Forecasting Methods and Applications

Applied Data Analysis School

Lecture 2

Regression Analysis and Forecasting

November 2021

Cristina Amado
University of Minho

Outline of the lecture:

- Some simple forecasting methods
- Evaluating forecast accuracy
- Linear regression and forecasting
- Applications

References:

- Diebold, F. X. (2007), Elements of Forecasting, 4th edition, South-Western College Publishing.
- Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2015), Introduction to Time Series Analysis and Forecasting, 2nd edition, Wiley.
- Brockwell, P. J. and Davis, R. A. (2002), Introduction to Time Series and Forecasting, 2nd edition, Springer-Verlag, New York.

Average method:

- This method can be used both for time series and cross-sectional data.
- The forecasts of all future values are equal to the mean of historical data:

$$\{y_1, ..., y_T\}$$

- Forecasts:

$$\widehat{y}_{T+h|T} = \bar{y} = \frac{\sum_{t=1}^{T} y_t}{T} = \frac{y_1 + ... + y_T}{T}$$

Some simple forecasting methods: Average method

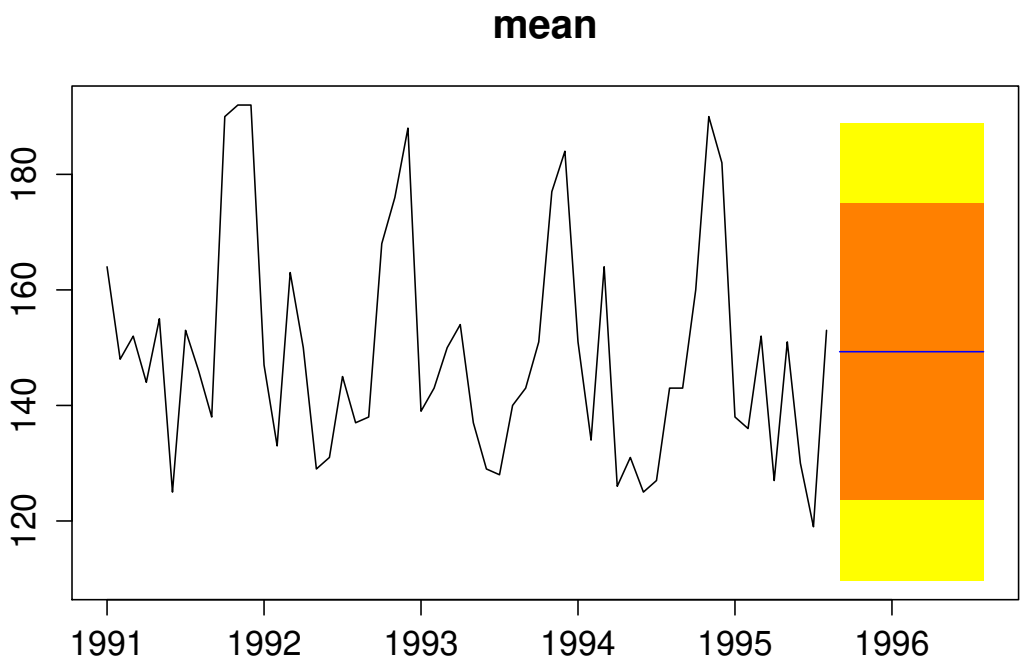

Figure 1: Forecasts of the monthly Australian beer production for twelve-step-ahead horizon using the average method. The forecast intervals at 80% and 95% are in orange and yellow color, respectively.

Some simple forecasting methods: Naïve method

Naïve method:

- This method is only appropriate for time series data.
- The forecasts of all future values are set to be equal to the last observed value.
- Forecasts:

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

Some simple forecasting methods: Naïve method

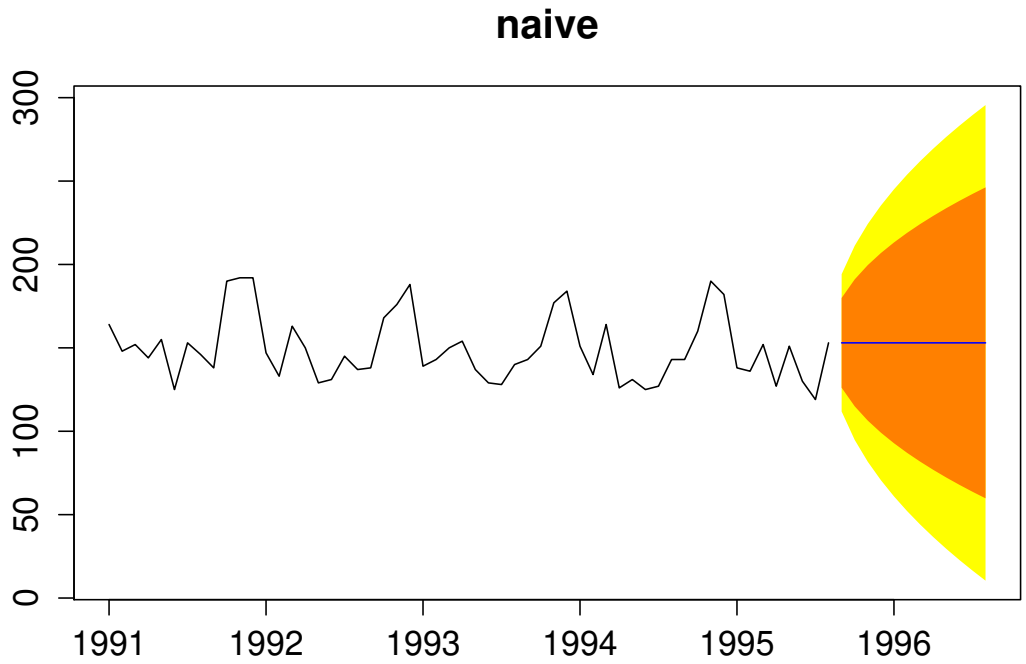

Figure 2: Forecasts of the monthly Australian beer production for twelve-step-ahead horizon using the naïve method. The forecast intervals at 80% and 95% are in orange and yellow color, respectively.

