Definitions

- Times series plot is a scatterplot of the data with time on the x-axis in, typically, equally spaced intervals, and the observations on the y-axis
  - For visual clarity adjacent observations are connected by lines
  - The equally spaced intervals can be in units of years, months, days, hours,… and these units are called the period of the time series

- The level is the local mean of the observations, around which we see random noise
- When the level varies with time we say there is a trend.
- A seasonal effect is a systematic and calendar related effect which repeats with a given period

- Terms like level, trend and seasonality can only be defined precisely in terms of a modelling exercise.

- A change point is a time at which at least of the of the following changes
  - The data generation process
  - The way that the data is measured
  - The way that the observation is defined

- Stationarity informally means that the underlying random process does not change in time.
  - Seasonality, trends, non-constant variance and change points are all examples of non-stationarity.
  - If we have a stationary process we would be able to assume that, at least statistically, the future is similar to the past

Time Series Models

- Markov Chain is a form of stochastic iteration. It is a sequence X_i od dependent random variables where the distribution of X_{n+1} depends on X_n and only on X_n. Which means
$$P(X_{n+1}|X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1}|X_n = x_n)$$

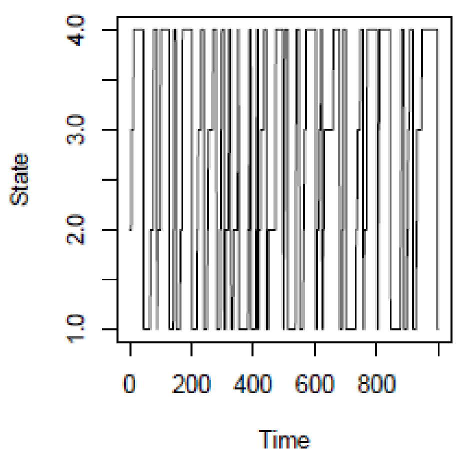
- If X_n takes values in the finite set S = {1, …, K}, it is a discrete space Markov chain

- If defines the K * K transition matrix, the n-step probabilities $p_n(i,j)$ are defined as
$$p_n(i,j) = P(X_n = j|X_0 = i) = P(X_{n+k} = j|X_k = i)$$

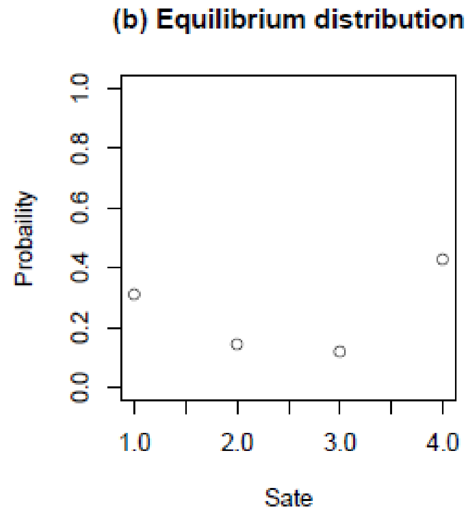
- For large n the n-step transition probability often converge to equilibrium distribution which is independent of the past

- The following figure shows a realization of a 4-state Markov chain which has the transition matrix
$$P = \begin{pmatrix} 0.9 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.8 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.9 & 0.1 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix}$$

(a) Realisation of Markov chain



- And the following is its equilibrium distribution



- (random walk) Consider a continuous state space Markov chain. Let Z_t, $t \in \mathbb{Z}$, be an i.i.d sequence of random variables. The series defined by

