the error ϵ_t can be computed by the recurrence

$$\epsilon_t = y_t - a_1 y_{t-1} - \ldots - a_p y_{t-p} + b_1 \epsilon_{t-1} + \ldots + b_q \epsilon_{t-q}$$

- Since the ϵ_t are unknown we consider the residuals

$$e_t(\theta) = y_t - a_1 y_{t-1} - \ldots - a_p y_{t-p} + b_1 e_{t-1}(\theta) + \ldots + b_q e_{t-q}(\theta)$$

- Note $e_t(\theta^0) = \epsilon_t$. The derivatives of $e_t(\theta)$ with respect to θ_j, $j = 1, \ldots, m$, are less easy to obtain because we have products for the MA part
- There is one case which is easy to treat: an AR(1) process defined by $y_t = a_1 y_{t-1} + \epsilon_t$, hence $\theta^0 = a_1$, the model is $y_t = \theta y_{t-1} + e_t(\theta)$, and

$$\begin{aligned} e_t(\theta) &= y_t - \theta y_{t-1} = \epsilon_t + (a_1 - \theta) y_{t-1} \\ &= \epsilon_t + (a_1 - \theta)\epsilon_{t-1} + a_1(a_1 - \theta)\epsilon_{t-2} + a_1^2(a_1 - \theta)\epsilon_{t-3} + \ldots \qquad (5) \end{aligned}$$

- For the derivatives $\partial e_t(\theta)/\partial\theta$, since the y_t's and ϕ don't depend on θ:

$$\frac{\partial e_t(\theta)}{\partial \theta} = -\epsilon_{t-1} - a_1 \epsilon_{t-2} - a_1^2 \epsilon_{t-3} - \ldots \qquad (6)$$

and we have an **exponential decrease** for both (5) and (6) if $|a_1| < 1$

And multivariate time series?

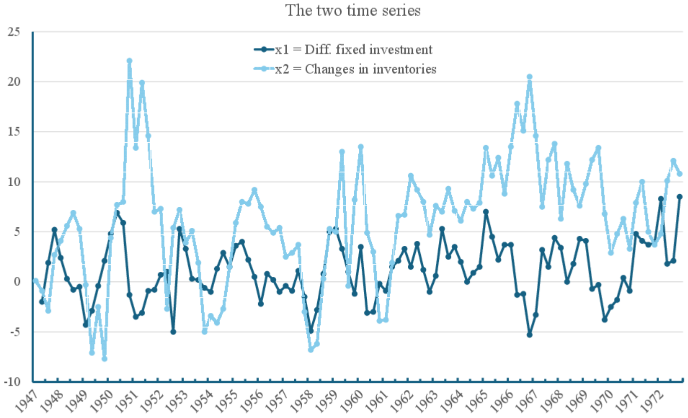
- Each time series can be studied individually but a collective/multivariate analysis can improve the results
- In particular, for related time series like housing sales & housing starts
- We will not analyze them, simply extend the ARMA models to vector ARMA or VARMA models
- If the observations at time t are a $r \times 1$ vector y_t, we replace the scalar AR coefficients $a_1, a_2, \ldots, a_p$ and the scalar MA coefficients $b_1, b_2, \ldots, b_q$ by $r \times r$ matrices (and the constant 1 by I_r, the $r \times r$ identity matrix)
- It is intentional that we use the same notations a_j, $j = 1, \ldots, p$, and b_j, $j = 1, \ldots, q$, either for scalar or **matrix AR and MA**, respectively
- Of course, the ϵ_t will be independent $r \times 1$ vectors with now a covariance matrix Σ (symmetric and strictly positive definite)
- Let θ be the set of parameters (the entries in the a_j's and the b_j's), the residual $e_t(\theta)$ at time t can be obtained by recurrence

$$e_t(\theta) = y_t - a_1 y_{t-1} - \ldots - a_p y_{t-p} + b_1 e_{t-1}(\theta) + \ldots + b_q e_{t-q}(\theta)$$

- Again, the derivative of $e_t(\theta)$ with respect to θ_j, $j = 1, \ldots, m$ are less easy to obtain because we have products for the MA part

An example - 1

The U.S. business investment (in diff.) and variations of inventories data first studied by Lütkepohl (2005, Section 3.8.2) are quarterly seasonally adjusted data over the period from the first quarter of 1947 to the fourth of 1972. They are used for several purposes, including the illustration of a **VARMA(1,1) model** by Reinsel (1998).



