Introduction to Dynamic Linear Models

- Dynamic Linear Models (DLMs) or state space models define a very general class of non-stationary time series models.
- DLMs may include terms to model trends, seasonality, covariates and autoregressive components.
- Other time series models like ARMA models are particular DLMs.
- The main goals are short-term forecasting, intervention analysis and monitoring.
- We will not focus as much on concepts like autocorrelation or stationary process.

- A Normal DLM is defined with a pair of equations

$$\begin{aligned} Y_t &= F_t^{'}\theta_t + \epsilon_t; \quad \text{(observation equation)} \\ \theta_t &= G_t\theta_{t-1} + \omega_t; \quad \text{(evolution equation)} \end{aligned}$$

  $t = 1, 2, \ldots, T$

  - Y_t is the observation at time t. We assume this is to be a scalar but could also be a vector.
  - $\theta_t = (\theta_{t,1}, \ldots, \theta_{p,1})^{'}$ is the vector of parameters at time t and of dimension $p \times 1$.
  - $F_t^{'}$ is the row vector (dimension $1 \times p$) of covariates at time t
  - G_t is a matrix of dimension $p \times p$ known as *evolution* or *transition* matrix.
  - Usually F_t and G_t are completely specified and

$F_t = F$, $G_t = G$.

- ϵ_t is the observation error at time t and ω_t is the evolution error ($p \times 1$ vector).
- For a Normal DLM, $\epsilon_t \sim N(0, V_t)$ and $\omega_t \sim N(0, W_t)$.
- ϵ_t is independent of ϵ_s, ω_t is independent of ω_s for $t \neq s$. $\epsilon' s$ independent of $\omega' s$.

- Mostly we will discuss the Bayesian analysis of these models the counterpart being the *Kalman Filter*.
- The main reference on Bayesian DLMs, West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models, 2nd ed.* Springer Verlag, New York.
- In general Dynamic Models are given by two pdfs:

$$f(Y_t|\theta_t) \quad \text{and} \quad g(\theta_t|\theta_{t-1})$$

which define a conditional dependence structure between observations (Y_t) and parameters (θ_t)

- DLMs have a sequential nature and from a Bayesian approach, one of its main targets, is

$$p(\theta_t|D_t) \quad , t = 1, 2, \ldots, T$$

the posterior distribution of θ_t given all the information available at time t, i.e., $D_t = \{Y_1, Y_2, \ldots, Y_t\}$

- The simplest "dynamic model" is well known:

$$Y_t = \mu + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

- In DLM notation $\theta_t = \mu$, $F_t' = 1$, $\omega_t = 0$ and $G_t = 1$.

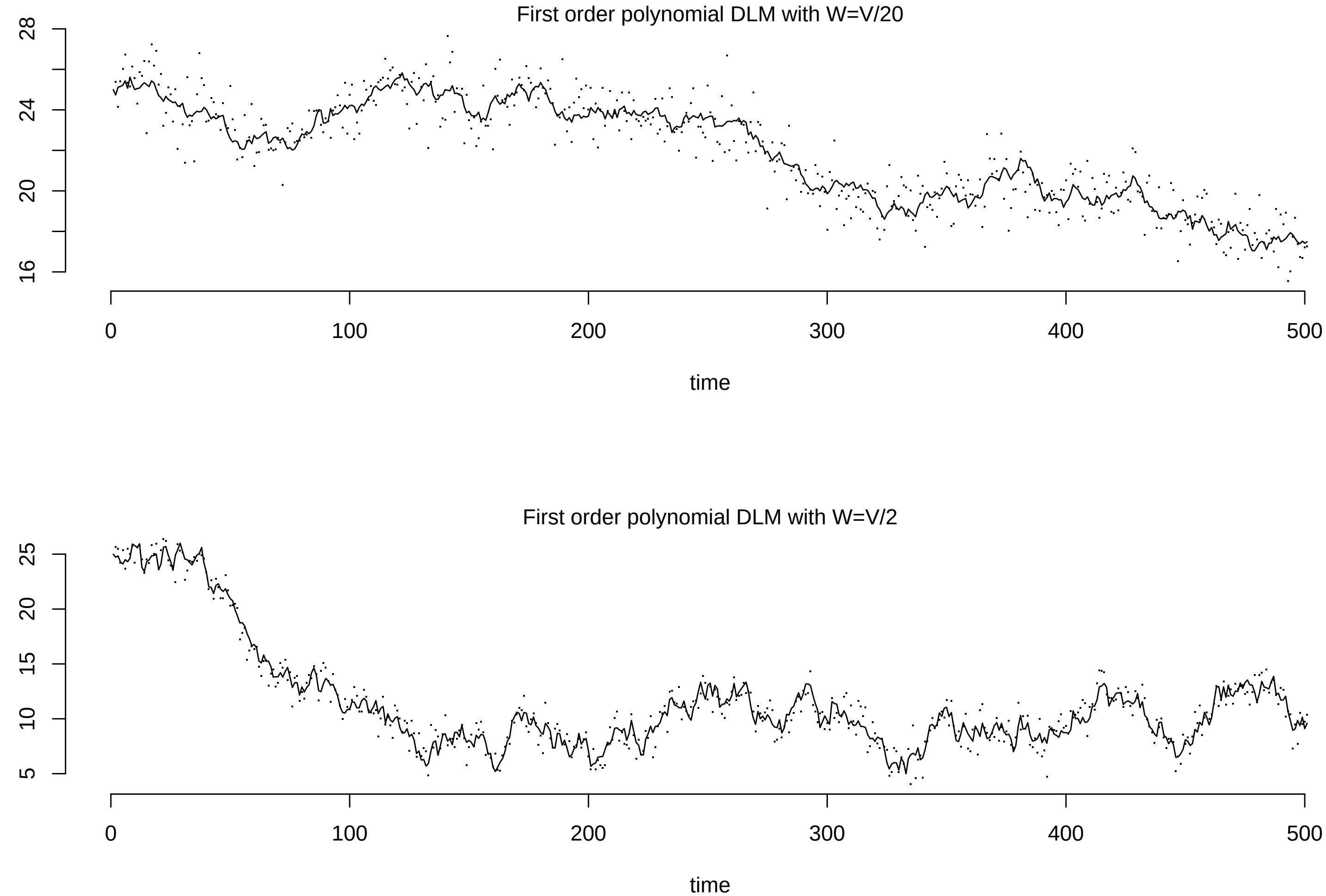
- A generalization is the *First order polynomial* DLM

