ST 337 / ST 405: Bayesian Forecasting and Intervention

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

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

CONTENTS

CHAPTER 3

A MORE COMPLEX DLM

Derived from the principle that 'any linear combination of independent linear models is a linear model', the **superposition principle** gives us a means of constructing a complex DLM, with basic blocks which are in turn simpler DLMs. The reverse process, that of identifying simpler components from a complex DLM is referred to as **decomposition**.

3.1 Model building by superposition

Sometimes a complex time series can be decomposed in two or more simpler series. For instance, the car sales series of Figure 3.1 presents a more or less linear increasing trend, and then some seasonal variations around this trend. Thus, we could represent this series as

$$y_t = y_{Lt} + y_{St} + \nu_t$$

where

$$\begin{aligned} y_{Lt} &= F_{Lt}\theta_{Lt} \\ \theta_{Lt} &= G_{Lt}\theta_{Lt-1} + \omega_{Lt} \end{aligned}$$

represents the linear block and

$$\begin{aligned} y_{St} &= F_{St}\theta_{St} \\ \theta_{St} &= G_{St}\theta_{St-1} + \omega_{St} \end{aligned}$$

stands for the seasonal part.

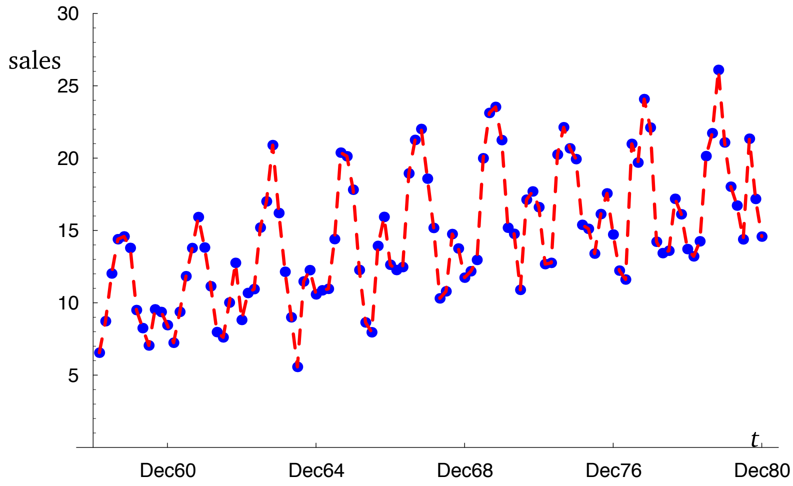


Figure 3.1. *Monthly car sales in Quebec.*

The observation equation is then a linear combination of these components

$$\begin{aligned} y_t &= F_{Lt}\theta_{Lt} + F_{St}\theta_{St} + \nu_t \\ &= F_t\theta_t + \nu_t \end{aligned}$$

where $F_t = (F_{Lt}, F_{St})$ and $\theta_t' = (\theta_{Lt}', \theta_{St}')$. Now, for the system equation we can write

$$\theta_t = G_t\theta_{t-1} + \omega_t$$

where $\omega_t' = (\omega_{Lt}', \omega_{St}')$. Thus, the evolution and variance matrices have the diagonal form:

$$G_t = \begin{pmatrix} G_{Lt} & 0 \\ 0 & G_{St} \end{pmatrix} \qquad W_t = \begin{pmatrix} W_{Lt} & 0 \\ 0 & W_{St} \end{pmatrix} ;$$

clearly indicating that the state variables for each component evolve independently.

A host of complex models can be built in this simple way and still remain a DLM.

Indeed, generally the following holds

Theorem 3.1 (Principle of superposition).
Consider h time series Y_{it}, generated by DLM*'s*

$$M_i = \{F_{it}, G_{it}, V_{it}, W_{it}\} \qquad \textit{for} \qquad i = 1, \ldots, h\,.$$

In M_i, the state vector $\boldsymbol{\theta}_t$ is of dimension n_i and the observation and evolution error series are respectively $\nu_{i\,t}$ and $\omega_{i\,t}$. The state vectors are distinct, and for all $i \neq j$, the series $\nu_{i\,t}$ and $\omega_{i\,t}$ are mutually independent of $\nu_{j\,t}$ and $\omega_{j\,t}$.

Then the series

$$Y_t = \sum_{i=1}^{h} Y_{i\,t}$$

follows the n-dimensional DLM $\{F_t, G_t, V_t, W_t\}$ *where $n = n_1 + \cdots + n_h$ and the state vector $\boldsymbol{\theta}_t$ and the quadruple are given by*

$$\boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{1\,t} \\ \boldsymbol{\theta}_{2\,t} \\ \vdots \\ \boldsymbol{\theta}_{h\,t} \end{pmatrix} \qquad F_t = \left(F_{1\,t} | F_{2\,t} | \ldots | F_{h\,t} \right) ,$$

$$G_t = \text{block diag}\left[G_{1\,t}, \ldots, G_{h\,t} \right] ,$$

$$W_t = \text{block diag}\left[W_{1\,t}, \ldots, W_{h\,t} \right]$$

and

$$V_t = \sum_{i=1}^{h} V_{i\,t}.$$

Using the fact that

$$\prod_{i=0}^{k-1} G_{t+k-i} = \begin{pmatrix} \prod_{i=0}^{k-1} G_{1\,t+k-i} & 0 & \ldots & \\ 0 & \prod_{i=0}^{k-1} G_{2\,t+k-i} & 0 & \\ & 0 & \ddots & \ddots \\ & 0 & \ldots & \prod_{i=0}^{k-1} G_{h\,t+k-i} \end{pmatrix} ,$$

we can easily show that the forecast function for Y_t, $f_t(k)$, is simply the sum of the forecast functions for the component series $Y_{i\,t}$, say $f_{i\,t}(k)$; i.e.

