Pre-adjustment step with Reg-Arima models

Anna Smyk (Insee) - Daniel Ollech (Deutsche Bundesbank)

Necessity of pre-treatment (1/2)

Assumptions of the seasonal adjustment kernels :

- absence of outliers
- absence of calendar effects
- linearized series

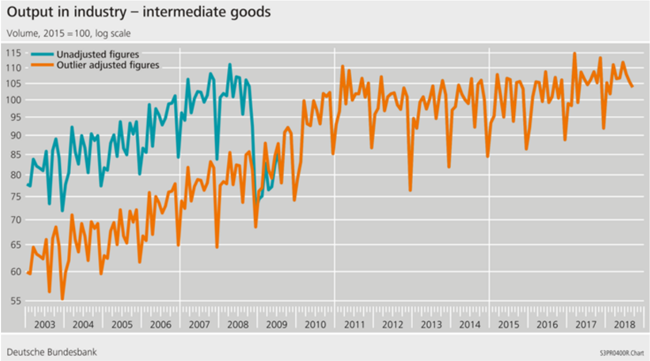
In the real world :

- outliers and calendar effects are present
- data not linearized

Pre-treatment is necessary to remove these effects in advance of the computation of the seasonal component

This step is performed with a **Reg-Arima** method, which will be outlined in this sequence. (Tramo preceding Seats is very similar)

Necessity of pre-treatment (2/2)



Reg-Arima Method

In X13-Arima and in Tramo-Seats (Tramo part) pre-treatment is performed with Reg-Arima models

- **Reg** : stands for regression and allows to correct for deterministic effects
- **Arima** is a class of models allowing to model the residuals in this regression

A bit of knowledge about Arima models is necessary to understand the Pre-adjustment step and later on the different revision policies.

After this sequence, you will know :

- why we use Reg-Arima models and how they are structured
- how to identify the models used by JDemetra+
- how to manually modify their specifications to improve Pre-treatment quality

Linear regression

Objective : remove deterministic effects by linear regression :

- outliers
- calendar

$$Y_t = \sum \hat{\alpha}_i O_{it} + \sum \hat{\beta}_j C_{jt} + X_t$$

Linearized series : $X_t = Y_t - \sum \hat{\alpha}_i O_{it} - \sum \hat{\beta}_j C_{jt}$

The regression's <u>BIG residual</u> is modeled with an Arima. (If no deterministic effects, this BIG residual is the raw series)

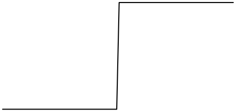
(Here, $\hat{\alpha}_i$ and $\hat{\beta}_j$ cannot be estimated by Ordinary Least Squares (OLS) because the residuals are autocorrelated.
$\rightarrow$ the Arima model formalizes this autocorrelation and allows us to use a generalized least squares estimator (GLS)

Decomposition is performed by the x11 and SEATS on the linearized series.

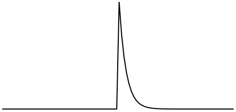
Most frequently used outliers

Additive outlier (AO)
Allocated at the end, after the decomposition, to the Irregular component.

Level Shift (LS)
Allocated at the end, after the decomposition, to the Trend.

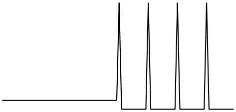


Transitory Change (TC)
Allocated at the end, after the decomposition, to the Irregular component.



Seasonal Outlier

Seasonal Outlier (SO)



Very rare, not automatically detected by default in JDemetra+.

Allocated at the end, after the decomposition to the Seasonal component.

Reg-Arima modelling

The Reg-Arima model can be formalized as follows :

$$\left(Y_t - \sum \alpha_i X_{it}\right) \sim Arima(p, d, q)(P, D, Q)$$

where the variables (regressors) X_i represent the deterministic effects. Note fundamental difference between Outliers (re-allocated) and Calendar effects (removed).

The Arima model represents the autocorrelation structure of the regression model : it describes the series temporal information. The residual of the Arima model is a white noise.