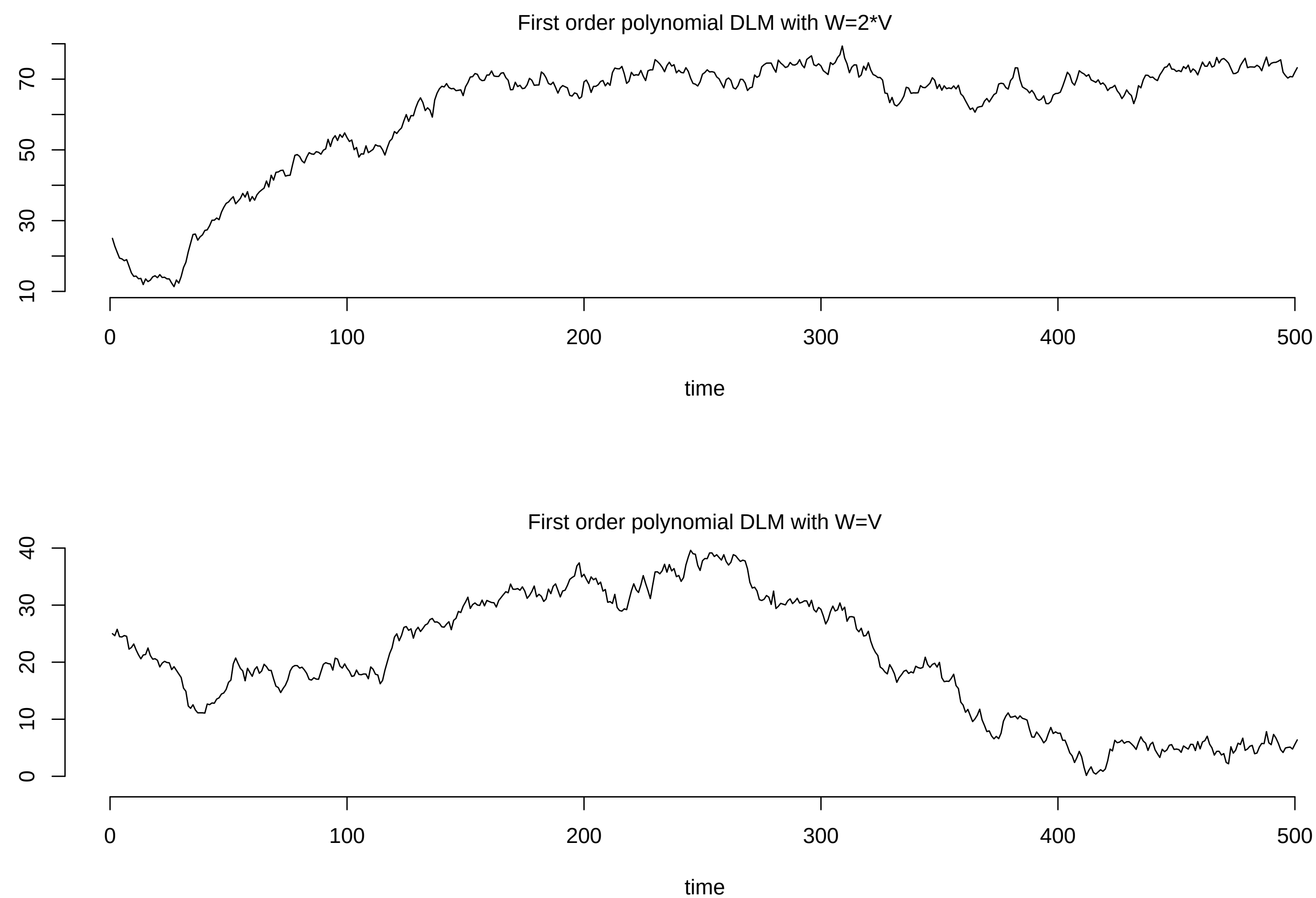
$$\begin{aligned} Y_t &= \mu_t + \epsilon_t; \quad \epsilon_t \sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \omega_t; \quad \omega_t \sim N(0, W_t) \end{aligned}$$

- The evolution equation allows smooth changes of the mean level. W_t is a scalar greater than zero.
- West and Harrison use ν_t instead of ϵ_t or $\nu_t \sim N(0, V_t)$.
- Equivalently the first order polynomial DLM is written as

$$(Y_t|\mu_t) \sim N(\mu_t, V_t); (\mu_t|\mu_{t-1}) \sim N(\mu_{t-1}, W_t)$$

- Lets consider the constant case with $V_t = V$, $W_t = W$, $V = 1$, $\mu_0 = 25$. Plots show simulations for $W = 2V$, $W = V$, $W = V/20$ and $W = V/2$.