$$X_t := \sum_{i=1}^{t} Z_i$$

  - For t = 1, 2, …, is called a <u>random walk</u>

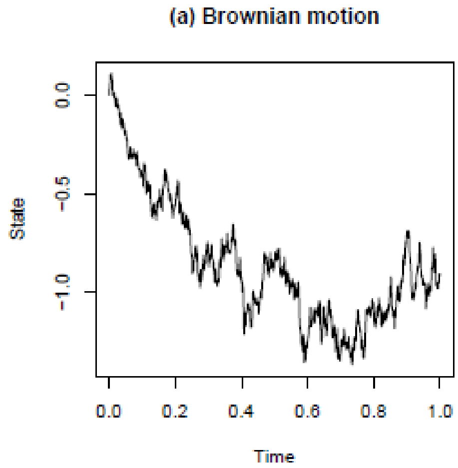
- A related Markov process has the following definition
-
- (Autoregressive model) Let Z_t, $t \in \mathbb{Z}$, be an i.i.d sequence of $N(0, \sigma^2)$ random variables. The series defined by

$$X_t = \phi X_{t-1} + Z_t$$

  - For all $t \in \mathbb{Z}$, is called a <u>first order autoregressive process (AR(1))</u> process.

- Note: not all time series have to be on discrete time points. If the time period of observation is arbitrary then we talk about <u>continuous time stochastic processes</u>

- (Brownian motion) A <u>Brownian motion</u> or Wiener process is a continuous-time stochastic process which is the unique process W_t which satisfies the following
  - $W_0 = 0$
  - The function $t \to W_t$ is almost surely everywhere continuous
  - The increments $W_{t_1} - W_{s_1} W_{t_2} - W_{s_2}$ are independent when $0 \le s_1 < t_1 \le s_2 < t_2$
  - The increment $W_t - W_s$ has a N (0, t-s) distribution for $0 \le s < t$
    - We can think of the Brownian motion as a limit of the random walk when time intervals shrink to zero, and we illustrate a realized path in the following figure



- (Non-constant variance process) if we want models where the underlying variance of the process changes with time then one possibility is to use an Auto-Regressive Conditional Heteroscedasticity (ARCH) model.
- We define the model by a hierarchically: first define $X_t = \sigma_t Z_t$ where $Z_t \sim N(0,1)\ i.i.d$, but treat σ as being random such that

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_q X_{t-q}^2$$

  - ➢ So the variance is time dependent – a large value of X_t will result in period of high volatility.
  - ➢ We illustrate an example of a realization in the following figure

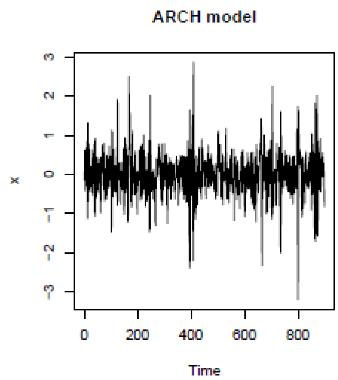


Figure 1.11: Realisation of ARCH process

- (Gaussian process) let T = ℝ, then $\{X_t\}$ is a discrete Gaussian process if any finite subset $\{X_{t_1}, \dots, X_{t_n}\}$ has an n-dimensional multivariate normal distribution.
  - ➢ This model is completely determined when the mean and variance-covariance structures are known
- The following figure shows three realizations from three different one dimensional Gaussian processes.
  - ➢ They all share the fact that the realization gives a continuous graph but they differ in the

“smoothness” of the realization.

✧ This smoothness is controlled by a single parameter ν and the amount of “smoothness” increases as we go from left to right