Some simple forecasting methods: Seasonal naïve method

Seasonal naïve method:

- This method is useful for highly seasonal data.
- The forecasts are equal to the last observed value from the same season of the year (or the same month of the previous year).
- Forecasts:

$$\widehat{y}_{T+h|T} = y_{T+h-kS}$$

where $S =$ seasonal period, $k = \lfloor (h-1)/m \rfloor + 1$, and $\lfloor u \rfloor$ denotes the integer part of u.

Some simple forecasting methods: Seasonal naïve method

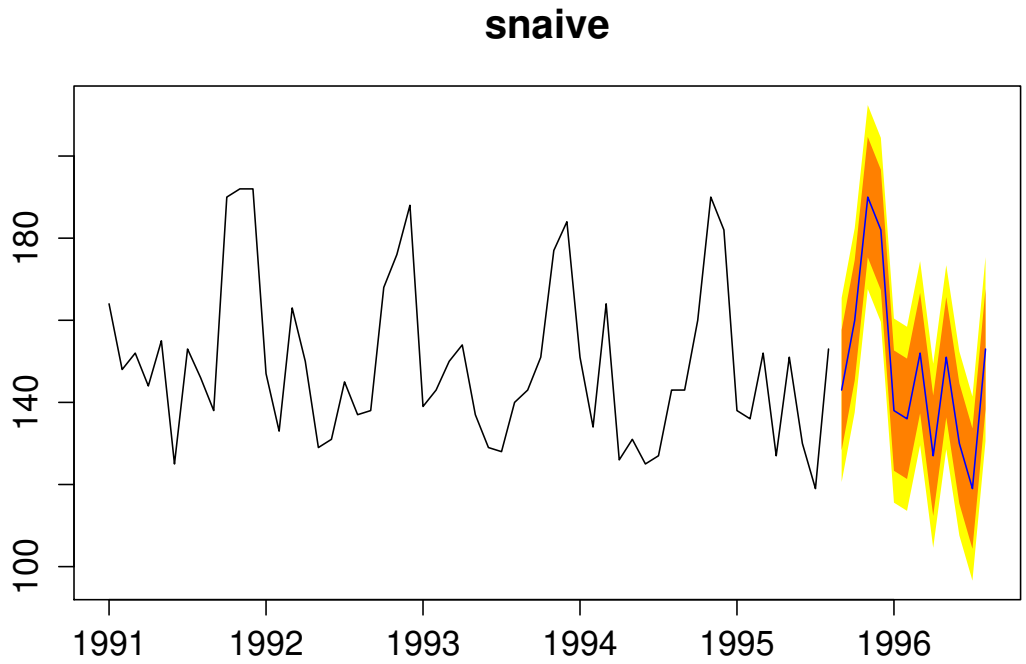

Figure 3: Forecasts of the monthly Australian beer production for twelve-step-ahead horizon using the seasonal naïve method. The forecast intervals at 80% and 95% are in orange and yellow color, respectively.

Some simple forecasting methods: Drift method

Drift method:

- The forecasts are equal to the last value plus an average change.
- The drift method allows the forecasts to increase or decrease over time, where the amount of change over time (or drift) is the average change observed in the historical data.
- Forecasts:

$$\begin{aligned}\widehat{y}_{T+h|T} &= y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) \\ &= y_T + \frac{h}{T-1}(y_T - y_1)\end{aligned}$$

- This is equivalent to drawing a line between the first and last observation, and extrapolating it into the future.