First order polynomial DLM with W=V/20
16
20
24
28
0
100
200
300
400
500
time
First order polynomial DLM with W=V/2
5
10
15
20
25
0
100
200
300
400
500
time



First order polynomial DLM with W=2*V
10
30
50
70
0
100
200
300
400
500
time
First order polynomial DLM with W=V
0
10
20
30
40
0
100
200
300
400
500
time

```
# Function to simulate 1st order polynomial DLM
genfoDLM=function(V,n,del)
{
yt=rep(NA,n+1)
W=V/del
mut=rep(NA,(n+1))
mut[1]=25
for(i in 2:(n+1))
{
mut[i]=mut[i-1]+rnorm(1,0,sqrt(W))
yt[i]=mut[i]+rnorm(1,0,sqrt(V))
}
return(mut,yt)
}
```

- Notice that for the first order polynomial DLM

$$\begin{aligned} E(Y_{t+k}|\mu_t) &= E(\mu_{t+k}|\mu_t) = \mu_t \\ E(Y_{t+k}|D_t) &= E(\mu_t|D_t) \equiv m_t \end{aligned}$$

which is useful for short term forecasting.

Inference for the First order Polynomial DLM

- Suppose the sequences V_t and W_t are known for all time t.
- At time 0 the prior for μ_0 is $N(m_0, C_0)$ and denoted by $(\mu_0|D_0) \sim N(m_0, C_0)$.
- We want to find $(\mu_t|D_t)$, the posterior for μ_t given D_t and we will proceed sequentially.
  - We start from the posterior at time $t-1$, $(\mu_{t-1}|D_{t-1})\ N(m_{t-1}, C_{t-1})$

- From this posterior we can get the prior at time t, $(\mu_t|D_{t-1})\ N(m_{t-1}, C_{t-1})$ where $R_t = C_{t-1} + W_t$.
- We can obtain the predictive at time $t-1$, $(Y_t|D_{t-1}) \sim N(f_t, Q_t)$ where $f_t = m_{t-1}$ and $Q_t = R_t + V_t$.
- Using Bayes' theorem we can get the posterior at time t, $(\mu_t|D_t) \sim N(m_t, C_t)$ and the recursive equations:

$$\begin{aligned} m_t &= m_{t-1} + A_t e_t; \\ C_t &= A_t V_t \\ A_t &= R_t / Q_t \\ e_t &= Y_t - f_t \end{aligned}$$

- *Proof* (by induction)

- We start from the posterior at time $t-1$; $(\mu_{t-1}|D_{t-1}) \sim N(m_{t-1}, C_{t-1})$.
- Using the evolution equation and Normal linear theory, $\mu_t = \mu_{t-1} + \omega_t$, we get $(\mu_t|D_{t-1}) \sim N(m_{t-1}, C_{t-1} + W_t)$ (initial at time t). $R_t \equiv C_{t-1} + W_t$
- From the observation equation $Y_t = \mu_t + \epsilon_t$, $E(Y_t|\mu_t, D_{t-1}) = \mu_t$ and $Var(Y_t|\mu_t, D_{t-1}) = R_t + V_t \equiv Q_t$, then $(Y_t|D_{t-1}) \sim N(m_{t-1}, Q_t)$
- By Bayes' Theorem,

$$p(\mu_t|D_{t-1}) \propto f(y_t|\mu_{t-1})p(\mu_t|D_{t-1})$$

- Then

$$p(\mu_t|D_{t-1}) \propto \exp\left\{-\frac{1}{2V_t}(Y_t - \mu_t)^2\right\} \exp\left\{-\frac{1}{2Q_t}(\mu_t - m_{t-1})^2\right\}$$

- After some algebra and completing a square for μ_t

$$p(\mu_t|D_{t-1}) \propto \exp\left\{-\frac{Q_t}{2R_tV_t}\left(\mu_t - \left(\frac{R_ty_t + V_tm_{t-1}}{Q_t}\right)\right)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2A_tV_t}\left(\mu_t - \left(\frac{R_t(y_t - m_{t-1}) + (V_t + R_t)m_{t-1}}{Q_t}\right)\right)^2\right\}$$

- $e_t = Y_t - f_t$ is known as the one-step predictive forecast.
- Since $A_t = R_t/(R_t + V_t)$ (adaptive coefficient), then $0 \leq A_t \leq 1$.
- Given that $m_t = A_t(e_t) + m_{t-1} = A_tY_t + (1 - A_t)m_{t-1}$ as $At \rightarrow 1$, $m_t \approx Y_t$. As $A_t \rightarrow 0$, $m_t \approx m_{t-1}$.

