Time Series Analysis via Kalman Filtering

Kristína Grolmusová, Supervisor: Maddie Smith

STOR-i, Lancaster University

Lancaster 26/08/2022

Aims

- To apply Dynamic Linear Models (DLMs) to time series, that do not show a particular trend or seasonal variation.
- To use the local level model to generate observed data.
- To apply the Kalman filter to perform inference on the unknown state vector and forecast.
- To test the performance of the Kalman filter on multiple observation sequences and in the case of missing data.

Theory

Dynamic Linear Models (DLMs)

Observation (1) and state (2) equation:

$$Y_t = F_t\theta_t + v_t, v_t \sim \mathcal{N}(0, V_t) \tag{1}$$

$$\theta_t = G_t\theta_{t-1} + w_t, w_t \sim \mathcal{N}(0, W_t) \tag{2}$$

Assumption: $\theta_0 \sim \mathcal{N}(m_0, C_0)$

Kalman filter

1) Prior at t: $p(\theta_t|D_{t-1})$

2) One-step forecast: $p(Y_t|D_{t-1})$

3) Posterior at t: $p(\theta_t|D_t)$

Generating Data

Local Level Model

$$Y_t = \theta_t + v_t, v_t \sim \mathcal{N}(0, V_t) \tag{3}$$

$$\theta_t = \theta_{t-1} + w_t, w_t \sim \mathcal{N}(0, W_t) \tag{4}$$

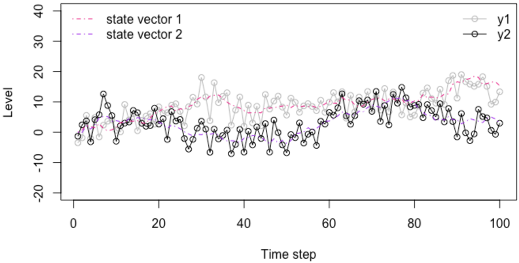
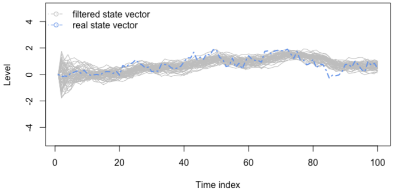
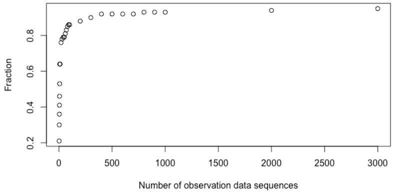


Figure: Generated state vector sequence by the random walk and the observation data sequence by local level model.

Generating Multiple Observation Sequences



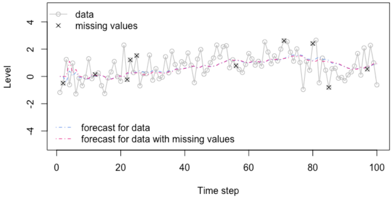
(a) 50 filtered state vector sequences together with the true state vector sequence.



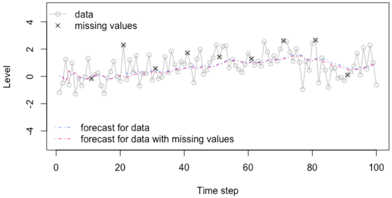
(b) The fraction of the true state vector sequence laying within the range of the filtered state vector sequences.

