Lecture 1 - Part 3: ARIMA models & residual bootstrap

Giancarlo Camarda
Institut national d'études démographiques

Ugofilippo Basellini
Max Planck Institute for Demographic Research

IDEM 117
Advances in Mortality Forecasting

International Advanced Studies in Demography
17 – 21 January, 2022

Introduction

- A *time-series* is a collection of observations made sequentially through time. Let us focus on discrete time series recorded at equal intervals of time
- Suppose we have an observed time series $y_1, y_2, \ldots, y_T$ and we wish to forecast future values such as y_{T+h}. The integer h is called *forecasting horizon*, and we denote by $\hat{y}_{T+h}$ the forecast made at time T for h steps ahead
- Forecasting methods may be broadly identified into three types:
  - *Judgemental forecasts*: based on subjective judgement and intuition
  - *Univariate methods*: forecasts depend on present and past values of a single series
  - *Multivariate methods*: forecasts depend, at least partly, on values of one or more additional variables (predictors)

$\Rightarrow$ here univariate time-series methods

Descriptive techniques

- In order to *forecast*, description and modelling of data is a prerequisite
- Always start from plotting your data!!
- This can help to identify two main sources of variation in many time series: i) trend, and ii) seasonal variation
- These variations are typically removed before time-series modelling via differencing $\Rightarrow$ helps to stabilize the **mean**
- The time plot may also help to decide whether a variable needs to be transformed $\Rightarrow$ helps to stabilize the **variance**
- One general class of transformations is the *Box-Cox transformation*:

$$\tilde{y}_t = \begin{cases} (y_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \ln y_t & \lambda = 0 \end{cases}$$

Stationary stochastic processes

- A *stochastic* time series is one where future values can only partly be determined by past values
- A process is defined *stationary* if its properties do not change through time
- More formally, let y_t be the realization of the underlying random variable Y_t, and the observed time-series $\boldsymbol{y} = [y_1, y_2, \ldots, y_T]$ be a realization of the stochastic process
- A stochastic process is second-order stationary if its first and second moments are finite and do not change over time, i.e.:

$$\mathbb{E}[Y_t] = \mu$$
$$\mathrm{COV}[Y_t, Y_{t+k}] = \mathbb{E}[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$$

for all t and for $k = 0, 1, 2, \ldots$ (note, for $k = 0$, $\gamma_0 = \sigma^2$)

- In simpler words, a stationary series has constant mean, constant variance and no predictable patterns in the long-term

The ACF & the correlogram

- The set of coefficients γ_k constitutes the autocovariance function, which is standardized to give the the autocorrelation function (ACF):

$$\rho_k = \gamma_k / \gamma_0 \tag{1}$$

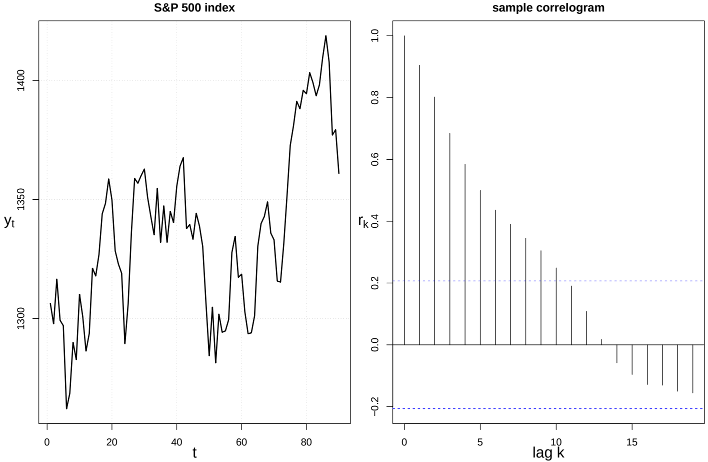
- The sample autocovariance coefficient at lag k is given by

$$c_k = \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) / T \tag{2}$$

  for $k = 0, 1, 2, \ldots$, and the sample autocorrelation coefficient at lag k is computed as $r_k = c_k / c_0 \Rightarrow r_k$ is an estimate of the theoretical ρ_k
- The **correlogram** is the graph of r_k against k
- Plotting the correlogram is very useful:
  - the ACF of non-stationary data decreases slowly, with r_1 often large and positive (see example in next slide)
  - the ACF of stationary data drops to zero relatively quickly
  - useful to identify orders of AR(p) and MA(q) models
- The **partial** ACF (PACF) measures the excess correlation at lag k which has not already been accounted for by autocorrelations at lower lags

Non-stationary time-series: an example

- non-constant mean
- ACF decreases slowly, r_1 large and positive



Left: Standard&Poor 500 index for US stock market for 90 trading days starting on March 16, 1999. Right: sample correlogram of the time-series. Source: data from Chatfield (2000).