Forecasting k-steps ahead

- Determine the distribution $(Y_{t+k}|D_t)$. From the observation equation we have that,

$$Y_{t+k} = \mu_{t+k} + \epsilon_{t+k}$$

- From the evolution equation,

$$\begin{aligned}
\mu_{t+k} &= \mu_{t+k-1} + \omega_{t+k} \\
&= \mu_{t+k-2} + \omega_{t+k-1} + \omega_{t+k} \\
&\vdots \\
&= \mu_t + \sum_{j=1}^{k} \omega_{t+j}
\end{aligned}$$

- Then,

$$Y_{t+k} = \mu_t + \sum_{j=1}^{k} \omega_{t+j} + \epsilon_{t+k}$$

- Since the posterior for μ_t at time t is $(\mu_t|D_t) \sim N(m_t, C_t)$ then $(Y_{t+k}|D_t) \sim N(m_t, Q_t(k))$ where

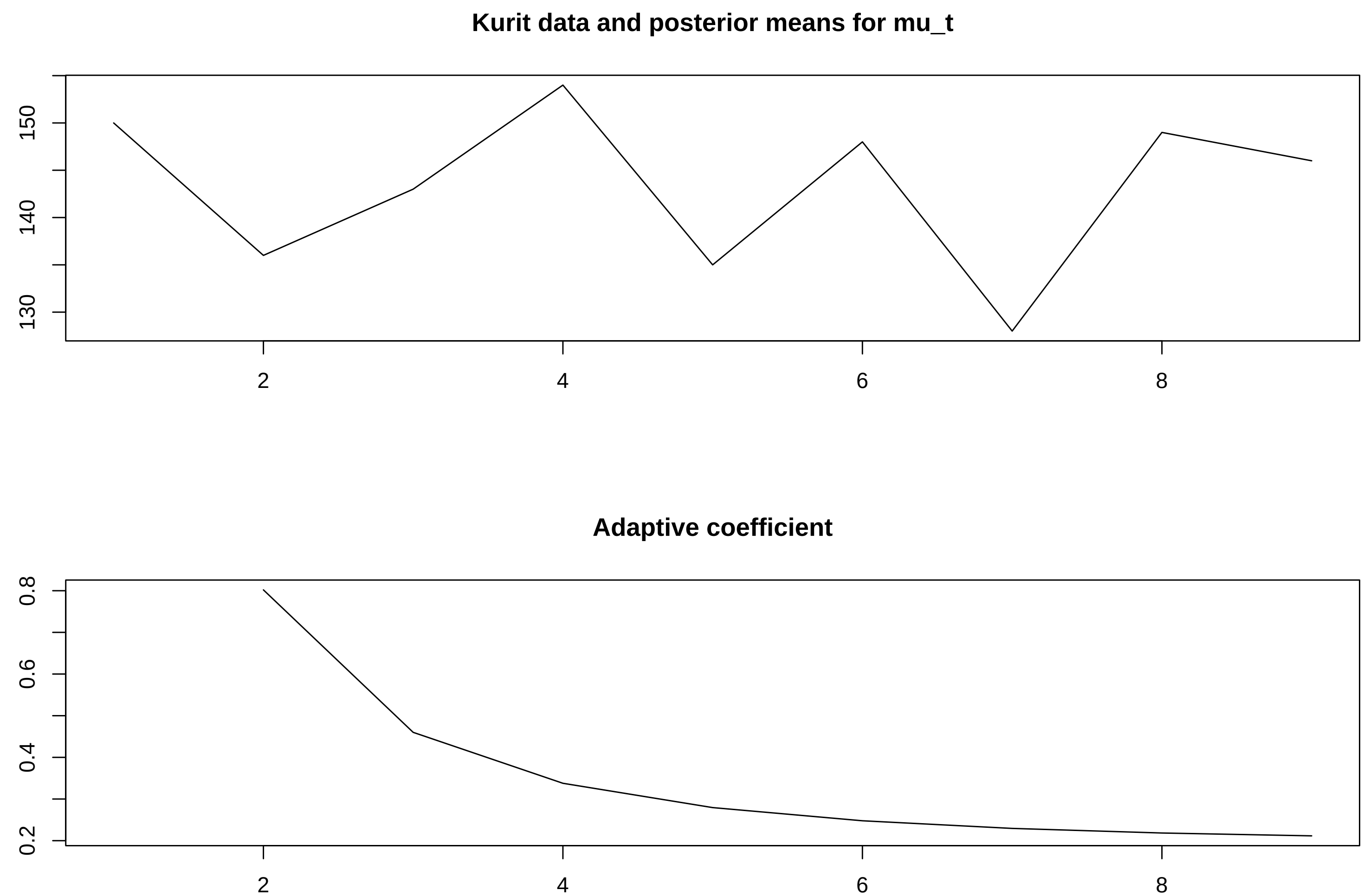
$$Q_t(k) = C_t + \sum_{j=1}^{k} W_{t+j} + V_{t+k}$$

- m_t is the predictive mean and $Q_t(k)$ is the predictive variance.