The Arima model is used to :

- estimate the regression model (OLS are no longer valid because of the autocorrelations)
- forecast the linearized series, so that symmetrical filters can be used by the decomposition algorithms, up to the last raw data point available (deterministic effects are known in the future)

Why are Arima models suitable ?

Arima models can be used to model any kind of series, provided it is made stationary first.

The notion of **stationarity** is important because it is a necessary condition to use Arma modelling. (more on the missing *i* later).

A series is said to be (weakly) stationary when :

- its moments of order 1 and 2 do not vary over time $E(X_t) = m$, $V(X_t) = \sigma^2$
- the covariance between t and $t - h$ depends not on time t but on the distance h : $cov(X_t, X_{t+h}) = \gamma(h)$

Informally : when examined through a “mobile window” of fixed length, the series always “looks the same”.

White Noise

Important example of stationary process : a white noise usually denoted ε_t :

- $E(\varepsilon_t) = 0$
- $\forall t, V(\varepsilon_t) = \sigma^2$
- non-correlated : $cov(\varepsilon_t, \varepsilon_{t'}) = 0$

Stationarity and ARMA modelling

The Wold theorem says that stationary series can be modeled with an *ARMA* model.

If the series is (weakly) stationary :

- there is an ARMA model that reasonably fits the series ;
- forecasting errors behave like the model's residuals : a white noise series uncorrelated with the series past.

Thus, such a model choice for modelling the series autocorrelation structure is a good one, as long as the series is (made) stationary.

First, let's assume that the series is stationary and outline the structure of these models

Construction of an ARMA model : Auto-Regressive models (AR)

An *ARMA* model has an Auto-Regressive part (*AR*) and a Moving Average part (*MA*).

To formally write these models, we use a lag operator often noted B (Backwards) or L (Lag) :
$B(X_t) = X_{t-1}$ and $B^p(X_t) = X_{t-p}$

Autoregressive model of order p, $AR(p)$:

$$\begin{aligned} & X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \\ \Longleftrightarrow\ & (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = \varepsilon_t \\ \Longleftrightarrow\ & \Phi(B) X_t = \varepsilon_t \end{aligned}$$

ε_t is the innovation process : a white noise independent of X's past

ε_t isn't correlated with $X_1, X_2 ......, X_{t-1}$

An $AR(p)$ models the influence of the p past series values on the value at current date t : it's a memory effect.

Moving Average models (MA)

Moving average model of order q, $MA(q)$:

$$\begin{aligned} & X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \iff & X_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\varepsilon_t \\ \iff & X_t = \Theta(B)\varepsilon_t \end{aligned}$$

Property : an MA process is always stationary.

Is is the result of a non-persistent accumulation of “q” independent shocks.

Common example : a phenomenon that fluctuates around its average value is modeled by an MA(1) with a constant.

ARMA models

$ARMA(p, q)$ models : a combination of $AR(p)$ and $MA(q)$, with or without constant

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

$$\Phi(B)X_t = \mu + \Theta(B)\varepsilon_t$$

An *ARMA* process is the result of a "memory" effect and a non-persistent accumulation of independent randomized chocs.

(S)ARMA models : adding suitable parameters when modeling seasonal series

To model seasonal series, we use **S**ARMA models that enable us to highlight lags of order equal to the series periodicity (s). Most of the times these models will be called ARMA anyway...

$SARMA(P, Q)$ model : an $ARMA$ with a polynomial term of order s (4 for quarterly series, 12 for monthly series) :

$$\Phi(B^s)X_t = \Theta(B^s)\varepsilon_t \text{ or } \Phi_s(B)X_t = \Theta_s(B)\varepsilon_t$$

Advantages :

- the autocorrelations of order s are immediately visible
- factorization simplifies the formula

This new way of writing the model doesn't impact the $ARMA$ model itself, it only clarifies it.

$ARMA(p, q)(P, Q)$

An $ARMA(p, q)(P, Q)$ model combines both regular and seasonal parts : $ARMA(p, q) \times SARMA(P, Q)$.

