Time Series Analysis

Descriptive analysis of a time series

Andrés M. Alonso   Carolina García-Martos

Universidad Carlos III de Madrid

Universidad Politécnica de Madrid

June – July, 2012

2. Descriptive analysis of a time series

Outline:

- Introduction
- Analysis of deterministic trends
- Smoothing methods
- Decomposition methods for seasonal series
- Seasonality and seasonal adjustment
- Exploration of multiple cycles

Recommended readings:

▷ Chapter 1 of Brockwell and Davis (1996).

▷ Chapter 1 of Peña, Tiao and Tsay (2001).

Introduction

▷ In this topic we present some descriptive procedures developed for time series analysis.

▷ The purpose of those procedures is to explain the past evolution of a series in terms of simple patterns and to forecast its future values.

- Models with deterministic trend which constitute an extension of regression methods.
- Smoothing methods which carry out predictions by imposing a structure where the dependency between the observations diminishes over time.
- The extension of these methods for seasonal series.

Analysis of deterministic trends

▷ We begin with series without seasonality, such as those in Examples 1 to 5.

▷ We assume that the observed series, z_t, is represented by

$$z_t = \mu_t + a_t, \tag{1}$$

where the first component

$$\mu_t = f(t, \beta)$$

is the **level of the series** which is a known deterministic function of the time that depends on the instant being studied and on a parameter vector, β. These parameters must be estimated from the data, as we will see next.

The second component, a_t, is usually called as the **innovation** and is a random component which contains the rest of the effects that affect the series.

It is assumed that the random variables a_t's have zero mean, constant variance, normal distribution and a_t and a_s are independent when $t \neq s$.

Analysis of deterministic trends

▷ The **prediction** of the series with model (1) for future period, $T + k$, is obtained by extrapolating the level of the series μ_t, since the prediction of the innovation is its expectation, which is always zero.

▷ Letting $\widehat{z}_T(k)$ be the prediction carried out from the origin T for k periods ahead, that is, the prediction of the value z_{T+k} with the information available until the moment T, we have

$$\widehat{z}_T(k) = \mu_{T+k} = f(T + k, \beta). \tag{2}$$

▷ The form that we establish for the evolution of the level of the series over time determines the specific model to be used.

Analysis of deterministic trends

Example 13

The simplest model assumes that the level of the series is constant in time, that is $\mu_t = \mu_{t-1} = \mu$, *and it is known as the model of constant level or* **detrended** *series.*

Thus the equation (1) is reduced to:

$$z_t = \mu + a_t \tag{3}$$

and the series moves around its mean, μ, *which is constant.*

The series in Examples 1 and 2 are detrended and could be explained by this model.

Since the level of the series is constant and it does not depend on t, *and the expected value of the innovation is zero, the prediction with this model for any horizon will be the mean,* μ.

Example 14

A more general model, which is applied to series with upward or downward trends is the **linear trend** *model. Example 3 shows a series that might have this property.*

The model for μ_t in (1) is:

$$\mu_t = \beta_0 + \beta_1 t, \tag{4}$$

where β_1 now represents the slope of the line that describes the evolution of the series. This slope corresponds to the expected growth between two consecutive periods.

The prediction with this model of the value of the series at time $T + k$ with information up to T, will be

$$\widehat{z}_T(k) = \beta_0 + \beta_1(T + k) \tag{5}$$

Analysis of deterministic trends

▷ These two models are special cases of **polynomial trends**, where the level of the series evolves according to a polynomial of order r:

$$\mu_t = \beta_0 + \beta_1 t + \dots + \beta_r t^r \tag{6}$$

▷ Fitting these models to a time series requires the estimation of the parameter vector, $\beta = (\beta_0, \dots, \beta_r)$. The estimations are obtained using the **least squares** criterion, that is by minimizing the differences between the observed values and those predicted up to horizon one by the model:

$$\text{Minimize } \sum_{t=1}^{T} (z_t - \mu_t)^2. \tag{7}$$

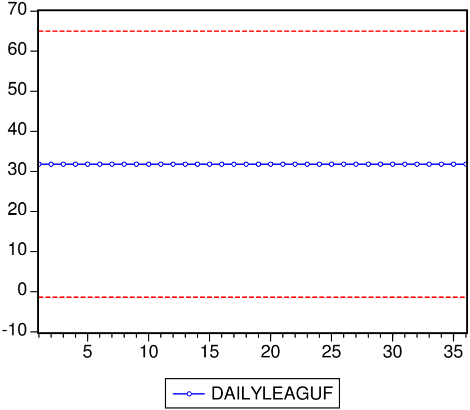
Analysis of deterministic trends - Example 1

Dependent Variable: DAILYLEAGUES
Method: Least Squares
Date: 01/25/08 Time: 17:22
Sample: 1 34
Included observations: 34
DAILYLEAGUES = C(1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	31.82353	2.844286	11.18858	0.0000

R-squared	0.000000	Mean dependent var	31.82353
Adjusted R-squared	0.000000	S.D. dependent var	16.58490
S.E. of regression	16.58490	Akaike info criterion	8.483833
Sum squared resid	9076.941	Schwarz criterion	8.528726
Log likelihood	-143.2252	Durbin-Watson stat	0.869676