An example - 2

We considered the series up to 1971, quarter 4.
We start from the VARMA(1, 1) model fitted by Box *et al.* (2015, Section 14.7.2) using the MTS package in R:

$$(y_t - \mu) = a_1(x_{t-1} - \mu) + \epsilon_t - b_1\epsilon_{t-1}.$$

Instead of a conditional estimation method, we preferred to use an exact maximum likelihood estimation method and obtained the following estimates **(with *t* statistics)**:

$$\widehat{a}_1 = \begin{pmatrix} \underset{(2.38)}{0.438} & \underset{(-2.92)}{-0.196} \\ \underset{(3.01)}{0.645} & \underset{(9.52)}{0.765} \end{pmatrix}, \quad \widehat{b}_1 = \begin{pmatrix} \underset{(-0.21)}{-0.041} & \underset{(-3.69)}{-0.311} \\ \underset{(1.13)}{0.328} & \underset{(1.56)}{0.205} \end{pmatrix},$$

with the estimate of Σ

$$\widehat{\Sigma} = \begin{pmatrix} 5.4660 & 1.8857 \\ 1.8857 & 18.4219 \end{pmatrix}.$$

The (not mentioned) mean vector μ is taken as the sample average $(0.9737, 6.0232)^T$.
Let the vector of parameters (of interest) as θ, here the entries in a_1 and b_1, plus μ.
The 3 entries in Σ are nuisance parameters, estimated a posteriori

Time-dependent ARMA model

- This is part of our own research (since 1981!), first if $r = 1$
- We replace the constant coefficients with **time-dependent (td) coefficients**, for instance for an autoregressive polynomial: $A_t(L) = 1 - a_{t1}L - a_{t2}L^2$
- That leads to **tdARMA** and tdARIMA models with an estimation algorithm proposed in 1982
- These coefficients can even depend on the **series length** n, for instance $a_{t1} = a'_1 + (t/n)a''_1$, leading to tdARMA$^{(n)}$. The vector of parameters θ includes a'_1 and a''_1
- I don't show the PRODIN time series: the results are not conclusive
- In the VARMA(1, 1) example, the results are more interesting: replacing the constant coefficients with linear functions of time, e.g., $a_t = a' + a''(t - 50)/98$, we obtained the estimates (with t statistics[1]):

$$\widehat{a}'_1 : \begin{pmatrix} \underset{(2.272)}{0.421} & \underset{(-2.946)}{-0.199} \\ \underset{(2.903)}{0.571} & \underset{(10.590)}{0.792} \end{pmatrix}, \quad \widehat{a}''_1 : \begin{pmatrix} \underset{(---)}{0} & \underset{(---)}{0} \\ \underset{(-2.181)}{-0.780} & \underset{(---)}{0.} \end{pmatrix}, \quad \widehat{b}'_1 : \begin{pmatrix} \underset{(0.281)}{0.060} & \underset{(3.689)}{0.318} \\ \underset{(-0.639)}{-0.178} & \underset{(-2.027)}{-0.298} \end{pmatrix}$$

[1] But is it valid? Introducing b''_1 also failed

Time-dependent ARMA model: implications

- Big problem: even after differences, the underlying process is **not stationary** hence the whole **asymptotic theory** has to be reinvented using deeper probability results of nonstandard
  - a **law of large numbers** (one published in 2009) and
  - a **central limit theorem**
- It was done in steps: Azrak thesis for tdAR(p) in 1996, Azrak-M (2006) for tdARMA(p, q), Alj, Azrak, Ley, & M (2017) for tdVARMA, Alj, Azrak, & M (2024) for tdVARMA$^{(n)}$
- The main tool was already mentioned for an AR(1) model with a constant coefficient: the infinite MA form in the ϵ_{t-k}, writing a development of the residual $e_t(\theta)$ and its derivative with respect to the parameter θ_j, $j = 1, ..., m$, where m is the number of parameters in the model:

$$\begin{aligned} e_t(\theta) &= \epsilon_t + \psi_{t1}\epsilon_{t-1} + \psi_{t2}\epsilon_{t-2} + \ldots + \psi_{tk}\epsilon_{t-k} + \ldots \\ \frac{\partial e_t(\theta)}{\partial \theta_j} &= \psi_{tj1}\epsilon_{t-1} + \psi_{tj2}\epsilon_{t-2} + \ldots + \psi_{tjk}\epsilon_{t-k} + \ldots \end{aligned}$$

- We should have **td coefficients decreasing exponentially** with k
- We said it is already difficult for ARMA(p, q) models, not to say VARMA(p, q) and tdVARMA(p, q) models

