Module 6 - Linear Innovations State Space Models

MATH1307 Forecasting

Prepared by: Dr. Haydar Demirhan based on the textbook by Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

Introduction

Six out of the fifteen exponential smoothing methods were involving linear relationships. So, these models are **linear innovations state-space models**. We will consider those six models and any other models of the same form. We will create a general framework for the estimation and prediction of the models, which will handle the six models and others.

We will cover

- the generalisation of state-space model for all exponential smoothing methods,
- model selection,
- the general linear innovations state-space model,
- a simple algorithm for computing the one-step prediction errors (or innovations),
- the properties of the models, including stationarity and stability,
- some basic innovations state-space models, and
- variations of the basic innovations state-space models.

Generalisation of State Space Model for All Exponential Smoothing Methods

The general model involves a state vector $X_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})'$ and state space equations of the form

$$Y_t \quad = w(X_{t-1}) + r(X_{t-1})\epsilon_t,$$
$$X_t \quad = f(X_{t-1}) + g(X_{t-1})\epsilon_t.$$

where $\{\epsilon_t\}$ is a white noise process with $N(o, \sigma^2)$ and $\mu_t = w(X_{t-1})$. We have the following generalizations:

- The model with additive errors has $r(X_{t-1}) = 1$, so that $Y_t = \mu_t + \epsilon_t$.
- The model with multiplicative errors has $r(X_{t-1}) = \mu_t$, so that $Y_t = \mu_t(1 + \epsilon_t)$.

Thus, the relative error for the multiplicative model is $\epsilon_t = (Y_t - \mu_t)/\mu_t$. Any value of $r(X_{t-1})$ yields identical point forecasts for Y_t.

All the models for the methods considered in Table 1 are given in Table 2.1. of the Textbook shown below (Hyndman et al., 2008, p. 18)

Table 2.1. Formulae for recursive calculations and point forecasts.

Trend	Seasonal		
	N	A	M
N	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$
A	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t + hb_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$
A_d	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$
M	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$
M_d	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^\phi$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^\phi) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^\phi$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^\phi$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^\phi)) + (1-\gamma)s_{t-m}$ $\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$

In each case, ℓ_t denotes the series level at time t, b_t denotes the slope at time t, s_t denotes the seasonal component of the series at time t, and m denotes the number of seasons in a year; α, β^*, γ and ϕ are constants, $\phi_h = \phi + \phi^2 + \cdots + \phi^h$ and $h_m^+ = [(h-1) \bmod m] + 1$.

* Taken from: Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

So, any of the methods in Table 2.1 of the Textbook can be formulated with equations in (19). Multiplicative error models are given in Table 2.2 of the Textbook shown below (Hyndman et al., 2008, p. 21)

Table 2.2. State space equations for each additive error model in the classification.

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$\mu_t = \ell_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$\mu_t = \ell_{t-1} + b_{t-1}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1} + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M_d	$\mu_t = \ell_{t-1}b_{t-1}^\phi$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$\mu_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

* Taken from: Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

The resulting multiplicative error equations are given in Table 2.3 of the Textbook shown below (Hyndman et al., 2008, p. 22)

Table 2.3. State space equations for each multiplicative error model in the classification.

Trend	Seasonal		
	N	**A**	**M**
N	$\mu_t = \ell_{t-1}$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$	$\mu_t = \ell_{t-1}+s_{t-m}$ $\ell_t = \ell_{t-1}+\alpha(\ell_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}(1+\alpha\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
A	$\mu_t = \ell_{t-1}+b_{t-1}$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1}+b_{t-1}+s_{t-m}$ $\ell_t = \ell_{t-1}+b_{t-1}+\alpha(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+b_{t-1}+s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1}+b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1}+b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}+\beta(\ell_{t-1}+b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
A$_d$	$\mu_t = \ell_{t-1}+\phi b_{t-1}$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$	$\mu_t = \ell_{t-1}+\phi b_{t-1}+s_{t-m}$ $\ell_t = \ell_{t-1}+\phi b_{t-1}+\alpha(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$ $s_t = s_{t-m}+\gamma(\ell_{t-1}+\phi b_{t-1}+s_{t-m})\varepsilon_t$	$\mu_t = (\ell_{t-1}+\phi b_{t-1})s_{t-m}$ $\ell_t = (\ell_{t-1}+\phi b_{t-1})(1+\alpha\varepsilon_t)$ $b_t = \phi b_{t-1}+\beta(\ell_{t-1}+\phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
M	$\mu_t = \ell_{t-1}b_{t-1}$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1}+s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}+\alpha(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}+\beta(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}+s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$
M$_d$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}+s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}+\alpha(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi}+\beta(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m}+\gamma(\ell_{t-1}b_{t-1}^{\phi}+s_{t-m})\varepsilon_t$	$\mu_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1+\alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1+\beta\varepsilon_t)$ $s_t = s_{t-m}(1+\gamma\varepsilon_t)$

* Taken from: Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

> Some of the combinations of trend, seasonality and error can occasionally lead to numerical difficulties; specifically, any model equation that requires division by a state component could involve division by zero. This is a problem for models with additive errors and either multiplicative trend or multiplicative seasonality, as well as the model with multiplicative errors, multiplicative trend and additive seasonality. These models should therefore be used with caution.

Notice that if the time series is not strictly positive, only the six fully additive models may be applied.

Model Selection

The forecast accuracy measures can be used to select a model from a set of candidate models, provided the errors are computed from data in a hold-out set and not from the same data as were used for model estimation. However, there are often too few out-of-sample errors to draw reliable conclusions. Consequently, a penalized method based on in-sample fit is usually better. Frequently used penalized measures are AIC, its corrected version AICc, and BIC. We can use AIC and BIC for the models with additive and multiplicative errors.