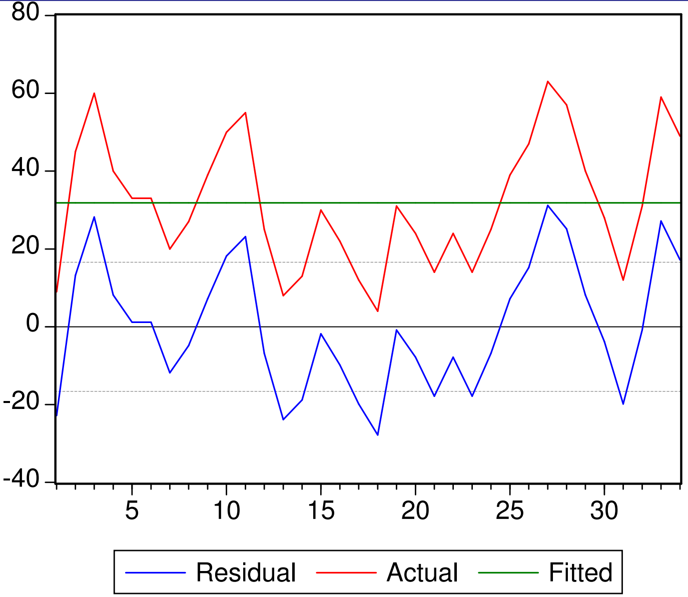
Analysis of deterministic trends - Example 1



Forecast: DAILYLEAGUF
Actual: DAILYLEAGUES
Forecast sample: 1 36
Included observations: 34

Root Mean Squared Error	16.33918
Mean Absolute Error	13.62630
Mean Abs. Percent Error	79.11366
Theil Inequality Coefficient	0.241716
Bias Proportion	0.000000
Variance Proportion	1.000000
Covariance Proportion	0.000000

Analysis of deterministic trends - Example 1



Analysis of deterministic trends - Example 1

▷ We fit the constant level or mean model (3) to the series of leagues sailed daily by Columbus's fleet. The mean of the observed data of the series is

$$\bar{z}_t = \frac{9 + ... + 49}{34} = 31.82$$

which is also, the prediction of the distance to be sailed the following day.

▷ The errors of this forecast within the sample are the residuals, $\widehat{a}_t$, estimated as the difference between the value of the series and its mean.

▷ The dispersion of these residuals measures the expected forecast error with this model.

$$\widehat{\sigma}_a = \sqrt{\frac{(-22.82)^2 + ... + (17.18)^2}{34}} = 16.6$$

which indicates that the average forecasting error with this model is 16.6 daily leagues.

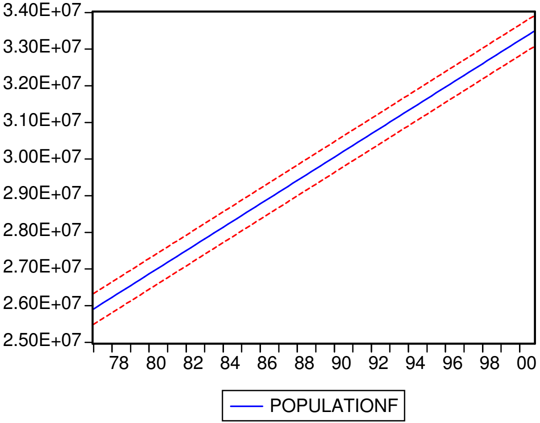
Analysis of deterministic trends - Example 3

Dependent Variable: POPULATIONOVER16
Method: Least Squares
Date: 01/25/08 Time: 18:07
Sample: 1977Q1 2000Q4
Included observations: 96
POPULATIONOVER16 = C(1) +C(2)*TIME

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	29693292	21446.05	1384.558	0.0000
C(2)	79696.88	773.9096	102.9796	0.0000

R-squared	0.991214	Mean dependent var	29693292
Adjusted R-squared	0.991120	S.D. dependent var	2229916.
S.E. of regression	210127.5	Akaike info criterion	27.36943
Sum squared resid	4.15E+12	Schwarz criterion	27.42285
Log likelihood	-1311.733	Durbin-Watson stat	0.011995

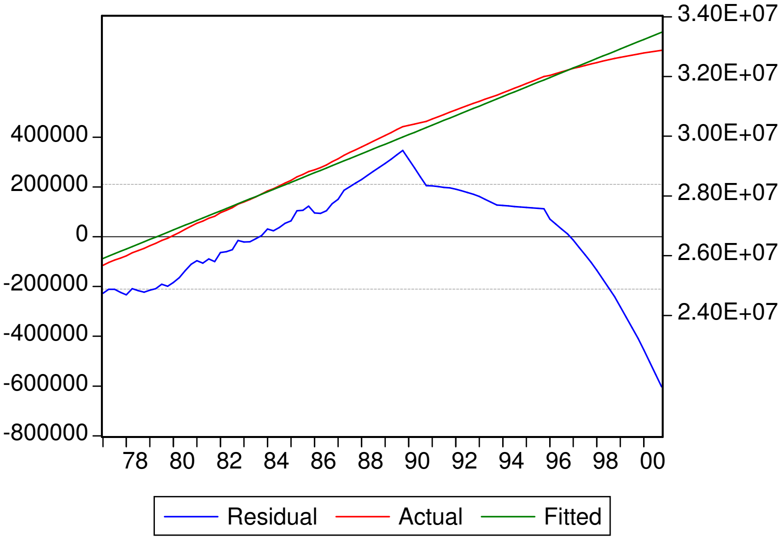
Analysis of deterministic trends - Example 3



Forecast: POPULATIONF
Actual: POPULATIONOVER16
Forecast sample: 1977Q1 2001Q4
Adjusted sample: 1977Q1 2000Q4
Included observations: 96

Root Mean Squared Error	207927.2
Mean Absolute Error	172321.7
Mean Abs. Percent Error	0.576364
Theil Inequality Coefficient	0.003492
Bias Proportion	0.000000
Variance Proportion	0.002206
Covariance Proportion	0.997794

Analysis of deterministic trends - Example 3



Analysis of deterministic trends - Example 3

▷ We fit the linear trend model to the population data of people in Spain over 16 years of age. We have the prediction equation

$$\widehat{z_t} = 29693292 + 797696.88t$$

This line indicates that, in the period being studied, the number of people over 16 is, on average, 29,693 million people and that each year this number increases by approximately 79700 people.

▷ The model seems to have good fit, since the correlation coefficient is .995.