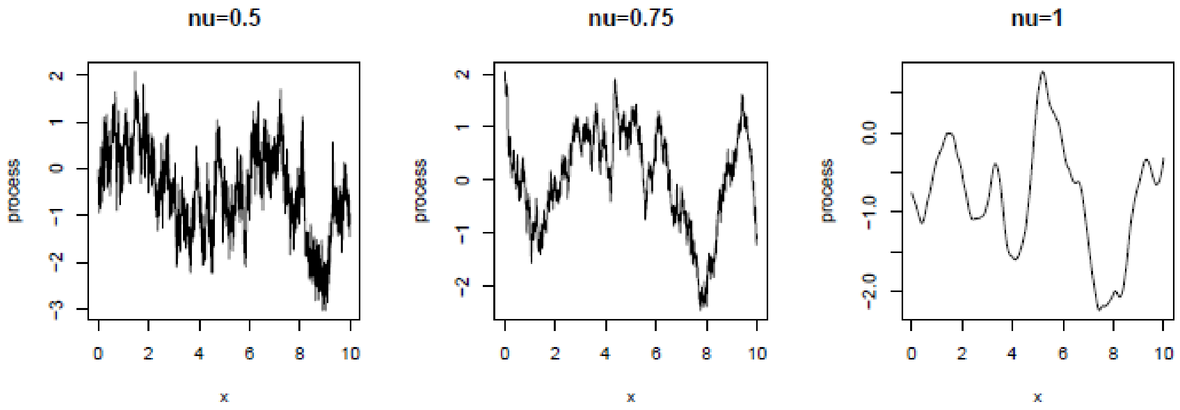


Figure 1.9: Three realisation of one dimension Gaussian processes

Forecasting, Prediction and Control Problems

- (**Housing Example**) The following figure (a) shows a summary statistic for the property market in San Diego, and the percentage change in price in (b)

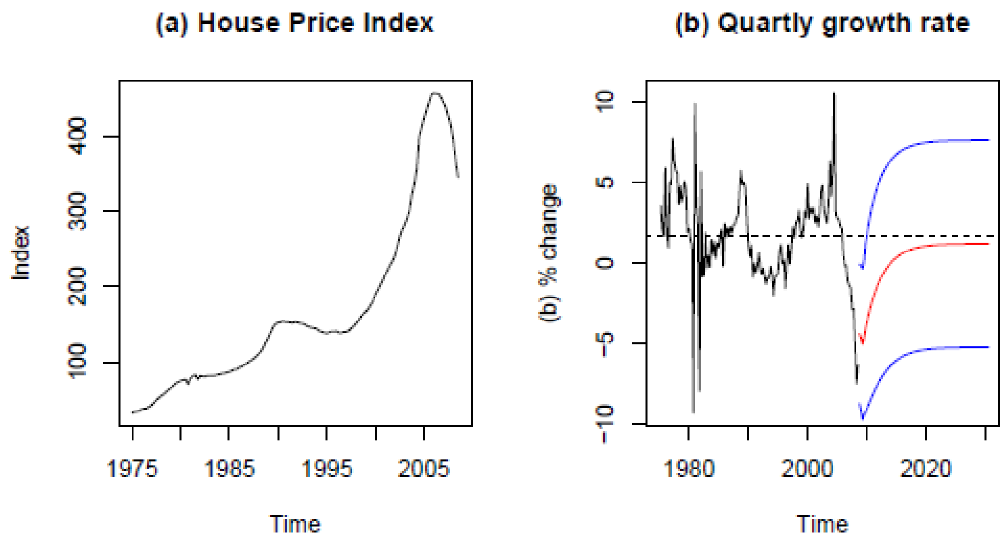


Figure 1.12: San Diego housing market

- We see that the index itself looks non-stationary but there is some stability in growth rate.

- Using one of the models proposed by Gonzalez-Rivera, we show a point estimate of the growth rate (red) and prediction range (blue), the dash horizontal line is the sample mean

- Different agents (property owners, real estate agents, ...) may have different loss functions thus even with the same model we may have different forecasts.
  - E.g. the loss associated with overestimating the index may have, for some agents, a more sever effect than underestimating it, so asymmetric loss functions may be appropriate
- We can also think about the length of time that different agents may be interested in forecasting.
  - Someone thinking about selling a house in the next year would have four quarters as their forecast window, whereas a real-estate company may be interested in a much longer time scale.
  - In (b), it is qualitatively typical of many forecasts.
    - Short term forecasts – here a few quarters – are close to the current values,
    - But as the forecast window lengthens, the forecast converges to the (unconditional or marginal) mean of the sample
    - The variance of the forecast is smaller for short term forecast and grow until it convergences to (unconditional or marginal) variance of the sample.
  - For this to be valid we have to assume stationarity in the data.
    - In particular that the marginal mean and variance in the observed data are good representative for longer time periods

