

W271-2 – Spring 2016 – HW 8

Juanjo Carin, Kevin Davis, Ashley Levato, Minghu Song

April 6, 2016

Build an univariate linear time series model (i.e AR, MA, and ARMA models) using the series in `hw08_series.csv`.

- Use all the techniques that have been taught so far to build the model, including date examination, data visualization, etc.
- All the steps to support your final model need to be shown clearly.
- Show that the assumptions underlying the model are valid.
- Which model seems most reasonable in terms of satisfying the model's underlying assumption?
- Evaluate the model performance (both in- and out-of-sample)
- Pick your “best” models and conduct a 12-step ahead forecast. Discuss your results. Discuss the choice of your metrics to measure “best”.

First we load the series:

```
hw08 <- read.csv('hw08_series.csv', header = TRUE)
str(hw08)

## 'data.frame':    372 obs. of  2 variables:
## $ X: int  1 2 3 4 5 6 7 8 9 10 ...
## $ x: num  40.6 41.1 40.5 40.1 40.4 41.2 39.3 41.6 42.3 43.2 ...

all(hw08$X == 1:dim(hw08)[1]) # check if 1st column is just an incremental index

## [1] TRUE

hw08 <- hw08[, -1]
```

The file has two columns but the first one is just an incremental index so we discard it. The second column (that is stored in a numeric vector called `hw08`) contains 372 observations. 372 is a multiple of 12 ($372/12 = 31$) so we'll assume that the series contains monthly observations from 31 years (*for labelling purposes only, sometimes we'll also assume that the period goes from 1980 to 2010*).

Let's explore the main descriptive statistics of the series, as well as its histogram and time-series plot:

```
# See the definition of the function in ## @knitr Libraries-Functions-Constants
desc_stat(hw08, 'Time series', 'Descriptive statistics of the time series.')
```

Table 1: Descriptive statistics of the time series.

	Time series
Mean	84.83
St. Dev	31.95
1st Quartile	57.38
Median	76.45
3rd Quartile	111.53
Min	36.00
Max	152.60

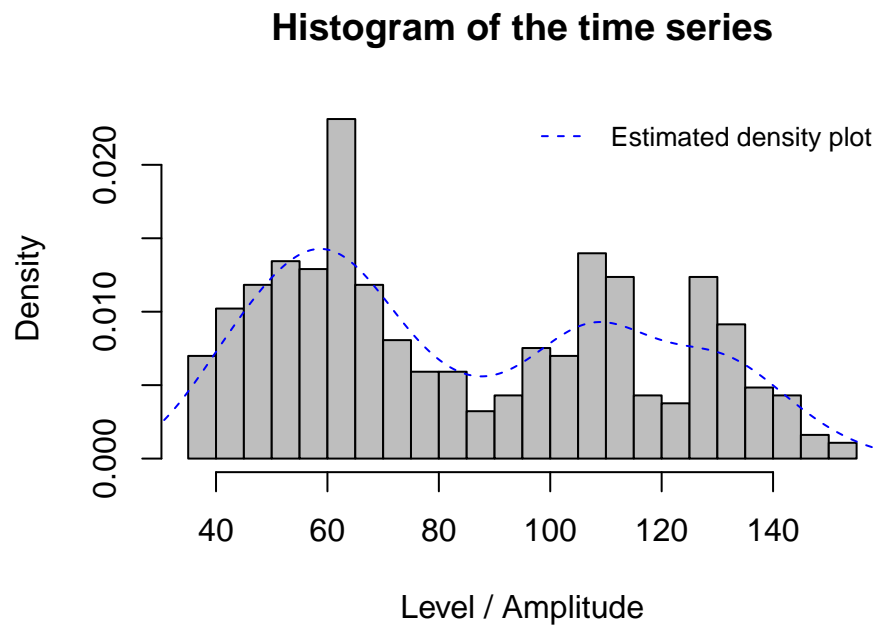


Figure 1: Histogram of the data.