MA form in the td case

- For a tdAR(1) model defined by $y_t = a_{t1}(\theta)y_{t-1} + \epsilon_t$:

$$y_t = \epsilon_t + a_{t1}(\theta)y_{t-1} = \epsilon_t + a_{t1}(\theta)\epsilon_{t-1} + a_{t1}(\theta)a_{t-1,1}(\theta)\epsilon_{t-2} + \ldots \quad (7)$$

  so no longer necessarily an exponential decrease but well products of coefficients for different lags: for ϵ_{t-k}, we have $a_{t1}a_{t-1,1}\ldots a_{t-k+1,1}$
- Similarly for the residuals $e_t(\theta)$ and (but more complex) their derivatives $\partial e_t(\theta)/\partial\theta_j$ with respect to parameter θ_j (non longer a coefficient)
- Of course, it is even more complex for tdARMA(p, q) models
- Note, however, that (7) still holds for matrix coefficients of a tdVAR(1) model, simply replacing products of scalars by products of matrices
- The key is putting a tdARMA(p, q) model into a **tdVAR(1) form**
- Strangely, the idea came when trying to handle VARMA(p, q) models as special cases of tdARMA(p, q)
- For that purpose, we need to introduce the companion matrix of a polynomial

Companion matrix

- Let us go back first to the scalar case, $r = 1$
- Let us consider a **monic polynomial** of degree d: $p(x) = x^d + c_{d-1}x^{d-1} + \ldots + c_1 x + c_0$. Its **companion matrix** is (at least one form, others can be considered)

$$C(p) = \begin{pmatrix} -c_{d-1} & \cdots & -c_1 & -c_0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \tag{8}$$

- The **roots** of p are the **eigenvalues** of that matrix
- The **reciprocal polynomial** of p is defined by $p^*(x) = x^d p(1/x) = 1 + c_{d-1}x + \ldots + c_1 x^{d-1} + c_0 x^d$
- Note that it is not monic: it has the form of an AR or MA polynomial
- The roots of p^* are the inverse of those of p
- We will need the **Frobenius norm** of a matrix M: $||M||_F = \sqrt{\mathrm{tr}(M^T M)}$, where ^{T} indicates transposition. For example, $||M||_F = \sqrt{c_0^2 + c_1^2 + \ldots + c_{d-1}^2 + d - 1}$

ULB

Companion matrix: example and generalization

- The AR and MA polynomials are not monic so we work with the reciprocal polynomials $p^*(x) = 1 + c_{d-1}x + \ldots + c_0 x^d$
- For an **AR polynomial of degree 1**, $A(x) = 1 - a_1 x$ with root $1/a_1$, and its reciprocal polynomial $A^*(x) = x - a_1$: $C(A) = a_1$, with eigenvalue $1/a_1$
- Let an **AR polynomial of degree 2**, $A(x) = 1 - a_1 x - a_2 x^2$, and its reciprocal polynomial $A^*(x) = x^2 - a_1 x - a_2$: $C(A) = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix}$
- The eigenvalues of $C(A)$ are the solutions of

$$\det\left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix}\right\} = \det\left\{\begin{pmatrix} 1-\lambda a_1 & -\lambda a_2 \\ -\lambda & 1 \end{pmatrix}\right\} = 0$$

  or $1 - \lambda a_1 - \lambda^2 a_2 = 0$. These are $-(a_1 \pm \sqrt{a_1^2 + 4a_2})/2a_2$ with sum $-a_1/a_2$ and product $-1/a_2$
- If these solutions are complex, the condition for roots greater than 1 in modulus is that $|a_2| < 1$
- We can generalize the companion matrix to **matrix polynomials**, e.g., $A(x) = I_r - a_1 x - a_2 x^2$ or $A^*(x) = x^2 - a_1 x - a_2$ but not the roots of matrix polynomials: we need to consider the roots of $\det(A(x))$ or those of $\det(A^*(x))$ and they are called the **eigenvalues of these matrix polynomials**