$$\begin{aligned} f_t(k) &= F_{t+k} \prod_{i=0}^{k-1} G_{t+k-i} \, \mathrm{E}[\boldsymbol{\theta}_t \mid D_t] \\ &= \sum_{j=1}^{h} F_{j\,t+k} \prod_{i=0}^{k-1} G_{j\,t+k-i} \, \mathrm{E}[\boldsymbol{\theta}_{j\,t} \mid D_t] = \sum_{j=1}^{h} f_{j\,t}(k) \, . \end{aligned}$$

So, the best point predictor for the superposition model is simply the sum of the corresponding predictors of the component models. This means that we can build up a DLM to have all the features we think the data need, such as the simple example of Figure 3.1.

We will now examine some of the most useful candidates for component models, to be used as 'building blocks' in DLM modelling.

3.1.1 Trend Components

The simplest trend model is the linear, or first order polynomial trend:

$$\begin{aligned} y_t &= \mu_t + \nu_t \\ \mu_t &= \mu_{t-1} + \omega_t \,. \end{aligned}$$

For this simple case, it is easy to verify that $F_t = F = 1$ and $G_t = G = 1$.

This model (also called the steady model) is extremely simple, with μ_t serving as the underlying level of the series, and is characterised by a simple constant forecast function: $f_t(k) = m_t = \mathrm{E}[\mu_t \mid D_t]$.

A second order polynomial trend model allows for a systematic change in the level, which is achieved by including an extra parameter in the state vector, i.e.

$$\begin{aligned} y_t &= \mu_t + \nu_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t} \,. \end{aligned}$$

Now, the state vector is $\theta_t = (\mu_t, \beta_t)'$, the first element determining the level of the series at time t and the second one representing the current rate of change in level (growth). Again, it is straightforward to identify the forms of F_t and G_t:

$$F_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \qquad \text{and} \qquad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \,.$$

This model is also called the linear growth model, as it leads to a linear forecast function

$$f_t(k) = F\, G^k\, \mathrm{E}[\boldsymbol{\theta}_t \mid D_t] = \mathrm{E}[\mu_t \mid D_t] + k\, \mathrm{E}[\beta_t \mid D_t], \quad \text{since} \quad G^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

It is now clear that μ_t still represents the underlying level, while β_t captures the (linear) growth of the series.

Higher order polynomial models can be easily defined in a similar fashion. However, in practice, first and second order polynomial models typically suffice.

Generally, a polynomial growth (or polynomial trend) model of order m is any TSDLM with k-step ahead forecast function of order $m-1$ in k.

3.1.2 Seasonality

A seasonal component incorporates periodic changes in the series (yearly, quarterly, monthly, weekly,...). This seasonal effects model defines parameters to measure seasonal departures from a trend (linear, quadratic,...).

The terms 'seasonality' is used for any periodic behaviour, whether or not it corresponds to well-defined seasons.

Definition 1 (Periodic function).
*A general function $g(t)$ is called **cyclical** or **periodic** if for some integer $p > 1$ and all integers t and $n \geq 0$, we have*

$$g(t + np) = g(t)\,.$$

*The smallest p for which this holds is called the **period** of $g(\cdot)$.*

We shall examine two types of seasonal models: the first are called **form-free seasonal** models, since they do not restrict the seasonal pattern. The second type are **seasonal harmonic** models, which represent the seasonal factors using trigonometric functions.

Form-free seasonal effects

Assume that series y_t exhibits seasonal quarterly effects. If the current quarter is quarter three, then the state is ordered as

$$\theta_t = \begin{pmatrix} qr_3 \\ qr_4 \\ qr_1 \\ qr_2 \end{pmatrix},$$

where qr_i is the seasonal effect associated with the i-th quarter. Note that these are simply seasonal deviations from a common mean and thus should add up to zero. Thus, $\mathbf{1}'\boldsymbol{\theta}_t = 0$ for all t, with $\mathbf{1}' = (1, 1, \ldots, 1)$. Naturally, the next quarter the state vector is rotated as

$$\theta_{t+1} = \begin{pmatrix} qr_4 \\ qr_1 \\ qr_2 \\ qr_3 \end{pmatrix}.$$

This periodic pattern is induced in the system equation by choosing the evolution matrix to be the cyclic form

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

So, the whole form-free seasonal effects model can be completed by setting $F_t = (1, 0, 0, 0)$.

Note that

$$G = \begin{pmatrix} 0 & I_{p-1} \\ 1 & \mathbf{0} \end{pmatrix}$$

is a p-cyclic matrix, so that for any integer $n \geq 0$, $G^{np} = I_p$ and $G^{k+np} = G^k$ for $k = 1, 2, \ldots, p$.

If we are in quarter 3 at time t, then the one-step ahead forecast function,

$$f_t(1) = FG\,\mathrm{E}[\boldsymbol{\theta}_t \mid D_t] = \mathrm{E}[qr_4 \mid D_t] ,$$

which corresponds to the appropriate (fourth) quarter. Clearly,

$$f_t(2) = FG^2\,\mathrm{E}[\boldsymbol{\theta}_t \mid D_t] = \mathrm{E}[qr_1 \mid D_t] ,$$
$$f_t(3) = \mathrm{E}[qr_2 \mid D_t] ,$$

while

$$f_t(4) = \mathrm{E}[qr_3 \mid D_t] ,$$

i.e. the mean of the seasonal factor for the current (third) quarter, which is the appropriate quarter for forecasting one year ahead.

