Chapter 4

Modes of intervention

Intervention, which can be understood as the action of an external operator on the parameter of an existing model, can be divided into two main categories:

Feed-forward: that is when additional information is received by the operator which indicates that the model needs to be tuned in order to face changes that are expected to arise and that are not covered by the existing model. This mode of intervention is anticipatory by nature.

Feedback: that is when the current model displays bad forecasting performance, either with respect to an adversarial model or in the absolute. In this case, the operator has to identify when the model most likely started to misbehave and, if possible, what is the cause of the decrease in forecasting performance.

These modes of intervention will be studied in turn in the next two sections. In both cases, the intervention can be local, e.g. removing an observation that is considered as an outlier, or it can have long-lasting effects, e.g. when acting directly on the predicted distribution of the parameter or when changing the model. Note that an outlier is not necessarily an erroneous observation, it can also be a true observation that differs significantly from the observed trend before and after the time step when it occurs; in this sense, an outlier might be a useful piece of information in itself, but not for the purpose of forecasting future trends.

4.1 Feed-forward intervention

There are three main ways of performing feed-forward intervention:

1. If it is known in advance that the observation at time step $k \geq 0$ will most-likely be an outlier, then it can simply be ignored for estimation purposes. The update step for that particular observation can be skipped so that the posterior at time step k will be defined as being equal to the prior. Another way of achieving the same result while still using the update equation is to set the variance V_k of this observation to infinity; indeed, in this case, the covariance of the innovation $S_k = H_k P_k H_k^\intercal + V_k$ will also be infinite so that the Kalman gain $K_k = P_k H_k^\intercal S_k^{-1}$ will be zero and the posterior mean and variance will simply be expressed as $\hat{m}_k = m_k$ and $\hat{P}_k = P_k$.

2. If some changes known to the operator are expected to affect the evolution of some or all of the components of the parameter of interest, then the operator can decide to add some extra noise at one or several time steps in order to bring more flexibility to the model. If there is some prior knowledge about the direction in which the change will affect the different components then one can introduce a control variable c_k corresponding to the shift in predicted mean that is expected. The prediction equation is then changed to

$$\boldsymbol{\theta}_k = F_k \boldsymbol{\theta}_{k-1} + c_k + \boldsymbol{u}_k,$$

   where the variance U_k of $\boldsymbol{u}_k$ can be increased to model that this given time step might be more uncertain than others. The only difference in the corresponding Kalman filter recursion is that the predicted mean becomes

$$m_k = F_k \hat{m}_{k-1} + c_k.$$

3. In general, the operator can simply change the predicted mean and variance to the desired values. However, acting directly on these quantities introduces inconsistencies in the sequential estimation of the predicted and posterior statistics of the parameter and extra care needs to be taken when performing operations involving past means and variances such as smoothing. This can be done as follows: if we want the predicted mean to be equal to a given vector $\breve{m}_k$ and the predicted variance to be equal to a given matrix $\breve{P}_k$ then we can modify the state equation to

$$\boldsymbol{\theta}_k = \breve{F}_k \boldsymbol{\theta}_{k-1} + c_k + \breve{\boldsymbol{u}}_k$$

where the modified transition matrix $\breve{F}_k$, modified noise $\breve{\boldsymbol{u}}_k \sim \mathrm{N}(\cdot\,; 0, \breve{U}_k)$ and control variable c_k are defined through

$$\begin{aligned} \breve{F}_k &= M_k F_k \\ c_k &= \breve{m}_k - M_k F_k \hat{m}_{k-1} \\ \breve{U}_k &= M_k U_k M_k^{\intercal} \end{aligned}$$

with $M_k = \breve{L}_k L_k^{-1}$, where $\breve{L}_k$ and L_k are the unique upper-triangular matrices such that $\breve{P}_k = \breve{L}_k \breve{L}_k^{\intercal}$ and $P_k = L_k L_k^{\intercal}$ (Cholesky decomposition).

4.2 Feedback intervention

In some cases, there is no external information available prior to the deterioration of the predictive performance which could have allowed for feed-forward intervention, so that only corrective measures are applicable. The ways of performing feedback intervention are very similar to the ones introduced in the feed-forward situation and will not be covered in details in this section. The main difficulty with feedback intervention is elsewhere and can be summarised in three questions:

i) how to detect a significant decrease in predictive performance?

ii) how to characterise the cause of this loss of performance?

iii) how to determine when corrective measures should be applied?

We will first consider a general way of answering these questions (in the sense that it applies to more general models than DLMs). This will be followed by a simpler error analysis that is restricted to DLMs.

