# ST 337 / ST 405: Bayesian Forecasting and Intervention

Dr Bärbel Finkenstädt[1], Department of Statistics, University of Warwick

[1]B.F.Finkenstadt@warwick.ac.uk.

# CONTENTS

CHAPTER 5

# CLASSICAL TIME SERIES

Here we will briefly describe classical modelling of time series as autoregressive moving-average models (ARMA), and explore its relation to the DLM.

## 5.1 Autoregressive models

Autoregressive models relate current values of the observed time series to values of the same series observed in the past, the further in the past, the higher is the order of the model. The simplest is the first order autoregressive, AR(1), model defined as

$$y_t = \phi\, y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is white noise (a series of uncorrelated, zero mean, constant variance stochastic quantities). More generally, the AR(p), is written as

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}\left(\varepsilon_t \mid 0, \sigma^2\right)$$

where $\mu$ is the mean of the series, $\left\{\phi_j\right\}$ is the set of autoregressive coefficients and we are adopting the usual assumption of normality. Note that the parameters do not vary over time.

## 5.2 Moving average models

Moving average models (MA) relate the current value of the observed series to past values of the unobservable error terms. Again the simplest of these models is the first order moving average MA (1)

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1}$$

with $\{\varepsilon_t\}$ white noise.

## 5.3 ARMA models

Combining these two basic models yields the ARMA class of models. In general an ARMA (p,q) model is written as

$$y_t - \mu = \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{j=1}^{q} \psi_j \, \varepsilon_{t-j} \; + \; \varepsilon_t .$$

## 5.4 Stationarity

Due to classical modelling of time series, some restrictions should be imposed in order to be able to apply the methodology. The main assumption is that $y_t$ must be stationary.

**Definition 1 (Stationarity).**
*A time series $\{y_t \, : \, t = 1, 2, \ldots\}$ is said to be stationary if the joint distribution of any collection of k values is invariant under arbitrary shifts of the time axis, i.e. iff for $k \geq 1$ and $s > 0$,*

$$p(y_{t_1}, y_{t_2}, \ldots, y_{t_k}) = p(y_{t_1+s}, y_{t_2+s}, \ldots, y_{t_k+s}) \, .$$

Sometimes a weaker condition suffices,

**Definition 2 (Weak stationarity).**
*A time series $\{y_t \, : \, t = 1, 2, \ldots\}$ is said to be weakly stationary (second order stationary) if its mean, variance and covariances are invariant under time shifts, i.e. iff for $t > 0$ and $s < t$*

$$\begin{aligned} \mathrm{E}[y_t] &= \mu \\ \mathrm{Var}[y_t] &= \sigma_y^2 \\ \mathrm{Cov}[y_{t-s}, y_t] &= \gamma_s \end{aligned}$$

The importance of stationarity for classical ARMA modelling is highlighted by the following result. First, remember that a **white noise process** $\{\epsilon_t\}$ is such that for all $t$ we have

$$\begin{aligned} \mathrm{E}[\epsilon_t] &= 0 \\ \mathrm{Var}[\epsilon_t] &= \sigma^2 \\ \mathrm{Cov}[\epsilon_t, \epsilon_{t-s}] &= 0 \qquad \forall\, s \neq 0\,. \end{aligned}$$

Such a process is a basic building block from the following

**Theorem 5.1 (Wold decomposition).**
*Every weakly stationary process $\{y_t\}$ can be written as the sum of a deterministic part and an infinite moving average of white noise random variables as*

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \, \epsilon_{t-i}\,, \qquad \psi_0 = 1\,, \qquad \text{and} \sum_{i=0}^{\infty} \psi_i^2 < \infty.$$

The stationarity assumption restricts the range of possible values for the parameters. In general, the autoregressive coefficients of an ARMA model must be such that all the roots of the autoregressive characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

are greater than one in absolute value (complex roots should lie outside the unit circle).

An ARMA model with such an AR part will be expressible as an infinite order MA model and will be such that the $\psi_i$ decrease with $i$: the influence of shocks a long time ago will eventually taper off. This gives the system a certain amount of stability; even large shocks $\epsilon_t$ will lose their effect after some time.