VAR(1) for of an ARMA(p, q) model

- This is well known (*). Let us consider, for instance, an **ARMA(2, 2) model** for the process $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \epsilon_t - b_1 \epsilon_{t-1} - b_2 \epsilon_{t-2}$:
  $y_t = a_1(\theta) y_{t-1} + a_2(\theta) y_{t-2} + e_t(\theta) - b_1(\theta) e_{t-1}(\theta) - b_2(\theta) e_{t-2}(\theta)$
- We define $Y_t(\theta) = (y_t, y_{t-1}, e_t(\theta), e_{t-1}(\theta))^T$, $E_t(\theta) = (e_t(\theta), 0, e_t(\theta), 0)^T$, and $E_t = E_t(\theta^0) = (\epsilon_t, 0, \epsilon_t, 0)^T$
- We can write $Y_t(\theta) = \mathcal{A}(\theta) Y_{t-1}(\theta) + E_t(\theta)$, more precisely

$$\begin{pmatrix} y_t \\ y_{t-1} \\ e_t(\theta) \\ e_{t-1}(\theta) \end{pmatrix} = \left(\begin{array}{cc|cc} a_1(\theta) & a_2(\theta) & -b_1(\theta) & -b_2(\theta) \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ e_{t-1}(\theta) \\ e_{t-2}(\theta) \end{pmatrix} + \begin{pmatrix} e_t(\theta) \\ 0 \\ e_t(\theta) \\ 0 \end{pmatrix} \quad (9)$$

- For an **ARMA(p, q)**, $\mathcal{A}(\theta)$ is a $(p+q) \times (p+q)$ matrix
  $\mathcal{A}(\theta) = \begin{pmatrix} C(A(\theta)) & \tilde{C}(-B(\theta)) \\ 0_{qp} & S_{qq} \end{pmatrix}$ where $C(A(\theta))$ is the companion matrix of the (reciprocal) AR polynomial, $\tilde{C}(-B(\theta))$ is a zero matrix with a first row like the companion matrix of minus the (reciprocal) MA polynomial or $-B(\theta)$, 0_{qp} is a zero matrix, and S_{qq} is a **lower-shifted identity matrix**
- Note that (9) remains valid for matrix instead of scalar coefficients but using **block matrices**, then $\mathcal{A}(\theta)$ is a $(p+q) \times (p+q)$ block matrix with $r \times r$ blocks

(*)References for VAR(1) form for ARMA(p, q) models

Taken from M (2022, p. 99)

- A representation of an ARMA model in tdVAR(1) form is not new, see Lütkepohl (1991, 2005, pp. 616-617)
- It was used by Francq and Gautier (2004) for tdARMA models and was detailed in a working paper by Francq and Gautier (2003). It was described there using a state-space representation
- Note that Francq and Zakoïan (2001) propose a similar technique for building a Markovian representation for Markov-switching VARMA models. See also Boubacar Maïnassara and Rabehasaina (2020)
- The purpose of these authors was to obtain a unique strictly stationary solution.
- In a sense, we combine features from those two papers by Francq and Gautier (2004) and Francq and Zakoïan (2001)

tdVAR(1) for a tdARMA(p, q) model

- **It is the same!** Let us consider, for instance, a **tdARMA(2, 2) model** for the process $y_t = a_{t1}y_{t-1} + a_{t2}y_{t-2} + \epsilon_t - b_{t1}\epsilon_{t-1} - b_{t2}\epsilon_{t-2}$:
  $y_t = a_{t1}(\theta)y_{t-1} + a_{t2}(\theta)y_{t-2} + e_t(\theta) - b_{t1}(\theta)e_{t-1}(\theta) - b_{t2}(\theta)e_{t-2}(\theta)$
- We can write $Y_t(\theta) = \mathcal{A}_t(\theta)Y_{t-1}(\theta) + E_t(\theta)$, more precisely

$$\begin{pmatrix} y_t \\ \dots \\ y_{t-p+1} \\ e_t(\theta) \\ \dots \\ e_{t-q+1}(\theta) \end{pmatrix} = \begin{pmatrix} C(A_t(\theta)) & \tilde{C}(-B_t(\theta)) \\ & \\ 0_{qp} & S_{qq} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \dots \\ y_{t-p} \\ e_{t-1}(\theta) \\ \dots \\ e_{t-q}(\theta) \end{pmatrix} + \begin{pmatrix} e_t(\theta) \\ 0 \\ \dots \\ e_t(\theta) \\ 0 \\ \dots \end{pmatrix} \quad (10)$$

  where $C(A_t(\theta))$ is the companion matrix of the (reciprocal) tdAR polynomial $A_t(\theta)$, $\tilde{C}(-B_t(\theta))$ is a zero matrix with a first row like the companion matrix of minus the (reciprocal) tdMA polynomial or $-B_t(\theta)$, 0_{qp} is a zero matrix, and S_{qq} is a lower-shifted identity matrix
- Note again that (10) remains valid for matrix instead of scalar coefficients but using block matrices, and $\mathcal{A}_t(\theta)$ is a $(p+q) \times (p+q)$ block matrix with $r \times r$ blocks