- Why are we only using the housing index itself as the sole source of information to make forecasts?
- The Efficient market hypothesis states that in an efficient market the price of an object always incorporates and reflects all relevant information
  - ➢ Should we look for other variables with predictive power? Or rather, is the housing market efficient?

- (**Prostate cancer Example**) Here the prediction problem is to estimate the level of a clinically important, but hard to measure directly, antigen value, using a number of variables which are easier to measure
- The problem can be thought of as a regression problem with the antigen being the response and the easy to measure variables possible explanatory variables.
  - ➢ There are many variable which have some correlation with the response – sharing much of the same information
  - ➢ We know from regression courses that the best prediction model is not the one that contains all possible explanatory variables
  - ➢ Rather we need to balance the simplicity of the model with goodness-of-fit.

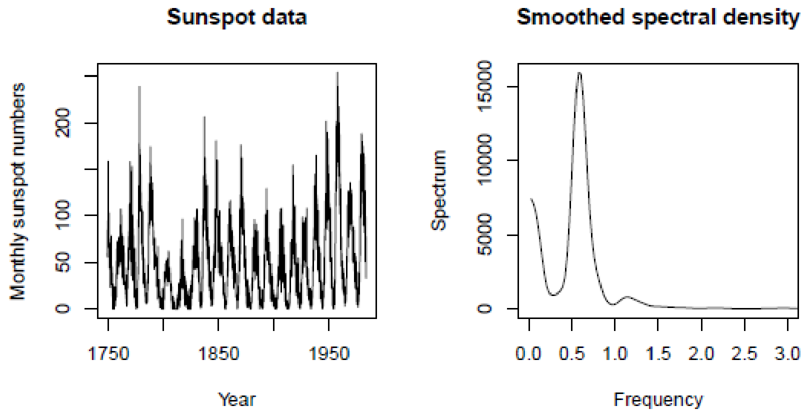
- (**GDP Example**) The use of the forecast of GDP by a central bank was part of a control problem
  - ➢ i.e. deciding what value to set interest rates
- there is a strong relationship between statistical forecasting and control problems

- (**Spacecraft Example**) Forecasting questions are often part of a larger problem of how to control complex and noisy systems using feedforward and feedback loops.
- An example of this would be a navigation system on a robotic spacecraft.
  - ➢ The probe needs to be able to control its navigation using sensor data – which can be noisy – in a time sensitive way.
  - ➢ The forecast comes in answering the question:
    - ✧ What is the future position of the craft if current settings of the controls are kept stable?
    - ✧ And a related question of: how should be change the controls to ensure that the position o the spacecraft stays on target?
- State space methods including the Kalman Filter are a powerful tool in such control problems
  - ➢ The filter is a recursive method which uses statistical models to combine new measurements from the sensors relative to past information. It also determines up-to-date uncertainties of the estimates for real-time quality assessments.

- We have seen that in order to make forecasts we need to understand and model different forms of non-stationary. With complex systems it is not always clear if there is periodic behavior or what the period is.

- We can get insights into the periodic structure by representing the time series in the frequency domain

- (**Sunspot Example**) The left panel of the following figure shows the sunspot numbers from 1749 to 1983. We see a complex pattern of periodic behavior.
- The spectral decomposition of a time series decomposes the series into a sum of sinusoidal components with different frequencies. The right plot shows the size of these components by the frequency and we see a peak at between $2\pi/11$ and $2\pi/10$ which corresponds to an approximate cycle with period around 10 to 11 years.