Using the forecast accuracy measures and penalized measures, we can create a model selection algorithm which will run automatically. Out Textbook gives the following algorithm for auto-exponential smoothing:

> 1. For each series, apply all models that are appropriate, optimizing the parameters of the model in each case.
> 2. Select the best of the models according to the AIC.
> 3. Producepointforecastsusingthebestmodel(with optimized parameters) for as many steps ahead as required.
> 4. Obtain prediction intervals for the best model either using the analytical results or by simulating future sample paths for $\{Y_{n+1}, \ldots, Y_{n+h}\}$ and finding the $\alpha/2$ and $1-\alpha/2$ percentiles of the simulated data at each forecasting horizon. If the simulation is used, the sample paths may be generated using the Gaussian distribution for errors (parametric bootstrap) or using the resampled errors (ordinary bootstrap).

Implementation of this algorithm is done by the `ets()` (https://www.rdocumentation.org/packages/forecast/versions/7.3/topics/ets) function of the `forecast` package. To trigger auto-search, we set the argument `model` to `"zzz"`. We can choose the penalised measure to use for auto-selection using the argument `ic=c("aicc","aic","bic")`.

We will illustrate the implementation of this algorithm over monthly US 10-year bond yields series, annual US net electricity generation series, quarterly UK passenger vehicle production series and monthly Australian overseas visitors series.

```
fit.auto.bonds = ets(bonds,model="ZZZ",ic="bic") # Use BIC
fit.auto.bonds$method # Simple exponential smoothing with additive errors is selecte
  d
```

```
## [1] "ETS(A,N,N)"
```

```
fit.auto.elec = ets(usnetelec,model="ZZZ",ic="aic") # Use AIC
fit.auto.elec$method # Holt's linear model with multiple errors is selected
```

```
## [1] "ETS(M,A,N)"
```

```
fit.auto.cars = ets(ukcars,model="ZZZ",ic="aicc") # Use AICc
fit.auto.cars$method # The method includes not trend component and additive seasonal
  component with additive errors is selected
```

```
## [1] "ETS(A,N,A)"
```

```
fit.auto.visitors = ets(visitors,model="ZZZ") # Use the default
fit.auto.visitors$method # Holt-Winters' multiplicative method with multiple errors
  is selected
```

```
## [1] "ETS(M,A,M)"
```

We can force the auto-search algorithm to consider, for example, damped trend models by setting the `damped=TRUE`.

```
fit.auto.damped.cars = ets(ukcars,model="ZZZ",damped = TRUE, ic="aicc") # Use AICc a
  nd restrict to damped trend models
fit.auto.damped.cars$method # The method includes a damped trend component and addit
  ive seasonal component with additive errors is selected
```

```
## [1] "ETS(A,Ad,A)"
```

The following visualisation displays the 5-time steps ahead forecasts for each series.

```
par(mfrow=c(2,2))
plot(forecast(fit.auto.bonds), ylab="US bond yields (percent per annum)",plot.conf=F
  ALSE, type="l", xlab="Year")
plot(forecast(fit.auto.elec), ylab="US electricity generation (billion kwh)",plot.co
  nf=FALSE, type="l", xlab="Year")
plot(forecast(fit.auto.cars), ylab="Vehicles in the UK (thousands of cars)",plot.con
  f=FALSE, type="l", xlab="Year")
plot(forecast(fit.auto.visitors), ylab="Number of visitors to Australia",plot.conf=F
  ALSE, type="l", xlab="Year")
```
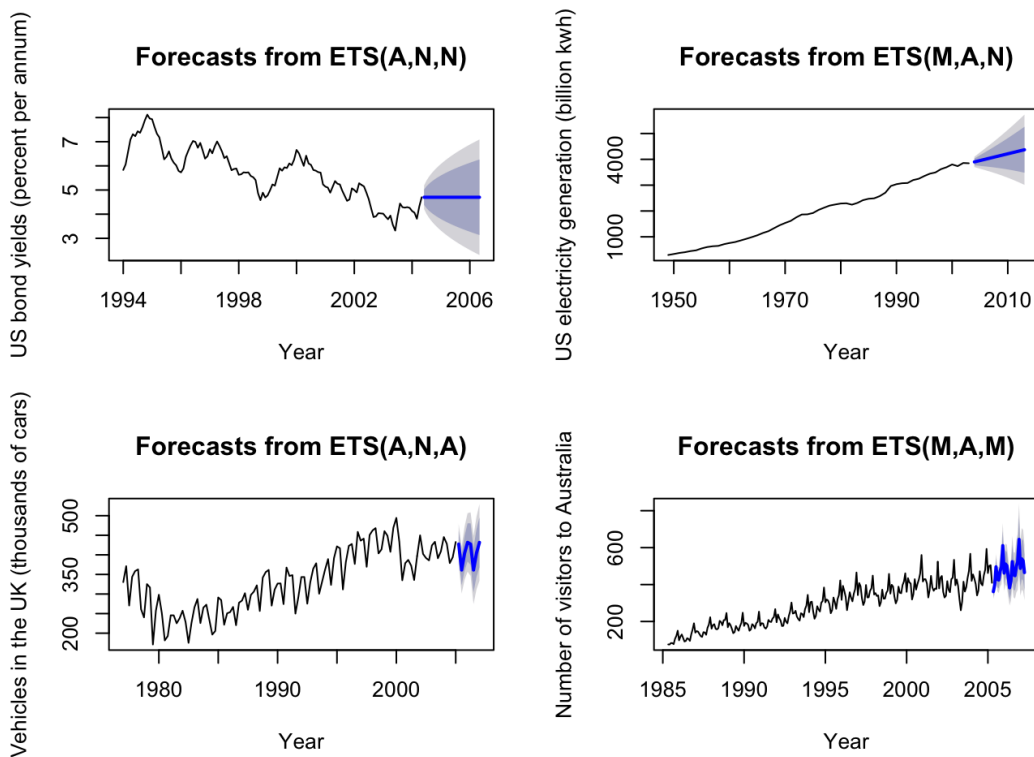
The General Linear Innovations State Space Model

The general characteristic of the state-space model is that the observed time series (Y_t) is supplemented by unobserved auxiliary variables called *states* represented by *state vector X_t*. The general1 linear innovations state-space model is written as

$$Y_t = \boldsymbol{\omega}' X_{t-1} + \epsilon_t,$$
$$X_t = \boldsymbol{F} X_{t-1} + \boldsymbol{g}\epsilon_t.$$

In exponential smoothing, the state vector contains information about the level, growth and seasonal patterns. The state variables help to define large complex models by first breaking them into smaller, more manageable parts, thus reducing the chance of model specification errors and the components of the state vector enable us to gain a better understanding of the structure of the series.