Treatment of ARMA models

- We start from the VAR(1) form for $Y_t(\theta)$: $Y_t(\theta) = \mathcal{A}(\theta)Y_{t-1}(\theta) + E_t(\theta)$
- Our aim: a tdMA form for $e_t(\theta)$. Let $E_t = E_t(\theta^0) = (\epsilon_t, 0, \dots, \epsilon_t, 0, \dots 0)^T$
- For an ARMA(2,2), we define two constant matrices

$$J = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- So $\mathcal{A}(\theta) = \begin{pmatrix} C(A(\theta)) & \tilde{C}(-B(\theta)) \\ 0_{qp} & S_{qq} \end{pmatrix}$, $J\mathcal{A}(\theta) = \begin{pmatrix} S_{pp} & 0_{pq} \\ \tilde{C}(-A(\theta)) & C(B(\theta)) \end{pmatrix}$, with **symmetric notations** for $\tilde{C}(A(\theta))$ and $C(B(\theta))$. It can be shown that

$$Y_t(\theta) = \sum_{k=0}^{t-1} \Psi_k(\theta)E_{t-k}, \text{ where } \Psi_k(\theta) = \sum_{s=0}^{k} (J\mathcal{A}(\theta))^{k-s} K(\mathcal{A})^s$$

- But we have

$$(\mathcal{A}(\theta))^k = \begin{pmatrix} C^k(A(\theta)) & \tilde{C}_k(-B(\theta)) \\ 0_{qp} & S_{qq}^k \end{pmatrix}, (J\mathcal{A}(\theta))^k = \begin{pmatrix} S_{pp}^k & 0_{pq} \\ \tilde{C}_k(-A(\theta)) & C^k(B(\theta)) \end{pmatrix}$$

where we do not detail $\tilde{C}_k(-B(\theta))$ nor $\tilde{C}_k(-A(\theta))$ but $S_{qq}^k = 0_{qq}$ and $S_{pp}^k = 0_{pp}$ for $k > p + q$. Note the presence of $C^k(A(\theta))$ and $C^k(B(\theta))$

VARMA models and derivatives of $\Psi_k(\theta)$

- Let $\mathcal{A} = \mathcal{A}(\theta^0)$. Then, if all the roots of $A(\theta^0)$ and $B(\theta^0)$ are greater than $1/\Phi$ in modulus, with $\Phi < 1$, then the Frobenius norm of $(\mathcal{A})^k$ and $(J\mathcal{A})^k$ are smaller than $c\Phi^k$ (c constant); the same can be deduced for $\Psi_k(\theta^0)$
- Once the $\Psi_k(\theta^0)$ are obtained, there is no problem with obtaining an MA form for $e_t(\theta)$ since $Y_t(\theta) = (y_t, y_{t-1}, e_t(\theta), e_{t-1}(\theta))^T$, with the same exponentially decreasing property for the coefficients $\psi_k(\theta^0)$
- First, there is no problem to have an ARMA(p, q) model
- Second, it works also for a VARMA(p, q) model, by replacing 1 with I_r, K, J, and polynomial roots with eigenvalues
- Third, the **exponential decrease** holds also for the derivatives of $\Psi_k(\theta)$ and those of $\psi_k(\theta)$, at $\theta = \theta^0$
- This, together with other conditions that cannot be detailed here (identifiability, see later; existence of fourth-order moments for the ϵ_t; ...), it is possible to prove the asymptotic properties of convergence and normality of the estimator
- It is worth noting that the present proof is based on similar arguments but for time-dependent VARMA models, see next slide or M (2022)
- Note that the standard proof is somewhat longer than mine

Time-dependent VARMA or tdVARMA models

- Indeed, nearly everything in the last two slides could be done for tdVARMA(p, q) models on the basis of a tdVAR(1) form
- We replace everywhere the matrix polynomials $A(\theta)$ and $B(\theta)$ with $A_t(\theta)$ and $B_t(\theta)$, respectively, and of course, Ψ_k with Ψ_{tk}, ψ_k with ψ_{tk}
- Then, we use the theory developed by Alj, Azrak, Ley, & M (2017)
- The coefficients can also depend on the length n of the series but then we have to rely to another paper by Alj, Azrak, and M (2024) and M (2024)
- Instead of the complex assumptions given there, we suppose a **sufficient condition**: the **roots** of $\det(A_t(\theta^0))$ and $\det(B_t(\theta^0))$ are **greater than** 1
- We cannot prove all the assumptions in that paper so some of them are still there but are not interesting given on our focus to the polynomials
- One can wonder if the time-dependent models are really **useful in practice**? The answer: yes, but the improvement is not always sensible
- For univariate models, M (2023) has analyzed a dataset of industrial production in the US and obtained that about one-half of the series benefit from time-dependent ARMA models although the forecasts obtained are rarely much better
- There is presently no analog study for multivariate time series
- We show now some other aspects of polynomials: (i) the equality of roots or eigenvalues, (ii) the information matrix, and (iii) the theoretical ACF