The histogram shows that the distribution of the data is multimodal, and far from normal. But as usual, it tells us nothing about the dynamics of the time series. To label the time-series plot, we will assume (as mentioned) that the data were collected on a monthly basis and will use 1980 as an arbitrary starting point.

```
hw08.ts <- ts(hw08, start = c(1980,1), frequency = 12)
```

Time-series plot of the data

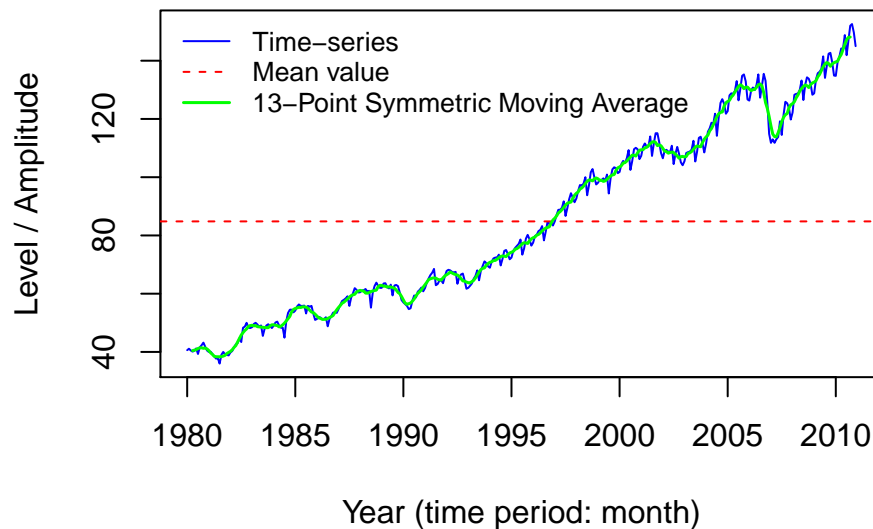


Figure 2: Time-series plot (assuming monthly data, from 1980 until 2010).

Our assumption that the data corresponds to a monthly time series seems reasonable after noticing that there seems to be some seasonality every 12 time periods (see Figure 3 below: the level increases over the first 6 months—especially in February and June—, goes down in July, up from August to October, and down again the last 2 months of the year).

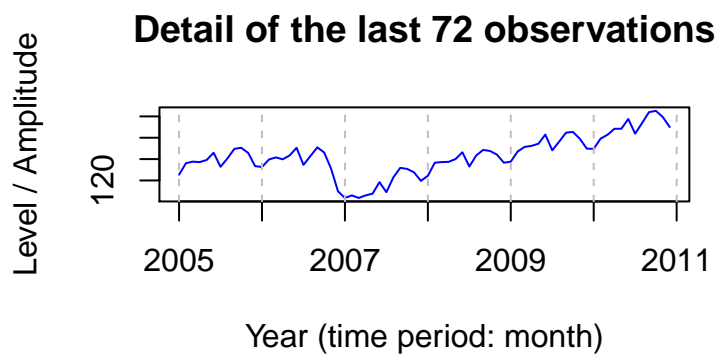


Figure 3: Time-series plot of (the last 72 observations—6 years?—of) the data.

Apart from showing that the time series is **not mean stationary** (the mean depends on time, with an increasing trend), Figure 2 in the previous page shows that the time is also **not variance stationary**: the variance is not constant but changes with time (increasing in the last years, especially the last 7); see Table 2 and Figure 4 in the next page.

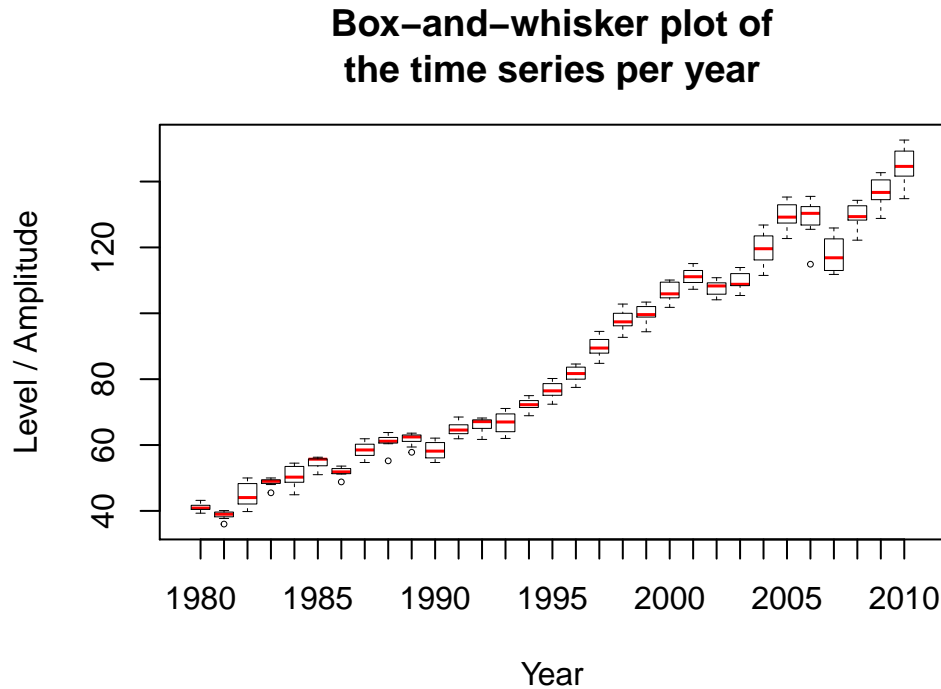


Figure 4: Boxplot of the series, per year (every 12 observations).

