Time Series Analysis via Kalman Filtering

Kristína Grolmusová[1], Maddie Smith[2]

University of Manchester[1], STOR-i Lancaster University[2]

STOR-i excellence with impact

1. Motivation

Time series analysis allows one to study phenomena that evolve with time. The aim of the project was to apply Dynamic Linear Models (DLMs) to time series, that do not show a particular trend or seasonal variation. In particular, the local level model was used to generate observed data and the Kalman filter was used to perform inference on the unknown state vector and forecast. The performance of the Kalman filter was tested on multiple observation sequences and in the case of missing data in the sequences.

2. Dynamic Linear Models (DLMs)

DLMs belong to a general class of time series models. They are described by the observation equation (1) and the state equation (2) specified as follows,

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

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

- Y_t ... the observed time series.
- θ_t ... the unobserved state vector.
- v_t, w_t ... the independent noise sequences drawn from normal distribution with mean $\mathbf{0}$ and covariance matrices V_t and W_t.
- F_t, G_t ... the known matrices chosen according to the desired model.

Assumption: θ_0 is normally distributed with mean m_0 and variance C_0.

Local Level Model

For a univariate model,

$$F_t = G_t = 1.$$

For a multivariate model,

$$F_t = G_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

3. Kalman Filter

- In a filtering problem, the data arrives sequentially in time.
- The aim is to estimate the current state vector and forecast of the next observation based on the previous observations and estimates of the state vector. This process is known as filtering.
- The Kalman filter updates the estimates as new observation becomes available.

4. Generating Multiple Observation Sequences

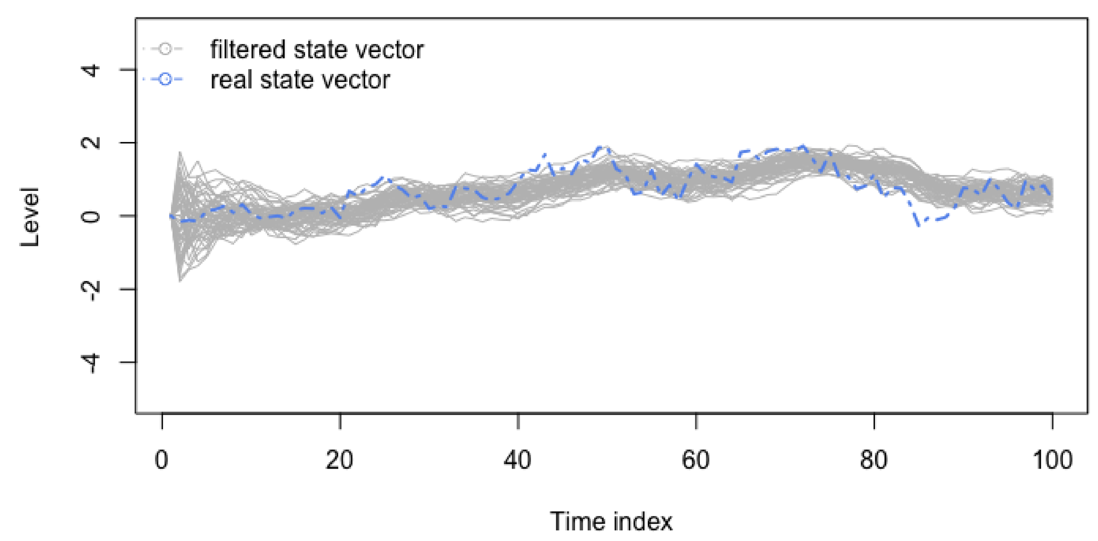

Figure 1: Graph showing 50 filtered state vector sequences together with the true state vector sequence.

When applying a DLM to simulated data, the true state vector is not always captured by the filtered estimates. For that reason, a sequence of state vectors was generated by a random walk, and 50 corresponding observation sequences were generated according to the univariate local level model.

