11. Time series and dynamic linear models

Objective

To introduce the Bayesian approach to the modeling and forecasting of time series.

Recommended reading

- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, (2'nd ed.). Springer.

- Pole, A.,West, M. and Harrison, J. (1994). *Applied Bayesian forecasting and time series analysis.* Chapman and Hall.

- Bauwens, L., Lubrano, M. and Richard, J.F. (2000). *Bayesian inference in dynamic econometric models.* Oxford University Press.

Dynamic linear models

West

The first Bayesian approach to forecasting stems from Harrison and Stevens (1976) and is based on the *dynamic linear model*. For a full discussion, see West and Harrison (1997).

The general DLM

Definition 29
The general (univariate) dynamic linear model is

$$\begin{aligned} Y_t &= \mathbf{F}_t^T \boldsymbol{\theta}_t + \nu_t \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

where ν_t and $\boldsymbol{\omega}_t$ are zero mean measurement errors and state innovations.

These models are linear *state space models*, where $x_t = \mathbf{F}_t^T \boldsymbol{\theta}_t$ represents the signal, $\boldsymbol{\theta}_t$ is the state vector, $\mathbf{F}_t$ is a regression vector and $\mathbf{G}_t$ is a state matrix. The usual features of a time series such as trend and seasonality can be modeled within this format.

In some cases, $\mathbf{F}$ and $\mathbf{G}$ are supposed independent of t. Then the model is a *time series DLM*. If V and $\mathbf{W}$ are also time independent then the DLM is *constant*.

Examples

Example 80

A slowly varying level model is

$$\begin{aligned} y_t &= \theta_t + \nu_t \\ \theta_t &= \theta_{t-1} + \omega_t \end{aligned}$$

The observations fluctuate around a mean which varies according to a random walk.

Example 81

A dynamic linear regression model is given by

$$\begin{aligned} y_t &= \mathbf{F}_t^T \boldsymbol{\theta}_t + \nu_t \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

Bayesian analysis of DLM's

If the error terms, ν_t and $\boldsymbol{\omega}_t$ are normally distributed, with known variances, e.g. $\nu_t \sim \mathcal{N}(0, V_t)$, $\boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$, then a straightforward Bayesian analysis can be carried out.

Example 82

In Example 80, suppose that at time $t-1$, the current accumulated information is $D_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$ and assume that the distribution for θ_{t-1} is $\theta_{t-1}|D_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$. and that the error distributions are $\nu_t \sim \mathcal{N}(0, V_t)$ and $\omega_t \sim \mathcal{N}(0, W_t)$. Then, we have:

1. The prior distribution for θ_t is:

$$\begin{aligned} \theta_t|D_{t-1} &\sim \mathcal{N}(m_{t-1}, R_t) \quad \text{where} \\ R_t &= C_{t-1} + W_t \end{aligned}$$

2. The one step ahead predictive distribution for y_t is:

$$\begin{aligned} y_t|D_{t-1} &\sim \mathcal{N}(m_{t-1}, Q_t) \quad \text{where} \\ Q_t &= R_t + V_t \end{aligned}$$

3. The joint distribution of θ_t and y_t is

$$\begin{array}{c|c} \theta_t & \\ y_t & \end{array} D_{t-1} \sim \mathcal{N}\left(\begin{array}{c} m_{t-1} \\ m_{t-1} \end{array}, \left(\begin{array}{cc} R_t & R_t \\ r_t & Q_t \end{array} \right)\right)$$

4. The posterior distribution for θ_t given $D_t = \{D_{t-1}, y_t\}$ is

$$\begin{aligned} \theta_t|D_t &\sim \mathcal{N}(m_t, C_t) \quad \text{where} \\ m_t &= m_{t-1} + A_t e_t \\ A_t &= R_t/Q_t \\ e_t &= y_t - m_{t-1} \\ C_t &= R_t - A_t^2 Q_t. \end{aligned}$$