- Since this is a statistics course we focus our attention on methods which are fundamentally statistical – where information about the forecast is extracted from observed data, maybe through the use of a model.
    - ➢ For completeness we note that other methods are often used in practice

- (**Non-statistical forecasting methods**) <u>Scenario analysis</u> looks at the forecasting by considering a (small) finite number of alternative possible outcomes, by asking the question: "What if?" The probability of events are not considered.
- The <u>Delphi method</u> is defined as "a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem"
    - ➢ It is used when analytical techniques are not available and subjective judgments need to be used.
    - ➢ These judgments though come from groups of individuals with diverse backgrounds with respect to experience or expertise

Simple Statistical Tools

- This section looks at some simple statistical tools which can be used with the time series.
  - These typically use minimal modelling assumptions and are often used for exploratory or descriptive purposes.
  - However, they can be competitive when compared to much more complex models

- (Model with <u>Trend</u>) A trend model for the time series X_t is a decomposition

$$X_t = m_t + Y_t$$

  - Where m_t is a slowly varying function and Y_t has zero mean
    - Note that slowly varying means the change in the function is slow enough that we get a good estimate using the observed data.
    - An example could be

$$m_t = \alpha_0 + \alpha_1 t \quad or$$
$$m_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

    - Here there are only a few parameters which are needed to be estimated so with a reasonable sample size we might make estimates which are good enough for purpose
  - Note that this definition is weaker than a regression models since we are not assuming anything about dependence structure of Y_t nor its variance

- (Model with <u>Seasonal</u> Component) A model with a seasonal component with period d for X_t is a decomposition

$$X_t = s_t + Y_t$$

    - Where s_t satisfies $s_t = s_{t+d}$ for all t
  - Simple examples would be monthly data with d = 12, weekly data with d = 52

- We can put trend and seasonal component together to get a linear decomposition model

- (Linear Decomposition Model) A linear decomposition model for the time series X_t is a decomposition

$$X_t = m_t + s_t + Y_t$$

  - Where $E(Y_t) = 0$, m_t is a slowly varying function, s_t is periodic with period d and, for identification reasons we further assume

$$\sum_{t=1}^{d} S_t = 0$$

- We can construct a multiplicative decomposition model by applying a linear decomposition to $\log(X_t)$

- In R we can estimate the components of the linear decomposition model by using the function *decompose (x, type = c("additive", "multiplicative"))*.
- (Example) we can see the result of this on the Ontario gas data.

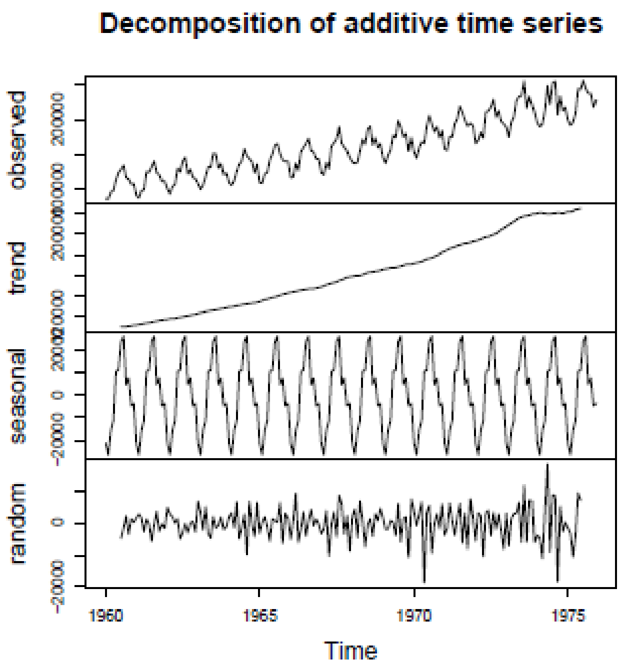
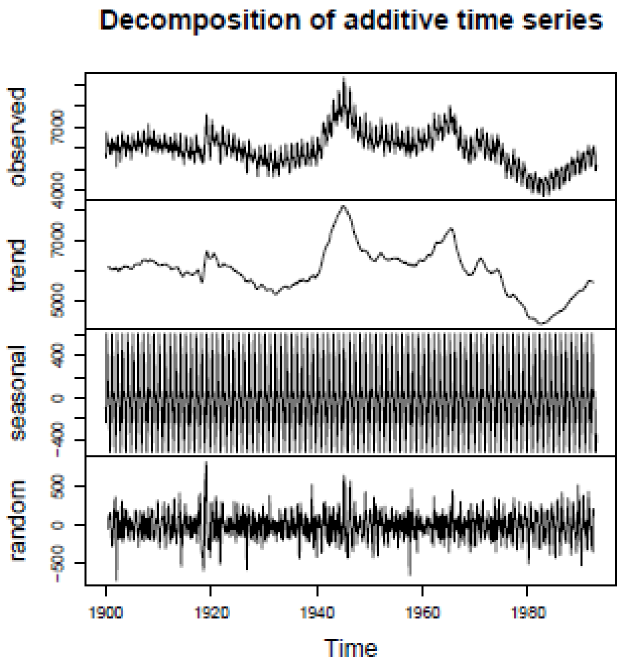