Some simple forecasting methods: Drift method

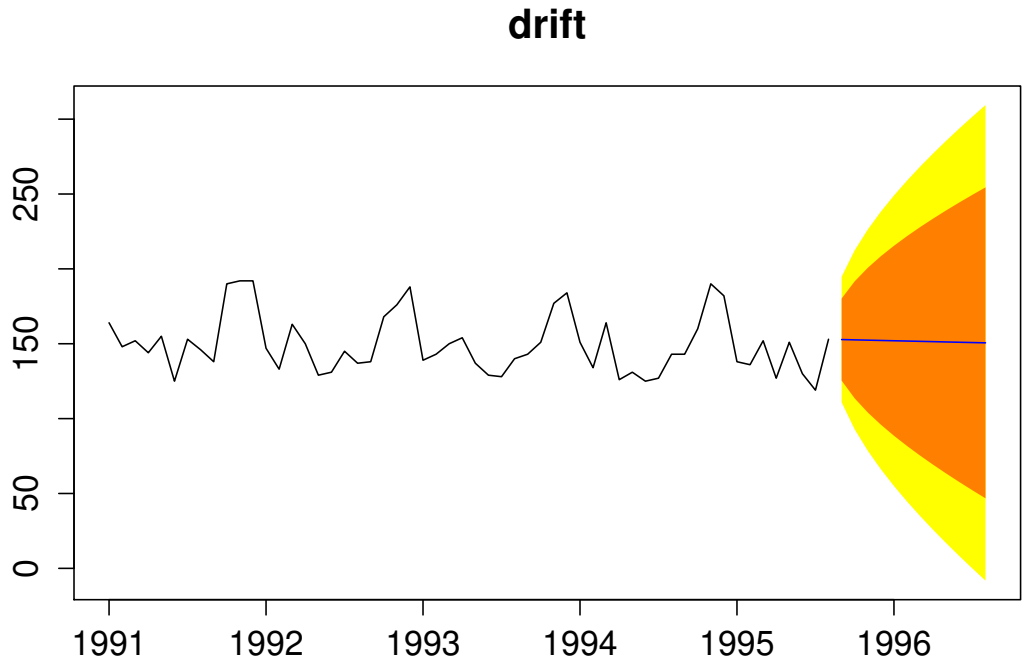

Figure 4: Forecasts of the monthly Australian beer production for twelve-step-ahead horizon using the drift method. The forecast intervals at 80% and 95% are in orange and yellow color, respectively.

Evaluating forecast accuracy

Question: How do we know whether a forecast is good? How do we assess a historical forecast?

We shall discuss:

1. Evaluating forecasts
2. Optimal forecasts
3. Measures of forecast accuracy

Note: Measures of forecast accuracy should always be part of a model validation and they can also be used to discriminate between competing models.

Evaluating forecast accuracy

- When making forecastings, the user must be concerned about the accuracy of future forecasts, and not how well the model fits the data.

Difference between **residuals** and **forecast errors**:

- Residuals are the errors from the model-fitting process (in-sample)
- Forecast errors reflect the capability of the model to successfully predict future observations (out-of-sample).

Residual: $\widehat{e}_t = y_{t+h} - \widehat{y}_{t+h}$

Forecast error: $e_t = y_{t+h} - E(y_{t+h}|\Omega_t)$

Evaluating forecast accuracy

Optimal forecasts:

- One has to check whether the single forecast has the properties expected of an optimal forecast
- Given a model, the optimal forecast is the conditional mean $\widehat{y}_{T+h|T} = \mathrm{E}(y_{T+h}|\Omega_T)$

Key properties of an optimal forecast:

1. Unbiased
2. 1-step-ahead errors are white noise
3. h-step-ahead errors are $\mathrm{MA}(h-1)$
4. Variance of h-step-ahead error is increasing in h
5. Unforecastable errors

Evaluating forecast accuracy

Unbiased:

- The h-step-ahead optimal forecast error is

$$\begin{aligned} e_{T+h|T} &= y_{T+h} - y_{T+h|T} \\ &= \varepsilon_{T+h} + b_1\varepsilon_{T+h-1} + b_2\varepsilon_{T+h-2} + \ldots + b_{h-1}\varepsilon_{T+1} \end{aligned}$$

- It has expectation

$$\mathrm{E}(e_{T+h|T}) = 0.$$

  Hence, the optimal forecast is unbiased.

1-step-ahead errors are white noise:

- 1-step-ahead forecast error:

$$e_{T+1|T} = y_{T+1} - y_{T+1|T} = \varepsilon_{T+1}$$

  which is an unforecastable white noise.

Evaluating forecast accuracy

h-step-ahead errors are MA$(h-1)$:

- h-step-ahead forecast error:

$$e_{T+h|T} = \varepsilon_{T+h} + b_1\varepsilon_{T+h-1} + b_2\varepsilon_{T+h-2} + ... + b_{h-1}\varepsilon_{T+1}$$

  is a MA$(h-1)$.

- Thus, two h-step-ahead optimal forecast errors are correlated if the distance between them is less than h.

Variance of h-step-ahead error is increasing in h:

- The variance of h-step-ahead forecast error:

$$var(e_{T+h|T}) = (1 + b_1^2 + ... + b_{h-1}^2)\sigma^2$$

Thus, the variance of optimal forecasts increases with the forecast horizon h.

Evaluating forecast accuracy

Unforecastable errors:

- The forecast errors should be unforecastable from all information available at the time of the forecast.
- Thus, the coefficients should be zero in the regression:

$$e_{T+h|T} = \alpha_0 + \alpha_1 y_{T+h|T} + u_t$$

- Since $e_{T+h|T} = y_{T+h} - y_{T+h|T}$, this means that in the regression:

$$y_{T+h} = \alpha + \beta y_{T+h|T} + u_t$$

the coefficients should be $\alpha = 0$ and $\beta = 1$.

Evaluating forecast accuracy

Unforecastable errors:

- **"Mincer-Zarnowitz" test:** Run the regression of the actual value on the ex-ante forecast and test the joint hypothesis

$$H_0 : \alpha = 0, \ \beta = 1$$

- If the hypothesis is rejected, it indicates systematic bias in the historical forecasts.

Evaluating forecast accuracy

Process:

- Form the historical sequence of forecasts and actual values.
- Construct the forecast error as the difference.
- Use various measures of forecast performance to compare forecast accuracy.