The first line of Eq. (1) is called the *measurement equation*. The term $\boldsymbol{\omega}' X_{t-1}$ describes the effect of the past on Y_t. It is usually assumed to be from a $N(0, \sigma^2)$ distribution, so it is also called *Gaussian innovations state space model*. ϵ_t represents what is new and unpredictable, and celled the *innovation*. The innovations are the only source of randomness for the observed time series.

The second line of Eq. (2) is called the *transition equation*. It is a first-order recurrence relationship that describes how the state vectors evolve over time. $\boldsymbol{F}$ is the *transition matrix* and $\boldsymbol{F} X_{t-1}$ shows the effect of past on the current state X_t. The term $\boldsymbol{g}\epsilon_t$ shows the unpredictable change in X_t. The vector $\boldsymbol{g}$ determines the extent of the effect of the innovation on the state. It is referred to as a *persistence vector*. The transition equation is the mechanism for creating the inter-temporal dependencies between the values of a time series.

The vectors $\boldsymbol{\omega}, \boldsymbol{g}$, and the matrix $\boldsymbol{F}$ all include the parameters to be estimated. The starting value X_0 for the transition equation may be fixed or random. If it must be random, the *infinite start-up assumption* applies. It is fixed the *finite start-up assumption* applies.

Innovations and One-Step-Ahead Forecasts

If the value for X_0 is known, the innovation ϵ_t corresponds to a one-step-ahead prediction error. If we denote the prediction by $\hat{Y}_t | t - 1$, the innovations can be computed recursively from the series values using the relationships

$$\hat{Y}_t | t - 1 = \omega' X_{t-1},$$
$$\epsilon_t = Y_t - \hat{Y}_t | t - 1,$$
$$X_t = F X_{t-1} + g\epsilon_t.$$

This transformation will be called *general exponential smoothing*. If we substitute the first and second lines of Eq. (2) into the last line, we get

$$X_t = D X_{t-1} + g Y_t,$$

where $D = F - g\omega'$. If we solve it back, we get

$$X_t = D^t X_0 + \sum_{j=0}^{t-1} D^j g Y_{t-j}.$$

If we substitute Eq. (4), lagged by one period, into the first line of Eq. (2), we get

$$\hat{Y}_t | t - 1 = a_t + \sum_{j=1}^{t-1} c_j Y_{t-j},$$

where $a_t = \omega' D^t X_0$ and $c_j = \omega' D^{j-1} g$ Thus, the forecast is a linear function of the past observations and the seed state vector.

The importance of these equations is that these equations demonstrate an important feature of the state space models.

> When a new observation becomes available, the state vector is updated using Eq. (3), and the new one-step-ahead forecast is immediately available.

Model Properties

Stability and Forecastability

As usual there are some condition for the stability and forecastability of the state space models. If the forecasts of Y_t are unaffected by observations in the distant past, we describe the model as *forecastable*. The following are the properties of a *forecastable* model:

As T approaches infinity, we get a constant

$$\sum_{j=1}^{\infty} |c_j| < \infty \text{ and } \lim_{t \leftarrow \infty} a_t = a.$$

The *stability* condition is that the eigenvalues of D lie inside the unit circle. In a stable model, the coefficients of the observations in (5) decay exponentially.

Stationarity

To get stationarity conditions, we can represent the first line of (2) as follows:

$$Y_t = d_t + \sum_{j=0}^{t-1} k_j \epsilon_{t-j},$$

where $d_t = \omega' F^{t-1} X_0$, $k_0 = 1$ and $k_j = \omega' F^{j-1} g$, for $j = 1, 2, \ldots$. Then, the model is *stationary_2* if

$$\sum_{j=0}^{\infty} |k_j| < \infty \text{ and } \lim_{t \leftarrow \infty} d_t = d.$$

Stationarity is a rare property in exponential smoothing state space models. None of the models discussed in Module 5 are stationary.

Basic Special Cases

The linear innovations state-space model effectively contains an infinite number of special cases that can potentially be used to model a time series. However, in practice, we use only a handful of special cases that possess the capacity to represent commonly occurring patterns such as trends, seasonality and business cycles.