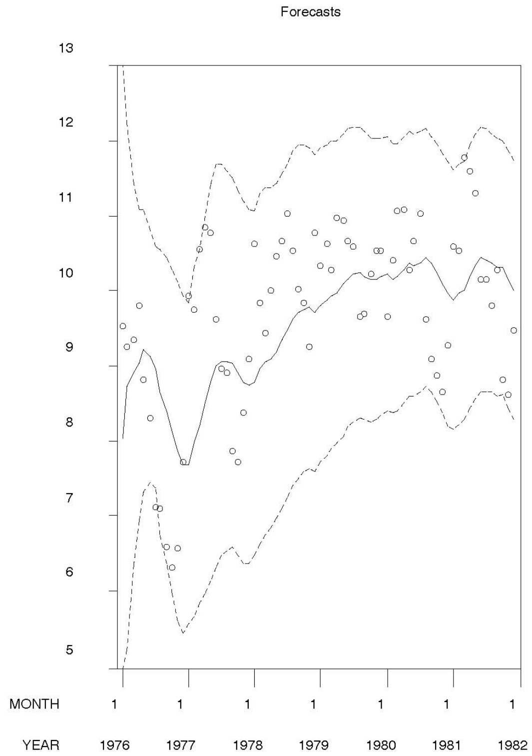
Proof and observations

Proof The first three steps of the proof are straightforward just by going through the observation and system equations. The posterior distribution follows from property iv) of the multivariate normal distribution as given in Definition 22. ■

In the formula for the posterior mean, e_t is simply a prediction error term. This formula could also be rewritten as a weighted average in the usual way for normal models:

$$m_t = (1 - A_t)m_{t-1} + A_t y_t.$$

The following diagram illustrates the one step ahead predictions for the sales data from Pole et al (1994) assuming a model with constant observation and state error variances and a non-informative prior.



An interesting feature to note is that the predictive variance approaches a fixed constant for this model as the number of observed data increases. See West and Harrison (1997) for more details.

Example 83

In Example 81, suppose that we have $\nu_t \sim \mathcal{N}(0, V_t)$ and $\omega_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$ with distribution $\boldsymbol{\theta}_t | D_{t-1} \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$. Then:

1. The prior distribution for $\boldsymbol{\theta}_t$ is:

$$\begin{aligned} \boldsymbol{\theta}_t | D_{t-1} &\sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{R}_t) \quad \text{where} \\ \mathbf{R}_t &= \mathbf{C}_{t-1} + \mathbf{W}_t. \end{aligned}$$

2. The one step ahead predictive distribution for y_t is:

$$\begin{aligned} y_t | D_{t-1} &\sim \mathcal{N}(f_t, Q_t) \quad \text{where} \\ f_t &= \mathbf{F}_t^T \mathbf{m}_{t-1} \\ Q_t &= \mathbf{F}_t^T \mathbf{R}_t \mathbf{F}_t + V_t \end{aligned}$$

3. The joint distribution of $\boldsymbol{\theta}_t$ and y_t is

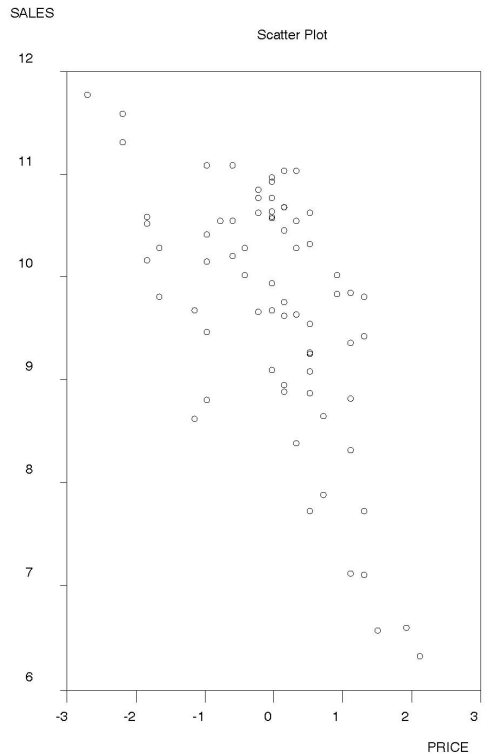
$$\left. \begin{array}{c} \boldsymbol{\theta}_t \\ y_t \end{array} \right| D_{t-1} \sim \mathcal{N}\left(\begin{array}{c} \mathbf{m}_{t-1} \\ f_t \end{array}, \left(\begin{array}{cc} \mathbf{R}_t & \mathbf{F}_t^T \mathbf{R}_t \\ \mathbf{R}_t \mathbf{F}_t & Q_t \end{array} \right) \right)$$