Evaluating forecast accuracy

Common forecast evaluation measures based on N h-step ahead forecasts:

- The forecast errors bias is measured by the **mean error**:

$$ME = \frac{1}{N} \sum_{j=t+1}^{t+N} e_{j+h|j}$$

- The dispersion of the forecast errors is measured by the **error variance**:

$$EV = \frac{1}{N} \sum_{j=t+1}^{t+N} \left(e_{j+h|j} - ME\right)^2$$

Evaluating forecast accuracy

- The most common measure of overall accuracy is the **mean squared error**:

$$MSE = \frac{1}{N}\sum_{j=t+1}^{T} e_{j+h|j}^2$$

- The **root mean squared error** is

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{j=t+1}^{t+N} e_{j+h|j}^2}$$

Evaluating forecast accuracy

- Another measure is the **mean absolute error**:

$$MAE = \frac{1}{N} \sum_{j=t+1}^{t+N} \left| e_{j+h|j} \right|$$

- The **mean absolute percentage error** (scale independent) is:

$$MAPE = 100 \times \frac{1}{N} \sum_{j=t+1}^{t+N} \left| \frac{y_{j+h} - y_{j+h|j}}{y_{j+h}} \right|$$

- **Theil's U Statistic**
- We can evaluate different historical forecasts by comparing the bias, EV, RMSE or MAE.

Evaluating forecast accuracy

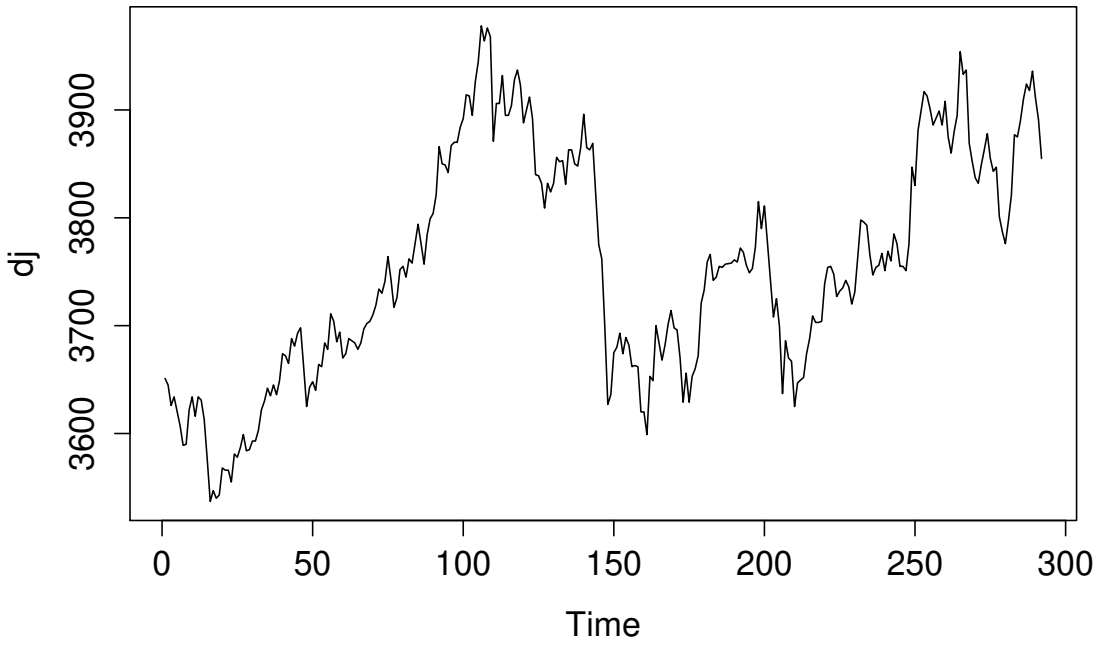

Figure 5: Dow-Jones index on 292 trading days ending 26 Aug 1994.

Evaluating forecast accuracy

Table 1: Forecast accuracy measures for the 42-step ahead forecasts for the Dow-Jones Index

Method	RMSE	MAE	MAPE	MASE	Theil's U
Average	148.24	142.42	3.66	8.70	6.07
Naïve	62.03	54.44	1.40	3.32	2.55
Drift	53.70	45.73	1.18	2.79	2.20

Evaluating forecast accuracy

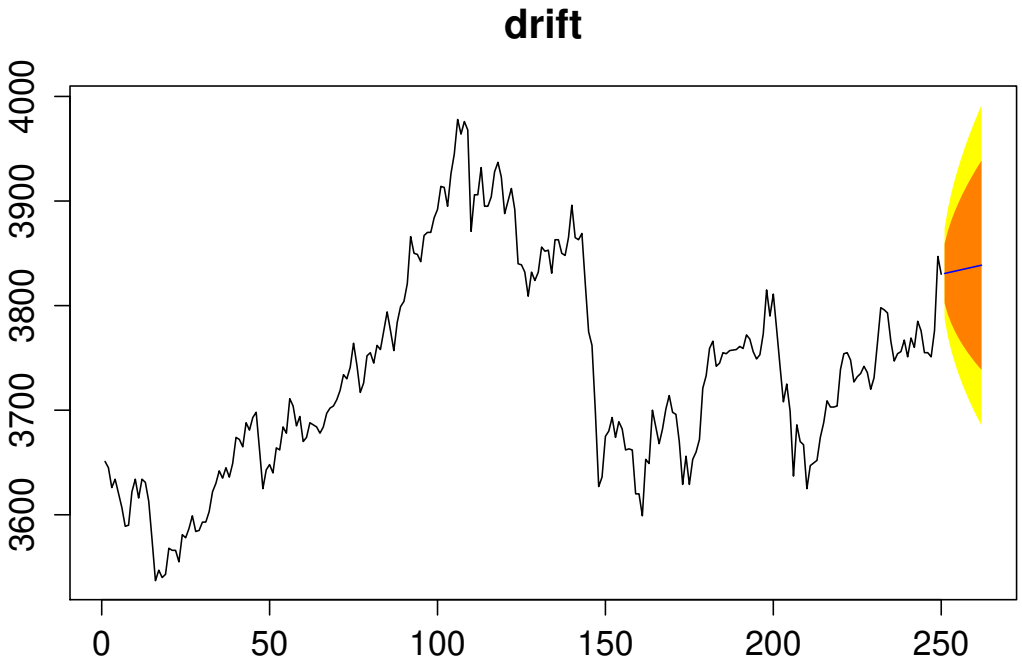

Figure 6: Forecasts of the daily Dow-Jones index for 42-step-ahead horizon using the drift method. The forecast intervals at 80% and 95% are in orange and yellow color, respectively.

