Notes on Signal extraction and prediction of pure trend series using state-space models - theory

Sonia Mazzi

13/05/2019

Suppose that we have $y = \{y_t\}_{t=1}^n$ and $y_t = \mu_t + \epsilon_t$, where μ_t is a stochastic process (with linear dynamic) and ϵ_t's are iid with mean zero and constant variance. Note that y_t can be a vector. This is a signal-plus-noise model. $\mu = \{\mu_t\}$ is the signal, which is unobserved and the feature that we wish to extract from $y = \{y_t\}$. $\epsilon = \{\epsilon_t\}$ is the noise, which has no significant stochastic structure, such as autocorrelation, and has expected value zero. We always assume that $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0, \sigma^2)$.

The stochastic specification of μ_t allows to model y in different ways. We will use a state-space model to determine the stochastic structure of μ, and therefore of y.

For more details on the state-space model and signal-extraction algorithms see

de Jong, P.(1991). The diffuse Kalman filter. *The Annals of Statistics.* Vol 19, No.2, pp 1073 - 1083.

The local linear trend model

The "emerging trends" data is what we call pure-trend data. The main source of variation of y is trend. If we do not consider covariates for y we can say that the trend is the only source of variation of y. Note that for the "emerging trends" data we do not wish to explain the source of variation of the data, we do not wish to explain what makes y vary. We wish to extract a smooth signal indicating the local trend and predict its behaviour in terms of trend.

In the LLM it is assumed

$y_i = \mu_i + \epsilon_i$, $\mu_{i+1} = \mu_i + d_i + \nu_i$, $d_{i+1} = \delta d_i + \eta_i$

where $\{\epsilon_i\}$, $\{\nu_i\}$, $\{\eta_i\}$ are mutually independent white noise processes with variances σ^2, σ_ν^2 and σ_η^2.

The rationale for the model stems from initially assuming that the signal, μ_i, follows a fully deterministic specification: $\mu_i = a + d\,i$, with a and d constant. Whilst this specification would be clearly unsatisfactory for most time series, if we assume that $\mu = \{\mu_i\}$ is "smooth", μ could be described locally by a linear model (Taylor's theorem). The idea is to let a and d vary with time to allow μ_i to adapt to the evolution of the series.

If $\mu_i = a + d\,i$ then $\mu_{i+1} = \mu_i + d$.

To allow for flexibility in the model we could establish that

$$\begin{aligned} \mu_{i+1} &= \mu_i + d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

where $\nu = \{\nu_i\}$ and $\eta = \{\eta_i\}$ are independent white noise processes with mean zero and variance σ_ν^2 and σ_η^2 respectively. ϵ is independent of ν and η.

So, the full model would be

$$\begin{aligned} y_i &= \mu_i + \epsilon_i, \\ \mu_{i+1} &= \mu_i + d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

Note that

$$\begin{aligned} y_{i+1} - y_i &= \mu_{i+1} - \mu_i + \epsilon_{i+1} - \epsilon_i, \\ &= d_i + \nu_i + \epsilon_{i+1} - \epsilon_i. \end{aligned}$$

So the model is ARIMA(0,1,1) plus a stochastic term, d_i, which has an AR(1) specification. In ARIMA(0,1,1) d_i would be a constant.

For the model above we have assumed equal spacing between time points at which y is observed. However, it is very easy to generalise the specification above for series observed at any time points, equally or unequally spaced, e.g. equally spaced with missing data, or high-frequency data which comes at irregularly spaced times.

Suppose y is observed at times $t_1, \ldots, t_n$ and let $h_i = t_{i+1} - t_i$. Then, if $\mu_i = \mu_{t_i}$

$$\begin{aligned} \mu_{i+1} &= \mu_i + d_i h_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

So, the complete model would be

$$\begin{aligned} y_i &= \mu_i + \epsilon_i, \\ \mu_{i+1} &= \mu_i + d_i h_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

To cast the model into state-space form we write

$$\begin{aligned} y_i &= Z\alpha_i + Gu_i, \\ \alpha_{i+1} &= T_i\alpha_i + Hu_i. \end{aligned}$$

where $u_0, u_1, \ldots, u_n$ are i.i.d with $Var(u_i) = \sigma^2 I_3$, where I_n denotes the $n \times n$ identity matrix, and

$$\alpha_i = \begin{pmatrix} \mu_i \\ d_i \end{pmatrix}, \;\; Z = (1\ 0), \;\; T_i = \begin{pmatrix} 1 & h_i \\ 0 & \delta \end{pmatrix}, \;\; u_i = \begin{pmatrix} \epsilon_i \\ \eta_i \\ \nu_i \end{pmatrix}, \;\; G = (1\ 0\ 0), \;\; H = \begin{pmatrix} 0 & \sigma_\nu/\sigma & 0 \\ 0 & 0 & \sigma_\eta/\sigma \end{pmatrix},$$

with $\alpha_1 = W_0\beta$, $\beta = b + B\gamma$, $\gamma \sim (c, \sigma^2 C)$ and in this case $W_0 = B = I_2$, $b = c = 0_2$, where 0_n denotes a vector of zeroes of length n.