4.2.1 Bayes' factor

In this section, we consider a model monitoring technique which involves comparing the marginal likelihood of two different models in order to determine which model best fits the data. The same technique can be used more generally for problems like model selection but we restrict ourselves to the case where one standard model $\mathcal{M}_0$, i.e. the one used for inference so far, is challenged by another model $\mathcal{M}_1$. The *Bayes' factor* at time step k for Model $\mathcal{M}_0$ vs. $\mathcal{M}_1$ is defined as

$$B_k = \frac{p^{(0)}_{\boldsymbol{y}_k | \boldsymbol{y}_{0:k-1}}(y_k \mid y_{0:k-1})}{p^{(1)}_{\boldsymbol{y}_k | \boldsymbol{y}_{0:k-1}}(y_k \mid y_{0:k-1})}$$

where $p^{(0)}(\cdot)$ and $p^{(1)}(\cdot)$ are probability density functions under Model $\mathcal{M}_0$ and $\mathcal{M}_1$ respectively. A small Bayes' factor, typically $\log B_k < -1$, indicates evidence against $\mathcal{M}_0$ when compared to $\mathcal{M}_1$ and $\log B_k < -2$ indicates strong evidence. Note that the value of the Bayes' factor depends on which model is monitored versus which other model; for instance, the Bayes' factor for Model $\mathcal{M}_1$ vs. $\mathcal{M}_0$ is simply $1/B_k$. The Bayes' factor B_k can be useful to detect a significant decrease in predictive performance, however, it might be more difficult to understand when corrective measures should be applied. For this purpose, we introduce a slightly more general version of it as follows

$$B_{k,\delta} = \prod_{i=k-\delta+1}^{k} B_i = \frac{p^{(0)}_{\boldsymbol{y}_{k-\delta+1:k} | \boldsymbol{y}_{0:k-\delta}}(y_{k-\delta+1:k} \mid y_{0:k-\delta})}{p^{(1)}_{\boldsymbol{y}_{k-\delta+1:k} | \boldsymbol{y}_{0:k-\delta}}(y_{k-\delta+1:k} \mid y_{0:k-\delta})}$$

for some lag $\delta \in \{1, \ldots, k+1\}$, and we note that $B_{k,1} = B_k$ and $B_{k,\delta} = B_k B_{k-1,\delta-1}$ when $\delta > 1$. This does not completely address the difficulties raised earlier since fixing δ to some value will only allow for detecting discrepancies between Model $\mathcal{M}_0$ and the received observations that are consistent in duration with the lag δ. Since we are interested in detecting when $\mathcal{M}_1$ is a better model than $\mathcal{M}_0$, even locally, we can consider a measure that gives an advantage to $\mathcal{M}_1$ such that

$$L_k = \min_{1 \leq \delta \leq k} B_{k,\delta}$$

which minimises the evidence in favour of $\mathcal{M}_0$. The associated quantity which answers the question about when to apply corrective measures is the optimum lag

$$\delta^*_k = \operatorname*{argmin}_{1 \leq \delta \leq k} B_{k,\delta},$$

so that, in particular, it holds that $L_k = B_{k,\delta^*_k}$. The lag δ^*_k is called the *run-length at* k and gives the number of consecutive observations contributing against $\mathcal{M}_0$. It indicates that the intervention should be implemented at time step $k - \delta^*_k + 1$. The following theorem shows how these two quantities can be computed recursively.

Theorem 4.1. *For any* $k > 1$, *it holds that*

$$L_k = B_k \min\{1, L_{k-1}\}$$

and

$$\delta^*_k = \begin{cases} 1 + \delta^*_{k-1} & \text{if } L_{k-1} < 1 \\ 1 & \text{otherwise.} \end{cases}$$

The recursion for the run-length can be interpreted as follows: if the evidence favours $\mathcal{M}_0$ at time step $k-1$, i.e. $L_{k-1} \geq 1$, then we focus on the Bayes' factor at time step k, otherwise we consider the accumulated evidence against $\mathcal{M}_0$. In practice, one can define a threshold τ, e.g. $\tau = 0.1$, and trigger the intervention as soon as $L_k < \tau$.

Figure 4.1 illustrates the use of the Bayes' factor for assessing a constant velocity model $\mathcal{M}_0$ with a Brownian motion $\mathcal{M}_1$ as a reference, that is a 1-dimensional first order model.