Regression analysis

What is regression analysis?

- Regression analysis is the principal method of causal forecasting.
- It is concerned with estimating and evaluating the relationship between y (dependent variable, regressand, effect variable, or predicted variable) and x (independent variable, explanatory variable, causal variable, or predictor).

Regression analysis

Regression vs. correlation:

- **Correlation:** if y and x are correlated, it means that y and x are being treated in a completely symmetrical way.
- **Regression:** the dependent variable y and the independent variable(s) $x's$ are treated very differently. The y variable is assumed to be random or "stochastic" such that it has a probability distribution. The x variables are, however, assumed to have fixed ("non-stochastic") values in repeated samples.
- **Correlation and forecasting:** A variable x may be useful for predicting y, but that does not mean x is causing y. It is important to note that correlations are useful for forecasting, even when there is no causal relationship between x and y. Example: Ice-creams and drownings.

Regression analysis

In this framework, the forecast and predictor variables are assumed to be related by the simple linear model:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, ..., T$$

where u_t is a stochastic (or random) error term.

Examples of a relationship between two (or more variables):

- How private consumption depends on real disposable income.
- How asset returns vary with their level of market risk.
- How wages are related to education levels.

Regression analysis

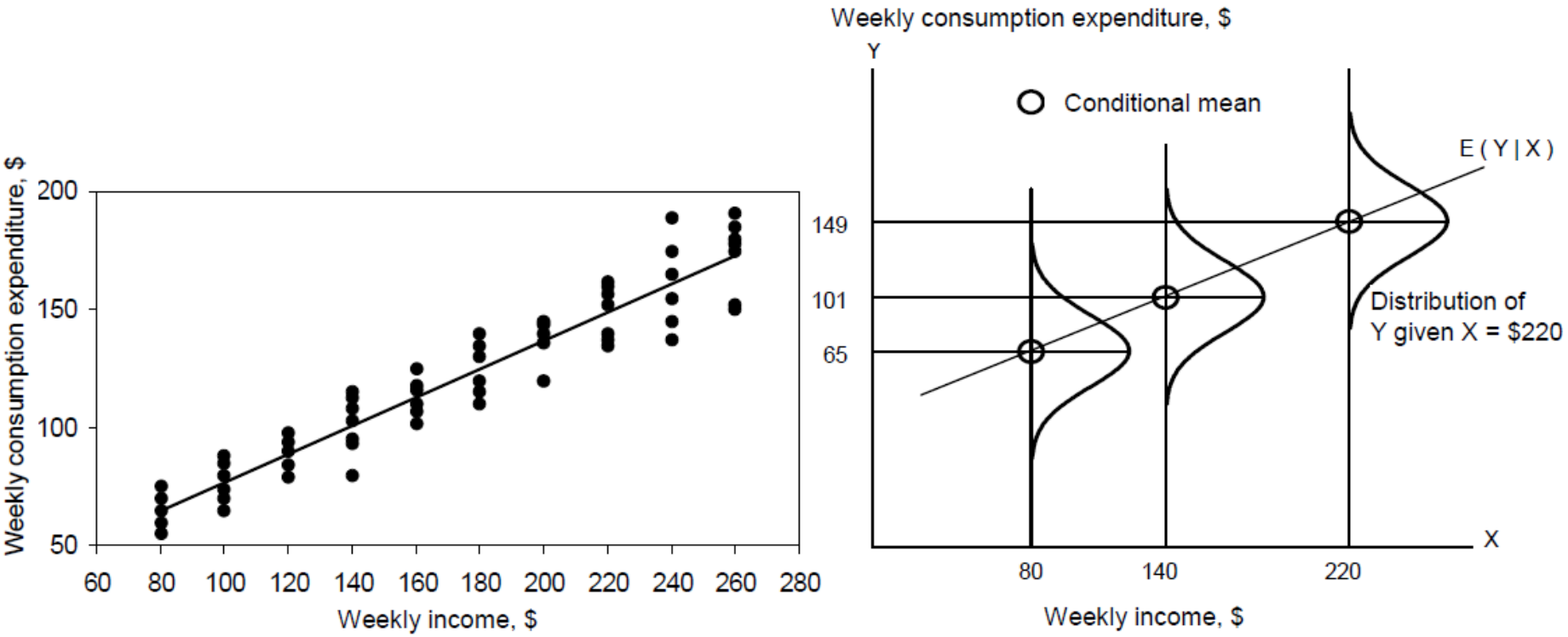

Figure 7: Population regression line between consumption expenditures and income.

Regression analysis

- The parameters β_0 and β_1 are the intercept and the slope of the line, respectively:
  - ▶ The intercept β_0 represents the predicted value of y when $x = 0$.
  - ▶ The slope β_1 represents the predicted increase in y resulting from a one unit increase in x.