▷ Nevertheless, if we look at the data and the fitted model, we can see that the fit is not good because the trend is not exactly constant and has changed slightly over time. In particular, the predictions generated in the year 2000 for the two following years are considerably higher than the observed data.

Analysis of deterministic trends - Example 3

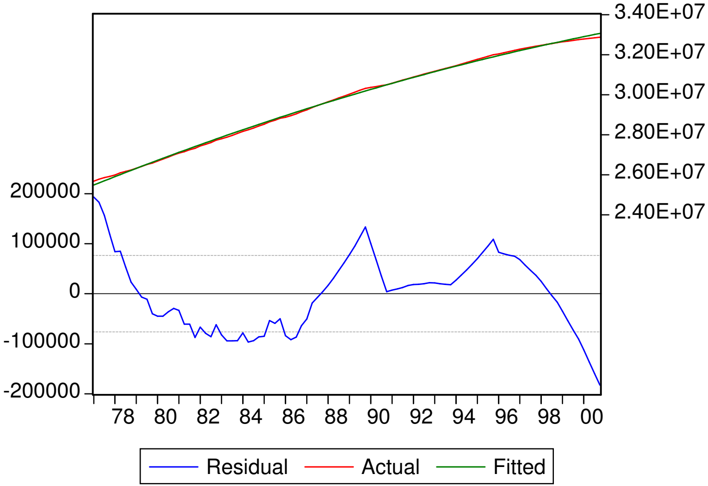
▷ One might think that the problem is that the trend is following a second degree polynomial and that the shape of the residuals indicates the need for a second degree equation in order to reflect this curvature.

Dependent Variable: POPULATIONOVER16
Method: Least Squares
Date: 01/25/08 Time: 18:38
Sample (adjusted): 1977Q1 2000Q4
Included observations: 96 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	29910092	11686.80	2559.306	0.0000
TIME	79696.88	281.1305	283.4872	0.0000
TIME^2	-282.3216	11.34427	-24.88671	0.0000

R-squared	0.998853	Mean dependent var	29693292
Adjusted R-squared	0.998828	S.D. dependent var	2229916.
S.E. of regression	76330.93	Akaike info criterion	25.35430
Sum squared resid	5.42E+11	Schwarz criterion	25.43443
Log likelihood	-1214.006	Durbin-Watson stat	0.050897

Analysis of deterministic trends - Example 3



▷ A second degree of the polynomial does not solve the problem. The prediction errors for the last data in the sample are quite high. The problem is the lack of flexibility of the deterministic models.

Limitations in deterministic trends

▷ As we saw in the above example, the main limitation of these methods is that, although series of constant levels are frequent, it is unusual for a real series to have a linear deterministic trend, or in general, a polynomial trend with $r \geq 1$.

▷ One possibility is to try to fit linear trends by intervals, that is divide the series into parts that have, approximately, a constant trend and to fit a linear or constant model in each part.

▷ Although these models signify a clear advance in explaining the historic evolution of some series, they are less useful in predicting future values, since we do not know how many past observations to use in order to fit the future level of the series.

▷ An additional difficulty of fitting a linear trend by intervals is that the implicit growth model is unreasonable.

Limitations in deterministic trends

▷ Let us look at the implications of assuming that a series follows a deterministic trend model in an interval.

Example 15

Let us assume a series z_t with 5 data points at times $t = 0, \pm 1, \pm 2$. Let $z_{-2}, z_{-1}, z_0, z_1, z_2$ be the values of the series. Applying the formulas for the estimation of the slope gives us

$$\begin{aligned}
\sum_{t=-2}^{2} t^2 &= (-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 = 10 \\
\sum_{t=-2}^{2} t^2 t z_t &= -2z_{-2} - z_{-1} + z_1 - 2z_2 = 2\left(z_{-1} - z_{-2}\right) + 3\left(z_0 - z_{-1}\right) \\
&\quad +3\left(z_1 - z_0\right) + 2\left(z_2 - z_1\right)
\end{aligned}$$

and therefore:

$$\widehat{\beta}_1 = 0.2\nabla z_{-1} + 0.3\nabla z_0 + 0.3\nabla z_1 + 0.2\nabla z_2.$$

Limitations in deterministic trends

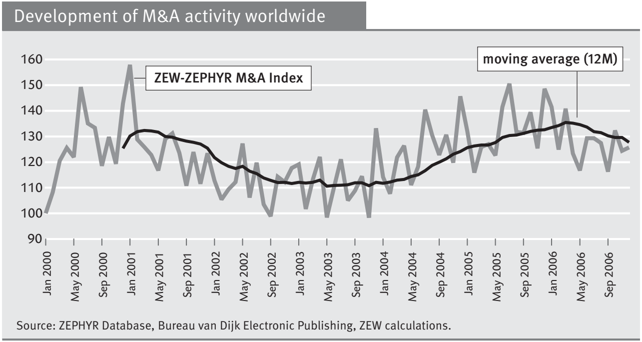
▷ This results indicates that the prediction of future growth is a weighted mean of observed growth in each one of the periods, with symmetric weights with respect to the center such that the minimum weight corresponds to growth in the extreme periods.

▷ If we accept that a series has a linear deterministic trend in an interval we are saying that the future prediction of its growth must be done weighting the observed past growths, but giving minimum weight to the last growth observed. This weight is also equal to that which is attributed to growth farthest back in time, that is, to the first growth observed in the sample.

▷ Moreover, if we increase the size of the interval, the weight of last observed growth diminishes, but always remains equal to the weight of the first growth observed. For instance, with a sample of 100 yearly data points, we fit the model in this sample, the weight of the last growth observed becomes a very small value, and equal to the growth observed 100 years ago!

Smoothing methods

▷ To find a solution to the limitations of deterministic trend models, smoothing methods were introduced in the 1960s. The idea behind these methods is to allow that the last data points in the series have more weight in the forecasts than the older values.