4. The posterior distribution for $\boldsymbol{\theta}_t$ given $D_t = \{D_{t-1}, y_t\}$ is

$$\begin{aligned} \boldsymbol{\theta}_t | D_t &\sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t) \quad \text{where} \\ \mathbf{m}_t &= \mathbf{m}_{t-1} + \mathbf{A}_t e_t \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^T Q_t \\ \mathbf{A}_t &= \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \\ e_t &= y_t - f_t. \end{aligned}$$

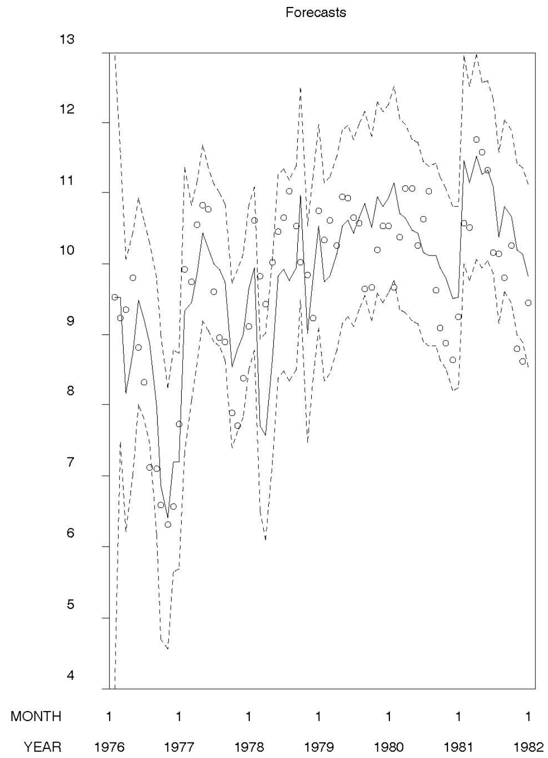
Proof Exercise. ■

The following plot shows sales against price.



Thus, a dynamic, simple linear regression model would seem appropriate.

The following diagram, assuming a constant variance model as earlier illustrates the improved fit of this model.



The general theorem for DLM's

Theorem 43

For the general, univariate DLM,

$$
\begin{aligned}
Y_t &= \mathbf{F}_t^T \boldsymbol{\theta}_t + \nu_t \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t
\end{aligned}
$$

where $\nu_t \sim \mathcal{N}(0, V_t)$ and $\omega_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$, assuming the prior distribution $\boldsymbol{\theta}_{t-1} | D_{t-1} \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$, we have

1. Prior distribution for $\boldsymbol{\theta}_t$:

$$
\begin{aligned}
\boldsymbol{\theta}_t | D_{t-1} &\sim \mathcal{N}(\mathbf{a}_t, \mathbf{R}_t) \quad \text{where} \\
\mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\
\mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^T + \mathbf{W}_t
\end{aligned}
$$

2. One step ahead prediction:

$$
\begin{aligned}
y_t|D_{t-1} &\sim \mathcal{N}(f_t, Q_t) \quad \text{where} \\
f_t &= \mathbf{F}_t^T \mathbf{a}_t \\
Q_t &= \mathbf{F}_t^T \mathbf{R}_t \mathbf{F}_t + V_t.
\end{aligned}
$$

3. Posterior distribution for $\boldsymbol{\theta}_t|D_t$:

$$
\begin{aligned}
\boldsymbol{\theta}_t|D_t &\sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t) \quad \text{where} \\
\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \\
\mathbf{C}_t &= \mathbf{R}_t \mathbf{R}_t^T Q_t \\
\mathbf{A}_t &= \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \\
e_t &= y_t - f_t.
\end{aligned}
$$