- How to determine the values of the parameters β_0 and β_1?
  - ▶ Choose the parameters β_0 and β_1 such that the vertical distances from the data points to the fitted line are minimised.

Regression analysis

Example: Using the data set hprice1.dta we want to predict the housing prices in millions of dollars ($price_i$) as a function of the house dimension in square feet ($sqrft_i$), number of bedrooms ($bdrms_i$), size of the lot in square feet ($lotsize_i$) and others characteristics of 88 houses in the Boston area.

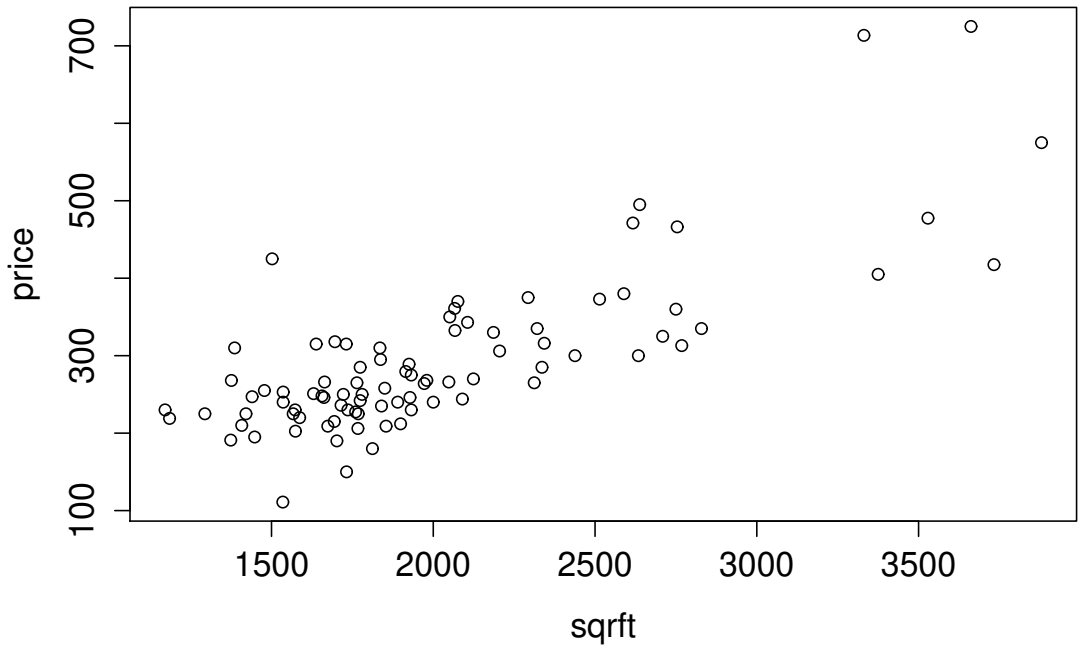

Figure 8: Scatterplot between *price* and *sqrft*

Regression analysis

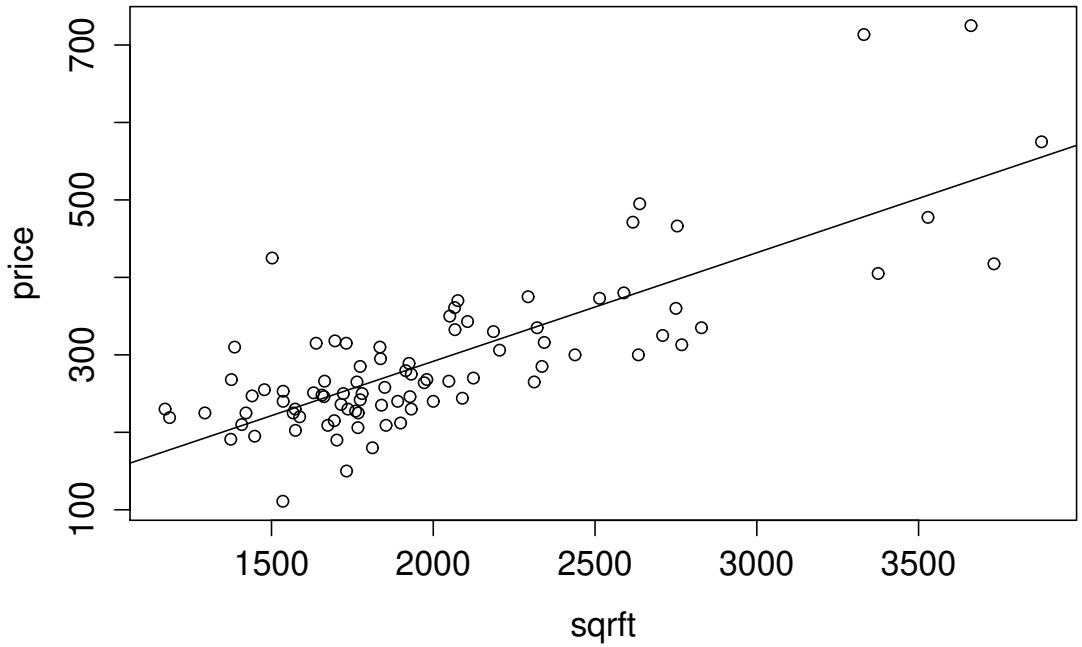

Figure 9: Regression line between the house price and the dimension of the house.