Figure 1.14: Ontario gas demand (gallons)

- The time series plot is on the top panel, an estimate of the trend m_t in the second panel, an estimate of the seasonal effect, s_t in the third panel and the "residual" random term in the lower panel
  - We note that these models do not assume the usual regressive model conditions on the random term – in particular we see that we do not have constant variance and we say nothing about the dependence structure

- (Example) we can generate the result in R by implementing *plot (decompose (birth, type = "additive"))*.



gure 1.15: Monthly birth data in Denmark 1900-1!

- (Simple moving average filter) For linear decomposition model mentioned above we can estimate m_t with <u>moving average filter</u>.
- Assuming that the period is even d = 2q, then the filter is defined on the realization x_t by

$$\widehat{m_k} = \frac{0.5x_{t-q} + x_{t-q+1} + \cdots + x_{t+q-1} + 0.5x_{t+q}}{d}$$

- Since we have that $\sum_{t=1}^{d} S_t = 0$, the seasonal component will not be part of this estimate
- We can then estimate the seasonal components. For each k = 1, …, d compute the average w_k of

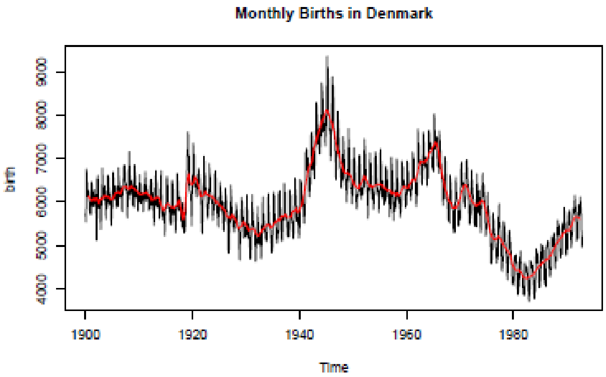
$$\{x_{k+jd} - m_{k+jd} | q < \widehat{k+j}d \leq n-q\}$$

- We then normalize to get

$$\widehat{S_k} = w_k - \frac{\sum_1^d w_j}{d}$$

- That is for monthly data we average over all January residuals, all February residuals etc.