**Example 1.**
An AR(1) model ($p = 1, q = 0$) with $\mu = 0$.

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$

so that

$$\begin{aligned} y_1 &= \phi_1 y_0 + \epsilon_1 \\ y_2 &= \phi_1 \left(\phi_1 y_0 + \epsilon_1\right) + \epsilon_2 \\ \vdots &= \vdots \\ y_t &= \phi_1^t y_0 + \sum_{i=0}^{t-1} \phi_1^i \epsilon_{t-i} \end{aligned}$$

- If $\left|\phi_1\right| < 1$ the root of $1 - \phi_1 z = 0$ will be larger than one in absolute value (stationarity) and $y_t$ will tend to $\sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}$, where the effect of a shock will die out with time. $y_t$ is stationary and mean-reverting (mean is zero).

- If $\phi_1 = 1$ then $y_t = y_0 + \sum_{i=0}^{\infty} \epsilon_{t-i}$, and we have a random walk without mean reversion, so shocks are permanent. We often say such a series has a unit root and we need to take a first difference $\{y_t - y_{t-1}\}$ for modelling, which will simply be a white noise process.
- If $|\phi_1| > 1$ we have an explosive process, which means that shocks are amplified over time. This occurs only very rarely. ◁

An important tool for identifying ARMA models from data is the **autocorrelation function** of the data.

**Definition 3 (Autocorrelation function (ACF)).**
*The autocorrelation at lag s of a time series $y_t$ is*

$$\rho_s = \frac{\gamma_s}{\gamma_0} \qquad \textit{where} \qquad \gamma_0 = \mathrm{Var}[y_t]\,.$$

Once a stationary series is obtained by differencing, the autocorrelation structure may be modeled by fitting it to known structures within the ARMA class.

In order to make sure the parameters are uniquely identified by the ACF (which is typically all the information we have to fit the AR and MA parameters), we need the condition of **invertibility**, which restricts the MA parameters. In particular, it states that the roots of the moving average characteristic polynomial

$$\psi(z) = 1 + \psi_1 z + \psi_2 z^2 + \cdots + \psi_q z^q = 0$$

are greater than one in absolute value (lie outside the unit circle for complex roots).

**Example 2.**
Consider an MA (1) process, i.e. $p = 0, q = 1$, with $\mu = 0$.

$$y_t = \epsilon_t + \psi_1 \epsilon_{t-1}\,, \quad \epsilon_t \sim \mathrm{N}\left(\epsilon_t \mid 0, \sigma^2\right) \text{ (white noise)}$$

Then $\mathrm{E}[y_t] = 0$ and

$$\mathrm{Var}[y_t] = \gamma_0 = \mathrm{Var}[\epsilon_t] + \psi_1^2 \mathrm{Var}[\epsilon_{t-1}] + 2\psi_1 \mathrm{Cov}[\epsilon_t, \epsilon_{t-1}] = \sigma^2(1 + \psi_1^2)$$
$$\mathrm{E}[y_t\, y_{t-1}] = \gamma_1 = \psi_1 \sigma^2$$

and thus

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\psi_1}{1+\psi_1^2}.$$

Note that $\left|\rho_1\right| < 1/2$ and $\rho_s = \gamma_s/\gamma_0 = 0$ for all $s > 1$.

Also note that we get exactly the same ACF if we take $\psi_1^{-1}$ rather than $\psi_1$. So we need to make a choice through the invertibility condition $\left|\psi_1\right| < 1$.

Under invertibility, we can rewrite the MA (1) process as an AR($\infty$) process. ◁

## 5.5 DLM, again

Classical ARMA models can be written in DLM form as follows. Let $m = \max\{p, q+1\}$ and $\phi_j = 0$ for $j = p+1, p+2, \ldots, m$; and, similarly $\psi_k = 0$ for $k = q+1, \ldots, m$. Thus (taking $\mu = 0$),

$$y_t = \sum_{i=1}^{m}\Big(\phi_i y_{t-i} + \psi_i \varepsilon_{t-i}\Big) + \varepsilon_t\,.$$

Now, define the following DLM

$$F = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad G = \begin{pmatrix} \phi_1 & 1 & 0 & \ldots & 0 \\ \phi_2 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{m-1} & 0 & \ldots & 0 & 1 \\ \phi_m & 0 & \ldots & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{\omega}_t = \begin{pmatrix} 1 \\ \psi_1 \\ \ddots \\ \psi_{m-2} \\ \psi_{m-1} \end{pmatrix} \varepsilon_t$$

and $v_t = 0$.