Regression analysis

- The most common method used to fit a line to the data is known as OLS (ordinary least squares).
- What we actually do is take each distance and square it (i.e. take the area of each of the squares in the diagram) and minimise the total sum of the squares (hence least squares) $\Longrightarrow$ Minimisation problem
- The OLS method minimises

$$RSS = \sum_{t=1}^{T} \widehat{u}_t^2 = \sum_{t=1}^{T} (y_t - \widehat{y}_t)^2$$

which is a function of the estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

Regression analysis

The first order conditions imply

$$\sum_{t=1}^{T} \widehat{u}_t = 0 \quad \text{and} \quad \sum_{t=1}^{T} \widehat{u}_t x_t = 0.$$

The estimators for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

and

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \quad \text{or} \quad \widehat{\beta}_1 = \hat{\rho}\frac{\widehat{\sigma}_y}{\widehat{\sigma}_x}.$$

Regression analysis and forecasting

- Observed or actual values of y_t : $y_t = \beta_0 + \beta_1 x_t + u_t$
- Predicted values of y_t: $\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_t$
- Residual: $\widehat{u}_t = y_t - \widehat{y}_t$
- The regression line can be used for forecasting.
- For each value of x we can forecast a corresponding value of y using

$$\widehat{y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 x_t$$

Numerical properties of the OLS:

- [P1] The regression line passes through the sample means of x and y :

$$\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}$$

- [P2] The predicted mean of y equals the sample mean of y :

$$\frac{\sum_{t=1}^{T} \widehat{y}_t}{T} = \bar{y}$$

- [P3] The mean (and sum) of prediction errors is zero:

$$\sum_{t=1}^{T} \widehat{u}_t = 0$$

- [P4] The prediction errors are uncorrelated with the predicted y values:

$$\sum_{t=1}^{T} \widehat{u}_t \widehat{y}_t = 0$$

- [P5] The prediction errors are uncorrelated with the sample values of x :

$$\sum_{t=1}^{T} \widehat{u}_t x_t = 0$$

Regression analysis

Evaluating the regression model

- **Residual plots:** Since each residual is an unpredictable random term, we expect the residuals to be randomly scattered without showing any systematic patterns.
- **Goodness-of-fit:** The most common goodness of fit statistic is the R^2 or coefficient of determination. It is a measure of how well the regression line fits the data.
- Total variability in the dependent variable = variability explained by the regression + variability that cannot be explained:

$$\sum_{t=1}^{T}(y_t-\bar{y})^2=\sum_{t=1}^{T}(\widehat{y}_t-\bar{y})^2+\sum_{t=1}^{T}\widehat{u}_t^2$$

- Our goodness of fit statistic is

$$R^2=\frac{\sum_{t=1}^{T}(\widehat{y}_t-\bar{y})^2}{\sum_{t=1}^{T}(y_t-\bar{y})^2}=1-\frac{\sum_{t=1}^{T}\widehat{u}_t^2}{\sum_{t=1}^{T}(y_t-\bar{y})^2},\quad 0\leq R^2\leq 1$$

Regression analysis

- Larger values of R^2 imply that x has more explanatory power for y:
  - $R^2 = 1$ means perfect fit. This means that all data points are exactly on the regression line (i.e. $RSS = 0$).
  - $R^2 = 0$ means that x does not have any explanatory power for y (i.e. x has no influence on y).
- Simple regression model: $R^2 = \widehat{\rho}^2_{yx}$.
- **Important:** A "high" R^2 does not always indicate a good model for forecasting. When making forecasts, checking the out-of-sample forecasting performance of the model is more important than measuring the value of the R^2 in-sample.

Regression analysis

- **Standard error of the regression:** Another measure of how well the model has fitted the data is the standard deviation of the residuals and it is given by

$$\widehat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T} \widehat{u}_t^2}{T-2}}$$

- The standard error is related to the size of the average error that the model produces.

Regression analysis

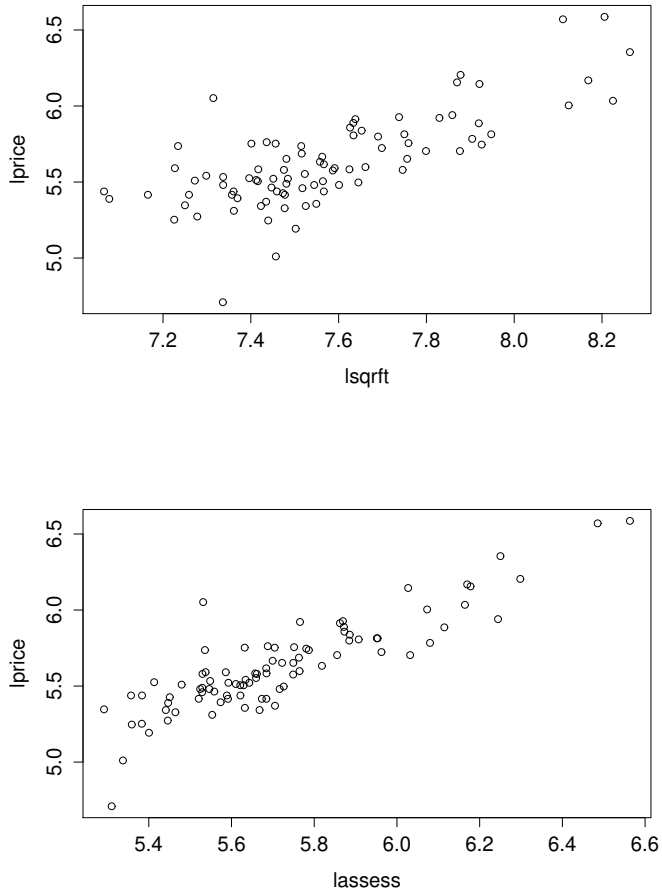

Figure 10: Scatterplots between lprice and lsqrft (upper panel) and lprice and lassess (lower panel).