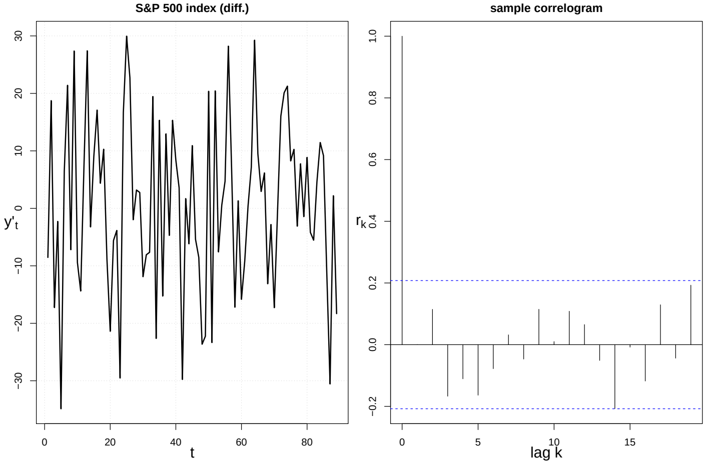
Differencing

- Powerful tool to stabilize the mean and obtain stationary time-series
- First-order differencing: $y'_t = y_t - y_{t-1}$, y'_t is the *change* between observations of y_t (composed by $T - 1$ values)
- If y'_t still non-stationary, second-order differencing: $y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_t - 2$ (composed by $T - 2$ values)
- Almost never necessary to go beyond y''_t
- For seasonal data, seasonal differencing: $y'_t = y_t - y_{t-m}$, where m is the number of seasons, y'_t is the change between one year to the next
- In addition to correlogram, two main tests for determining the required order of differencing:
  - Augmented Dickey Fuller test: H_0 data are non-stationary and non-seasonal
  - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test : H_0 data are stationary and non-seasonal

First-order differencing of the S&P500 time-series:

- ∼constant mean and variance, no predictable patterns
- ACF drops to zero quickly, tiny value of KPSS test (cannot reject H_0)

$\Rightarrow$ stationary time-series



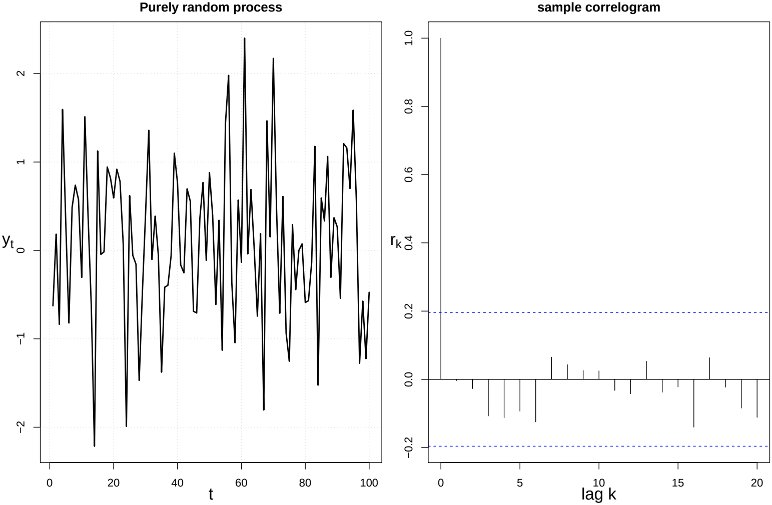
Purely random process (white noise)

- A sequence of uncorrelated, identically distributed random variables Z_t with zero mean and constant variance σ_z^2
- A stationary process with ACF equal to:

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

- Often used to model the random disturbances in more complicated processes

Purely random process: an example



Simulated time-series of a purely random process

Random walk

- Random walk models are widely used for non-stationary data (e.g. economic and financial data):