**Example 3 (AR (1)).**
In this case $p = 1$ and $q = 0$, so that $m = 1$ and $F = 1$, $G = \phi_1$ and $\omega_t = \varepsilon_t$.

Indeed, then $y_t = \theta_t$ and $\theta_t = \phi_1\,\theta_{t-1} + \varepsilon_t$, which gives the usual AR (1) model. ◁

**Example 4 (MA(1)).**
Now, $p = 0$, $q = 1$ and thus $m = 2$. Therefore,

$$F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad G = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \omega_t = \begin{pmatrix} 1 \\ \psi_1 \end{pmatrix} \varepsilon_t$$

Thus,

$$y_t = \theta_{1\,t}$$

and

$$\theta_{1\,t} = \theta_{2\,t-1} + \varepsilon_t$$
$$\theta_{2\,t} = \psi_1 \varepsilon_t$$

so that $y_t = \theta_{1\,t} = \varepsilon_t + \psi_1 \varepsilon_{t-1}$, the MA (1) model. ◁

### 5.5.1 DLM representation for AR (p)

For an AR (p) model one can alternatively define

$$F_t = \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix}, \qquad \boldsymbol{\theta}_t = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}, \qquad G = I_p\,, \quad \omega_t = 0\,, \quad \text{and} \quad \nu_t = \varepsilon_t\,.$$

The fact that $\theta_t = \theta_{t-1}$ highlights the parameter constancy of ARMA models.

With this representation it is readily seen that the forecast function requires the future value of the series being forecasted. The need of knowing future values of series is a feature of all DLM s with a regression component, when forecasting.

## 5.6 Classical Modelling

Many observed time series are not stationary. Thus, in practice, original series must be transformed before applying the methodology. The most common of such transformation is differencing. For instance, a series exhibiting a constant drift in trend can be transformed to a constant level (no mean drift) series by taking first differences $z_t = y_t - y_{t-1}$.

When a series needs to be differenced in this way to achieve stationarity, the original series is said to be integrated. An ARMA model applied to a differenced series is referred to as an autoregressive integrated moving average (ARIMA) model. An ARMA (3,2) model applied to a first differenced series is an ARIMA(3,1,2) model, for the original series.

Let $B$ denote the back-shift operator i.e. $Bx_t = x_{t-1}$, $B^p x_t = x_{t-p}$. Thus we can write $z_t = (1-B)y_t$. A quadratic trend growth may be removed by differencing the series twice, i.e. $w_t = (1-B)^2 y_t = y_t - 2y_{t-1} + y_{t-2}$. When attempting to achieve stationarity this way, it is important that the original series is non-seasonal. Seasonality may be removed by seasonal differencing; for instance, for monthly data exhibiting annual seasonality, the twelfth seasonal difference $(1-B^{12})y_t$ would remove such seasonality. So, the classical strategy is to construct a series of the form

$$z_t = (1-B)^k (1-B^p) y_t\,.$$

Plot of the differenced series are typically useful tools to help us decide which order of differencing is more appropriate. In particular, we consider the autocorrelation function. Once we feel that the ACF of a transformed series mimics one of the (known) ACF patterns of stationary ARMA series, we can use that transformation to estimate an ARMA model.

## 5.7 Forecasting

As described in Section 5.5, any ARIMA model is a TSDLM, i.e. the matrices $F$ and $G$ are constant. Recall that for a constant DLM, $C_t^* = G\,C_{t-1}G' + W$, $q_t = F'C_t^*F + v$ and $A_t = C_t^*F/q_t$. For a constant DLM we can also show that $\lim_{t\to\infty} C_t = C$, for any $D_0$. This leads to the following convergence results:

$$\lim_{t\to\infty} C_t^* = G\,C\,G' + W = C^*$$
$$\lim_{t\to\infty} q_t = F'C^*F + v = q$$
$$\lim_{t\to\infty} A_t = C^*F/q = A\,.$$