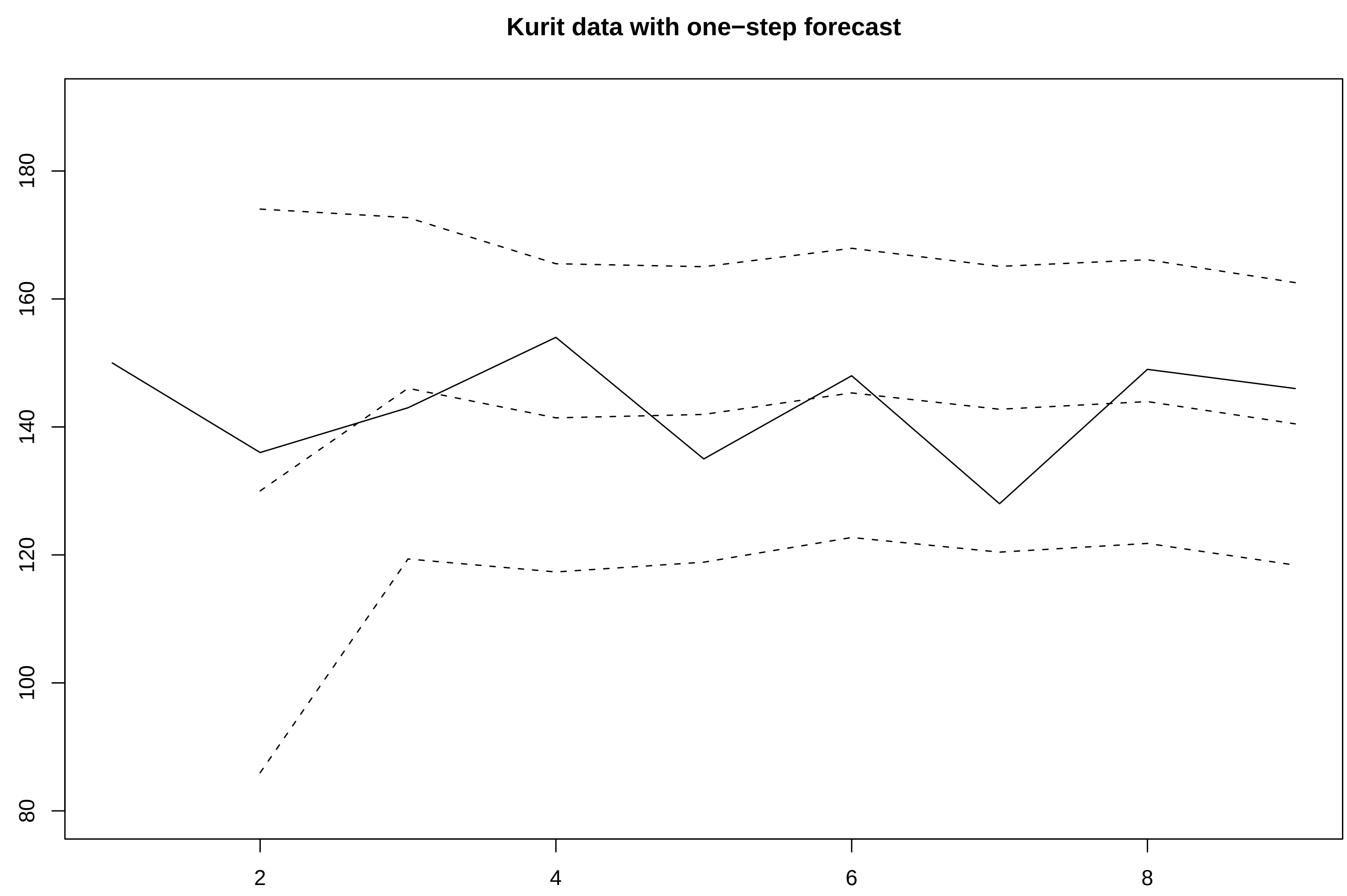
- *Example* (W&H pag. 40) Data is monthly sales of a pharmaceutical company of a product "Kurit". Approximately 100 units are sold every month.

- It is expected that a new advertisement campaign leads to increases in demand of this product.
- At $t = 0$, a 30% increase in sales is expected. A range of 80 units is equivalent to $4\sqrt{C_0}$.
- Then $C_0 = 400$ and the prior at time 0 is $(\mu_0|D_0) \sim N(130, 400)$.
- The proposed model is:

$$
\begin{aligned}
Y_t &= \mu_t + \epsilon_t; \quad \epsilon_t \sim N(0, 100) \\
\mu_t &= \mu_{t-1} + \omega_t; \quad \omega_t \sim N(0, 5) \\
(\mu_0|D_0) &\sim N(130, 400)
\end{aligned}
$$



Kurit data and posterior means for mu_t
130
140
150
2
4
6
8
Adaptive coefficient
0.2
0.4
0.6
0.8
2
4
6
8



Kurit data with one−step forecast
180
160
140
120
100
80
2
4
6
8

Kurit example

t	Q_t	f_t	A_t	Y_t	e_t	m_t	C_t
0						130.0	400
1	505	130.0	0.80	150	20.0	146.0	80
2	185	146.0	0.46	136	-10.0	141.4	46
4	139	141.9	0.28	154	12.1	145.3	28
6	130	142.6	0.23	148	5.3	143.9	23
9	126	142.2	0.21	146	3.9	143.0	20
10	125	143.0	0.20				

- One of the main advantages of the Bayesian approach for DLMs is the ease to incoporate external information.
- For the Kurit example, at $t = 9$, the posterior and one-step forecast distributions are

$$\begin{aligned} (\mu_9|D_9) &\sim N(143, 20) \\ (Y_{10}|D_9) &\sim N(143, 125) \end{aligned}$$

- Suppose that a competitive product (BURNIT) is withdrawn from the market.
- At $t = 10$, patients that were prescribed BURNIT will switch to a competitor. This information is denoted as S_9.
- Our company estimates a 100 % increase in KURIT

demand which translates to $E(\mu_{10}|D_9, S_9) = 286$

- There is a large uncertainty about this figure which is expressed as $(\omega_{10}|D_9, S_9) = N(143, 900)$ leading to the following revised distributions,

$$\begin{aligned}
\mu_{10} &= \mu_9 + \omega_{10}; \quad (\mu_{10}|D_9, S_9) \sim N(286, 920) \\
Y_{10} &= \mu_{10} + \epsilon_{10}; \quad (Y_{10}|D_9, S_9) \sim N(286, 1020)
\end{aligned}$$

- Then, $A_{10} = 920/1020$ increases from 0.2 to 0.9 providing a faster adaptation to the inmediately forthcoming data.