The restriction $\mathbf{1}'\boldsymbol{\theta}_t = 0,\ \forall t$ has to be imposed throughout and implies that both C_0 (in the prior) and W_t will no longer have full rank.

Harmonics

The above formulation requires $(p-1)$ parameters to model a p-periodic series (recall that the seasonal effects must sum up to zero). We may use instead periodic, trigonometric functions. The cosine function $\cos(w\,t)$ has a period of $2\pi/w$, e.g. when $w = \pi/6$, the function has a period of 12, useful in monthly series with annual seasonal patterns. So, instead of having to define 11 effect parameters, only one is required

$$y_t = a_t \cos\left(\frac{\pi\, t}{6}\right) + \nu_t ;$$

where the parameter a_t models the amplitude, i.e. the seasonal peak. By adding a sine term with the same frequency, the seasonal peak can be translated to any position in the cycle

$$y_t = a_t \cos\left(\frac{\pi\, t}{6}\right) + b_t \sin\left(\frac{\pi\, t}{6}\right) + \nu_t \ .$$

Thus, a simple DLM with one harmonic function is defined by the regression vector and evolution matrix

$$F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{and} \qquad G = \begin{pmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{pmatrix} \ .$$

This model is p-cyclic, with $p = 2\pi/w$.

Including a second harmonic in the model allows for modelling more complex seasonal patterns. For instance, the function $\cos(2w)$ completes a cycle twice as fast as $\cos w$. Thus, the model

$$y_t = \Big[a_{1t} \cos(w\, t) + b_{1t} \sin(w\, t) \Big] + \Big[a_{2t} \cos(2w\, t) + b_{2t} \sin(2w\, t) \Big] + \nu_t \ .$$

can accommodate both semestral and annual seasonality.

Actually, the *Fourier representation theorem* assures that any cyclical function of even period p, can be expressed as a linear combination of $p/2$ sine and cosine terms on (frequency) $w_r = 2\, r\, \pi/p$, $r = 1, \ldots, p/2$.

The so-called r-th harmonic will be associated with frequency w_r and will have a cycle length of p/r. thus, in the example above $r = 1$ ($w_1 = 2\,\pi/12 = \pi/6$) corresponds to an annual cycle and $r = 2$ ($w_2 = \pi/3$) induces a six-monthly cycle. Combining such cycles can be done through the principle of model superposition. From the Fourier representation theorem, any set of p seasonal effects can be represented through the full set of $p/2$ harmonics ($r = 1, \ldots, p/2$) which (when we use superposition) leads to a state vector of $p - 1$ free elements, just like in the form-free case.[1]

The forecast function for the r-th seasonal harmonic is, for $r = 1, \ldots, p/2 - 1$

$$f_t(k) = a_t \cos(w_r k) + b_t \sin(w_r k) \qquad \text{where} \qquad \mathrm{E}[\boldsymbol{\theta}_t \mid D_t] = \begin{pmatrix} a_t \\ b_t \end{pmatrix}$$

and for $r = p/2$

$$f_t(k) = (-1)^k\, \mathrm{E}[\theta_t \mid D_t] \qquad \text{where now } \theta_t \text{ is a scalar.}$$

Example 1 (Gas sales).

To exemplify the concepts introduced above, let us briefly analyse the GAS data set in **?**, pp. 257–263. The 65 observations are monthly totals of inland UK natural gas consumption over the period of May-1979 to September-1984, both inclusive. The first salient feature

[1] Note that for $w_p = \pi$ we simply have $F_t = 1$, $G_t = -1$ and a scalar state.

of the plot is the seasonal pattern that it follows, with a peak in winter months (Dec to Feb) and falling to a minimum in the summer. The underlying trend appears to be more or less constant. Since gas consumption follows the natural annual temperature cycle, the first harmonic is expected to dominate the seasonal pattern. However, as we can see in Figure 3.2, the series is not perfectly sinusoidal and higher frequency harmonics are necessary to account for e.g. industrial demand patterns and holiday effects.

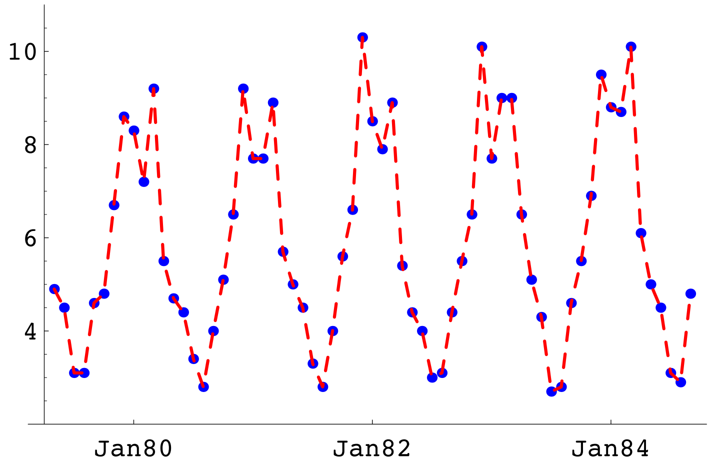


Figure 3.2. *Time plot of Gas consumption data*

As the level appears to be approximately constant, a first order polynomial DLM is used to model it, with a seasonal module with (full) 6 harmonics representing the 11 seasonal effects. After fitting the model (with a given prior), the one-step ahead forecast values are plotted in Figure 3.3 along with their corresponding 90% HPD regions. These forecasts appear to be good, given that, as expected, the regions are roughly covering 90% of the points.