From the 60s but they are still used in practice.

Simple exponential smoothing method

▷ Let us assume that we have made a prediction of the value of a variable for time period T that we will denote as $\widehat{z}_T$ and afterwards we observe its value, z_T.

How do we generate the next prediction?

▷ Holt (1956) proposed making a linear combination of the last prediction and the last observed value, such that the prediction of the next period, $T+1$, is given by:

$$\widehat{z}_{T+1} = \theta \widehat{z}_T + (1-\theta) z_T \tag{8}$$

where $0 < \theta < 1$ determines the weight that we give to each of the two components to generate the predictions.

▷ If we take θ near the unit, the predictions for both periods are very similar, and change little with the new information. However, if θ is small, near zero, the prediction adapts itself as a function of the last observed value.

Simple exponential smoothing method

▷ In order to better understand this model, suppose that, $\widehat{z}_T$, also follows equation (8) and we group terms, thus:

$$\widehat{z}_{T+1} = \theta(\theta\widehat{z}_{T-1} + (1-\theta)z_{T-1}) + (1-\theta)z_T = \theta^2\widehat{z}_{T-1} + (1-\theta)(z_T + \theta z_{T-1}),$$

repeating this substitution process, we have:

$$\widehat{z}_{T+1} = \theta^T\widehat{z}_1 + (1-\theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + ...)$$

▷ Assuming that T is large and $\theta < 1$ the first term will be very small, and we can write the prediction equation as:

$$\widehat{z}_{T+1} = (1-\theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + ...) \tag{9}$$

which is a weighted mean of all the previous observations with decreasing weights that add up to one, since $(1 + \theta + \theta^2 + ...) = \frac{1}{1-\theta}$.

Simple exponential smoothing method

▷ The predictions generated by the simple smoothing model are a weighted mean of the previous values of the series with geometrically decreasing weights.

▷ Using this model requires that the parameter θ be determined. In the first applications this parameter was set a priori, usually between .70 and .99, but progressively better results were obtained by permitting a wider range of possible values and estimating their size from the data of the series with the criterion of minimizing prediction errors.

▷ This can be done trying a value grid, like $0.1, 0.2, ..., 0.9$ for θ, calculating the prediction errors within the sample $\widehat{a}_t = z_t - \widehat{z}_t$ and taking the value of θ that leads to a smaller value of $\sum \widehat{a}_t^2$, the residual sum of squares or prediction errors.

▷ The book of Hyndman, Koehler, Ord and Snyder (2008), *Forecasting with exponential smoothing: the state space approach*, provides a very recent view of these methods.

Example 16

We can determine the "best" value for the smoothing parameter θ for the series of leagues sailed by Columbus's fleet.

Date: 01/27/08 Time: 11:12
Sample: 1 34
Included observations: 34
Method: Single Exponential
Original Series: DAILYLEAGUES
Forecast Series: DAILYLSM

Parameters:	Alpha	0.9990
Sum of Squared Residuals		8363.125
Root Mean Squared Error		15.68357
End of Period Levels:	Mean	49.00997

θ	1.0	0.9	0.8	0.7	0.6
SSE	8414.9	8478.6	8574.0	8686.9	8798.8
θ	0.5	0.4	0.3	0.2	0.1
SSE	8891.3	8957.1	9018.8	9141.6	9360.5

Holt's double exponential smoothing

▷ The above ideas can be applied to linear trend models. Instead of assuming that the parameters are fixed we can allow them to evolve over time and estimate them giving decreasing weight to the observations. Suppose the model:

$$z_t = \mu_t + a_t$$

but now instead of assuming a deterministic trend we allow the level to evolve linearly over time, but with a slope that may differ in different periods:

$$\mu_t = \mu_{t-1} + \beta_{t-1},$$

such that the difference between the levels of two consecutive times, $t-1$ and t, is β_{t-1} the slope at time $t-1$.

▷ Notice that if $\beta_{t-1} = \beta$, constant over time, this model is identical to that of the linear deterministic trend. By allowing the slope to be variable this model is much more flexible.

Holt's double exponential smoothing

▷ The prediction of z_t with information to $t-1$, that we denote as $\widehat{z}_{t-1}(1)$, is obtained as follows

$$\widehat{z}_{t-1}(1) = \widehat{\mu}_{t|t-1} = \widehat{\mu}_{t-1|t-1} + \widehat{\beta}_{t-1}$$

where the estimation of the level of the series at time t is the sum of the last estimations of the level and of the slope with information to $t-1$.

▷ The notation $\widehat{\mu}_{t|t-1}$ indicates that we are estimating the level at time t, but with information available up to time $t-1$, that is, data point z_{t-1}.

▷ The prediction is $\widehat{\mu}_{T+1|T} = \widehat{\mu}_{T|T} + \widehat{\beta}_T$, where $\widehat{\mu}_{T|T}$ and $\widehat{\beta}_T$ are the estimations of the level and growth with information up to time T.

Holt's double exponential smoothing

▷ By observing the value z_{T+1} we can calculate the forecasting error $(z_{T+1} - \widehat{\mu}_{T+1|T})$, and, as with the simple smoothing method, correct the previous estimation by a fraction of the error committed. As a result, the estimation $\widehat{\mu}_{T+1|T+1}$ with information up to $T+1$, will be

$$\begin{aligned} \widehat{\mu}_{T+1|T+1} &= \widehat{\mu}_{T+1|T} + (1-\theta)(z_{T+1} - \widehat{\mu}_{T+1|T}) = \\ &= \widehat{\mu}_{T|T} + \widehat{\beta}_T + (1-\theta)(z_{T+1} - \widehat{\mu}_{T|T} - \widehat{\beta}_T) \end{aligned}$$

where $\theta <1$ is a discount factor.