- If $Y_{10} = 326$, then $e_{10} = 326 - 286 = 40$ and $(\mu_{10}|D_{10}) \sim N(322, 90)$ where $m_t = 286 + 0.9(40) = 322$ and $C_t = 0.9(100) = 90$.

- For a first-order polynomial DLM with constant variances $V_t = V$ and $W_t = W$, as $t \to \infty$, $A_t \to A$ and $C_t \to C = AV$ where

$$A = \frac{r}{2}\left(\sqrt{1+\frac{4}{r}} - 1\right); r = \frac{W}{V}$$

- The one-step ahead forecast function $m_t = E(Y_{t+1}|D_t)$ takes the limit form:

$$m_t = (1-A)m_{t-1} + AY_t = m_{t-1} + Ae_t$$

- As part of the DLM recursive equations, we have that $R_t = C_{t-1} + W$ so in the limit $(t \to \infty)$,

$$R = C + W$$

- Also, $R_t = A_t(R_t + V)$ or $R_t(1 - A_t) = A_t V$ and in the limit this implies that $R(1-A) = AV$ or

$$R = \frac{AV}{1-A} = \frac{C}{1-A}$$

- Combining both equations for R, we obtain that

$$W = \frac{AC}{1-A}$$

and W is a fixed proportion of C.

- This is a natural way of thinking about the evolution variance, the addition of the error term ω_t leads to an increase uncertainty of $W = 100A/(1-A)\%$ of C.
- If $\delta = 1 - A$, it follows that $R = C/\delta$ and $W = (1-\delta)C/\delta$.
- Since for the first order polynomial DLM, the limiting

behavior is rapidly acheived, we can adopt a **discount factor** δ for all t by choosing

$$W_t = C_{t-1}(1-\delta)/\delta$$

- This DLM is not constant but quickly converges to a constant DLM with $V_t = V$ and $W_t = rV$ with $r = (1-\delta)^2/\delta$ since

$$\begin{aligned} C_t^{-1} &= V^{-1} + R_t^{-1} = V^{-1} + \delta C_{t-1}^{-1} \\ &= V^{-1}[1+\delta+\delta^2+\ldots+\delta^{t-1}] + \delta^t C_0^{-1} \end{aligned}$$

so the limiting case of C_t is $C = (1-\delta)V$.

Unknown observational variance

- In the case of a constant variance $V_t = V$ that is unknown and $W_t = VW_t^*$, where W_t^* is known , the Bayesian analysis leads to specific equations for the relevant posterior distributions.
- The DLM is given by

$$\begin{aligned} Y_t &= \mu_t + \epsilon_t; \quad \epsilon_t \sim N(0, V) \\ \mu_t &= \mu_{t-1} + \omega_t; \quad \omega_t \sim N(0, VW_t^*) \end{aligned}$$

- The prior is specified as
  - $(\mu_0 | D_0, V) \sim N(m_0, VC_0^*)$
  - For $\phi = 1/V$; $(\phi | D_0) \sim G(n_0/2, d_0/2)$.
- The values m_0, C_0^*, n_0 and d_0 are treated as known.

- For $t = 1, \ldots, T$ we have the recursive equations,

$$\begin{aligned} R_t^* &= C_{t-1}^* + W_t^*; \\ f_t &= m_{t-1}; \\ Q_t^* &= R_t^* + 1 \\ e_t &= Y_t - f_t \\ A_t &= R_t^*/Q_t^* \\ C_t^* &= R_t^* - A_t^2 Q_t^* \\ m_t &= m_{t-1} + A_t e_t \end{aligned}$$

- and the following distributions,

$$\begin{aligned} (\mu_{t-1}|D_{t-1}, V) &\sim N(m_{t-1}, VC_{t-1}^*); \text{(posterior at time } t-1) \\ (\mu_t|D_{t-1}, V) &\sim N(m_{t-1}, VR_t^*); \text{(prior at time } t) \\ (Y_t|D_{t-1}, V) &\sim N(f_t, VQ_t^*); \text{(one-step predictive)} \end{aligned}$$

$$(\mu_t|D_t, V) \sim N(m_t, VC_t^*); \text{(posterior at time } t)$$

- For the precision $\phi = V^{-1}$:
  - $(\phi|D_{t-1}) \sim Ga(n_{t-1}/2, d_{t-1}/2)$.
  - $(\phi|D_t) \sim Ga(n_t/2, d_t/2)$.

  where $n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + \epsilon_t^2/Q_t$.
- *Proof* At time $t-1$,

$$p(\phi|D_{t-1}) \propto \phi^{\frac{n_{t-1}}{2}-1} exp\left(-\frac{d_{t-1}}{2}\phi\right)$$