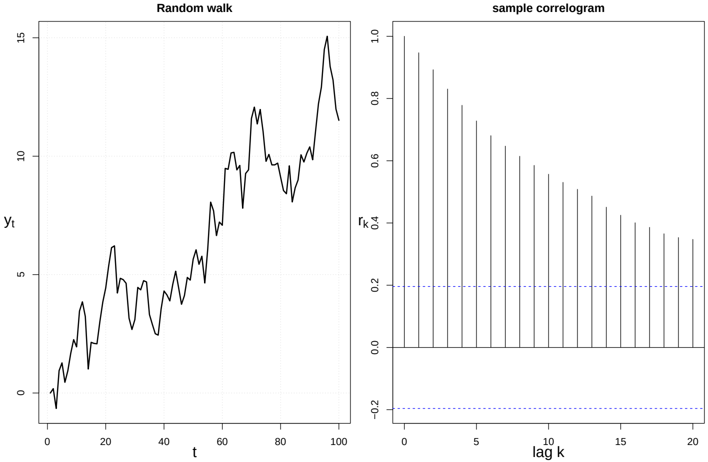
$$Y_t = Y_{t-1} + Z_t$$

where Z_t is a purely random process

- Random walk is non-stationary (variance increases through time), but first-order differences (purely random process) is stationary
- Typical features:
  - sudden and unpredictable changes of direction
  - long periods of apparent trends up/down
- Forecasts for the random walk model are simply given by the value of the last observation (i.e. naïve forecast), i.e.:

$$\hat{y}_{T+h|T} = y_T$$

Random walk: an example



Simulated time-series of a random walk process. Note the non-constant mean and the ACF slowly decreasing (indicating non-stationarity)

Random walk with drift

- A closely related model that allows first differences to have non-zero mean is the **random walk with drift**:

$$Y_t = c + Y_{t-1} + Z_t$$

where Z_t is a purely random process, and c is a constant
  - if $c > 0$, y_t will tend to drift upwards
  - if $c < 0$, y_t will tend to drift downwards
- The estimate of the drift c is given by the average of the changes between consecutive observations, i.e.

$$c = \frac{1}{T-1} \sum_{t=2}^{T} (y_t - y_{t-1}) = \frac{y_T - y_1}{T-1}$$

- Forecasts for the random walk model with drift are given by:

$$\hat{y}_{T+h|T} = y_T + c\,h$$

$\Rightarrow$ equivalent to drawing a line between the first and last observations, and extrapolating it into the future

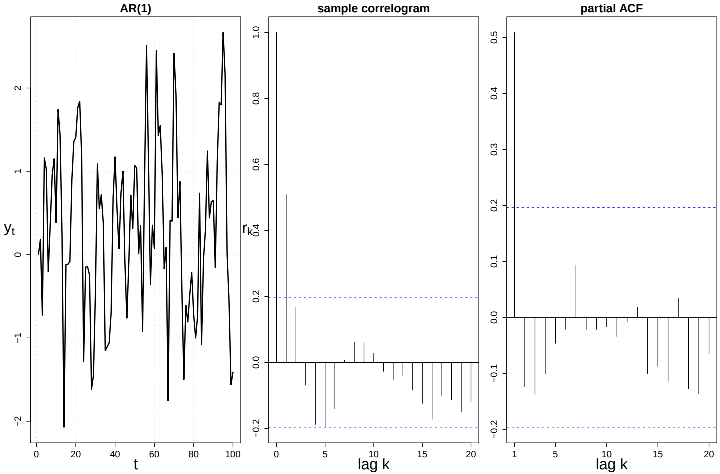
Autoregressive processes

- A process Y_t is autoregressive of order p (denoted AR(p)) if

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + Z_t$$

  where Z_t is a purely random process and c is a constant
- A multiple regression with lagged values of Y_t as predictors
- Simplest example: first-order case, i.e. AR(1), where $Y_t = c + \phi Y_{t-1} + Z_t$
  - if $c = 0$ and $\phi = 1$, then random walk
  - if $c \neq 0$ and $\phi = 1$, then random walk with drift
  - if $|\phi| > 1$, then the series is explosive and non-stationary
  - if $|\phi| < 1$, then the process is stationary, with ACF given by $\rho_k = \phi^k$ for $k = 0, 1, 2, \ldots$ ($\Rightarrow$the ACF decreases exponentially)
- A useful property of an AR(p) process is that the **partial** ACF is zero at all lags greater than p $\Rightarrow$ useful for model selection