Equality of roots

- Suppose we have two (scalar) polynomials $A(x)$ and $B(x)$ of respective degrees p and q, say $A(x) = a_0 + a_1 x + \ldots + a_{p-1}x^{p-1} + a_p x^p$ and $B(x) = b_0 + b_1 x + \ldots + b_{q-1}x^{q-1} + b_q x^q$. Obtaining the exact roots of polynomials can only be done for degrees at most 4. Otherwise, it has to be done numerically and it is often challenging
- On the contrary, checking if two polynomials have at least a common root is easy to do whatever their degrees
- It makes use of a **Sylvester matrix** associated with the two polynomials: a square $(p+q)$ matrix, obtained from the coefficients and shifts of them.

Example: $p = 3$, $q = 2$: $S_{pq}(A, B) = \begin{pmatrix} a_3 & a_2 & a_1 & a_0 & 0 \\ 0 & a_3 & a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 & 0 & 0 \\ 0 & b_2 & b_1 & b_0 & 0 \\ 0 & 0 & b_2 & b_1 & b_0 \end{pmatrix}$

- Its determinant is called the **resultant** of the two polynomials
- The resultant is zero if the polynomials have **at least a common root**
- More generally, the rank of the Sylvester matrix is related to the degree of the greatest common divisor of the two polynomials
- It can be used to check for **identifiability**: no common root for AR and MA

Generalization to matrix polynomials

- The purpose is to check if two **matrix polynomials** $A(x)$ and $B(x)$ of respective degrees p and q, and common dimension r, have at least a common eigenvalue (root of their determinant)
- It can be seen that the Sylvester matrix is then not useful
- It should be replaced by a so-called **tensor Sylvester matrix**: $S^{\otimes}_{pq}(A, B) = S_{pq}(I_r \otimes A, B \otimes I_r)$, where $\otimes$ represents the **Kronecker product** between matrices M_{ns} and N_{pq} which is the $np \times sq$ matrix

$$M \otimes N = \begin{pmatrix} m_{11}N & \dots & m_{1s}N \\ \vdots & \ddots & \vdots \\ m_{n1}N & \dots & m_{ns}N \end{pmatrix}$$

- Similarly as when $r = 1$, $S^{\otimes}_{pq}(A, B)$ is a resultant and is singular when there is at least one common eigenvalue between A and B
- Note, however, that identifiability for VARMA models is that the AR and MA polynomials do not have a common (non-constant) left factor
- No common eigenvalues guarantee identifiability but there can be common eigenvalues between the two matrix polynomials. So we have only a **sufficient condition of identifiability**
- Finally, this is not for time-dependent ARMA or VARMA models

Information matrix

- We don't have spoken of the (Fisher) **information matrix** until now
- We have well mentioned, under some assumptions (including on the roots of the AR and MA polynomials), asymptotic normality without details
- As a matter of fact, it is $\sqrt{n}(\widehat{\theta}_n - \theta^0) \rightarrow N(0, V^{-1})$ when $n \rightarrow \infty$ in distribution, where the **asymptotic covariance matrix** V^{-1} is the **inverse of the information matrix** V
- In practice, an estimate of V is obtained as a by-product of numerical estimation but here are alternative approaches
- The information matrix $V(\theta)$ is defined as a mathematical expectation at θ of the matrix defined by $(\partial e_t(\theta)/\partial\theta^T)^T \Sigma^{-1} (\partial e_t(\theta)/\partial\theta^T)$
- We consider here a **Gaussian stationary ARMA or VARMA**(p, q) **model**; then $V(\theta)$ is the same for all t
- Then, it is also possible to obtain the $V(\theta)$ as an **integral of a matrix composed of rational functions** with polynomials involving the AR and MA polynomials (or entries if $r > 1$)
- With co-author Klein, since 1989, I have developed algorithms for computing the information matrix for univariate models ($r = 1$)
- These integrals can be computed using **recurrences** with polynomials of **decreasing degrees**