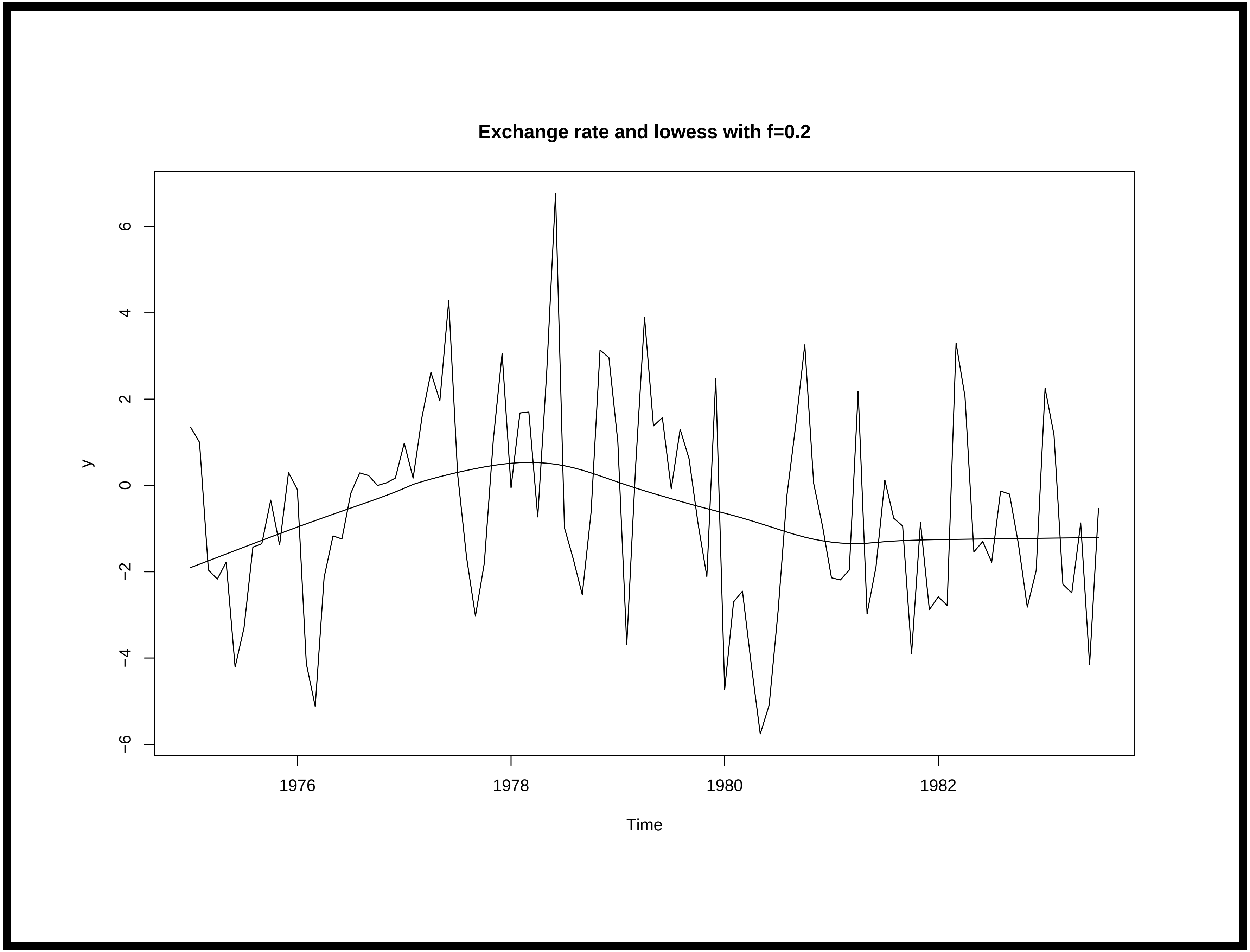
- Also,

$$f(Y_t|D_{t-1}, \phi) \propto \phi^{1/2} exp\left(-\phi(Y_t - m_{t-1})^2/2Q_t^*\right)$$