The first special case to be considered, the *local level model*. In this model, at any point along the data path, the values in the neighbourhood of the point are approximated by a short flat line representing what is referred to as a local level. This case is illustrated in part (a) of the following figure.
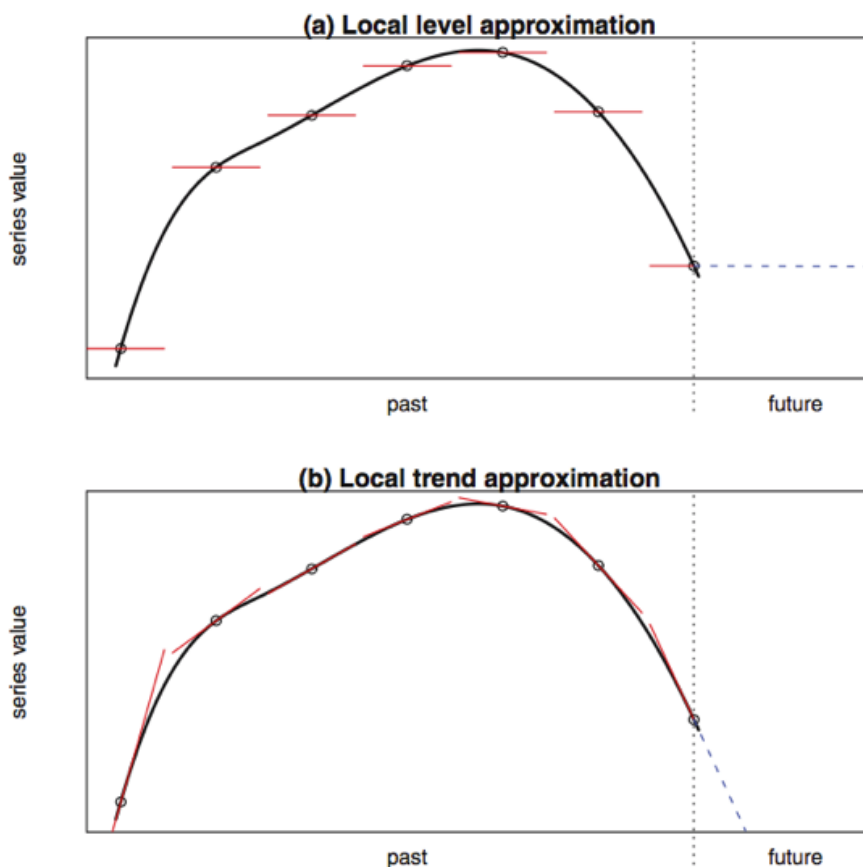
(a) Local level approximation

(b) Local trend approximation

* Taken from: Hyndman et al., Forecasting with Exponential Smoothing: The State Space Approach. Springer, 2008.

As its height changes over time, it is necessary to approximate the data path by many local levels. Thus, the local level effectively represents the state of a process generating a time series.

The gap between successive levels is treated as a Normal distributed random variable. The final local level is projected into the future to give predictions. As the approximation is only effective in a small neighbourhood, predictions generated this way are only likely to be reliable in the shorter term.

The second special case involves a *first-order polynomial approximation*. At each point, the data path is approximated by a tangential line in the deterministic case. In the stochastic case, it can only be said that the line has a similar height and a similar slope to the data path. Randomness means that the line is not exactly tangential. The approximating line changes over time, as seen in part (b) of the above figure, to reflect the changing shape of the data path. The state of the process is now summarized by the level and the slope at each point of the path. The stochastic representation is based on the assumption that the gaps between successive slopes are Normal distributed random variables with a zero mean. Note that the prediction is obtained by projecting the last linear approximation into the future.

Local Level Model: ETS(A,N,N)

The corresponding ETS(A,N,N) state space model is formulated as follows:

$$Y_t \quad = \ell_{t-1} + \epsilon_t,$$
$$\ell_t \quad = \ell_{t-1} + \alpha\epsilon_t,$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. When we set $X_t = \ell_t$, $\omega = 1$, $F = 1$, and $g = \alpha$, we obtain this model from the general form. Here the level is controlled by ℓ_t. Extreme values of α give special dependence cases.

- If $\alpha = 0$, the local levels do not change at all. Their common level is then referred to as the global level. Successive values of the series Y_t are independently and identically distributed. Its moments do not change over time.
- If $\alpha = 1$, the model reverts to a random walk $Y_t = Y_{t-1} + \epsilon_t$. Successive values of the time series Y_{t} are dependent.

This model corresponds to the simple exponential smoothing model with additive error terms. We use the `ets()` function to fit the model. Let's revisit the annual US new freight cars series to illustrate the model. We have fitted a simple exponential smoothing model to this series.

```
fit.ANN = ets(freight,model="ANN")
summary(fit.ANN)
```

```
## ETS(A,N,N)
##
## Call:
##   ets(y = freight, model = "ANN")
##
##    Smoothing parameters:
##      alpha = 0.4183
##
##    Initial states:
##      l = 3615.4904
##
##    sigma:  1451.738
##
##       AIC      AICc       BIC
## 869.2817 869.8398 874.8321
##
## Training set error measures:
##                      ME     RMSE     MAE       MPE     MAPE     MASE
## Training set -90.3855 1420.514 1123.33 -51.64303 81.33062 1.073404
##                    ACF1
## Training set 0.2182539
```

To compare the forecast performance of the ANN model to the simple exponential smoothing (SES) model, we can refer to RMSE and MASE values. We obtained RMSE=1634.026 and MASE=1.279911 from SES and RMSE = 1420.514 and MASE = 1.073496 from the ANN model. So, we can conclude that the inclusion of additive error terms improves the forecast performance of the model.

```
plot(forecast(fit.ANN), ylab="US new freight cars",plot.conf=FALSE, type="l",  xlab=
   "Year")
```