Proof Exercise. ∎

DLM's and the Kalman filter

The updating equations in the general theorem are essentially those used in the Kalman filter developed in Kalman (1960) and Kalman and Bucy (1961). For more details, see

http://en.wikipedia.org/wiki/Kalman_filter

Superposition of models

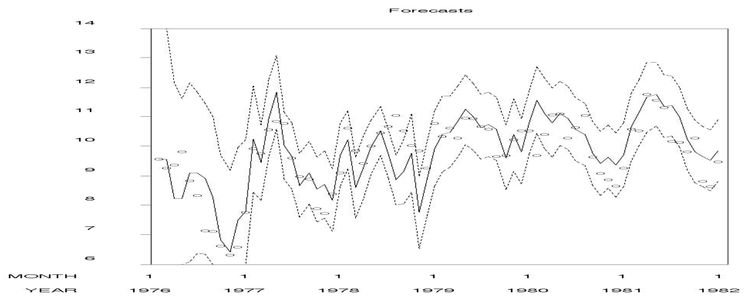
Many time series exhibit various different components. For example, as well as the regression component we have already fitted, it may well be that the sales series exhibits a seasonal component. In such cases, we may often wish to combine these components in a single model. In such cases, we may write

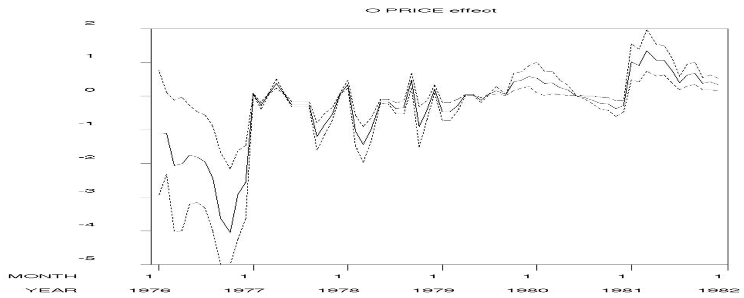
$$
\begin{aligned}
y_t &= y_{1t} + \ldots + y_{kt} + \nu_t \quad \text{where} \\
y_{jt} &= \mathbf{F}_{jt}^T \boldsymbol{\theta}_{jt} \quad \text{and} \\
\boldsymbol{\theta}_{jt} &= \mathbf{G}_{jt} \boldsymbol{\theta}_{j,t-1} + \boldsymbol{\omega}_{jt} \quad \text{for } j = 1, \ldots, k.
\end{aligned}
$$

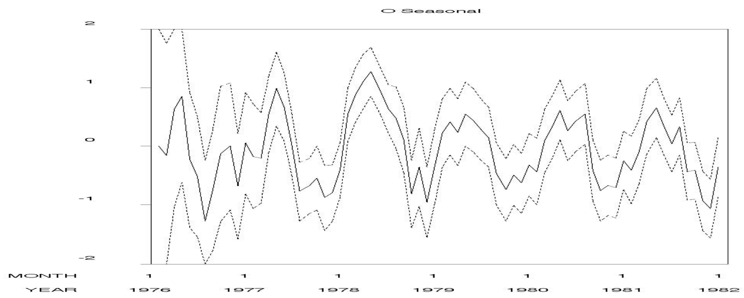
This leads to a combined model

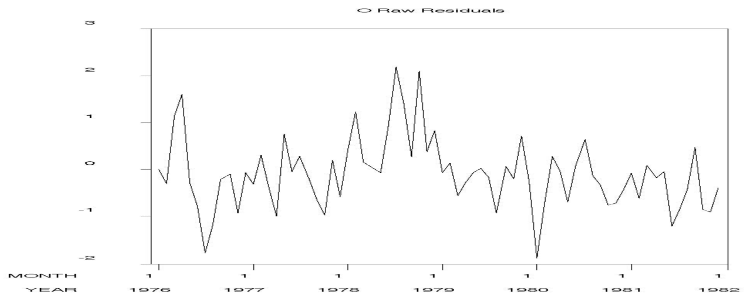
$$
\begin{aligned}
y_t &= \mathbf{F}_t^T \boldsymbol{\theta} + \nu_t \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \text{where}
\end{aligned}
$$

$$
\mathbf{F}_t = \begin{pmatrix} \mathbf{F}_{1t} \\ \vdots \\ \mathbf{F}_{kt} \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_{kt} \end{pmatrix}.
$$