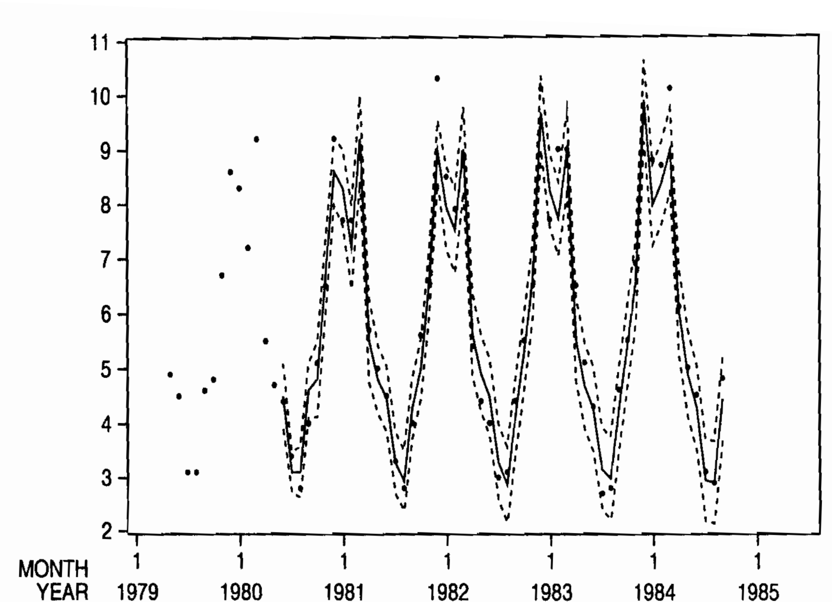


Figure 3.3. *One-step ahead forecast and 90% HPD's for gas consumption data.*

Now, Figure 3.4 depicts the overall estimated seasonal pattern of the series. There we can see how the model captures the more important features of the seasonal behaviour of the series. The vertical bars represent the 90% HPD intervals for the effects in each month.

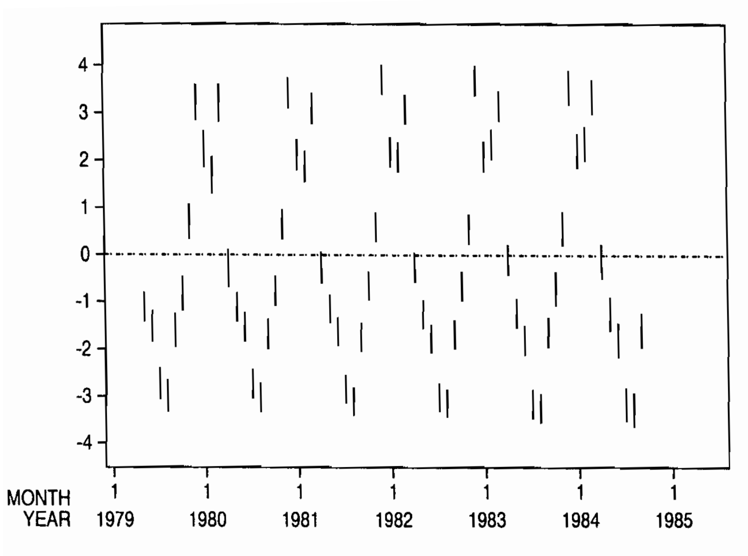


Figure 3.4. *Estimated seasonal pattern in gas consumption data.*

Finally, Figure 3.5 show the individual harmonic components 90% HPD's. There we can see how the first and fourth harmonic are the main contributors to the overall pattern.

3.1.3 Regression components

Covariates enter a DLM in the same way as into a regression model. Assume that $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q\}$ are observables related to series $\boldsymbol{y}$ and that are available for all $t = 0, 1, \ldots, T$; i.e. $\boldsymbol{x}_j' = \{x_{j1}, \ldots, x_{jT}\}$. Thus the corresponding DLM reads