If we think of mu_i as $\mu_i = a_i + d_i t_i$, α_1 represents $(a_1 + d_1 t_1 \quad d_1)^t$. v^t means the transpose of v.

The Diffuse Kalman Filter (DKF) is a generalisation of the Kalman Filter which allows to deal with initial conditions in the same paradigm that estimation, signal extraction and prediction take place.

The DKF in this case is the recursion

$$\left.\begin{aligned} E_i &= (0,\, y_i) - ZA_i, \\ D_i &= ZP_iZ^t + GG^t, \\ K_i &= T_iP_iZ^tD_i^{-1}, \\ A_{i+1} &= T_iA_i + K_iE_i, \\ P_{i+1} &= (T_i - K_iZ)P_iT_i^t + HH^t, \\ Q_{i+1} &= Q_i + E_i^tD_i^{-1}E_i, \end{aligned}\right. \qquad i = 1, \ldots, n$$

with starting conditions $A_1 = (-I_2, 0_2)$, $P_1 = 0$, $Q_1 = 0$.

After the DKF is run, obtain Q_{n+1} and partition it

$$Q_{n+1} = \begin{pmatrix} S & s \\ s^t & q \end{pmatrix},$$

so that if Q_{n+1} has c columns and r rows, S is a $(q-1)\times(c-1)$ matrix.

The MLE of α_1 is $S^{-1}s$, with covariance matrix $\sigma^2 S^{-1}$, the MLE of σ^2 is $(q - s^t S^{-1}s)/n$ and the log-likelihood (maximised with respect to σ^2 and α_1) is $-\frac{1}{2}\left[n\log(\hat{\sigma}^2) + \sum_{i=1}^n \log(|D_i|)\right]$.

The parameters δ, σ_η, σ_ν need to be estimated. We can do this via maximum likelihood.

Once the parameters are estimated and the DKF run, we can proceed to prediction (forecasting).

Prediction using a state-space model

We know that

$$
\begin{aligned}
y_i &= Z\alpha_i + Gu_i,\\
\alpha_{i+1} &= T_i\alpha_i + Hu_i.
\end{aligned}
$$

For the sake of simplicity, let us assume again that data is equally spaced. In this case $T_i = T$, $\forall i$.

Then $y_{n+k+1} = Z\alpha_{n+k+1} + Gu_{n+k+1}$, $k \geq 0$. If $y = \{y_1, \ldots, y_n\}$, the predictor of y_{n+k+1} given y is

$$
E(y_{n+k+1}|y) = ZE(\alpha_{n+k+1}|y),
$$

since $E(u_{n+k+1}|y) = 0$ if $k \geq 0$.

Now, $\alpha_{n+k+1} = T\alpha_{n+k} + Hu_{n+k} = \cdots = T^k\alpha_{n+1} + H\sum_{j=1}^k T^{j-1}u_{n+k+1-j}$. This can be easily shown by induction.

Therefore, $E(\alpha_{n+k+1}|y) = T^k E(\alpha_{n+1}|y)$ and so the predictor of y_{n+k+1} is

$$
E(y_{n+k+1}|y) = ZE(\alpha_{n+k+1}|y) = ZT^k E(\alpha_{n+1}|y).
$$

In order to obtain any prediction of the series we just need the one-step ahead predictor of the state vector.

The DKF gives us all what is needed to compute the one step-ahead predictor of the state vector, $E(\alpha_{n+1}|y)$.

$$
\begin{aligned}
\hat{\alpha}_{n+1} &= E(\alpha_{n+1}|y) = A_{n+1}(-S^{-1}s;\,1),\\
mse(\hat{\alpha}_{n+1}) &= \sigma^2[P_{n+1} + MS^{-1}M^t],
\end{aligned}
$$

where M denotes all but the last column of A_{n+1}.

So, the predictor of y_{n+k+1} given y and its mean squared error are

$$
\begin{aligned}
\hat{y}_{n+k+1} &= ZT^k\hat{\alpha}_{n+1},\\
mse(\hat{y}_{n+k+1}) &= ZT^k\hat{\alpha}_{n+1}(T^k)^t Z^t.
\end{aligned}
$$

Signal extraction

In the context of a pure trend model we wish to extract the smooth signal. That is we would like to predict $E(\mu_t|y)$, $t = 1, \ldots, n$.

The DKF gives us the predictions of $E(\alpha_{t+1}|y_1, \ldots, y_t)$. We use the smoothing filter, a backwards recursion, to obtain the smoothed valued of the series.

The Smoothing Filter (SF) is the recursion

$$
\begin{aligned}
N_{i-1} &= Z^t D_i^{-1} E_i + L_i^t N_i,\\
R_{i-1} &= Z^t D_i^{-1} Z + L_i^t R_i L_i,
\end{aligned}
$$

with $N_n = 0$, $R_n = 0$, $L_i = T_i - K_i Z$, $i = 1, \ldots, n$, and all other quantities as in the DKF.