Forecasts from ETS(A,N,N)
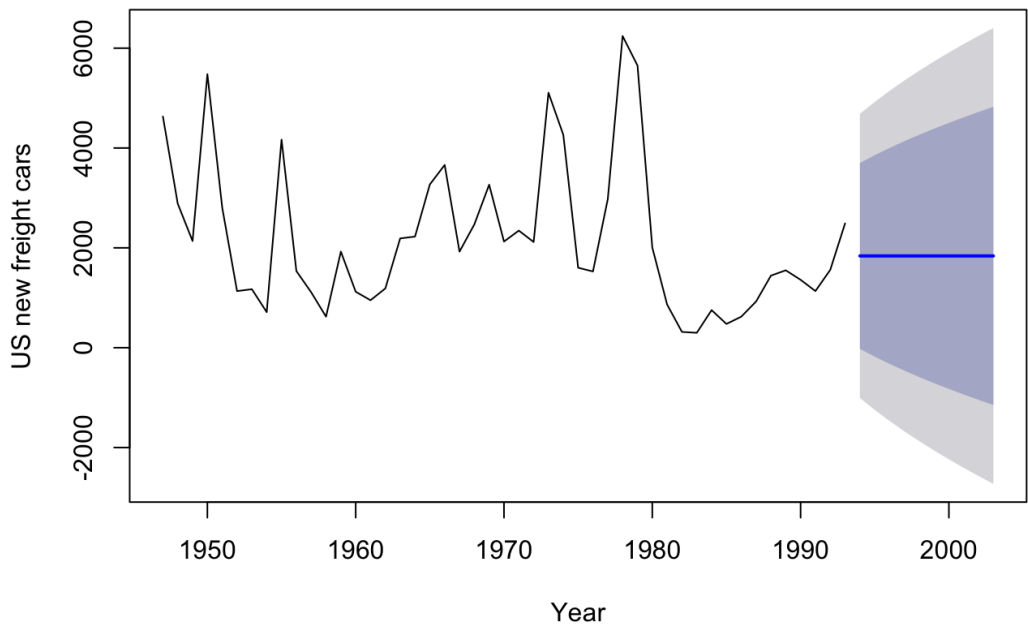
Forecasting is way too wide

Local Trend Model: ETS(A,A,N)

The local level model can be augmented by a growth rate of b_t to give

$$
\begin{aligned}
Y_t &= \ell_{t-1} + b_{t-1} + \epsilon_t, \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\epsilon_t, \\
b_t &= b_{t-1} + \beta\epsilon_t.
\end{aligned}
$$

In this model, we have two smoothing parameters α and β. The state space structure yielding this model is set by

$$
X_t = \begin{bmatrix} \ell_t & b_t \end{bmatrix}', \, \omega \begin{bmatrix} 1 & 1 \end{bmatrix}', \, F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}' \text{ and } g = \begin{bmatrix} \alpha & \beta \end{bmatrix}'.
$$

The size of the smoothing parameters reflects the impact of the innovations on the level and growth rate. As the smoothing parameters increase in size, there is a tendency for the series to become smoother.

This model is Holt's linear model with additive errors. We have fitted Holt's linear model to the annual US net electricity generation series. Let's see the effect of the inclusion of additive error terms in the model. We use the `ets()` function to fit the model.

```
fit.AAN = ets(usnetelec, model="AAN")
summary(fit.AAN)
```

```
## ETS(A,A,N)
##
## Call:
##   ets(y = usnetelec, model = "AAN")
##
##     Smoothing parameters:
##       alpha = 0.9985
##       beta  = 1e-04
##
##     Initial states:
##       l = 244.2092
##       b = 65.5768
##
##     sigma:  51.8636
##
##        AIC      AICc       BIC
## 660.5982 661.8227 670.6349
##
## Training set error measures:
##                            ME      RMSE      MAE       MPE      MAPE      MASE
## Training set -0.04060982 49.94203 36.44997 -0.8111995 2.440206 0.5164647
##                           ACF1
## Training set 0.1263413
```

```
fit.AAN.damped = ets(usnetelec, model="AAN",damped = TRUE)
summary(fit.AAN.damped)
```

```
## ETS(A,Ad,N)
##
## Call:
##   ets(y = usnetelec, model = "AAN", damped = TRUE)
##
##     Smoothing parameters:
##       alpha = 0.9999
##       beta  = 0.1565
##       phi   = 0.98
##
##     Initial states:
##       l = 247.3489
##       b = 60.1566
##
##     sigma:  54.3966
##
##        AIC      AICc       BIC
## 666.7543 668.5043 678.7983
##
## Training set error measures:
##                         ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 6.679144 51.86508 38.11248 0.238989 2.244731 0.540021
##                           ACF1
## Training set 0.04500634
```

The following table shows the values of model selection criteria for Holt's linear model and AAN model with and without a damped trend.

Model Comparison for Holt's method				
Model	Trend	RMSE	MASE	AIC
Holt's linear		50.05715	0.5201558	660.8515
Holt's linear	Damped	51.84055	0.5378807	666.7023
AAN		50.10032	0.5204496	660.9463
AAN	Damped	51.84056	0.5378808	666.7023

According to all criteria, Holt's linear model is slightly better than the one with additive error terms. The following displays the forecast and fits from four candidate models.

```
fit11 <- holt(usnetelec, h=5) # Holt's linear model
fit12 <- holt(usnetelec, h=5,damped = TRUE) # Holt's linear model with damped trend
plot(forecast(fit.AAN,h=5), type="l", ylab="US net electricity generation (billion k
  wh)", xlab="Year",fcol="cyan", plot.conf=FALSE)
lines(fitted(fit11), col="blue")
lines(fitted(fit12), col="red")
lines(fitted(fit.AAN.damped), col="green")
lines(fit11$mean, col="blue", type="l")
lines(fit12$mean, col="red", type="l")
lines(forecast(fit.AAN.damped,h=5)$mean, col="green", type="l")
legend("topleft", lty=1, col=c("black","blue","red","cyan","green"),
       c("Data","Holt's linear trend","Holt's linear damped trend","AAN", "AAN dampe
  d trend"))
```