Forecasts
MONTH
YEAR
1976
1977
1978
1979
1980
1981
1982


O PRICE effect
MONTH
YEAR
1976
1977
1978
1979
1980
1981
1982


O Seasonal
MONTH
YEAR
1976
1977
1978
1979
1980
1981
1982


O Raw Residuals
MONTH
YEAR
1976
1977
1978
1979
1980
1981
1982

Discount factors

Thus far, we have not considered how to model the uncertainty in the unknown variances. It is possible to model the uncertainty in the observation variances analytically in the usual way (via inverse gamma priors). However, the treatment of the system variances is more complex. In this case, *discount factors* can be used.

The idea is based on information discounting. As information ages, it becomes less useful and so its value should diminish. Thus, in our problem, with system equation

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$$

then given that $V[\boldsymbol{\theta}_{t-1}|D_{t-1}] = \mathbf{C}_{t-1}$, we have

$$\mathbf{R}_t = V[\boldsymbol{\theta}_t|D_{t-1}] = \mathbf{P}_t + \mathbf{W}_t$$

where

$$\mathbf{P}_t = V[\mathbf{G}_t\boldsymbol{\theta}_{t-1}|D_{t-1}] = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t^T \quad \text{and} \quad \mathbf{W}_t = \mathbf{R}_t - \mathbf{P}_t.$$

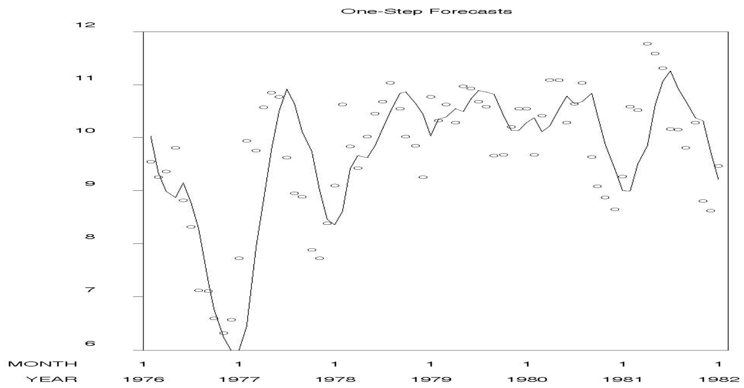
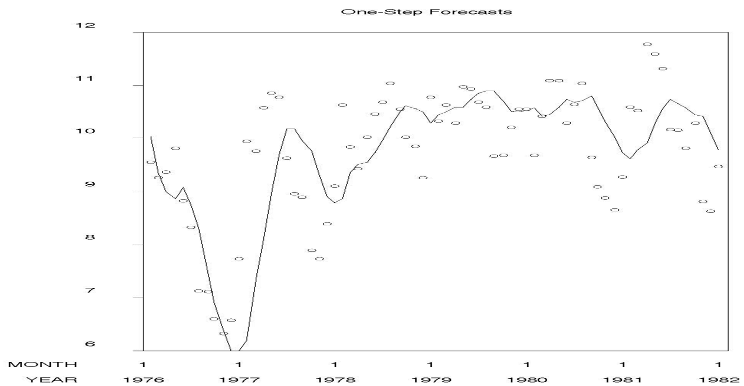
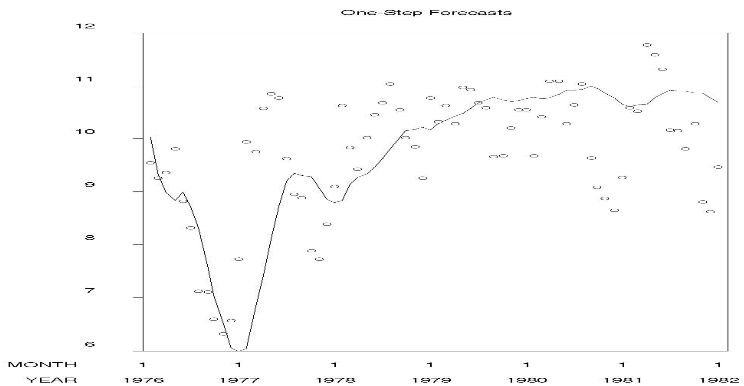
If we define δ such that $\mathbf{R}_t = \mathbf{P}_t/\delta$, then we can interpret δ as the percentage of information that passes from time $t-1$ to time t and in this case,

$$\mathbf{W}_t = \mathbf{P}_t \left(\delta^{-1} - 1\right).$$

Typical values for systems without abrupt changes are usually around $\delta = 0.9$. Small values of δ (below 0.8) imply large levels of uncertainty and lead to predictions with very wide bounds.