Then,

$$\tilde{\alpha}_i = E(\alpha_i|y) = (A_i + P_i N_{i-1})(-S^{-1}s;\, 1)$$
$$mse(\tilde{\alpha}_i) = \sigma^2 \left(P_i - P_i R_{i-1} P_i + N_{i-1,\gamma} S^{-1} N_{i-1,\gamma}^t\right),$$

where $N_{i-1,\gamma}$ denotes all but the last column of $A_i + P_i N_{i-1}$.

The state vector is $\alpha_i = (\mu_i \;\; d_i)^t$. The second entry of α_i can be interpreted as we usually interpret the first derivative. Negative values indicate a decline in the levels of μ, positive values of d_i indicate incremental values of μ_i, values of d_i near zero may indicate a change point in μ_i.

From the point of view of forecasting, it might be beneficial not just to forecast values of the series but also predict values of the first derivative to predict the general future trend: upwards, downwards and/or approaching a change point.

The Local quadratic model (LQM)

If the LLM is not flexible enough, one may fit a LQM.

In the LQM it is assumed

$y_i = \mu_i + \epsilon_i$, $\mu_{i+2} = 2\mu_{i+1} - \mu_i + d_i + \nu_i$, $d_{i+1} = \delta d_i + \eta_i$

where $\{\epsilon_i\}$, $\{\nu_i\}$, $\{\eta_i\}$ are mutually independent white noise processes with variances σ^2, σ_ν^2 and σ_η^2.

The rationale for the model stems from initially assuming that the signal, μ_i, follows a fully deterministic specification: $\mu_i = a + b\,i + c\,i^2$, with a, b and c constant. Whilst this specification would be clearly unsatisfactory for most time series, if we assume that $\mu = \{\mu_i\}$ is "smooth", μ could be described locally by a quadratic model (Taylor's theorem). The idea is to let a, b and c vary with time to allow μ_i to adapt to the evolution of the series.

If $\mu_i = a + b\,i + c\,i^2$ then $\mu_{i+1} = \mu_i + b + c + 2c\,i$. So,

$$\mu_{i+1} - \mu_i = b + c + 2c\,i,$$
$$\mu_{i+2} - \mu_{i+1} = b + c + 2c\,(i+1).$$

Therefore, subtracting one equation from the other $\mu_{i+2} - 2\mu_{i+1} + \mu_i = 2c$.

To allow for flexibility in the model we could establish that

$$\mu_{i+2} = 2\mu_{i+1} - \mu_i + d_i + \nu_i,$$
$$d_{i+1} = \delta d_i + \eta_i.$$

where $d_i = 2c_i$, and $\nu = \{\nu_i\}$ and $\eta = \{\eta_i\}$ are independent white noise processes with mean zero and variance σ_ν^2 and σ_η^2 respectively. ϵ is independent of ν and η.

So, the full model would be

$$y_i = \mu_i + \epsilon_i,$$
$$\mu_{i+2} = 2\mu_{i+1} - \mu_i + d_i + \nu_i,$$
$$d_{i+1} = \delta d_i + \eta_i.$$

To cast the model into state-space form we write

$$y_i = Z\alpha_i + Gu_i,$$
$$\alpha_{i+1} = T_i\alpha_i + Hu_i.$$

where $u_0, u_1, \ldots, u_n$ are i.i.d with $Var(u_i) = \sigma^2 I_3$, where I_n denotes the $n \times n$ identity matrix, and

$$\alpha_i = \begin{pmatrix} \mu_{i+1} \\ \mu_i \\ d_i \end{pmatrix}, \quad Z = (0\ 1\ 0), \quad T_i = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & \delta \end{pmatrix}, \quad u_i = \begin{pmatrix} \epsilon_i \\ \eta_i \\ \nu_i \end{pmatrix}, \quad G = (1\ 0\ 0), \quad H = \begin{pmatrix} 0 & \sigma_\nu/\sigma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_\eta/\sigma \end{pmatrix}.$$

As with the local linear model we can also write the model for unequally spaced data.

The initial conditions for the DKF are $\alpha_1 = W_0\beta$, $\beta = b + B\gamma$, $W_0 = B = I_3$, $b = 0$, $H_0 = 0$.

Now, note that if $f(i) = a + b\,i + c\,i^2$, $f'(i) = b + 2c\,i$, and $f''(i) = 2c$.

Also $f(i+1) - f(i) = b + c + 2c\,i = f'(i) + c$.

So, we interpret $\mu_{i+1} - \mu_i - \frac{1}{2}d_i$ as the derivative of the signal μ and d_i as the second derivative.

Let $M_1 = \begin{pmatrix} 1 & -1 & -\frac{1}{2} \end{pmatrix}$ and $M_2 = (0\ 0\ 1)$. Then, $Z\alpha$ is the signal, $M_1\alpha$ is the first derivative and $M_2\alpha$ is the second derivative.

The DKF and SF recursions are the same as before.