▷ The new estimation of future growth with information up to $T+1$, $\widehat{\beta}_{T+1}$ is made by modifying the last estimation by a fraction of the last error committed:

$$\widehat{\beta}_{T+1} = \widehat{\beta}_T + (1-\gamma)(\widehat{\mu}_{T+1|T+1} - \widehat{\mu}_{T|T} - \widehat{\beta}_T)$$

where $\gamma < 1$ is another discount factor over the previous error in the growth estimation.

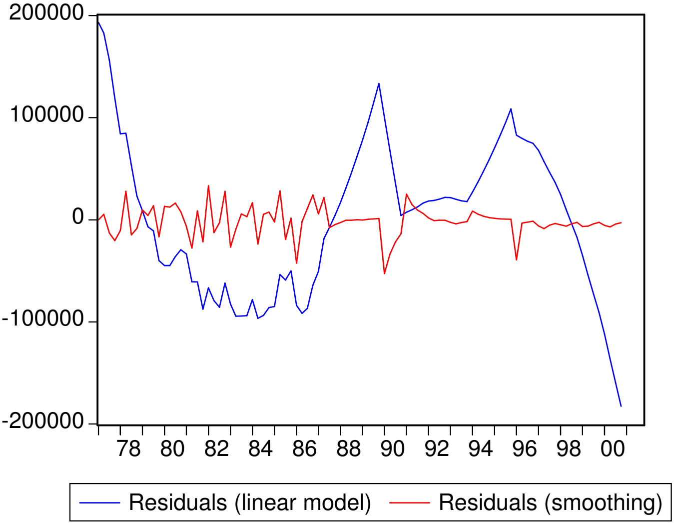
▷ The parameters θ and γ are determined as in the above case, testing with a grid of values and choosing those that minimize the sum of the squared prediction errors.

Example 17

We can determine the values for θ and γ for the series of Spanish population over 16 years.

Date: 01/27/08 Time: 11:47
Sample: 1977Q1 2000Q4
Included observations: 96
Method: Holt-Winters No Seasonal
Original Series: POPULATIONOVER16
Forecast Series: POPULASM

Parameters:	Alpha	1.0000
	Beta	0.3600
Sum of Squared Residuals		2.02E+10
Root Mean Squared Error		14500.58
End of Period Levels:	Mean	32876100
	Trend	31971.10



▷ If we compare these results with the linear model, we see that the forecasting errors are much smaller and that now the residuals do not show a marked trend and the predictions are fairly good in many periods.

Decomposition methods for seasonal series

▷ When a series has not only a trend and a random component but seasonality as well, the decomposition methods assume that the data are generated as a sum of these three effects:

$$z_t = \mu_t + S_t + a_t$$

where μ_t is the **level** of the series, S_t is the **seasonal component** and a_t is the purely **random component**.

▷ The classical decomposition methods assume that both the level as well as the seasonality are deterministic.

▷ The level μ_t is modelled using a deterministic time polynomial of order less than or equal to two.

▷ The seasonality is modelled as a periodic function, which satisfies the condition:

$$S_t = S_{t-s}$$

where s is the period of the function depending on the data's seasonality.

Decomposition methods for seasonal series

▷ The procedure for constructing the model for the series is carried out in the three following steps:

1. Estimate the level of the observed series as in the deterministic trend model. Next the estimated level , $\widehat{\mu}_t$, is subtracted from the series in order to obtain a residual series, $E_t = z_t - \widehat{\mu}_t$, which will contain seasonality plus the random component. This is called a **detrended series**.
2. The **seasonal coefficients**, $S_1, \ldots, S_s$, are defined as a set of coefficients that add up to zero and are repeated each year. They are estimated in the detrended series as the difference between the mean of the seasonal periods and the general mean.
$$\widehat{S}_j = \overline{E}_j - \overline{E}.$$
3. The series of **estimated innovations** is obtained by subtracting the seasonal coefficient of each observation from the detrended series. For example, for monthly data, using the above notation: $\widehat{a}_{12i+j} = E_{12i+j} - \widehat{S}_j$.

Decomposition methods for seasonal series

▷ The prediction of the series is done by adding the estimations of the trend and the seasonal factor that corresponds to each observation that month. If we subtract the seasonal coefficient of the month from the original series we obtain the **deseasonalized series**.

▷ As we have seen in the above sections, there are series that clearly have no constant trend and for which fitting a deterministic trend is not suitable.

▷ A possible alternative is to estimate the level of the series locally using a **moving average** of twelve months as follows:

$$\widehat{\mu}_t = \frac{z_{t-5} + \ldots + z_{t+5} + z_{t+6}}{12}$$

that is, we construct a mean of twelve observations.

▷ Applying this method we obtain an estimation of the level of the series at times $t = 6, \ldots, T-6$. Next we carry out the decomposition of the series, as explained above, using steps ② and ③.

Example 18

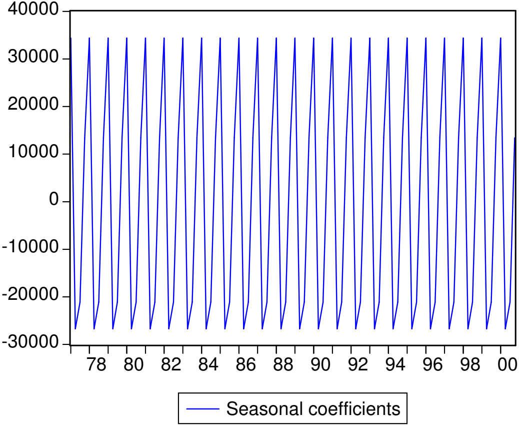
We are going to analyze the series of unemployment in Spain. We use the following MA expression for quarterly data:

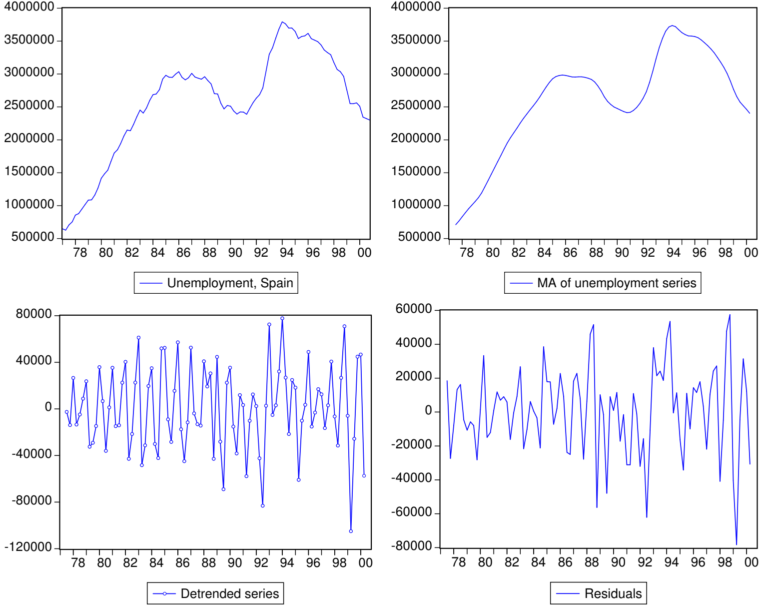
$$\widehat{\mu}_t = \frac{0.5z_{t-2} + z_{t-1} + z_t + z_{t+1} + 0.5z_{t+2}}{4}.$$

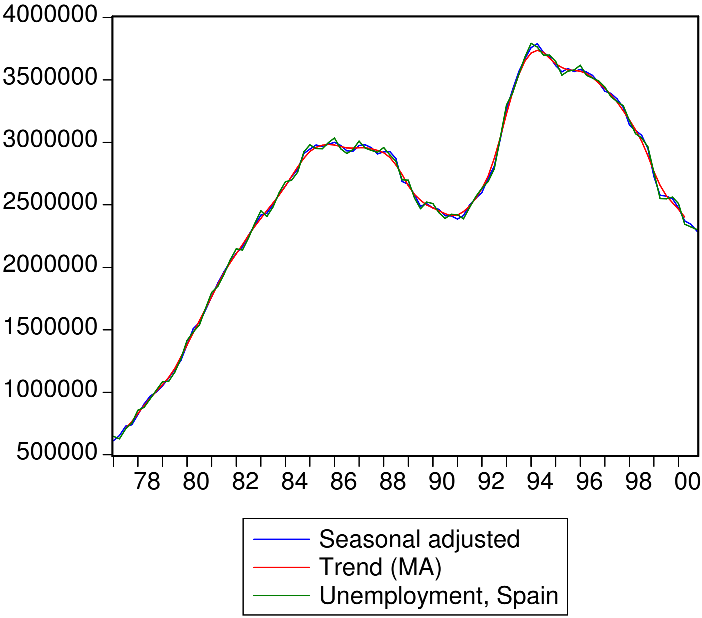
Date: 01/27/08 Time: 13:12
Sample: 1977Q1 2000Q4
Included observations: 96
Difference from Moving Average
Original Series: UNEMP
Adjusted Series: UNEMPSA

Scaling Factors:

1	34388.36
2	-26700.20
3	-21094.95
4	13406.79




4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
78 80 82 84 86 88 90 92 94 96 98 00
Unemployment, Spain
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
78 80 82 84 86 88 90 92 94 96 98 00
MA of unemployment series
80000
40000
0
-40000
-80000
-120000
78 80 82 84 86 88 90 92 94 96 98 00
Detrended series
60000
40000
20000
0
-20000
-40000
-60000
-80000
78 80 82 84 86 88 90 92 94 96 98 00
Residuals


4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
78
80
82
84
86
88
90
92
94
96
98
00
Seasonal adjusted
Trend (MA)
Unemployment, Spain

Seasonality and seasonal adjustment

▷ An alternative procedure for modelling seasonality is to represent it using a harmonic function with period s. Assuming that we have eliminated the trend, we consider series that have only a seasonal component, with structure:

$$z_t = S_t + a_t.$$

▷ The simplest alternative for representing S_t as a periodic function, with $S_t = S_{t-s}$, is to assume a harmonic function, such as the sine or cosine: $\sin(2\pi t/s)$ and $\cos(2\pi t/s)$

▷ s is called **period** of the periodic function; its inverse $f = 1/s$ is called **frequency** and $w = 2\pi f = 2\pi/s$ is the **angular frequency**.

Example 19

In a quarterly series ($s = 4$), the frequency is $f = 1/4 = .25$, indicating that between two observations, a quarter, .25 of the period of the function has gone by or 25% of a full cycle. The angular frequency is $w = 2\pi/4 = \pi/2$, indicating that in one quarter an angle of $\pi/2$ is covered with respect to the full cycle of 2π.

Representation of seasonality by a cycle

▷ We assume a series $(z_1, ..., z_T)$ that has cyclical seasonality of period s, and in which we observe j complete cycles, that is $T = js$, with j an integer. We are going to model the seasonality using a sine function with angular frequency $w = 2\pi/s$.

▷ The first observation of the series will not be, in general, the average of the cycle, as corresponds to the sine function. Instead, the sinusoidal wave which describes the seasonality will start in the first observation with a certain angle θ of difference, which is unknown, with relation to the start of the cycle.

▷ Furthermore, the cycle will also have an unknown amplitude that we will denote as R.

▷ The model for the series is:

$$z_t = \mu + R \sin(wt + \theta) + a_t. \tag{10}$$

Representation of seasonality by a cycle

▷ To fit the model (10) to the data we are going to write it in a more convenient form. We will take the sine of the sum of the two angles as the sum of the products of the sines plus the product of the cosines, with which we can write the above equation as

$$z_t = \mu + Rsin(wt) \sin[\theta] \, R\cos(wt) \cos[\theta] + a_t$$

and letting $A = R\sin\theta$ and $B = R\cos\theta$, we have:

$$z_t = \mu + A\sin(wt) + B\cos(wt) + a_t. \tag{11}$$

This expression is simpler than (10) since it represents the series as the sum of two sinusoidal functions of known angular frequency.