It is identical to a $ARMA(p + P * s, q + Q * s)$.

Example on a monthly series : $ARMA(1, 1)(1, 1) = ARMA(13, 13)$

But the formula $ARMA(13, 13)$ is more ambiguous than $ARMA(1, 1)(1, 1)$, unless it is mentioned that the lags 2 to 12 have null coefficients.

Dealing with non-stationary series : Integrated models

Only stationary processes can be correctly modeled by ARMA models. In real life, most series are non stationary (economic indicators often have a trend). Here, we will see where non-stationarity stems from and how to **stationarize processes**.

Let X be a process “linear trend” :

$$X_t = \alpha + \beta t + \varepsilon_t$$

To test if a series is stationary, we calculate its expectation and, if necessary, its variance and autocovariances. (NB. Variance is the autocovariance with lag $k = 0$.)

For a (deterministic) linear trend + (random) white noise :

$$E(X_t) = \alpha + \beta t + E(\varepsilon_t) = \alpha + \beta t$$

Here the expectation depends on time : the series is not stationary.

Linear trend and differentiation

Let's calculate the series variance.

$$V(X_t) = V(\varepsilon_t) = \sigma^2$$

It is constant, therefore it does not depend on time.

Let's **differentiate** this series :

Difference of order 1 :

$$(I - B)X_t = X_t - X_{t-1} = \alpha + \beta t + \varepsilon_t - \alpha - \beta(t-1) - \varepsilon_{t-1}$$

Which becomes :

$$(I - B)X_t = \beta + \varepsilon_t - \varepsilon_{t-1}$$

Stationarisation by differentiation

Let's calculate the expectation, variance and covariances of the differentiated series.

$$E((I-B)X_t) = \beta$$

$\rightarrow$ the expectation does not depend on time

$$V((I-B)X_t) = V(\varepsilon_t) + V(\varepsilon_{t-1}) = 2\sigma^2$$

because ε_t and ε_{t-1} aren't correlated.

For covariance between t and $t+h$, with $h > 0$:
$cov((I-B)X_t, (I-B)X_{t+h}) = cov(\beta + \varepsilon_t - \varepsilon_{t-1}, \beta + \varepsilon_{t+h} - \varepsilon_{t+h-1}) = 0$
because $\forall t \neq t' \ : \ cov(\varepsilon_t, \varepsilon_{t'}) = 0$

$\rightarrow$ differentiation (here of ordrer 1) stationarised the series.

Differentiation orders

If X is a “polynomial trend of order 2” differenciating twice will make it stationary :

$(I-B)^2X_t$ is stationary.

More generally, $(I-B)^d$ stationarizes a polynomial of order d.

$\rightarrow d$ is the differentiation order.

Therefore, an AR**I**MA model requires 3 orders : (p, d, q).

The sequence and choice of the letters are conventional : AR order appears first and is called p, and MA order appears last and is called q.

Stationarisation of a stable seasonal process

Let X be a "stable seasonal" process :

$$X_t = S_t + \varepsilon_t \quad \text{with} \quad \forall t,\, S_t = S_{t+s} \quad \text{and} \quad \varepsilon_t \text{ a white noise}$$

Is X_t a stationary process ?

Its expectation is : $E(X_t) = S_t$.

For a monthly series, $E(X_t) = E(X_{t+12})$ but $E(X_t) \neq E(X_{t+11})$, therefore the process isn't stationary.

$V(X_t) = E((X_t - E(X_t))^2) = E((S_t + \varepsilon_t - S_t - 0)^2) = E(\varepsilon_t^2) = \sigma^2$

Seasonal differentiation

To stationarize X_t, let's apply a **seasonal differentiation** of order 1 :

$$(I - B^s)X_t = S_t + \varepsilon_t - S_{t-12} - \varepsilon_{t-12} = \varepsilon_t - \varepsilon_{t-12}$$

The result is a linear combination of white noises, therefore it is stationary.