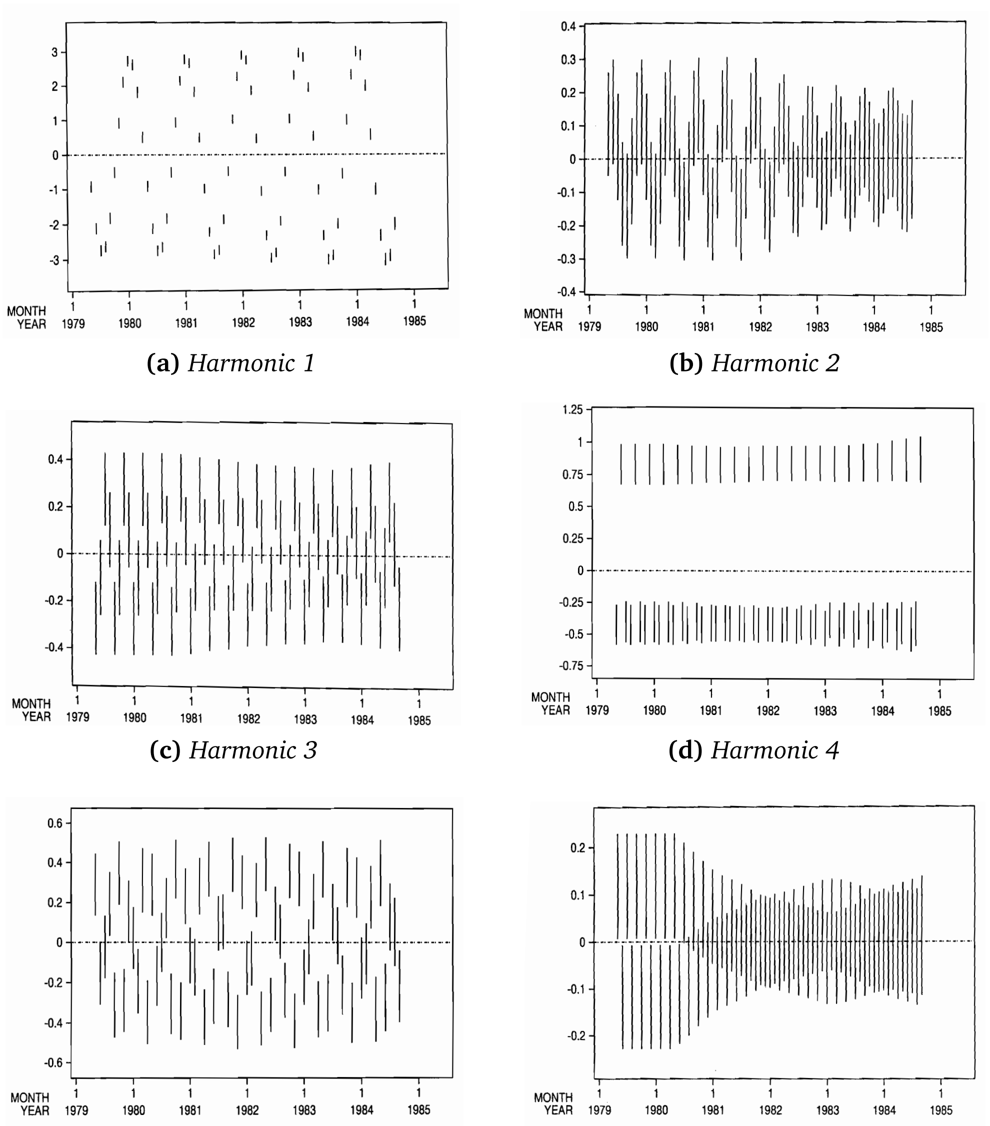
$$
\begin{aligned}
y_t &= \beta_{1t}x_{1t} + \cdots + \beta_{qt}x_{qt} + \nu_t\,, \\
\beta_{it} &= \beta_{it-1} + \omega_{it}\,, \quad i = 1, \ldots, q\,.
\end{aligned}
$$

It is readily seen that $F_t = X_t = \{\boldsymbol{x}_{1t}, \ldots \boldsymbol{x}_{qt}\}'$ and $G = I_q$. Also, the forecast function is simply

$$
f_t(k) = \sum_{i=1}^{q} x_{i\,t+k}\,\mathrm{E}[\beta_{i\,t} \mid D_t]\,.
$$

3.1.4 Discount factors

We have so far seen how to construct a DLM out of some basic components (where the form of F_t and G_t is determined by the properties that we want the DLM to have; i.e. the forecasting



(a) *Harmonic 1* **(b)** *Harmonic 2*

(c) *Harmonic 3* **(d)** *Harmonic 4*

(e) *Harmonic 5* **(f)** *Harmonic 6*

Figure 3.5. *Estimated 90% HPD's for the individual harmonic effects for the gas consumption data.*

behaviour we want to induce), and also, how to estimate a constant observational variance V, when it is unknown. To complete the model specification (barring the initial information D_0) we still have to specify the sequence of state evolution variance matrices, W_t. The values of W_t control the extent of the stochastic variation in the evolution of the model and hence determine the stability over time. In the system equation the addition of W_t leads to an increase in uncertainty or, equivalently, a loss of information about the state vector $\boldsymbol{\theta}$ from time $t-1$ to time t.

The often difficult task of proposing W_t can be facilitated if we note that

$$\mathrm{Var}[\boldsymbol{\theta}_t \mid D_{t-1}] = G_t C_{t-1} G_t' + W_t = P_t + W_t \,,$$

where P_t can be thought of as the appropriate variance for the DLM $\{F, G, V, 0\}$, i.e. a DLM where we know for sure that $\boldsymbol{\theta}_t = G\,\boldsymbol{\theta}_{t-1}$. As usually we do not know that this relation holds deterministically, we add to the ideal variance, P_t, a measure of our uncertainty about this relation to end up with the (more) realistic $C_t^* = P_t + W_t$. Thus, we can discount the actual precision, C_t^{*-1} by a factor $0 < \delta \le 1$, so that

$$\mathrm{Var}[\boldsymbol{\theta}_t \mid D_{t-1}] = C_t^* = \frac{1}{\delta}\, P_t\,.$$

This immediately lets us identify W_t, since $C_t^* = P_t + W_t$, and thus

$$W_t = \frac{1-\delta}{\delta}\, P_t\,.$$

Note that this is a slightly more general version of the discussion on discount factors for the steady model in the previous chapter. Also, this immediately generalises to superpositions, where the discounting can be done componentwise. In practice, different components (e.g. trend and seasonal) require different discount factors.