▷ Model (11) is linear in the three unknown parameters, μ, A and B, and we can estimate it using least squares.

Representation of seasonality by a cycle

▷ Assuming that T is an integer number of cycles, we can obtain the following expressions:

$$\widehat{\mu} = \frac{1}{T} \sum_{t=1}^{T} z_t$$

and

$$\widehat{A} = \frac{2}{T} \sum_{t=1}^{T} z_t \sin(wt) \tag{12}$$

$$\widehat{B} = \frac{2}{T} \sum_{t=1}^{T} z_t \cos(wt) \tag{13}$$

Then, an estimator of the amplitude, R, is:

$$\widehat{R}^2 = \widehat{A}^2 + \widehat{B}^2 \tag{14}$$

and, and estimator of the phase, θ, is:

$$\theta = \arctan A/B \tag{15}$$

Representation of seasonality by a cycle

- The residuals of the model are calculated using:

$$\widehat{a}_t = z_t - \widehat{\mu} + \widehat{A}\sin(wt) + \widehat{B}\cos(wt)$$

and they have zero mean and variance $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{a}_t^2$.

- The variance of the variable, z, is

$$\frac{1}{T}\sum_{t=1}^{T}(z_j - \widehat{\mu})^2 = \frac{1}{T}\sum_{t=1}^{T}(\widehat{A}\sin(wt) + \widehat{B}\cos(wt) + \widehat{a}_t)^2$$

and using the properties that the variables $\sin(wt)$ and $\cos(wt)$ have zero mean, variance $1/2$ and they are uncorrelated, gives us:

$$\frac{1}{T}\sum_{t=1}^{T}(x_j - \widehat{\mu})^2 = \frac{\widehat{A}^2}{2} + \frac{\widehat{B}^2}{2} + \widehat{\sigma}^2 = \frac{\widehat{R}^2}{2} + \widehat{\sigma}^2, \quad (16)$$

which can be interpreted as a decomposition of the variance into two orthogonal components of variability.

Example 20

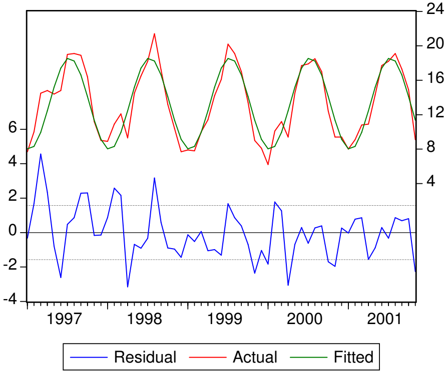
We fit a 12 period sinusoidal function, or angular frequency $2\pi/12$, to the average monthly temperature in Santiago de Compostela (Spain):

Dependent Variable: TEMPERATURE
Method: Least Squares
Date: 01/27/08 Time: 20:18
Sample (adjusted): 1997M01 2001M11
Included observations: 59 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	13.27567	0.204400	64.94950	0.0000
SINE	-3.335080	0.286562	-11.63823	0.0000
COSINE	-4.135446	0.291546	-14.18452	0.0000

R-squared	0.857379	Mean dependent var	13.34576
Adjusted R-squared	0.852285	S.D. dependent var	4.083834
S.E. of regression	1.569567	Akaike info criterion	3.788985
Sum squared resid	137.9582	Schwarz criterion	3.894623
Log likelihood	-108.7751	Durbin-Watson stat	1.428331

Datafile `tempsantiago.xls`



▷ The amplitude of the wave is $\widehat{R} = 5.36$. The initial angle is $\theta = -0.6727$ radians.

▷ The temperature in Santiago can be represented by a sine wave whose amplitude is 5.36 degrees centigrade and which begins in January with an angle difference of -0.6727 radians.

▷ Twice the amplitude of the wave indicates the average maximum difference between the coldest and warmest months, 10.7 degrees in this case.

Example 21

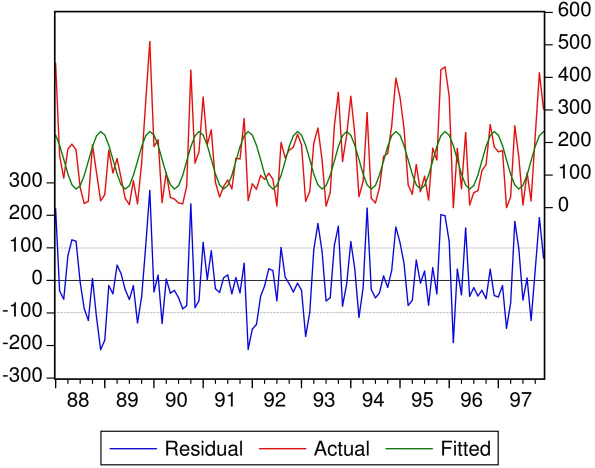
We fit a 12 period sinusoidal function, or angular frequency $2\pi/12$, to the average monthly rainfall in Santiago de Compostela (Spain):

Dependent Variable: RAINFALL
Method: Least Squares
Date: 01/27/08 Time: 21:48
Sample: 1988M01 1997M12
Included observations: 120
RAINFALL = C(1) + C(2)*SINE + C(3)*COSINE

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	145.3224	9.078518	16.00728	0.0000
C(2)	3.046136	12.84079	0.237223	0.8129
C(3)	88.27209	12.83713	6.876309	0.0000

R-squared	0.288063	Mean dependent var	145.3500
Adjusted R-squared	0.275893	S.D. dependent var	116.8703
S.E. of regression	99.45017	Akaike info criterion	12.06187
Sum squared resid	1157169.	Schwarz criterion	12.13156
Log likelihood	-720.7124	Durbin-Watson stat	1.578078

Datafile `rainfall.xls`



▷ Although the sinusoidal model explains part of the variability of the series, the fit now is not as good as in the temperature series in the above example.