Forecasts from ETS(A,A,N)
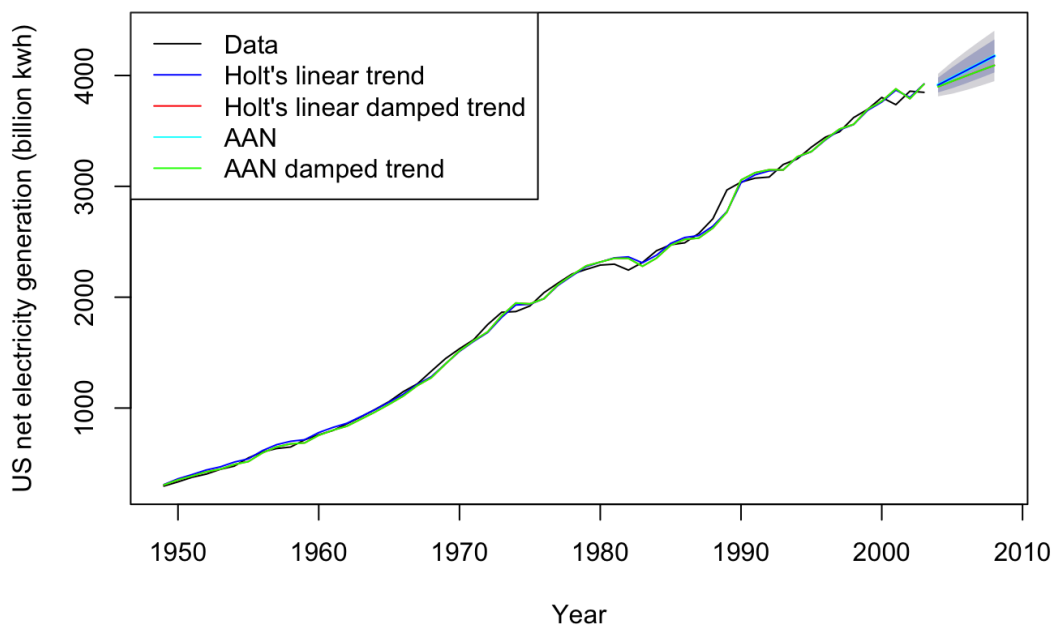
Local Additive Seasonal Model: ETS(A,A,A)

For time series that exhibit seasonal patterns, the local trend model can be augmented by seasonal effects, denoted by s_t. Often the structure of the seasonal pattern changes over time in response to changes in tastes and technology. For example, electricity demand used to peak in winter, but in some locations, it now peaks in summer due to the growing prevalence of air conditioning. Thus, the formulae used to represent the seasonal effects should allow for the possibility of changing seasonal patterns. The ETS(A,A,A) model is

$$\begin{aligned}
Y_t &= \ell_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t, \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\epsilon_t, \\
b_t &= b_{t-1} + \beta\epsilon_t, \\
s_t &= s_{t-m} + \gamma\epsilon_t.
\end{aligned}$$

This model corresponds to the general state-space model with

$$\omega = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 1 \end{bmatrix}',$$

$$X_t = \begin{bmatrix} \ell_t \\ b_t \\ s_{t-1} \\ s_{t-2} \\ \vdots \\ s_{t-m+1} \end{bmatrix}, \quad
F = \begin{bmatrix}
1 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & 1 \\
0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix} \quad \text{and } g = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This model corresponds to the Holt-Winters' additive exponential smoothing model with additive errors. Let's revisit the quarterly motor vehicle production in the UK series to illustrate the model. We use the `ets()` function to fit the model.

```
fit.AAA = ets(ukcars, model="AAA")
summary(fit.AAA)
```

```
## ETS(A,Ad,A)
##
## Call:
##   ets(y = ukcars, model = "AAA")
##
##    Smoothing parameters:
##      alpha = 0.5814
##      beta  = 1e-04
##      gamma = 1e-04
##      phi   = 0.9284
##
##    Initial states:
##      l = 343.6012
##      b = -5.3444
##      s = -1.1652 -45.1153 21.2507 25.0298
##
##    sigma:  26.2512
##
##        AIC      AICc       BIC
## 1283.319 1285.476 1310.593
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 2.009896 25.18409 20.44382 0.10939 6.683841 0.6662543
##                    ACF1
## Training set 0.03323651
```

```
fit.AAA.damped = ets(ukcars, model="AAA",damped = TRUE)
summary(fit.AAA.damped)
```

```
## ETS(A,Ad,A)
##
## Call:
##  ets(y = ukcars, model = "AAA", damped = TRUE)
##
##   Smoothing parameters:
##     alpha = 0.5814
##     beta  = 1e-04
##     gamma = 1e-04
##     phi   = 0.9284
##
##   Initial states:
##     l = 343.6012
##     b = -5.3444
##     s = -1.1652 -45.1153 21.2507 25.0298
##
##   sigma:  26.2512
##
##      AIC     AICc      BIC
## 1283.319 1285.476 1310.593
##
## Training set error measures:
##                    ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 2.009896 25.18409 20.44382 0.10939 6.683841 0.6662543
##                   ACF1
## Training set 0.03323651
```

The following table shows the values of model selection criteria for Holt-Winters' additive model and AAA model with and without a damped trend.

Model Comparison for Holt Winters' method

Model	Trend	RMSE	MASE	AIC
Holt-Winters' additive		25.63276	0.67813	1285.31
Holt-Winters' additive	Damped	25.16871	0.6689918	1283.181
AAA		25.16871	0.6689815	1283.181
AAA	Damped	25.16871	0.6689815	1283.181

According to all criteria, all the candidate models are very very close to each other in terms of predictive performance. Holt-Winters' additive model with additive errors is so slightly better than the others. The following displays the forecast and fits from four candidate models.

```
fit21 <- hw(ukcars,seasonal="additive", h=5*frequency(ukcars))
fit22 <- hw(ukcars,seasonal="additive",damped = TRUE, h=5*frequency(ukcars))
plot(forecast(fit.AAA,h=5), type="l", ylab="UK motor vehicles (thousands of cars)",
  xlab="Year",fcol="cyan", plot.conf=FALSE)
lines(fitted(fit21), col="blue")
lines(fitted(fit.AAA.damped), col="green")
lines(fitted(fit22), col="red")
lines(fit21$mean, col="blue", type="l")
lines(fit22$mean, col="red", type="l")
lines(forecast(fit.AAA.damped,h=5)$mean, col="green", type="l")
legend("topleft", lty=1, col=c("black","blue","red","cyan","green"),
      c("Data","Holt-Winters' additive","Holt-Winters' additive damped trend","AAA"
  , "AAA damped trend"))
```