Prediction of the mean value:

- Point forecast: $\mathrm{E}(Y|X = X_0)$ is estimated by

$$\widehat{y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$$

with

$$\widehat{var}(\widehat{y}_0) = \widehat{\sigma}^2 \left(\frac{1}{T} + \frac{(x_0 - \overline{x})^2}{\sum_{t=1}^{T} x_t^2} \right)$$

- Interval forecast: $100(1-\alpha)\%$ prediction interval for the mean

$$\widehat{y}_0 \pm t_{\alpha/2;T-2} \sqrt{\widehat{var}(\widehat{y}_0)}$$

Prediction of the individual value:

- Point forecast: $\mathrm{E}(Y|X = X_0)$ is estimated by

$$\widehat{y}_0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$$

with

$$\widehat{var}(y_0 - \widehat{y}_0) = \widehat{\sigma}^2 \left(1 + \frac{1}{T} + \frac{(x_0 - \overline{x})^2}{\sum_{t=1}^{T} x_t^2}\right)$$

- Interval forecast: $100(1-\alpha)\%$ prediction interval for the individual

$$\widehat{y}_0 \pm t_{\alpha/2;T-2} \sqrt{\widehat{var}(y_0 - \widehat{y}_0)}$$

Regression analysis

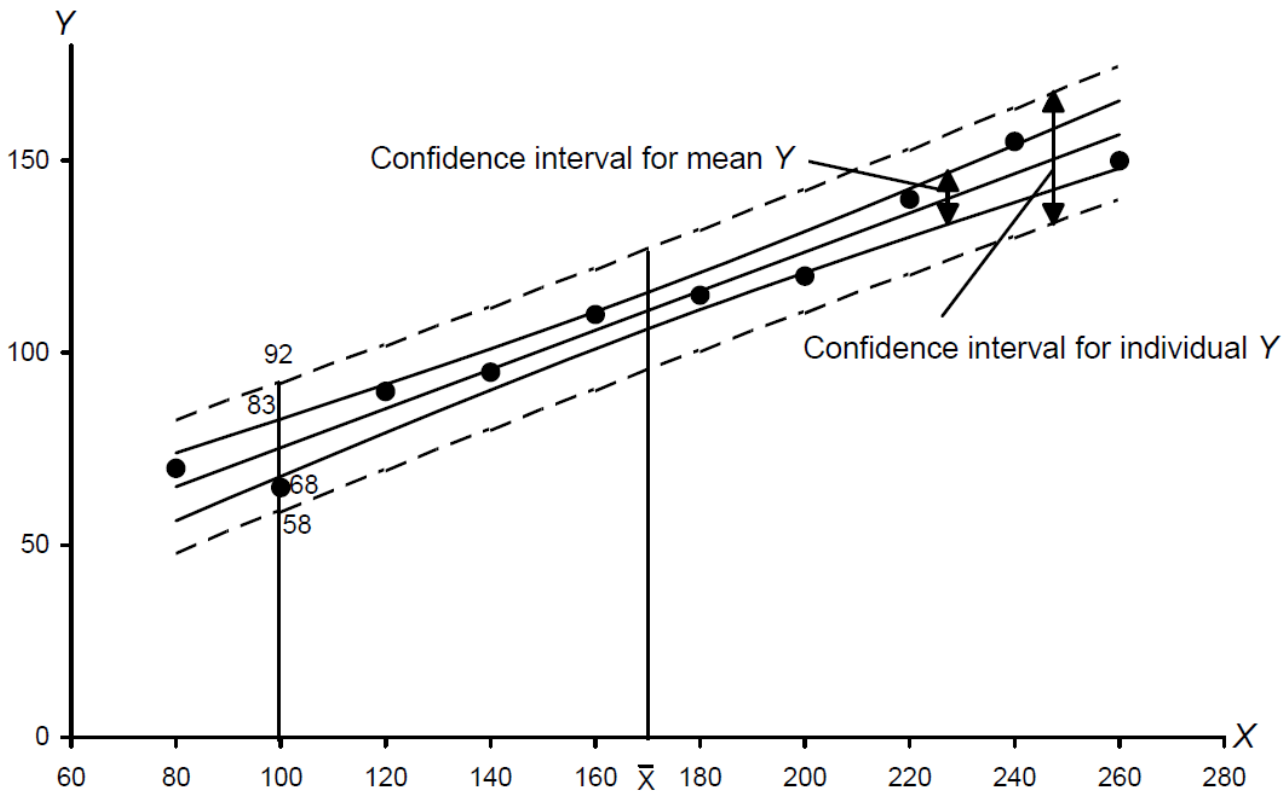

Figure 11: Confidence intervals for the mean and individual.

Regression analysis

Linear regression between lprice (log of the house price) and lassess (log of the assessed value of the house):

$$\widehat{lprice}_i = \underset{(0.346)}{-0.161} + \underset{(0.060)}{1.013}\, lassess_i, \;\; \widehat{\sigma} = 0.148$$

Compute confidence intervals when $lassess = 250$:

- **Point forecast:** $\widehat{lprice} = 253.19$.
- **Interval estimate on the mean at a specific point:**

  95% Prediction Interval: $[223.828, 282.553]$.
- **Interval estimate on a single future observation:**

  95% Confidence Interval: $[223.829, 282.551]$.