Lähde: Rob Hyndman, http://www.r-bloggers.com/constants-and-arima-models-in-r/ (jos aihe kiinnostaa laajemmin, ks. OTexts.com/fpp/).

A non-seasonal ARIMA model can be written as

(1) $$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d y_t = c+(1+\theta_1 B+\cdots+\theta_q B^q)e_t$$

or equivalently as

(2) $$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d(y_t-\mu t^d/d!) = (1+\theta_1 B+\cdots+\theta_q B^q)e_t,$$

where $B$is the backshift operator, $c=\mu(1-\phi_1-\cdots-\phi_p)$and $\mu$is the mean of $(1-B)^d y_t$. R uses the parametrization of equation (2).

Thus, the inclusion of a constant in a non-stationary ARIMA model is equivalent to inducing a polynomial trend of order $d$in the forecast function. (If the constant is omitted, the forecast function includes a polynomial trend of order $d-1$.) When $d=0$, we have the special case that $\mu$is the mean of y_t.

Including constants in ARIMA models using R

arima()

By default, the `arima()` command in R sets $c=\mu=0$when $d>0$and provides an estimate of $\mu$when $d=0$. The parameter $\mu$is called the "intercept" in the R output. It will be close to the sample mean of the time series, but usually not identical to it as the sample mean is not the maximum likelihood estimate when $p+q>0$.

The `arima()` command has an argument `include.mean` which only has an effect when $d=0$ and is `TRUE` by default. Setting `include.mean=FALSE` will force $\mu=0$.

Arima()

The `Arima()` command from the forecast package provides more flexibility on the inclusion of a constant. It has an argument `include.mean` which has identical functionality to the corresponding argument for `arima()`. It also has an argument `include.drift` which allows $\mu\neq 0$when $d=1$. For $d>1$, no constant is allowed as a quadratic or higher order trend is particularly dangerous when forecasting. The parameter $\mu$is called the "drift" in the R output when $d=1$.

There is also an argument `include.constant` which, if `TRUE`, will set `include.mean=TRUE` if $d=0$and `include.drift=TRUE` when $d=1$. If `include.constant=FALSE`, both `include.mean` and `include.drift` will be set to `FALSE`. If `include.constant` is used, the values of `include.mean=TRUE` and `include.drift=TRUE` are ignored.

When $d = 0$and `include.drift=TRUE`, the fitted model from `Arima()` is

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y_t - a - bt) = (1 + \theta_1 B + \cdots + \theta_q B^q)e_t.$$

In this case, the R output will label $a$as the "intercept" and $b$as the "drift" coefficient.

auto.arima()

The `auto.arima()` function automates the inclusion of a constant. By default, for $d = 0$or $d = 1$, a constant will be included if it improves the AIC value; for $d > 1$the constant is always omitted. If `allowdrift=FALSE` is specified, then the constant is only allowed when $d = 0$.

Eventual forecast functions

The eventual forecast function (EFF) is the limit of $\hat{y}_{t+h|t}$as a function of the forecast horizon h as $h \to \infty$.

The constant $c$has an important effect on the long-term forecasts obtained from these models.

- If $c = 0$and $d = 0$, the EFF will go to zero.
- If $c = 0$and $d = 1$, the EFF will go to a non-zero constant determined by the last few observations.
- If $c = 0$and $d = 2$, the EFF will follow a straight line with intercept and slope determined by the last few observations.
- If $c \neq 0$and $d = 0$, the EFF will go to the mean of the data.
- If $c \neq 0$and $d = 1$, the EFF will follow a straight line with slope equal to the mean of the differenced data.
- If $c \neq 0$and $d = 2$, the EFF will follow a quadratic trend.

Seasonal ARIMA models

If a seasonal model is used, all of the above will hold with $d$replaced by $d + D$where $D$is the order of seasonal differencing and $d$is the order of non-seasonal differencing.