Forecasts from ETS(A,Ad,A)
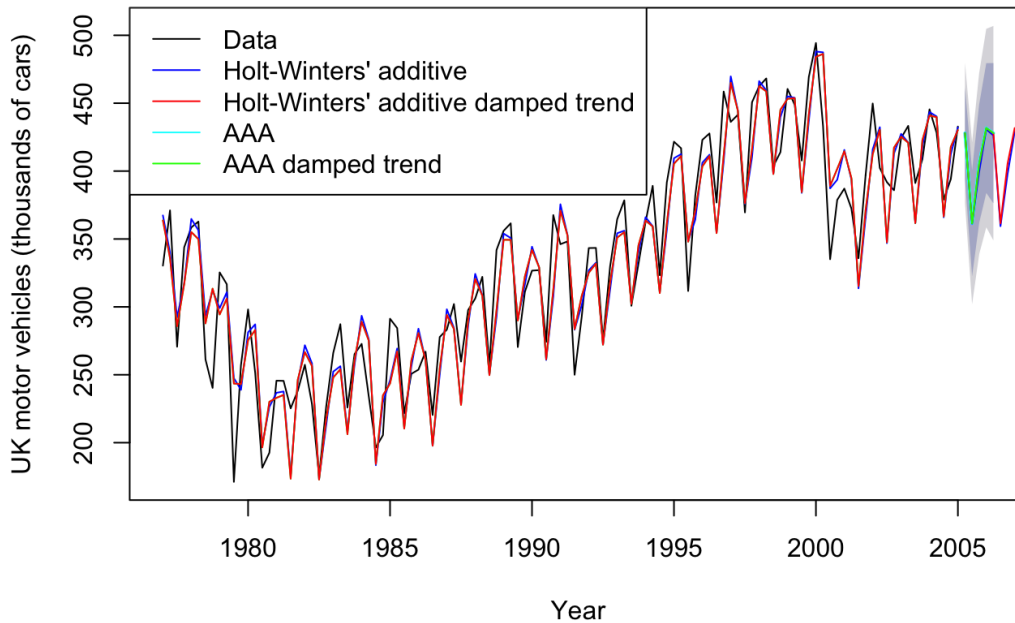
Variations on the Common Models

We will study some variations of the model discussed in the previous section.

Local Level Model with Drift

A local trend model allows the growth rate to change stochastically over time. If $\beta = 0$, the growth rate is constant and equal to a value that will be denoted by b. The local level model then reduces to

$$Y_t = \ell_{t-1} + b + \epsilon_t,$$
$$\ell_t = \ell_{t-1} + b + \alpha\epsilon_t,$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. This model is called *local level model with drift* and originates to the state space structure with

$$X_t = \begin{bmatrix} \ell_t & b \end{bmatrix}', \boldsymbol{\omega}\begin{bmatrix} 1 & 1 \end{bmatrix}', F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}' \text{ and } \boldsymbol{g} = \begin{bmatrix} \alpha & 0 \end{bmatrix}'.$$

This model can be applied to economic time series that display an upward (or downward) drift. The local level model with drift is also known as *simple exponential smoothing with drift*. We use `ets()` function with `beta=0` to fit this model.

Let's fit this model to the annual US net electricity generation (billion kWh) series. We set the argument `beta` to a value very close to zero.

```
fit.drift = ets(usnetelec, model="AAN", beta = 1E-4)
summary(fit.drift)
```

```
## ETS(A,A,N)
##
## Call:
##   ets(y = usnetelec, model = "AAN", beta = 1e-04)
##
##   Smoothing parameters:
##     alpha = 0.9999
##     beta  = 1e-04
##
##   Initial states:
##     l = 246.48
##     b = 65.974
##
##   sigma:  51.8699
##
##      AIC      AICc       BIC
## 658.6116 659.4116 666.6409
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -0.4796838 49.94809 36.48702 -0.858204 2.465155 0.5169896
##                  ACF1
## Training set 0.1254151
```

```
plot(forecast(fit.drift,h=5), type="l", ylab="US electricity generation (billion kw
  h)", xlab="Year", plot.conf=FALSE)
```


Forecasts from ETS(A,A,N)
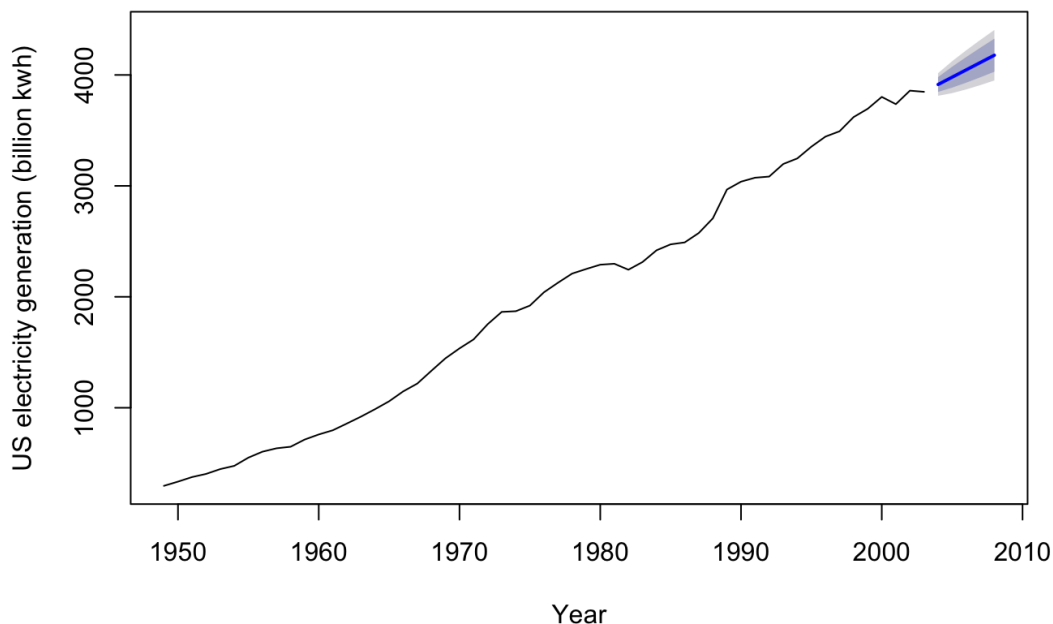
Damped Trend Model: ETS(A,Ad,N)