Consider the superposition model of Theorem 3.1 and define $P_{i\,t} = \mathrm{Var}[G_{i\,t}\boldsymbol{\theta}_{i\,t-1} \mid D_{t-1}]$, i.e. the uncertainty about $G_{i\,t}\boldsymbol{\theta}_{i\,t-1}$ (for the i-th component, $i = 1, \ldots, h$) before the addition of evolution noise. Then it is natural to choose

$$W_{i\,t} = \frac{1-\delta_i}{\delta_i}\, P_{i\,t}\,, \qquad i = 1, \ldots, h$$

where $0 < \delta_i \le 1$ is the discount factor for model component M_i. The resulting superposition model is called a **component discount** DLM.

3.2 Equivalent models

Throughout, we assume that the DLM is not overparameterised, i.e. that the observations provide information about all elements in the state vector. This is usually called **observability** of the DLM.

However, even among observable DLM's, there are still many ways to reparameterise the models. In particular, we will focus on TSDLM's in this section.

Two DLM's are **similar** if the system matrices have identical eigenvalues. This will lead to forecast functions with the same qualitative (algebraic) form (e.g. both linear in k). A stronger requirement is that two models produce *exactly* the same forecast functions. Such models are called **equivalent** and can be related through a reparametrisation of the state vector. In particular, consider the DLM $\{F, G, V_t, W_t\}$ in the state $\boldsymbol{\theta}_t$; i.e.

$$\begin{aligned} Y_t &= F\,\boldsymbol{\theta}_t + \nu_t \qquad & \nu_t &\sim \mathrm{N}\left(\nu_t \mid 0, V_t\right) \\ \boldsymbol{\theta}_t &= G\,\boldsymbol{\theta}_{t-1} + \omega_t \qquad & \omega_t &\sim \mathrm{N}\left(\omega_t \mid 0, W_t\right)\ . \end{aligned}$$

Given any nonsingular matrix H, this DLM can be reparameterised to $\{F_1, G_1, V_{1t}, W_{1t}\}$ in states $\boldsymbol{\phi}_t = H\,\boldsymbol{\theta}_t$ as

$$\begin{aligned} Y_t &= FH^{-1}\boldsymbol{\phi}_t + \nu_t = F_1\boldsymbol{\phi}_t + \nu_{1t} \\ \boldsymbol{\phi}_t &= H\,\boldsymbol{\theta}_t = H\,G\,H^{-1}\boldsymbol{\phi}_{t-1} + H\,\omega_t \\ &= G_1\boldsymbol{\phi}_{t-1} + H\,\omega_t\ ; \end{aligned}$$

so that

$$F_1 = FH^{-1}\,, \quad G_1 = H\,G\,H^{-1}\,, \quad V_{1t} = V_t\,, \quad W_{1t} = H\,W_t H'\ .$$

The resulting forecast functions are indeed exactly the same: for the model in terms of $\boldsymbol{\theta}$:

$$f_t^{\theta}(k) = FG^k\boldsymbol{m}_t$$

and for the one in $\boldsymbol{\phi}$

$$\begin{aligned} f_t^{\phi}(k) &= F_1 G_1^k\,\mathrm{E}[\boldsymbol{\phi}_t \mid D_t] = FH^{-1}HG^kH^{-1}H\boldsymbol{m}_t \\ &= FG^k\boldsymbol{m}_t \end{aligned}$$

since

$$\begin{aligned} \left[H\,G\,H^{-1}\right]^k &= HGH^{-1}HGH^{-1}\cdots HGH^{-1} \\ &= HG^kH^{-1}\ . \end{aligned}$$

Of course, for exact equivalence we also need the priors on $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ to be compatible.

Thus, if two DLM's, M and M' are equivalent we can obtain the same updating and forecast distributions as those derived with model M, with the (perhaps simpler) parameterisation used in model M'.

Theorem 3.2 (Equivalent DLM).
Let $\boldsymbol{y}_t \sim \mathrm{DLM}\{F, G, V_t, W_t\}$, *with parameter* $\boldsymbol{\theta}_t$ *such that* $\boldsymbol{\theta}_t \mid D_t \sim [\boldsymbol{m}_t, C_t]$. *Define the new parameterisation as* $\boldsymbol{\phi}_t = H\,\boldsymbol{\theta}_t$, *where* $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ *is a full rank matrix. Then*

$$\boldsymbol{y}_t \sim \mathrm{DLM}\left\{FH^{-1}, H\,G\,H^{-1}, V_t, H\,W_t H'\right\} \quad \textit{with parameter} \quad \boldsymbol{\phi}_t \mid D_t \sim [H\,\boldsymbol{m}_t, H\,C_t H'].$$

Remark 3.1. *Clearly, equivalent models are also similar. The eigenvalues of G and $H\,G\,H^{-1}$ are the same since*

$$\begin{aligned}\left|\lambda\,I_n - H\,G\,H^{-1}\right| &= \left|\lambda\,H\,H^{-1} - H\,G\,H^{-1}\right| = |H|\,\left|\lambda\,I_n - G\right|\,\left|H^{-1}\right| \\ &= \left|\lambda\,I_n - G\right|\ .\end{aligned}$$

3.3 Canonical models

Among equivalent models it is important to choose the simplest version. These provide **canonical** DLM's for a given required forecast function form.