High values represent more smoothly changing systems, and in the limit, when $\delta = 1$, we have a static system with no information loss.

The following diagrams show the effects of fitting a trend model with discount factors 0.8, 0.9 and 1 to the sales data. We can see that the higher the discount factor, the higher the degree of smoothing.


One-Step Forecasts
12
11
10
9
8
7
6
MONTH
YEAR
1
1976
1
1977
1
1978
1
1979
1
1980
1
1981
1
1982

One-Step Forecasts
12
11
10
9
8
7
6
MONTH
YEAR
1
1976
1
1977
1
1978
1
1979
1
1980
1
1981
1
1982

One-Step Forecasts
12
11
10
9
8
7
6
MONTH
YEAR
1
1976
1
1977
1
1978
1
1979
1
1980
1
1981
1
1982

The forward filtering backward sampling algorithm

This algorithm, developed in Carter and Kohn (1994) and Frühwirth-Schnatter (1994) allows for the implementation of an MCMC approach to DLM's.

The forward filtering step is the standard normal linear analysis to give $p(\boldsymbol{\theta}_t|D_t)$ at each t, for $t = 1, \ldots, n$.

The backward sampling step uses the Markov property and samples $\boldsymbol{\theta}_n^\star$ from $p(\boldsymbol{\theta}_n|D_n)$ and then, for $t = 1, \ldots, n-1$, samples $\boldsymbol{\theta}_t^\star$ from $p(\boldsymbol{\theta}_t|D_t, \boldsymbol{\theta}_{t+1}^\star)$. Thus, a sample from the posterior parameter structure is generated.

Example 84

The $AR(p)$ model with time varying coefficients takes the form

$$\begin{aligned} y_t &= \theta_{0t} + \theta_{1t} + \ldots + \theta_{pt} y_{t-p} + \nu_t \\ \theta_{it} &= \theta_{i,t-1} + \omega_{it} \end{aligned}$$

where we shall assume that the error terms are independent normals:

$$\nu_{it} \sim \mathcal{N}(0, V) \quad \text{and} \quad \omega_{it} \sim \mathcal{N}(0, \lambda_i V).$$

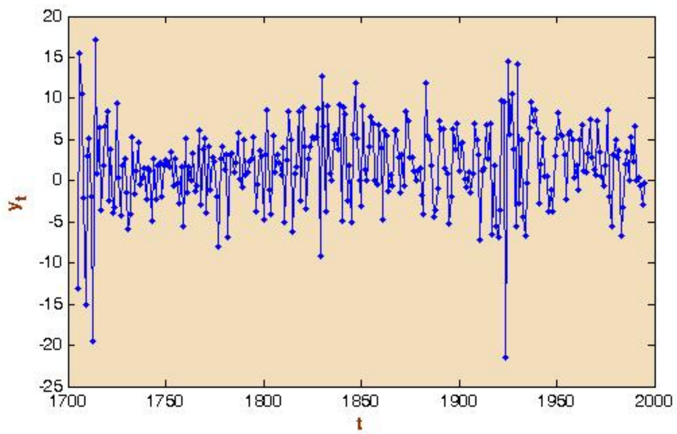
Then this model can be expressed in state space form by setting

$$\begin{aligned} \boldsymbol{\theta}_t &= (\theta_{0t}, \ldots, \theta_{pt})^T \\ \mathbf{F} &= (1, y_{t-1}, \ldots, y_{t-p})^T \\ \mathbf{G} &= \mathbf{I}_{p+1} \\ \mathbf{W} &= V \text{diag}(\boldsymbol{\lambda}) \end{aligned}$$