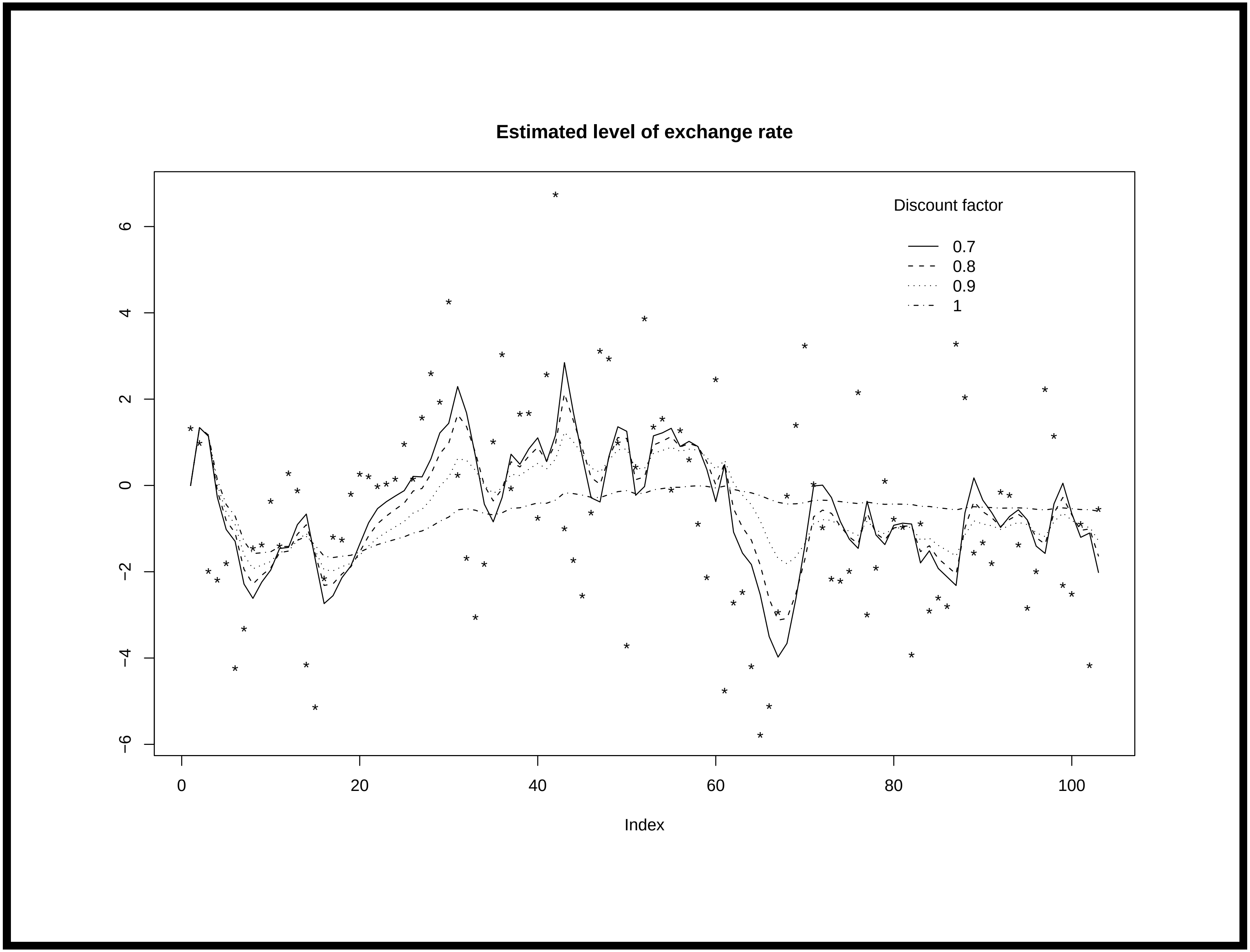
- By Bayes' theorem, $p(\phi|D_t) \propto f(Y_t|D_{t-1}, \phi)p(\phi|D_{t-1})$ which leads to the $Ga(n_t/2, d_t/2)$.

Exchange Rates Example

- USA/UK exchange rate data from January 1975 to July 1984.
- The data has short-term variation about a changing level.
- First order polynomial DLM was considered with evolution variances given by discount factors ($\delta = 0.7, 0, 8, 0.9$ and 1.0).
- The prior distribution is defined by $m_0 = 0$, $C_0 = 1$, $n_0 = 1$ and $d_0 = 0.01$.
- The degree of adaptation to data increases as δ decreases.



Exchange rate and lowess with f=0.2
y
6
4
2
0
−2
−4
−6
1976
1978
1980
1982
Time


Estimated level of exchange rate
Discount factor
0.7
0.8
0.9
1
Index
0
20
40
60
80
100
-6
-4
-2
0
2
4
6

- Each value of δ defines a different model.
- To compare between models, the 3 adopted criteria are:
  - The mean absolute deviation, $MAD = \sum_{t=1}^{T} |e_t|$.
  - The mean square error, $MSE = \sum_{t=1}^{T} e_t^2/115$
  - The third summary is the observed predictive density for all the data

$$p(Y_T, Y_{T-1}, \ldots, Y_1|D_0) = \prod_{t=1}^{T} p(Y_t|D_{t-1})$$

- This third measure is a likelihood function for δ.
- LLR is the log-likelihood ratio of the predictive density relative to the model $\delta = 0.1$

Exchange Rates example

δ	MAD	$\sqrt{\text{MSE}}$	LLR
1.0	0.019	0.024	0.00
0.9	0.018	0.022	3.62
0.8	0.018	0.022	2.89
0.7	0.018	0.023	0.96

Code for first order DLM in file 'code9.s'

```
update.dlm=function (Y,delta, m.0, C.0, n.0, S.0) {
  N <- length(y)
  m =n=C=R=Q=S=f=A=e=rep( NA,N )
  Y = c(NA,Y)
  C[1] <- C.0
  m[1] <- m.0
  S[1] <- S.0
  n[1] <- n.0
 for (t in 2:N ) {
  n[t] <- n[t-1] + 1
  W[t] <- C[t-1] * (1-delta) / delta
  R[t] <- C[t-1] + W[t]
  f[t] <- m[t-1]
```

```
    Q[t] <- R[t] + S[t-1]
    A[t] <- R[t] / Q[t]
    e[t] <- Y[t] - f[t]
    S[t] <- S[t-1]+(S[t-1]/n[t] )*(e[t]^2/Q[t] - 1)
    m[t] <- m[t-1] + A[t]*e[t]
    C[t] <- A[t]*S[t]
    }
   return (list(m=m,C=C, R=R,f=f,Q=Q,n=n,S=S))
  }
```