In general, the system matrix, G, has s real and distinct eigenvalues $\lambda_1, \ldots, \lambda_s$ of multiplicities $r_1, \ldots, r_s$, respectively; and v pairs of complex conjugate eigenvalues $\lambda_{s+k} e^{i\,w_k}$ and $\lambda_{s+k} e^{-i\,w_k}$, $k = 1, \ldots, v$, for real distinct $\lambda_{s+1}, \ldots, \lambda_{s+v}$ and $w_1, \ldots, w_v$ (in principle, these can also be repeated, but we abstract from that, in practice rather unlikely, possibility). Thus, the dimension of the state vector is

$$n = \sum_{k=1}^{s} r_k + 2\,v\ .$$

For any invertible square matrix G we can find a non-singular matrix H such that an equivalent model can be found with a simple block-diagonal system matrix J (a *canonical equivalent* model) since

$$G = H\,J\,H^{-1}$$

with

$$J = \text{block diag}\left[J_{r_1}(\lambda_1)\cdots J_{r_s}(\lambda_s)\,J_2(\lambda_{s+1}, w_1)\cdots J_2(\lambda_{s+v}, w_v)\right]$$

and we have defined the $r \times r$ Jordan block

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \ldots\ldots & 0 \\ 0 & \lambda & 1 & 0 & \ldots\ldots \\ 0 & 0 & \lambda & 1 \quad 0 & \ldots \\ \vdots & \ddots & \ddots & \ldots\ldots & \vdots \\ 0 & \ldots\ldots & \ldots\ldots & \ldots\ldots & 1 \\ 0 & \ldots\ldots & \ldots\ldots & \ldots\ldots & \lambda \end{pmatrix}$$

and

$$J_2(\lambda, w) = \lambda \begin{pmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{pmatrix}\ .$$

Since $G^k = HJ^kH^{-1}$, the forecast function is particularly simple and becomes

$$f_t(k) = FG^k \boldsymbol{m}_t = \sum_i f_t^{(i)}(k)$$

where $f_t^{(i)}(k)$ is the forecast function corresponding to the i-th component of the whole DLM. More precisely, it corresponds to the canonical component DLM $\{E_r, J_r(\lambda), \cdot, \cdots\}$, or DLM $\{E_2, J_2(\lambda, w), \cdot, \cdot\}$, where E_r, $r \geq 1$ is the r dimensional vector with unit in its first element and zeros elsewhere.

Familiarity with this canonical form and its forecast function is all that is needed to comprehend the forecast function of any complex DLM

Thus, we are really considering the entire DLM as a superposition of simple, canonical models. We now briefly examine the forecast functions of these canonical blocks, depending on their eigenvalue structure.

Real non-zero eigenvalue λ of multiplicity r

$$f_t(k) = \lambda^k \left[\sum_{l=0}^{r-1} a_{l,t} k^l \right]$$

where $a_{l,t}$ are linear functions of $\boldsymbol{m}_t = \mathrm{E}[\boldsymbol{\theta}_t \mid D_t]$. In the case $r = 1$, we get $f_t(k) = \lambda^k \boldsymbol{m}_t$, which corresponds to:

(a) A first order polynomial DLM, when $\lambda = 1$ (i.e. a constant forecast function).

(b) A dampened forecast if $0 < \lambda < 1$.

(c) An explosive forecast if $\lambda > 1$.

(d) An oscillating forecast if $\lambda < 0$.

Complex eigenvalues. For a real time series, complex eigenvalues must occur in pairs, typically of multiplicity $m = 1$. Write the eigenvalues as $\lambda \{\exp[iw], \exp[-iw]\}$, where $\lambda > 0$ is the modulus and the period is $p = 2\pi/w$. Then the forecast function is

$$\begin{aligned} f_k(t) &= \lambda^k [a_t \cos(kw) + b_t \sin(kw)] \\ &= \lambda^k r_t \cos(kw + \varphi_t) \end{aligned}$$

where $r_t = \sqrt{a_t^2 + b_t^2}$ is the amplitude and $\varphi_t = \arctan(-\frac{b_t}{a_t})$ is the phase.

The forecast function

(a) if $\lambda = 1$, is a constant cycle of period p;

(b) if $0 < \lambda < 1$, is a dampened cycle;

(c) is an explosive cycle if $\lambda > 1$.

3.3.1 Summary

A real non-zero eigenvalue λ of multiplicity r

$$\{E_r, J_r(\lambda)\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & \lambda \end{pmatrix} \right\}$$

with forecast function

$$f_t(k) = \lambda^k \sum_{l=0}^{r-1} a_{l,t} k^l$$

A complex pair of eigenvalues $\lambda(e^{iw}, e^{-iw})$ of multiplicity 1

$$\{E_2, J_2(\lambda, w)\} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{pmatrix} \right\}$$

with forecast function

$$f_t(k) = \lambda^k \left(a_t \cos(kw) + b_t \sin(kw) \right)$$

If $w = \pi$ the component corresponds to a real negative eigenvalue, $-\lambda$, and is $\{E_1, J_1(-\lambda)\}$=$\{1, -\lambda\}$.

The forecast function of the entire DLM corresponding to $\{F, G, V_t, W_t\}$ will simply be the sum of the relevant canonical component forecast functions.