AR(1): an example



Simulated time-series of an AR(1) process $y_t = 0.6y_{t-1} + z_t$. Note the ACF dies out in a sine-wave manner, PACF has all zero spikes beyond the 1st spike

Moving average processes

- A process Y_t is moving average of order q (denoted MA(q)) if

$$Y_t = c + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

  where Z_t is a purely random process and c is a constant

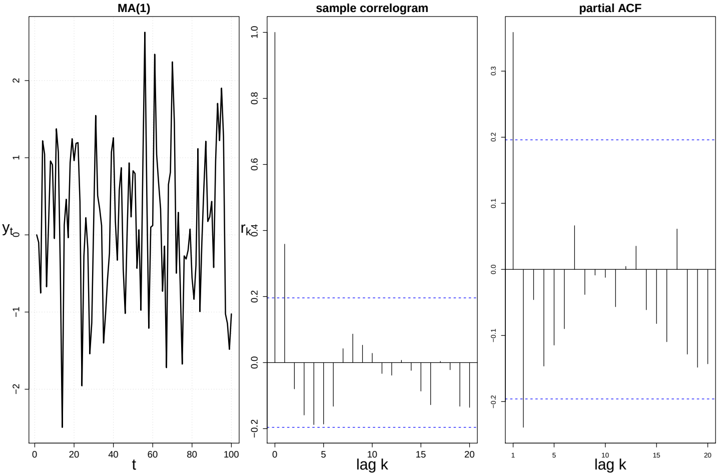
- A multiple regression with past errors as predictors
- Finite-order MA process is stationary for any values of $\boldsymbol{\theta}$
- Simplest example: first-order case, i.e. MA(1), where $Y_t = Z_t + \theta Z_{t-1}$
  - the process is stationary for any value of θ, with ACF given by:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \theta / \left(1 + \theta^2\right) & k = 1 \\ 0 & k > 1 \end{cases}$$

    i.e. the ACF "cuts-off" at lag 1.

- A useful property of an MA(q) process is that the ACF is zero at all lags greater than $q \Rightarrow$ useful for model selection

MA(1): an example



Simulated time-series of an MA(1) process $y_t = z_t + 0.45 z_{t-1}$. Note the ACF is zero at all lags greater than 1, while the PACF dies out in a sine-wave manner

ARMA & ARIMA processes

- Putting together AR(p) and MA(q), a process Y_t is autoregressive moving average (denoted ARMA(p,q)) if

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

  where Z_t is a purely random process and c is a constant
- Very important since many real data sets better approximated by mixed ARMA rather than pure AR or MA
- Moreover, most time-series are not stationary, and we cannot apply stationary AR, MA or ARMA processes $\Rightarrow$ ARIMA(p, d, q) models combine ARMA with **differencing** of order d, employed to make series stationary
- Autoregressive Integrated Moving Average (ARIMA(p, d, q)) is a very general class of time-series models:
  - purely random process: ARIMA(0,0,0)
  - random walk: ARIMA(0,1,0) without constant
  - random walk with drift: ARIMA(0,1,0) with constant
  - AR(p): ARIMA(p,0,0)
  - MA(q): ARIMA(0,0,q)

Modelling procedure with ARIMA

1. Plot the data and identify unusual observations
2. If necessary, transform the data to stabilize the variance
3. If data non-stationary: take first-order differences until data are stationary
4. Examine ACF/PACF: is an AR(p) or MA(q) model appropriate?
5. Try chosen model(s) and employ AICc to search for a better one
6. Check residuals of the model $\Rightarrow$ they should look like a purely random process
7. Forecast

Interpreting the ACF/PACF can be very challenging, and not suitable for ARMA processes $\Rightarrow$ points 3–5 with `auto.arima` function from `forecast` package