▷ This is because the sinusoidal function is not able to pick up the asymmetries or the peaks in the observed series.

Exploration of multiple cycles

▷ The representation of a seasonal series using equation (11) is adequate when the seasonality from period s is exactly sinusoidal, but it does not help us describe general periodic functions.

▷ A generalization of this analysis is to allow the periodic function to be the sum of various harmonic functions with different frequencies.

▷ Given a series of length T, we use the terms **basic periods** or **Fourier periods** to describe those that are exact period fractions of the sample size. That is, the basic periods are defined by:

$$s_j = \frac{T}{j}, \text{ for } j = 1, 2, ..., T/2$$

The maximum value of the basic period is obtained for $j = 1$ and is T, the sample size. Hence, we observe the wave only once. The minimum value of the basic period is obtained for $j = T/2$ and is 2, because we cannot observe periods that last fewer than two observations.

Exploration of multiple cycles

▷ In fitting the cycle we usually work with frequencies instead of periods, and the **basic frequencies**, or **Fourier frequencies**, are defined as the inverses of the basic periods:

$$f_j = \frac{j}{T}, \text{ for } j = 1, 2, ..., T/2,$$

which gives us $1/2 \geq f_j \geq 1/T$, and the maximum value of the frequency we can observe is $f = .5$.

▷ We can obtain a general representation of a periodic function as a sum of waves associated with all the basic frequencies, using:

$$z_t = \mu + \sum_{j=1}^{T/2} A_j \sin(w_j t) + \sum_{j=1}^{T/2} B_j \cos(w_j t). \tag{17}$$

This equation contains as many parameters as observations, thus it will always exactly fit any series being observed.

Exploration of multiple cycles

▷ Therefore, we have to find a procedure for selecting the frequencies that we must include in this equation in order to explain the evolution of the series. This is the purpose of the **periodogram**.

▷ The equation (17) allows us to decompose exactly an observed time series as a sum of harmonic components. According to equation (16), the contribution of a wave to the variance of a series is the square of the amplitude divided by two.

▷ Therefore, waves with a high estimated amplitude will be important in explaining the series, whereas those waves with low amplitude contribute little to its explanation.

▷ To select the important frequencies we can calculate the parameters A_j and B_j for all the basic frequencies and represent the contribution to the variance of the series, that is the amplitude of the wave squared and divided by two, as in a frequency function.

Exploration of multiple cycles

▷ Given the estimated coefficients $\widehat{A}_j$ and $\widehat{B}_j$ for frequency w_j we calculate $\widehat{R}_j = \widehat{A}_j^2 + \widehat{B}_j^2$, and using equation (17), we decompose the variance of the series in components associated with each one of the harmonic functions. Letting s_z^2 be the sample variance of the series, we write:

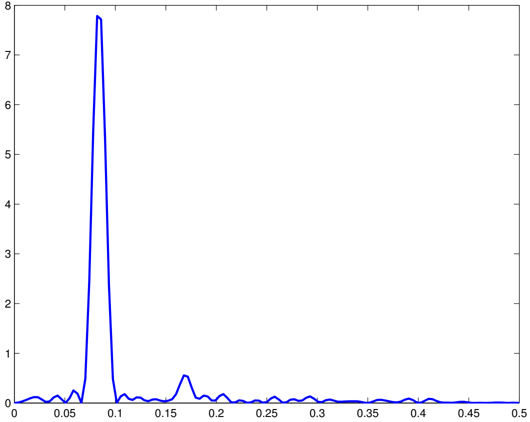
$$Ts_z^2 = \sum_{t=1}^{T}(z_t - \mu)^2 = \sum_{j=1}^{T/2} \frac{T}{2}\widehat{R}_j^{\,2} \tag{18}$$

▷ The **periodogram** is the representation of the contribution of each frequency, $T\widehat{R}_j^{\,2}/2$, as a function of the frequency w_j or f_j:

$$I(f_j) = \frac{T\widehat{R}_j^{\,2}}{2}, \quad \text{with } 1/T \leq f_j \leq .5. \tag{19}$$

Example 22

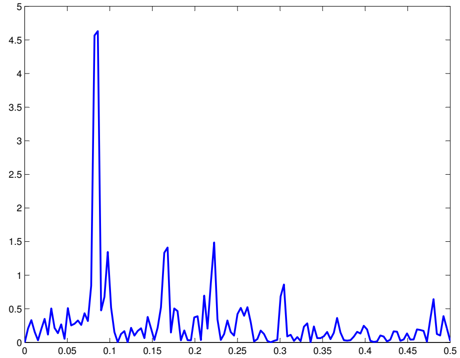
We are going to calculate the periodogram for the series of temperature in Santiago.



▷ As expected, a high and isolated peak is observed in the monthly seasonal frequency, $f = 1/12 = .083$.

Example 23

We are going to calculate the periodogram for the series of rainfall in Santiago.



▷ A peak is observed in the monthly seasonal frequency, $f = 1/12 = .083$.

▷ The graph also shows peaks in frequencies around $1/6$=.16 and $1/3$=.33, which are associated with seasonal frequencies.

▷ This result indicates that seasonality is modelled better as a sum of these three harmonic components than with a deterministic 12 period wave.

Summary

- We have sawn how to fit deterministic trend models and their limitations in representing real series.
- The smoothing methods based on giving decreasing weight to the past work better, but they involve a dependence structure that, while flexible, is not applicable to all real series.
- Decomposition methods are useful, but more flexible methods are needed for the components.
- The periodogram is a valuable tool for detecting deterministic sinusoidal components in a series, such as cyclical seasonal effects.