If X had a linear trend $X_t = S_t + a \times t + b + \varepsilon_t$:

$$(I - B^s)X_t = S_t + a \times t + b + \varepsilon_t - S_{t-12} - a \times (t - 12) - b - \varepsilon_{t-12} = 12 \times a + \varepsilon_t - \varepsilon_{t-12}$$

Therefore, seasonal differentiation stationarizes series that have a linear trend.

It can be noted that simple differentiation is included in seasonal differentiation :

$$(I - B^s)X_t = (I - B)(I + B + ... + B^{s-1})X_t$$

Differentiation order and polynomial trends

A simple differentiation of order d removes polynomial trends of order d :

$$(I - B)^d X_t$$

A seasonal differentiation removes stable seasonality (and also linear trends) :

$$(I - B^s) X_t$$

In JD+, $D \leq 1$ and $d \leq 1$

Arima models : full set of parameters

$Arima(p, d, q)(P, D, Q)$ models series with a trend and seasonality :

$$\Phi(B)\Phi_s(B)(I-B)^d(I-B^s)^D X_t = \Theta(B)\Theta_s(B)\varepsilon_t$$

Warning : there are in fact two kinds of parameters

- $(p, d, q)(P, D, Q)$ are called **orders**
- each p, q, P, Q has its corresponding **coefficients**

Let's have a look at a series in the GUI.

Arima models and seasonality (1/3)

Let's consider the seasonal part of an Arima model :

1 - Is a series with Arima $(p, d, q)(0, 0, 0)$ seasonal ?

2 - And with Arima $(p, d, q)(0, 0, Q)$?

3 - $(p, d, q)(0, 1, 0)$?

Arima models and seasonality (2/3)

Answers :

1 - No, no autocorrelation of order s.

2 - No, a MA represents non-persistent fluctuations, and seasonality is persistent.

3 - Yes, a stable seasonality.

Arima models and seasonality (3/3)

Two common cases :

- $(0,1,1)(0,1,1)$: stable seasonality on average, with occasional level fluctuations of θ_s (the greater θ_s, the greater the fluctuations)
- $(p,d,q)(1,1,1)$: evolving seasonality with drift + occasional level fluctuations θ_s

How does it work in practice (hint)

The algorithms Reg-Arima of the X13-Arima method and Tramo of Tramo-Seats :

- first provide a complete estimation (regression and Arima coefficients, by iteration of Maximum likelihood and GLS) using the most common Arima model, (0,1,1)(0,1,1), called the *Airline* model (it has historically been used to model airline passenger flows)
- Then look for competing models to compare to the airline default model. In the GUI you can specify that you would like to stick with this one, and it can be a good choice to get back to if the algorithm converged towards “a bad solution”.

Choosing the final model

Information criteria (to **minimize**) to compare models :

- The AIC (Akaike criterion) :

$$AIC(p,q) = -2\ln(L) + 2 * (p+q)$$

- The AICC (corrected for short time periods) :

$$AICC(p,q) = -2\ln(L) + 2(p+q)\left(1 - \frac{n+p+1}{N_{obs}}\right)^{-1}$$

- The BIC (Schwarz criterion) :

$$BIC(p,q) = -2\ln(L) + (p+q)\ln(N_{obs})$$

Pre-treatment with Reg-Arima method : take home message (1/2)

Economic series are not stationary : neither their level, nor their fluctuations are constant in time.

They can be stationarized by differentiation and modeled with $Arima(p, d, q)(P, D, Q)$ models

The Moving Average (MA) part represents the non-persistent fluctuations around a constant level : therefore, it is a stationary process.

An Auto-Regressive part (AR) highlights the influence of past realizations on the current one.

Pre-treatment with Reg-Arima method : take home message (2/2)

An Arima model gives the series autocorrelation structure and its degree of variability.

Models are validated via residual testing (white noise ?). Discrimination of valid models is done thanks to information criteria.

The pre-adjustment Reg-Arima modelling enables us to linearize and extend the series (allows the use of symmetrical filters). Note, that it does not isolate the seasonal component.