Missing Data - Univariate

Randomly and periodically missing data



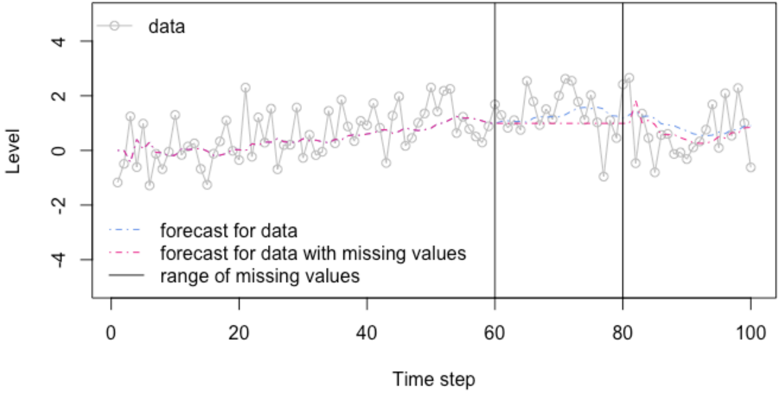
(a) Randomly missing data



(b) Periodically missing data

Missing Data - Univariate

Range of missing data



Missing Data - Multivariate

Range of missing data

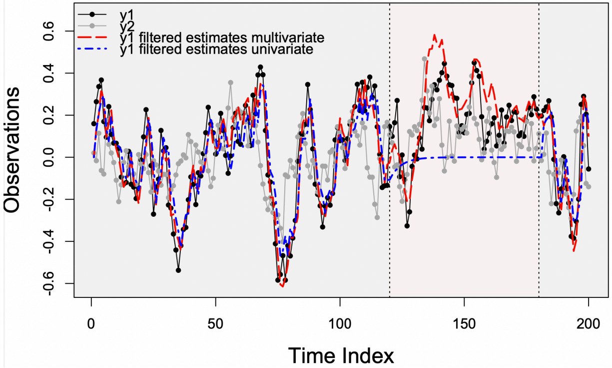
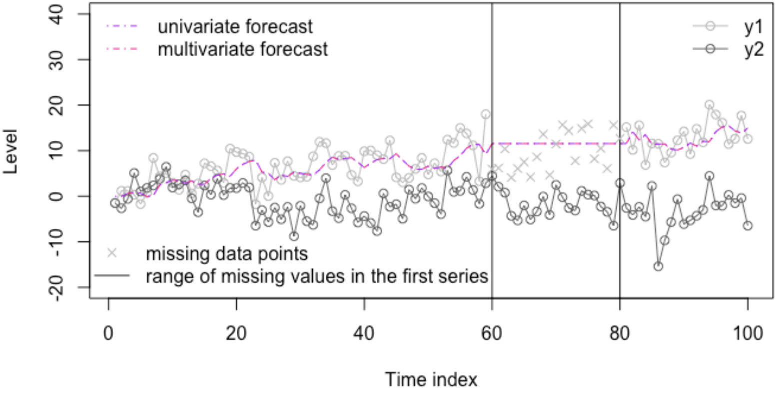


Figure: Univariate and multivariate method applied to a bivariate VAR model Smith et al. [2022]

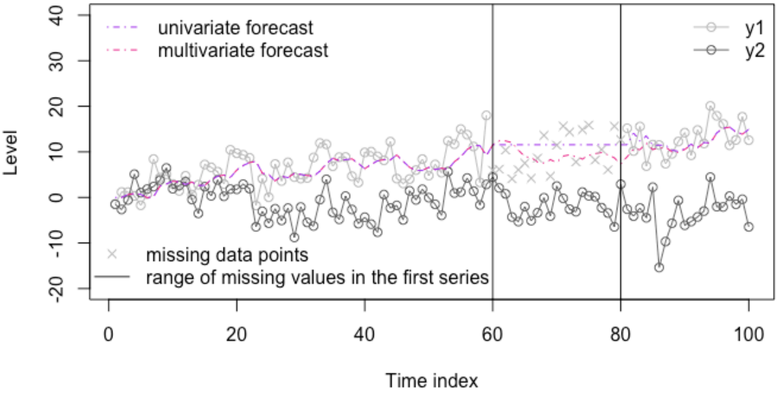
Missing Data - Multivariate

Range of missing data



Missing Data - Multivariate

Range of missing data - different method



Missing Data - Multivariate

Comparing the two multivariate methods

Correlation		Standard multivariate method (univariate)	Modified multivariate method
90%	MSE(Y)/MSE(θ)	15.873/4.331	14.259/2.263
80%	MSE(Y)/MSE(θ)	15.531/3.906	14.678/2.537
70%	MSE(Y)/MSE(θ)	15.317/3.626	15.194/2.937
60%	MSE(Y)/MSE(θ)	15.171/3.421	15.761/3.401
50%	MSE(Y)/MSE(θ)	15.070/3.266	16.359/3.907
40%	MSE(Y)/MSE(θ)	15.001/3.148	16.981/4.443

Table: Table comparing the two methods by computing the mean-square error for different correlations.

Conclusion

- Generating multiple observation sequences.
- Missing data (univariate) - random, periodic, range.
- Missing data (multivariate) - standard and modified method

Thank you!

References

Maddie Smith et al. *Dynamic Linear Models and Applications to Modelling Financial Data*. unpublished, 2022.

Giovanni Petris. *Dynamic Linear Models with R*. Springer, 2009.

Mike West and Raquel Prado. *Time Series: Modelling, Computation and Inference*. Chapman and Hall/CRC, 2010.

Kalman Filter in More Detail

$$\theta_{t-1}|D_{t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1}), t > 1$$

Prior at t:

$$\theta_t|D_{t-1} \sim \mathcal{N}(a_t, R_t),$$

$$a_t = G_t m_{t-1},$$

$$R_t = P_t + W_t, P_t = G_t C_{t-1} G_t'.$$

One-step forecast:

$$Y_t|D_{t-1} \sim \mathcal{N}(f_t, Q_t),$$

$$f_t = F_t a_t,$$

$$Q_t = F_t R_t F_t' + V_t.$$

Posterior at t:

$$\theta_t|D_t \sim \mathcal{N}(m_t, C_t),$$

$$m_t = a_t + R_t F_t' Q_t^{-1} e_t, e_t = y_t - f_t,$$

$$C_t = R_t - R_t F_t' Q_t^{-1} F_t R_t.$$

Missing data

Univariate

$$m_t = a_t,$$
$$C_t = R_t.$$

Multivariate

Missing data in the first row,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \quad M_t = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

New observation equation,

$$\tilde{y}_t = \tilde{F}_t \theta_t + \tilde{v}_t,$$
$$\tilde{F}_t = M_t F_t,$$
$$\tilde{v}_t = M_t v_t.$$