When we take the local trend model and dampen its growth rate with a factor ϕ in the region $0 \le \phi < 1$. The resulting model is

$$
\begin{aligned}
Y_t &= \ell_{t-1} + \phi b_{t-1} + \epsilon_t, \\
\ell_t &= \ell_{t-1} + phib_{t-1} + \alpha\epsilon_t, \\
b_t &= phib_{t-1} + \beta\epsilon_t.
\end{aligned}
$$

The characteristics of the damped local trend model are compatible with features observed in many business and economic time series. It sometimes yields better forecasts than the local trend model. Note that the local trend model is a special case where $\phi = 1$. The ETS(A,Ad,N) model performs remarkably well when forecasting real data.

We will fit this model to US net electricity generation (billion kWh) series as well.

```
fit.AAdN = ets(usnetelec, model="AAN", damped = TRUE)
summary(fit.AAdN)
```

```
## ETS(A,Ad,N)
##
## Call:
##   ets(y = usnetelec, model = "AAN", damped = TRUE)
##
##    Smoothing parameters:
##      alpha = 0.9999
##      beta  = 0.1565
##      phi   = 0.98
##
##    Initial states:
##      l = 247.3489
##      b = 60.1566
##
##    sigma:  54.3966
##
##       AIC      AICc       BIC
## 666.7543 668.5043 678.7983
##
## Training set error measures:
##                     ME     RMSE      MAE      MPE     MAPE      MASE
## Training set 6.679144 51.86508 38.11248 0.238989 2.244731 0.540021
##                    ACF1
## Training set 0.04500634
```

```
plot(forecast(fit.drift,h=5), type="l", ylab="US electricity generation (billion kw
  h)", xlab="Year", plot.conf=FALSE)
lines(fitted(fit.drift), col="blue")
lines(fitted(fit.AAdN), col="red")
lines(forecast(fit.AAdN,h=5)$mean, col="red", type="l")
legend("topleft", lty=1, col=c("black","blue","red"),
       c("Data","ES with drift","AAdN"))
```

Forecasts from ETS(A,A,N)
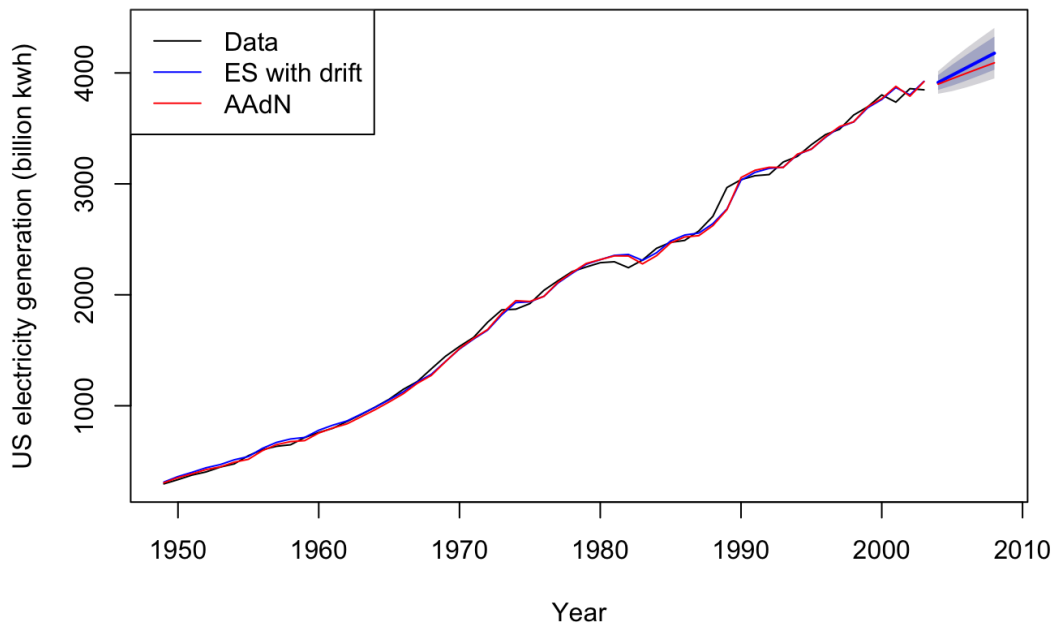
Exponential smoothing with drift model performs better for this series in terms of RMSE, MASA and AIC.

Summary

In this module, we introduced a general form for the state-space models and studied the state-space models in a comparative way to the exponential smoothing models. We illustrated the relationships between the state-space models and exponential smoothing methods in terms of their parameters.

References

Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D. (2008). Forecasting with exponential smoothing: the state space approach (http://www.exponentialsmoothing.net), Springer-Verlag.

Hyndman, R.J., Athanasopoulos, G. (2014). *Forecasting: Principles and Practice*. OTEXTS.