- (Example) We can look at the Danish birth data, we can estimate the level, m_t, using a 12 point moving average filter.
- The following figure shows both the raw data (black) and the estimated level $\widehat{m_t}$ which smooths out the seasonal component (red)



- We note that the change point around 1919 has not be well estimated by the filter. Rather than having a discontinuity we have estimated a linear slope in the year containing the discontinuity.
- The key point here is that smoothing methods can over-smooth real features in that data.
  - ➢ This can also be seen in the random component where there is an increase in the variance around the discontinuity.
- We also see other regions where the variance is inflated

- The moving average filter is not the only way of filtering the data.
- One disadvantage of it symmetric form is that it can only estimate the m_t in the middle of the data.
  - ➢ This means it can't be used directly for forecasting.
- There are other filtering methods which over come this. We look at a useful methods and related methods, exponential smoothing and, the more general, Holt-Winters filtering

- (Exponential smoothing) Exponential smoothing can be used for a time series to estimates the level of the process m_t which is assumed slowing varying.
  - ➢ It should not be used when there is a trend or seasonality
- Assume that we can observe $x_1, \ldots, x_n$. We select an initial value m_0, typically x_1, then we update the estimate recursively via the update equation

$$m_{t+1} = \alpha x_t + (1-\alpha)m_t = m_t + \alpha(x_{t+1} - m_t)$$

- The second form of the update is called the error correcting version and

$$e_{t+1} \coloneqq x_{t+1} - m_t$$

  - ➢ Is the one-step ahead forecast error.
- The tuning parameter α needs to be selected and this can be done by finding the α which minimizes

$$\sum_{t=1}^{n}(x_{t+1} - m_t)^2$$

- (Holt-Winters filtering) The Holt-Winters method generalizes exponential smoothing to the case where there is a trend and seasonality.

- We have three terms which depend on time t.
  - ➢ The first is a level a_t,
  - ➢ The second is the trend b_t
  - ➢ The third is the seasonal component s_t.
- The terms a_t, b_t are considered slowly varying and the mean of the time series at time t+h is given by

$$m_{t+h} = a_t + b_t h + s_{t+h}$$

- The Holt-Winters prediction function for h time periods ahead of current t is

$$\hat{x}_{t+h} = a_t + b_t h + s_{t+h}$$

- The error term for this prediction is then defined as

$$e_t \coloneqq x_t - \left(a_{t-1} + b_{t-1} + s_{t-p}\right)$$

  - ➢ Then starting with initial values for a, b and s and then update the parameters according to the size of the error over the range of t by using

$$a_t = a_{t-1} + b_{t-1} + \alpha e_t$$
$$b_t = b_{t-1} + \alpha\beta e_t$$
$$s_t = s_{t-p} + \gamma e_t$$

  - ➢ The method uses three tuning parameters α, β, γ and these are selected by minimizing the sum of squared errors.

- (Example) we can apply the Holt-Winters method on the Denmark data and predicting using the result with the functions
- *Birth.hw = HoltWinters (birth)*
- *Birth.hw.predict = predict (birth.hw, n.ahead = 12 * 8)*
- The following figure shows the fit (red) and the forecast (blue) for the next 8 years.

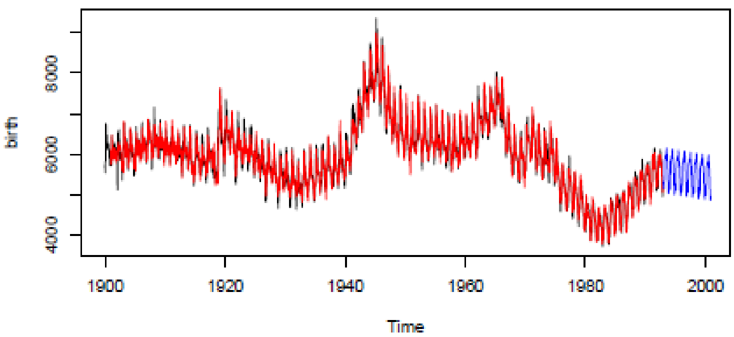


Figure 1.17: Monthly birth data in Denmark 1900-1992: fitted red and forecast blue

- The prediction models the seasonality and the prediction is based on a local linear approach.
- This is a strength for short term forecasting but, as the figures shows, may be unrealistic for longer term forecasts