Example 2.
In order to gain some further insight into the properties of these canonical models, assume G is a 2×2 nonsingular matrix. We then have three different combinations of eigenvalues and many different shapes for the forecasting function:

(i) Two distinct real eigenvalues, λ_1, λ_2; i.e.

$$J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \qquad \text{with} \qquad \lambda_1 > \lambda_2$$

implying that

$$f_t(k) = \lambda_1^k a_t + \lambda_2^k b_t$$

One important special case, illustrated in Figure 3.6, is when $\lambda_1 = 1$ and $0 < \lambda_2 < 1$.

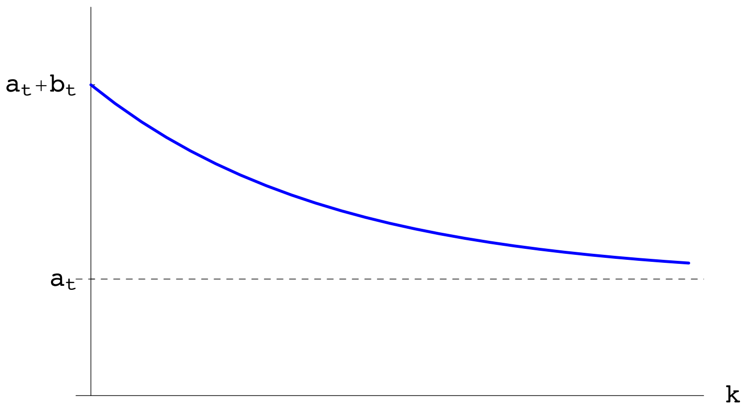


Figure 3.6. *Shape of the forecasting function with two distinct real eigenvalues.*

This may be useful when the series is reasonably steady but increases for some reason (e.g. advertisement might boost sales for a period, and then its effect fades away). Notice that the effect dampens exponentially.

(ii) One real eigenvalue, λ, of multiplicity two. This yields

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

then

$$f_t(k) = \lambda^k \left(a_t + k\, b_t\right)$$

One important special case is when $\lambda = 1$, which leads to a growth model.

Other important special case, illustrated in Figure 3.7, is when $0 < \lambda < 1$ and $a_t = 0$. In this case the forecast is quickly dampened, and is useful in modelling, for instance, sales of a product which quickly loses appeal, or emission of gas in an accident, etc. In practice, λ is fitted to the peak or the half life of the product.

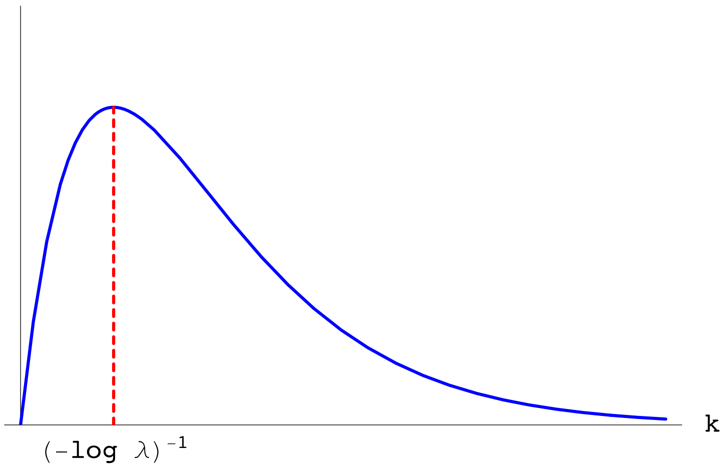


Figure 3.7. *Shape of the forecasting function with one real eigenvalue of multiplicity two.*

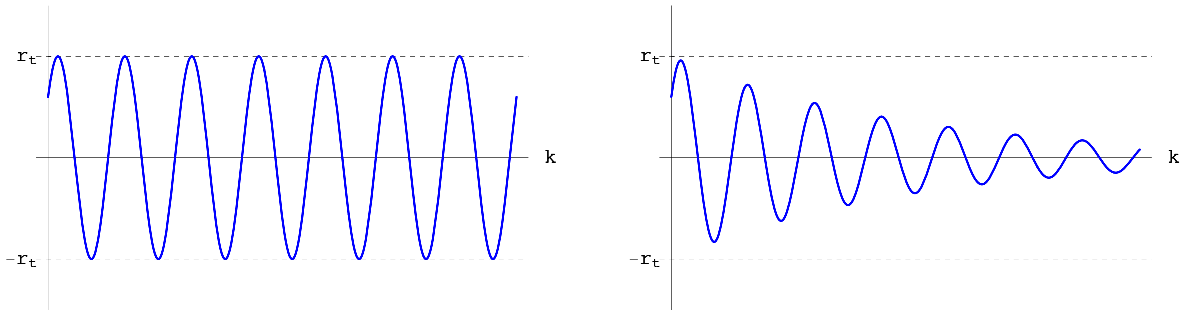
(iii) Complex conjugate eigenvalues, λe^{iw} and λe^{-iw}. Then,

$$J = \lambda \begin{pmatrix} \cos w & \sin w \\ -\sin w & \cos w \end{pmatrix}$$

which yields

$$\begin{aligned} f_t(k) &= \lambda^k \left[a_t \cos(k w) + b_t \sin(k w) \right] \\ &= \lambda^k r_t \cos\left(k w + \varphi_t\right) \end{aligned}$$

Two important special cases are $\lambda = 1$ which gives a harmonic seasonal model; and $\lambda < 1$, which gives a seasonal model with diminishing seasonal effects.



(a) $\lambda = 1$. **(b)** $\lambda < 1$.

Figure 3.8. *Shape of the forecasting function with two complex conjugate eigenvalues*

◁