An algorithm for VARMA models

- More recently, Klein and M (2023) have published an algorithm for **Mathematica**, the program for symbolic mathematical computation, see the next slide
- It is for ARMA and even VARMA models
- Advantages: it is short (see next slide) and **exact**; inconveniences: the entries need to be entered as rational numbers, not decimal numbers, and it takes much time
- Indeed, integration is performed **symbolically**, not numerically
- It even works with symbolic entries but then still slower
- Note that there is a generalization for VARMAX models (VARMA models with added regressors) with two matrix integrals instead of one
- The resulting information matrix is denoted Fcal on the next slide
- Like our other algorithms, it is not for time-dependent models
- I have also produced code for other (open-source) symbolic software packages like **Maxima** and **Octave** (a clone of Matlab), not using integration but well calculations of residues (Cauchy) or with the old Söderstrom (1984) algorithm we used in the 1990s but symbolically now

Mathematica program of the information matrix

```
In[126]:= Gcal[z_] =
           Together[ArrayFlatten[{{KroneckerProduct[up[z], -Inverse[A[z]].B[z]]},
              {KroneckerProduct[uq[z], In]}}]]

In[127]:= sigma[z_] = Together[Transpose[Inverse[B[z]]].Inverse[Sigma].Inverse[B[1/z]]]

In[128]:= Pcal[z_] = Together[Gcal[z].Sigma.Transpose[Gcal[1/z]]]

In[129]:= integrand1[z_] = Simplify[Together[KroneckerProduct[Pcal[z], sigma[z]]]]

In[130]:= integrand1expim[f_] = integrand1[Exp[I f]]

In[131]:= Fcal = Integrate[(1/(2 Pi)) * integrand1expim[f], {f, 0, 2 Pi}]

In[132]:= Fcal // MatrixForm

In[133]:= Det[Fcal]

In[139]:= N[In[133], 6]

In[141]:= N[Eigenvalues[Fcal], 6]
```

Figure: 1. Mathematica Notebook to compute the information matrix **Fcal** of a VARMA model defined by the matrix polynomials **A(z)** and **B(z)**; **up[z]** is $[1\ z \ldots z^{p-1}]$, **In** is the identity matrix of order r (6 lines!)

ULB

Theoretical ACF

- Here only the scalar case, $r = 1$
- **Stationarity** of a stochastic process $y_t; t \in \mathbb{Z}$ supposed of constant mean 0 and second-order moments:
  - $\text{var}(y_t) = E(y_t^2) = \sigma^2, \forall t$
  - $\text{cov}(y_t, y_{t-k}) = \gamma_k, \forall t$ (not mentioned until now)
- The **theoretical ACF** is defined by: γ_k/σ^2
- Let a **MA(q) process**, defined by $y_t = \epsilon_t - b_1\epsilon_{t-1} - ... - b_q\epsilon_{t-q}$ with $\text{cov}(\epsilon_t\epsilon_{t-k}) = 0, \forall k, \forall t$
- Then $\gamma_k = \text{cov}(y_t, y_{t-k}) = 0, \forall k > q$
- The **sample ACF** is an estimate of γ_k/σ^2, hence the ACF of a MA(q) process is (statistically) truncated for $k > q$
- The **theoretical PACF** of Partial ACF is more difficult to introduce (defined as partial correlations or as ratio between two $k \times k$ determinants) but the PACF of an AR(p) process is truncated for $k > p$
- The **sample PACF** of an AR(p) process is (statistically) truncated for $k > p$
- For an introduction to this, M (2007, Chapter 9) and on the method we used for the examples, M (2007, Chapter 10)
- The properties of the ACF can be extended to VMA models

Computation of the theoretical ACF

- For an ARMA(p, q) model, the **theoretical ACF** can be computed by solving a **system of p linear equations**, hence of the order of p^3 operations
- There are other, faster, algorithms with the order of p^2 operations, like Tunnicliffe-Wilson (1979) or Demeure & Mullis (1989)
- Moreover, these algorithms include a check of the condition related to the roots
- Indeed, they are related to the Lehmer-Schur algorithm for checking the **position of the roots with respect to the unit circle** of the Gauss plane using a sequence of **polynomials with decreasing degrees**, e.g. M (1985)
- The treatment of VARMA models is not comparable
- They are used in fast algorithms for computing the **Gaussian likelihood**, where a problem of inverting the $n \times n$ matrix covariance matrix of the y_t's is replaced by $n \max(p, q)^2$ operations, e.g., M (1985)