The Bootstrap

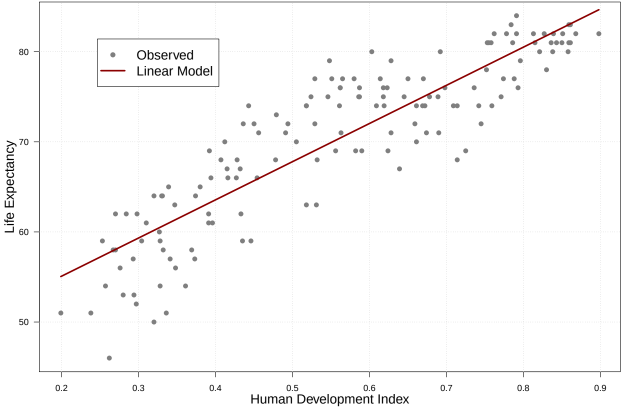
- A resampling method introduced in 1979 by Bradley Efron
- Since then, it has had a huge impact on statistics and every area of statistical application
- One of the first computer-intensive statistical techniques, replacing traditional algebraic derivations with data-based computer simulations
- A simulation-based approach to:
  - determine the accuracy of statistical estimates
  - estimate bias and variance of estimators
  - obtain approximate confidence intervals
- Bootstrap ideas of are at the heart of subsequent statistical methods such as random forests
- Here, we will focus on **residual bootstrap** as a technique to derive confidence intervals for the outcomes of a given model from its residuals
- For a thorough introduction, refer to Efron and Tibshirani (1993)

Residual bootstrap: the idea

1. Fit the model and derive the outcome of interest (say $\hat{\boldsymbol{\theta}}$)
2. Compute the fitted values $\hat{y}_i$ and the residuals r_i
3. Randomly resample r_i (with replacement) to derive a new set of residuals r_i^*
4. Refit the model using the fictitious responses y_i^* (computed from r_i^*) and retain the model's output of interest (say $\hat{\boldsymbol{\theta}}^*$)
5. Repeat 1,000 times points 3 and 4
6. Compute the pointwise confidence intervals (CI) of $\hat{\boldsymbol{\theta}}$ from the 1,000 estimates of $\hat{\boldsymbol{\theta}}^*$ (e.g. 80% CI from the lowest and upper deciles)

Residual bootstrap: a simple example

- Back to the linear model (Note: this is an illustrative example only, since CI for the linear model can be derived from the standard errors of the estimated parameters)



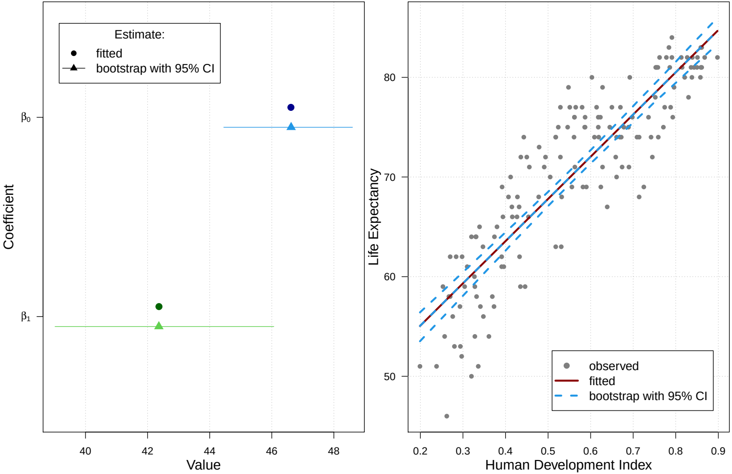
Life expectancy (both sexes) vs. Human Development Index in 2012.
Source: World Bank and World Health Organization.

Residual bootstrap in a linear model

- Residuals are the difference between observed and fitted data, i.e. $r_i = y_i - \hat{y}_i$
- Resample (with replacement) r_i to derive r_i^*
- Compute the fictitious responses $y_i^* = \hat{y}_i + r_i^*$
- Refit the model to y_i^*
- Repeat 1,000 times

$\Rightarrow$ here we will derive pointwise CI for β as well as for $\hat{y}_i$

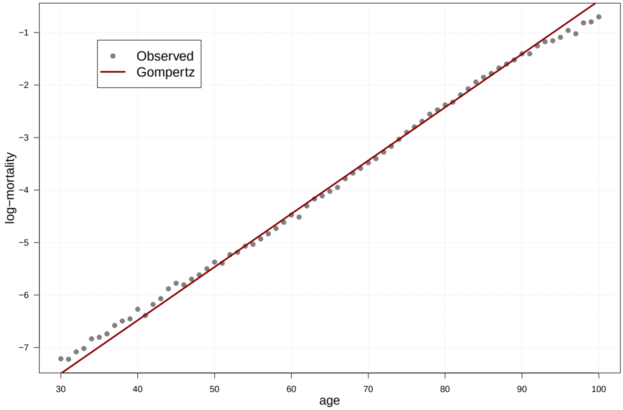
Residual bootstrap in a linear model



Estimated parameters and fitted values of a linear model with 95% bootstrap CI between life expectancy (both sexes) and Human Development Index in 2012.