Moreover, $m_t = G\,m_{t-1} + Ae_t = H\,m_{t-1} + AY_t$, where $H = C\,C^{*-1}G$. Let $\lambda_i$ denote the eigenvalues of $G$ and $\rho_i$ those of $H$ for $i = 1, \ldots, n$. Then the limiting representation of the series in terms of the forecasting errors is given by

$$Y_t = \sum_{j=1}^{n} \alpha_j Y_{t-j} + e_t + \sum_{j=1}^{n} \beta_j e_{t-j} \tag{5.1}$$

with

$$\alpha_1 = \sum_{i=1}^{n} \lambda_i\,, \qquad \alpha_2 = -\sum_{i=1}^{n}\sum_{k=i+1}^{n} \lambda_i\lambda_k\,, \qquad \alpha_n = (-1)^{n+1}\lambda_1\lambda_2\ldots\lambda_n$$
$$\beta_1 = -\sum_{i=1}^{n} \rho_i\,, \qquad \beta_2 = \sum_{i=1}^{n}\sum_{k=i+1}^{n} \rho_i\rho_k\,, \qquad \beta_n = (-1)^{n}\rho_1\rho_2\ldots\rho_n$$

This representation provides a link with the ARMA predictors. Further,

*i.–* Suppose that $p$ of the eigenvalues of $G$ satisfy $0 < \lambda < 1$; exactly $d$ are equal to one and the remaining $n - p - d$ are zero. Also, suppose that $q$ of the eigenvalues of $H$ satisfy $0 < \rho < 1$ with the remainder being zero. Then (5.1) is an ARIMA$(p, d, q)$ predictor.

*ii.–* The ARIMA predictor is just a limit result of a particular DLM.

**Theorem 5.2 (ARIMA DLM).**
*In the univariate constant* DLM$\{F, G, v, W\}$, *denote by $\lambda_i$ the eigenvalues of $G$ and by $\rho_i$ the eigenvalues of $H$, $i = 1, \ldots, n$. If the series $\{Y_t\}$ is indeed generated by this* DLM *then it can be*

*represented as*

$$\prod_{i=1}^{n}\Big(1-\lambda_i B\Big)Y_t=\prod_{i=1}^{n}\Big(1-\rho_i B\Big)a_t\,,$$

*where* $a_t \sim N\left(a_t \mid 0,q\right)$ *i.i.d.*

In this case, and when conditions $i$ above are met, $Y_t$ can be thought of as following an ARIMA$(p,d,q)$ process.

Thus, all ARIMA $(p,d,q)$ processes can be represented by a constant DLM with $n=\max\{p+d,q\}$. However, DLM's are, of course, more general as they do not need to be constant, and even constant DLM's can deal with $\lambda_i>1$ (explosive cases). Finally, DLM's are, of course, naturally adaptive and do not need to behave as their limiting forms.

**Example 5.**
For the first order polynomial model we have

$$\begin{aligned} e_t &= y_t - m_{t-1} \\ m_t &= m_{t-1} + A_t e_t \end{aligned}$$

And, therefore we may write $y_t - y_{t-1} = e_t - (1-A_{t-1})e_{t-1}$. If, in addition $v_t = v$ and $W_t = W$, then it is straightforward to show that

$$\lim_{t\to\infty} A_t = A = \frac{\sqrt{1+4v/W}-1}{2v/W}$$

and for the forecast variance

$$\lim_{t\to\infty} q_t = q = v/(1-A)$$

Leading to the limiting model

$$y_t = y_{t-1} + \varepsilon_t + \psi\varepsilon_{t-1} \quad \text{where} \quad \psi = -(1-A) \quad \text{and} \quad \varepsilon_t \sim \mathrm{N}\left(\varepsilon_t \mid 0,q\right) \text{ i.i.d.}\,.$$

Thus the limiting form of the forecast function is that of an ARIMA $(0,1,1)$. In a similar fashion, one can show that if a second order polynomial trend DLM is considered, then the limiting form is that of an ARIMA$(0,2,2)$. ◁

**Note:** for the model in Example 14 we can immediately write (without need for a limit) $y_t - y_{t-1} = \omega_t + v_t - v_{t-1}$ from which we see that the process $\{y_t\}$ has an ARIMA $(0,1,1)$ structure. In order to express it as a particular ARIMA process in terms of the forecast errors, we need the limiting argument in Example 14.