Appendix B: State Space Models

Contents

UBC Master of Data Science

DSCI 574
Spatial & Temporal Models

1. Introduction

This appendix builds on material presented on exponential smoothing algorithms in Lecture 2. The algorithms we looked at included simple exponential smoothing (SES), Holt's method, and Holt-Winter's method. These are all algorithms for generating point forecasts. It is desireable to extend these algorithms to statistical models that model distributions rather than points, thus giving us the ability to produce intervals around our forecasts.

We can reformulate the above algorithms as state space models. These state space models give the same point forecasts as the underlying algorithms, but are flexible frameworks that provide the added benefit of providing distributions (prediction intervals) and maximum likelihood estimation, amongst other benefits. The text Forecasting with Exponential Smoothing is an excellent resource for learning more about the state space approach to exponential smoothing. The details are well beyond the scope of this course and are not important for you to know, but I'll provide a quick summary in this appendix.

2. State space models

A state space model is a generic framework for modelling a dynamic system. The basic premise is that we can model the system based on two equations:

1. The observation equation - $y_t = Ax_{t-1} + w_t$ (describes the relationship between the observation y_t and the system's states)
2. The state equation - $x_t = Bx_{t-1} + Cw_t$ (describes the evolution of the system's states)

An example in the context of our exponential smoothing models might help. Recall that exponential smoothing models aim to forecast a series based on some combination of historically-weighted level, trend, and seasonal components. We can think of the level, trend, and seasonal components as the states of the system. They evolve over time as some function of their previous state plus prediction error. A forecast is produced as a combination of the states, and because our model is not perfect, an observed value can be defined as the forecast (a function of the states) plus some error. I think that seeing a worked example will help drive this point home.

3. Deriving exponential smoothing state space models

Recall from [Lecture 2]() that the simple exponential smoothing algorithm is:

$$\hat{y}_{t+1|t} = \alpha y_t + (1-\alpha)\hat{y}_{t|t-1}$$

We can re-write that into the following form of two component equations:

1. Forecast equation:

$$\hat{y}_{t+1|t} = \ell_t$$

2. Smoothing equation:

$$
\begin{aligned}
\ell_t &= \alpha y_t + (1-\alpha)\ell_{t-1} \\
&= \alpha y_t + \ell_{t-1} - \alpha\ell_{t-1} \\
&= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\
&= \ell_{t-1} + \alpha(y_t - \hat{y}_{t|t-1}) \\
&= \ell_{t-1} + \alpha e_t
\end{aligned}
$$

Where ℓ_t is called the "level" of the series and e_t is the residual at time t. While it looks like there's a lot going on there, the equation is fairly simple to explain in words. The level of the series at time t is simply the previous level adjusted by α multiplied by the previous error. For example, if e_t is positive, that means the level ℓ_{t-1} under-estimated the true value y_t, and so the new level ℓ_t will be adjusted in a positive direction by αe_t (i.e., $\ell_t > \ell_{t-1}$).

To account for the fact that our algorithm is not going to make perfect predictions, we can specify that the observed value can be represented by the previous level plus an error:

$$y_t = \ell_{t-1} + e_t$$

We'll call this the "observation equation". Sounds familiar right? Well we now almost have enough enough information to formulate our state space model. The final thing we need to do is assume a probability distribution for the errors - usually we assume the errors are Gaussian white noise (w_t), i.e., normal and i.i.d. Our two state space equations are therefore:

1. Observation equation:

$$y_t = \ell_{t-1} + w_t$$

2. State equation:

$$\ell_t = \ell_{t-1} + \alpha w_t$$

Okay so what? We re-formulated our algorithm into a state space model, what's the big deal? It doesn't look that different to our original algorithm. Well, that's exactly the point! Our state space model retains the "intuitiveness" of our original algorithm, but we also now have access to maximum likelihood estimation of model parameters, and the state space framework allows us to build produce prediction intervals (I'll talk about these more in the next section). If you're interested in the derivation of state space model for all exponential smoothing methods, check out Section 2.5 of [Forecasting with Exponential Smoothing](#).

4. Maximum likelihood estimation and prediction intervals

The parameters of our statistical state space models can be estimated via maximum likelihood. If you're interested in the details, take a look at Chapter 5 of Forecasting with Exponential Smoothing). MLE provides a more flexible approach to estimating model parameters than simple minimzing the sum of squared errors as we did with our exponential smoothing algorithms. Although, I'll note that MLE gives the same results as minimzing the SSE for several of the ETS models assuming normally distributed errors. That assumption of normality is discussed in detail in Chapter 15 of Forecasting with Exponential Smoothing but for the most part, it result in a "reasonable framework for parameter estimation".

Apart from MLE, the key advantage of the ETS framework (relevant to our studies) is the ability to produce prediction intervals. For many ETS models, there are analytical solutions for calculating prediction intervals, you can read more about those solutions in Chapter 6 of Forecasting with Exponential Smoothing.