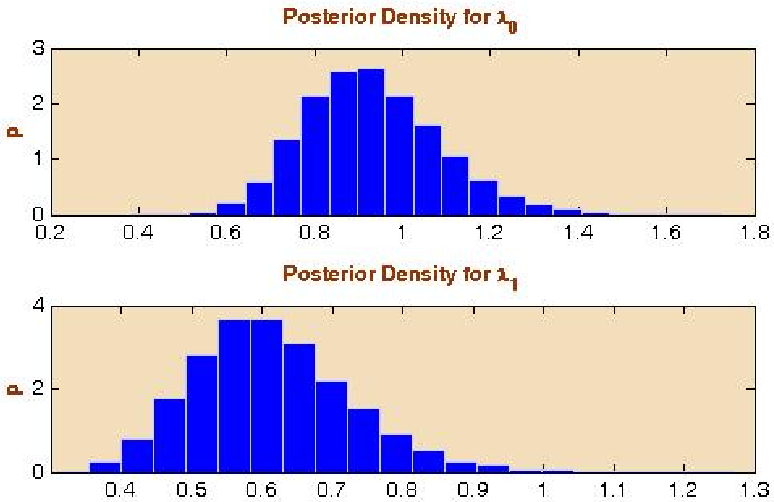
Here $\text{diag}(\boldsymbol{\lambda})$ represents a diagonal matrix with ii'th entry equal to λ_i, for $i = 1, \ldots, p + 1$.

Now, given gamma priors for V and for λ_i^{-1} and a normal prior for $\boldsymbol{\theta}_0$, then it is clear that the relevant posterior distributions are all conditionally conjugate.

Koop (2003) examines data on the annual percentage change in UK industrial production from 1701 to 1992.



Koop (2003) assumes a time varying coefficient AR(1) model and uses proper but relatively uninformative prior distributions. The posterior distributions of λ_0 and λ_1 estimated from running the Gibbs sampler are as follows.



The posterior means and deviations of both λ_0 and λ_1 suggest that there is quite high stochastic variation in both θ_{0t} and θ_{1t}.

Software for fitting DLM's

Two general software packages are available.

- `BATS`. Pole et al (1994). This (somewhat out of date) package can be used to perform basic analyses and is available from:

  `http://www.stat.duke.edu/~mw/bats.html`

- `dlm`. Petris (2006). This is a recently produced R package for fitting DLM's, including ARMA models etc. available from

  `http://cran.r-project.org/src/contrib/Descriptions/dlm.html`

Other work in time series

- ARMA and ARIMA models. Marriot and Newbold (1998).
- Non linear and non normal state space models. Carlin et al (1992).
- Latent structure models. Aguilar et al (1998).
- Stochastic volatility models. Jacquier et al (1994).
- GARCH and other econometric models. Bauwens et al (2000).
- Wavelets. Nason and Sapatinas (2002).

References

Aguilar, O., Huerta, G., Prado, R. and West, M. (1998). Bayesian inference on latent structure in time series. In *Bayesian Statistics 6* eds. Bernardo et al. Oxford: University Press.

Bauwens, L., Lubrano, M. and Richard, J.F. (2000). *Bayesian inference in dynamic econometric models.* Oxford University Press.

Carlin, B.P., Poison, N.G. and Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association*, **87**, 493–500.

Carter, C.K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **83**, 589–601.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society Series B*, **38**, 205-247.

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, **12**, 371–389.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering Series D*, **82**, 34-45.

Kalman, R.E. and Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering Series D*, **83**, 95-108.

Koop, G. (2003). *Bayesian Econometrics*. New York: Wiley.

Marriott, J. and Newbold, P. (1998). Bayesian Comparison of ARIMA and Stationary ARMA Models. *International Statistical Review*, **66**, 323–336.

Nason, G.P. and Sapatinas, T. (2002). Wavelet packet transfer function modelling of nonstationary time series. *Statistics and Computing*, **12**, 45–56.

Pole, A.,West, M. and Harrison, J. (1994). *Applied Bayesian forecasting and time series analysis.* Chapman and Hall.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, (2'nd ed.). Springer.