Conclusions

- This talk was devoted to the use of **polynomials**, including **matrix polynomials**, in the **statistical analysis of time series**
- The treatment that we have shown has used scalar and matrix polynomials to obtain the results needed for **parameter estimation** in **ARMA** and **VARMA** models
- We have mentioned that these polynomials can also help for **models with time-dependent coefficients**
- We have shown the aspects involving polynomials, **leaving aside** the other aspects like theorems proving convergence, point-wise and in distribution
- It was unfortunately not possible to give proofs and/or examples, see the references
- For **deeper references** on scalar and matrix polynomials, respectively, see Barnett (1983) and Gohberg et al. (1982)

Thank you very much for your attention

Reminder: the slides are available from me at Guy.Melard@ulb.be

The references follow

References I

- Alj, A., Azrak, R., Ley, C. & Mélard, G. (2017). Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients. *Scandinavian Journal of Statistics* **44**, 617-635. DOI: 10.1111/sjos.12268.
- Alj, A., Azrak, R. & Mélard, G. (2024). General estimation results for tdVARMA array models, *Journal of Time Series Analysis*, forthcoming.
- Azrak, R. and Mélard, G. (2006) Asymptotic properties of quasi-likelihood estimators for ARMA models with time-dependent coefficients, *Statistical Inference for Stochastic Processes* **9**, 279-330.
- Azrak, R. and Mélard, G. (2021) Asymptotic properties of conditional least-squares estimators for array time series, *Statistical Inference for Stochastic Processes* **24**, 525-547.
- Barnett, S. (1983). *Polynomials and Linear Control Systems*, Marcel Dekker, New York.
- Boubacar Maïnassara, Y. & Rabehasaina, L. (2019). Estimation of weak ARMA models with regime changes. *Statistical Inference for Stochastic Processes*, in press, DOI: 10.1007/s11203-019-09202-3.
- Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*, Holden-Day, San Francisco.
- Box, G. E. P., Jenkins, G. M., Reinsel G. C. & Ljung, G. M. (2015). *Time series analysis, forecasting and control*, 5th edn. Wiley, New York.
- Demeure, C.J., & Mullis, C.T. (1989). The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences. *IEEE Trans. Acoust., Speech and Signal Processing* **37**, 545–552.
- Francq, C., & Gautier, A. (2003). Estimation of time-varying ARMA models and applications to series subject to Markovian changes in regime. Working Paper, Université Lille 3, http://christian.francq140.free.fr/Christian-Francq/statistics-econometrics-papers/longversion.ps
- Francq, C., & Gautier, A. (2004). Estimation of time-varying ARMA models with Markovian changes in regime, Statistics & Probability Letters **70**, 243–251.

References II

- Francq, C. and Zakoïan, J.-M. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* **102**(2), 339-364.
- Gohberg, I., Lancaster, P., and Rodman, L. (1982). *Matrix Polynomials*. Academic Press, New York.
- Klein, A., & Mélard G. (2023). An algorithm for the Fisher information matrix of a VARMAX process. *Algorithms* **2023**, 16.
- Lutkepohl, H. (1991) *Introduction to Multiple Time Series*, Springer-Verlag.
- Lutkepohl, H. (2005) *New Introduction to Multiple Time Series*, Springer-Verlag.
- Mélard, G. (1985) *Analyse de données chronologiques*, Coll. Séminaire de mathématiques supérieures - Séminaire scientifique OTAN (NATO Advanced Study Institute) n° 89, Presses de l'Université de Montréal, Montréal.
- Mélard, G. (2007). *Méthodes de prévision à court terme*, Editions de l'Université de Bruxelles et Editions Ellipses.
- Mélard, G. (2022). An indirect proof for the asymptotic properties of VARMA model estimators, *Econometrics and Statistics* **21**, 96–111.
- Mélard, G. (2023). ARMA models with time-dependent coefficients: official statistics examples, in Rifaat Abdalla, Mohammed El-Diasty, Andrey Kostogryzov, Nikolay Makhutov (Eds.), *Time Series Analysis - New Insights*, IntechOpen, pp. 18-35.
- Mélard, G. (2024). Time-dependent processes and time series models: comments on Marc Hallin's early contributions and a pragmatic view on estimation, in *Recent Advances in Econometrics and Statistics, Festschrift in Honour of Marc Hallin*, Barigozzi, M., Hörmann, S., and Paindaveine, D., Springer Nature Switzerland, forthcoming.
- Reinsel (1997). *Elements of Multivariate Time Series Analysis*, Springer-Verlag.
- Söderström, T. (1984). Description of a program for integrating rational functions around the unit circle. Technical Report 8467R, Department of Technology, Uppsala University.

ULB

References III

- Tunnicliffe Wilson, G.T. (1979). Some efficient computational procedures for high order ARMA models. *Journal of Statistical Computation and Simulation* **8**, 303–309.