Residual bootstrap: a more useful example

- Back to the Gompertz model



Actual and estimated death rates (in log scale) using a Gompertz over age. England and Wales, females, 1960, ages 30-100. Source: Human Mortality Database (2021).

Residual bootstrap in a Poisson setting

- In a Poisson setting, we start from the deviance residuals

$$r_i = \text{sign}(y_i - \hat{y}_i)\sqrt{2\left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - y_i + \hat{y}_i\right]} \tag{3}$$

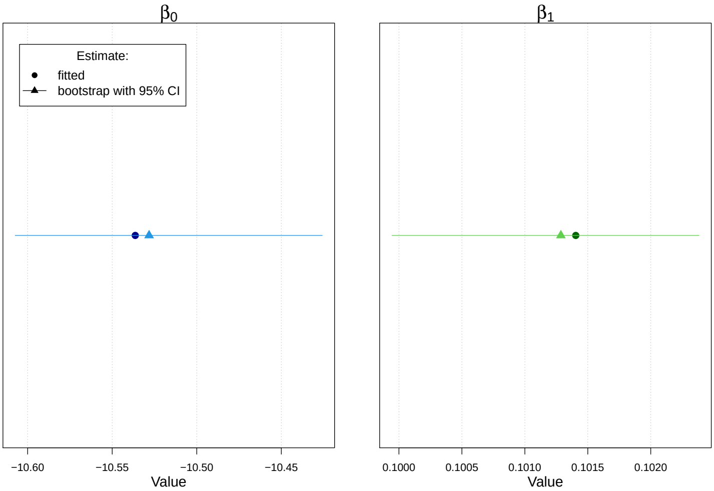
- Resample r_i (with replacement) to derive r_i^*
- To compute fictitious responses y_i^*, we substitute r_i^* in Eq. (3) and re-arrange the terms:

$$\frac{(r_i^*)^2}{2} - y_i \ln(y_i) + y_i \ln(\hat{y}_i) + y_i - \hat{y}_i = 0 \tag{4}$$

- Solving numerically (with Newton-Raphson!) Eq. (4) with respect to $\hat{y}_i$ leads to y_i^*
- Refit the model to y_i^* and repeat 1,000 times

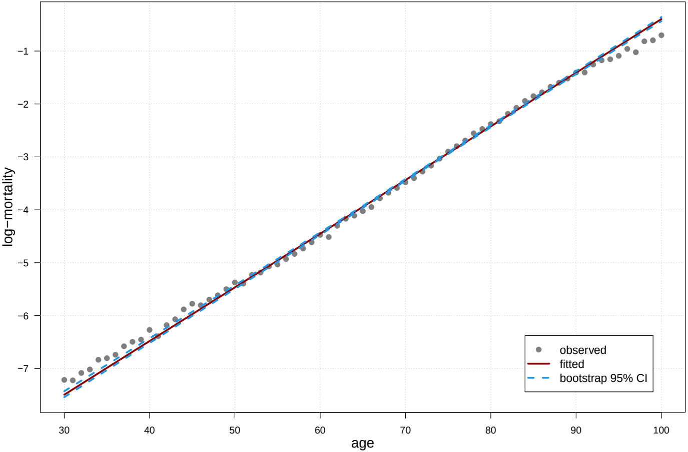
$\Rightarrow$ here we will derive pointwise CI for $\boldsymbol{\beta}$ as well as for $\hat{y}_i$

Residual bootstrap in a Poisson setting



Estimated and bootstrap coefficients with 95% CI for the Gompertz model. England and Wales, females, 1960, ages 30-100.

Residual bootstrap in a Poisson setting



Actual and estimated death rates (in log scale) with 95% bootstrap CI using a Gompertz model over age. England and Wales, females, 1960, ages 30-100. Source: Human Mortality Database (2021).

References

- Chatfield, C. (2000). *Time-series forecasting*. Chapman & Hall/CRC
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*, 2nd edition, OTexts: Melbourne, Australia. Available at https://otexts.com/fpp2/