4.2.2 Error analysis

A simpler approach that is more specific to DLMs is the monitoring of the consistency between the observations and the forecast distribution. The error term to consider is

$$\boldsymbol{z}_k = \boldsymbol{y}_k - H_k m_k.$$

Two important cases can be identified:

1. Known observational variance, in which case the error $\boldsymbol{z}_k$ is characterised by the normal distribution $\mathrm{N}(\cdot\,; 0, H_k P_k H_k^{\intercal} + V_k)$.
2. Unknown observational variance, in which case the error $\boldsymbol{z}_k$ has the generalised multivariate Student's t distribution $\mathrm{St}(\cdot\,; n_k, 0, \tilde{S}_k)$ with n_k and $\tilde{S}_k$ the corresponding number of degrees of freedom and shape matrix at time step k, respectively.

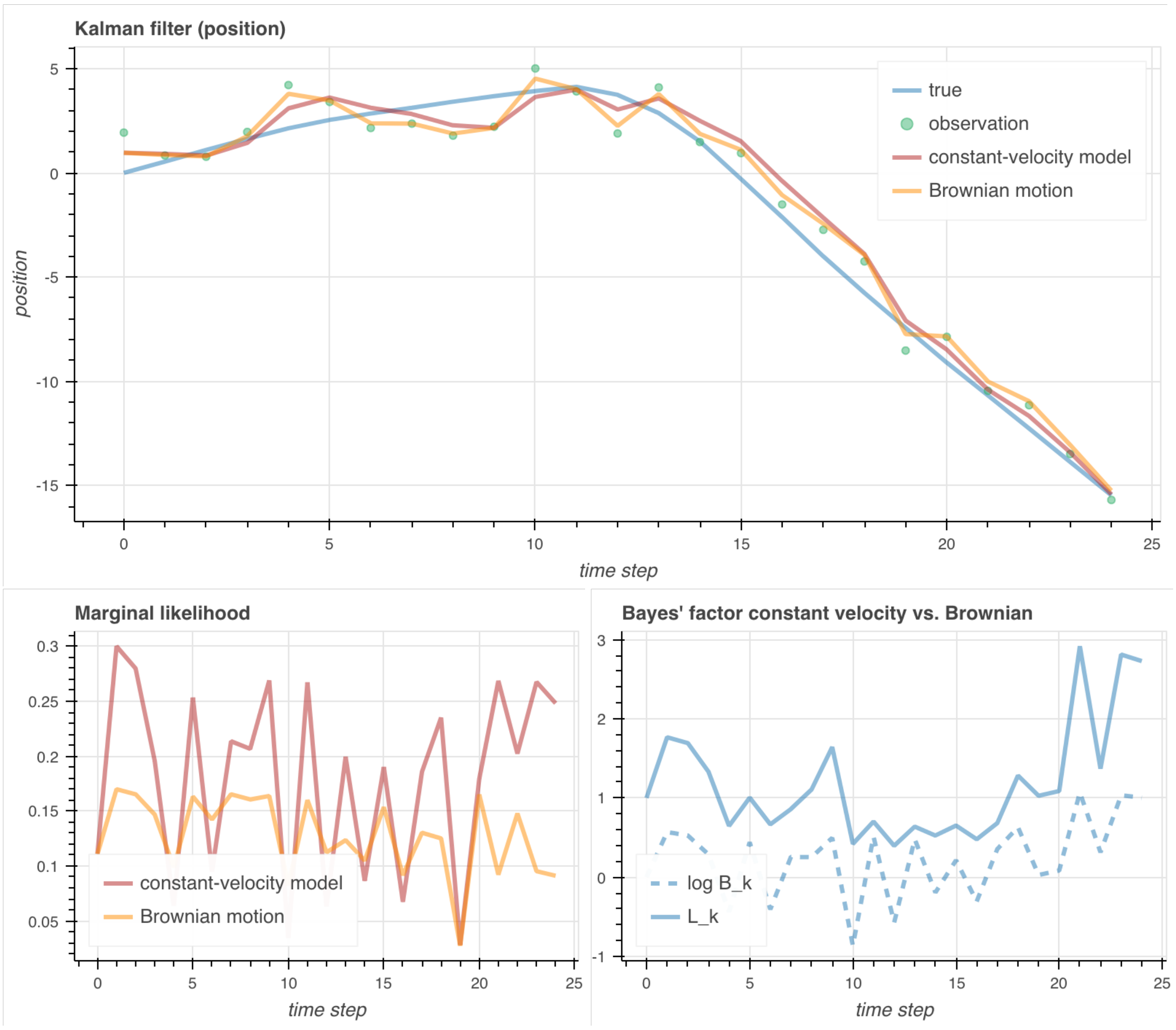

Figure 4.1: Model checking for a constant velocity model with a Brownian motion as a reference.