Table 2: Variance of the time-series amplitude per year (for the first 30 out of 31).

Year	Variance	Year	Variance	Year	Variance
1980	1.12	1990	6.61	2000	7.84
1981	1.36	1991	3.66	2001	7.23
1982	13.06	1992	4.21	2002	5.03
1983	1.37	1993	9.79	2003	7.79
1984	8.63	1994	3.31	2004	22.65
1985	2.59	1995	5.49	2005	13.84
1986	1.73	1996	5.59	2006	30.61
1987	4.94	1997	8.18	2007	29.39
1988	4.42	1998	9.03	2008	12.05
1989	3.17	1999	6.37	2009	16.68

Both results indicate that the data does not seem to be a realization of a stationary process, so **an ARMA model may not be a good fit for our data**.

To finish the Exploratory Data Analysis, let's decompose the time series to check the growing trend and seasonality:

Decomposition of additive time series

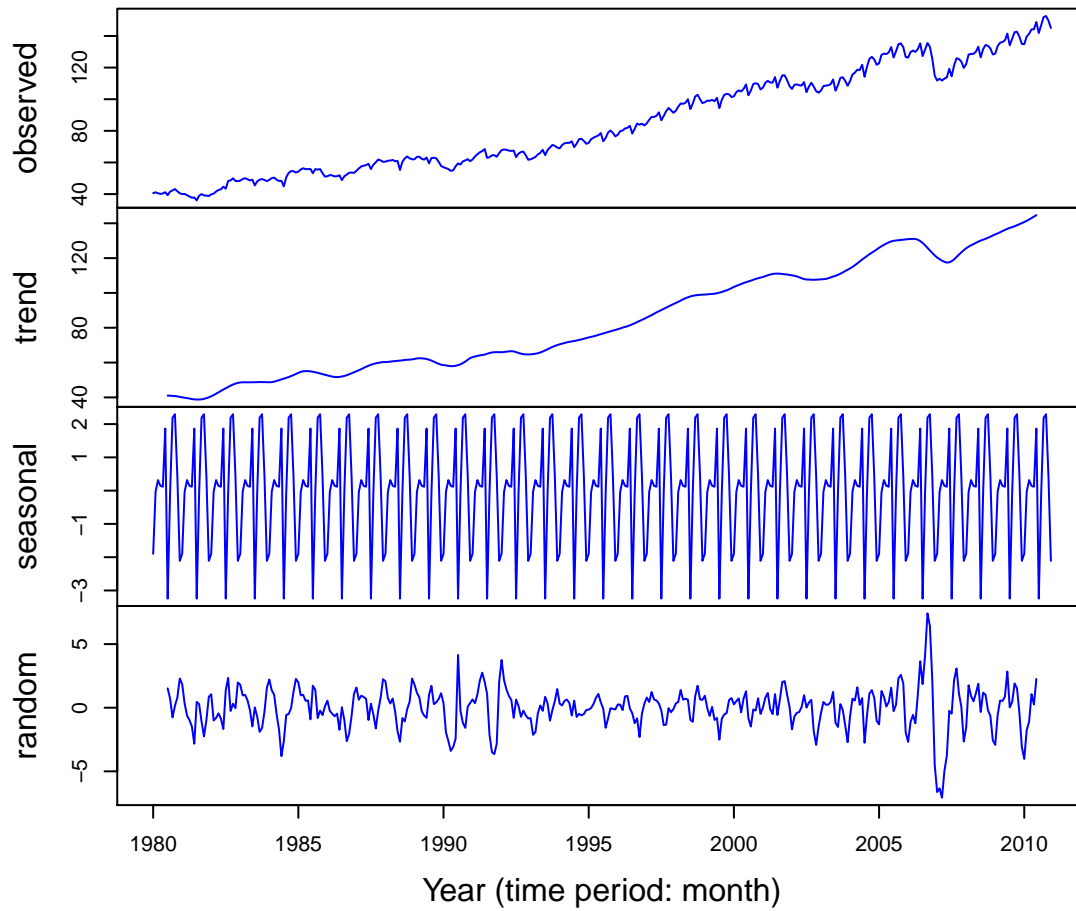


Figure 5: (Additive) decomposition of the time series.