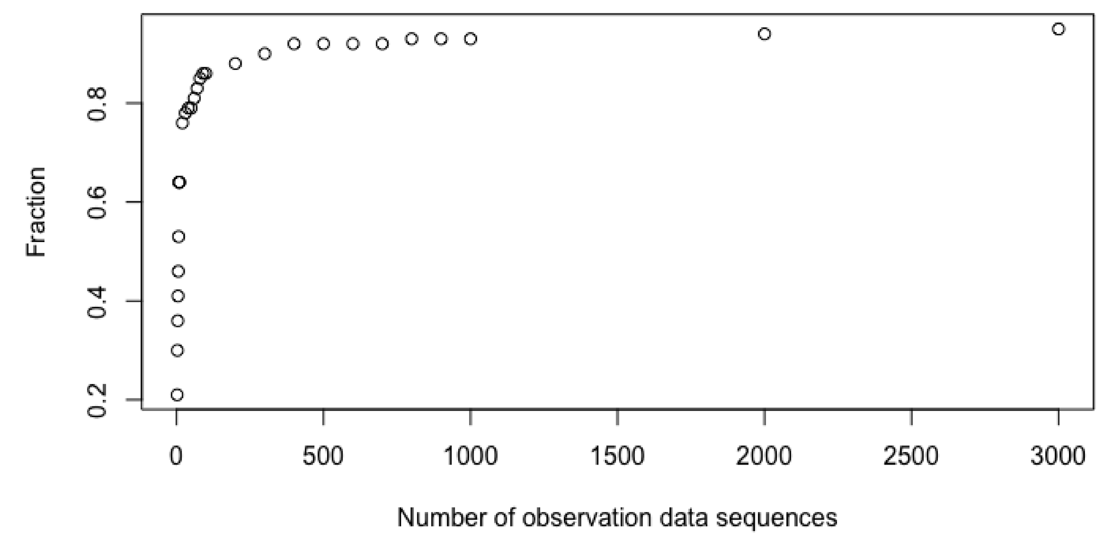

Figure 2: Graph showing the fraction of the generated state vector sequence laying within the range of the filtered state vector sequences, against the number of observation data sequences.

The Kalman filter was applied to these repetitions.
At each time step the range of the filtered values was computed, in order to determine the fraction of the true state vector sequence laying within the range of the filtered state vector sequences. It can be noticed empirically that a large number of repetitions increases the chance of the true state vector sequence lying within the range of filtering estimates, Figure 2.

5. Missing Data

? What happens if the observation sequence contains missing data?

Univariate method: The last state vector and forecast estimates are simply carried forward.

- Randomly and periodically missing data - works well.
- Range of missing data - the state vector and forecast estimates will not change for the duration of the whole range.

? Can we do better? A Yes!

Multivariate method: Two or more correlated observation data sequences, where one or more, but not all of the data sequences have data missing; use the data from the non-missing sequence to estimate the state vector and forecast.

- This has been done for a bivariate VAR model and showed significant improvements to the univariate case, Smith et al. (2022).
- However, it was found not to be the case for the multivariate local level model, Figure 3.

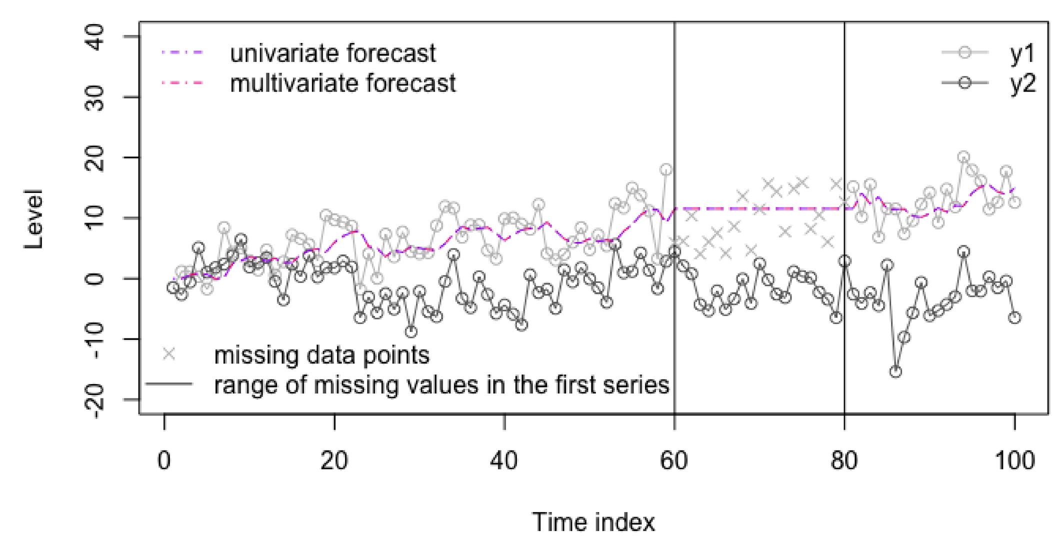

Figure 3: Graph showing the generated data using the multivariate local level model and the estimated forecasts for range of missing data using both univariate and multivariate methods.

6. Conclusions and Further Work

At around 250 sequences, the true state vector is captured within the range of filtered estimates for a significant fraction of time, after which increasing the number of repetitions does not have significant effect on the fraction of true state vector captured within the range of filtered estimates.

The multivariate method of dealing with missing data tends to perform better than the univariate. However, when applied to the multivariate local level model, the resulting state vector and forecast estimates were non-changing throughout the range of missing data. This is due to G_t being diagonal.

Further work could be done on a multivariate method for missing data for the case of diagonal G_t matrix.

References

Petris, G. (2009). *Dynamic Linear Models with R*. Springer.

Smith, M., et al. (2022). Approaching Modern Problems in Time Series Analysis.

West, M., & Prado, R. (2010). *Time Series: Modelling, Computation and Inference.* Chapman and Hall/CRC.

kristina.grolmusova@student.manchester.ac.uk