Lecture 2
Regression

CHEN Ying

FE5209 Financial Econometrics

Outline

- ➢ Regressions and linear models.
- ➢ Estimation of linear regressions.
- ➢ Troubleshooting.

Reading:
SDA chapter 4/5 and 12
FTS chapter 1
SFM chapter 3
http://cran.r-project.org/doc/manuals/R-intro.pdf
https://cran.r-project.org/doc/contrib/usingR.pdf

***[SDA]**Statistics and Data Analysis for Financial Engineering** (2010) by David Ruppert*
[FTS] Tsay, R.S. (2010) Analysis of Financial Time Series, Third Edition, Wiley.
[SFM] Franke, J., Härdle, W. K., Hafner, C. M.(2015) Statistics of Financial Markets An Introduction. Springer.

The concept of dependence

- Data: one ***response/dependent variable*** Y and p ***predictor variables*** $X_1, \ldots, X_p$ all measured on each of T observations. The predictors are also named ***independent variables, regressor variables, or explanatory variables.***

Dependent	Predictors		
Y_1	X_{11}	…	X_{1p}
		…	
Y_n	X_{n1}	…	X_{np}

- In regression analysis, we estimate the relationship between a random variable Y and one or more variables X_i .
- The goals of regression modeling include the investigation of how Y is related to $X_1, \ldots, X_p$, and prediction of future Y values when the corresponding values of $X_1, \ldots, X_p$ are already available.

Regression model

- The regression equation can be written as $Y = E\left[Y \,\middle|\, X_1, \ldots, X_p\right] + \epsilon$.

  The deterministic function $\mathrm{y} = f(z)$ where $\mathrm{y} = f(z) = E\left[Y \,\middle|\, X_1 = z_1, \ldots, X_p = z_p\right]$ is called the regression function.

- The following properties of regression equations hold.

  (1) The conditional mean of the error is zero: $E[\epsilon \,|\, X_1, \ldots, X_p] = 0$.

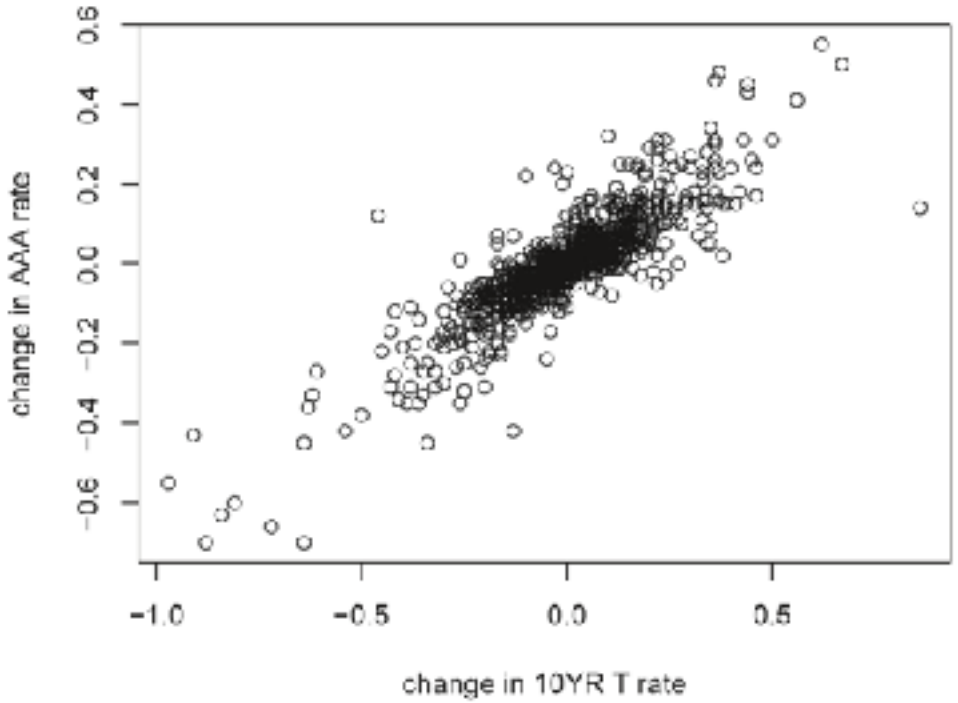
  (2) The unconditional mean of the error is zero: $E[\epsilon] = 0$.

  (3) The errors are uncorrelated with the variables $X_1, \ldots, X_p$: $E[\epsilon X] = 0$.

Example: Weekly interest rate vs. AAA yield

Weekly interest rates from February 16, 1977, to December 31, 1993, from the Federal Reserve Bank of Chicago. Below is a plot of changes in the 10-year Treasury constant maturity rate and changes in the Moody's seasoned corporate AAA bond yield.

- The dependence looks linear.
- There is only one predictor.
- Data: WeekInt.txt
- R code: WeeklyInterestRate.R



Simple linear regression

Simple linear regression is a linear model with a constant and one predictor:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \epsilon_t, \qquad t = 1, \ldots, T.$$

where parameter β_0 is the unknown intercept and $\beta_1 = \dfrac{\partial E\left[Y_t \mid X_{t1}\right]}{\partial X_{t1}}$. Therefore, β_1 is the unknown slope that tells change in the expected value of Y_t when X_{t1} changes one unit.

We observe the dependent variable Y_t and the independent variable X_t, but not the error ϵ_t.

We can formulate the regression problem in a matrix form that is standard in regression analysis as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\boldsymbol{\epsilon}$ are the noises.

What is $\boldsymbol{X}$?

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \vdots & \vdots \\ 1 & X_{T1} \end{bmatrix}$$

Estimation of linear regressions

- There is a key assumption of linear regression model, i.e. independent and identically noise: $\epsilon_1, \ldots, \epsilon_T$ are independent with mean zero and constant variance $Var(\epsilon_t) = \sigma_\epsilon^2$ for all t.
- We discuss how to estimate the regression parameters based on the assumption. We consider two main estimation techniques:

least squares: estimate the unknown parameters of a statistical model so that the particular parametric values make the sum of squared error minimal

maximum likelihood: estimate the unknown parameters of a statistical model so that the particular parametric values make the observed data the most probable given the model.

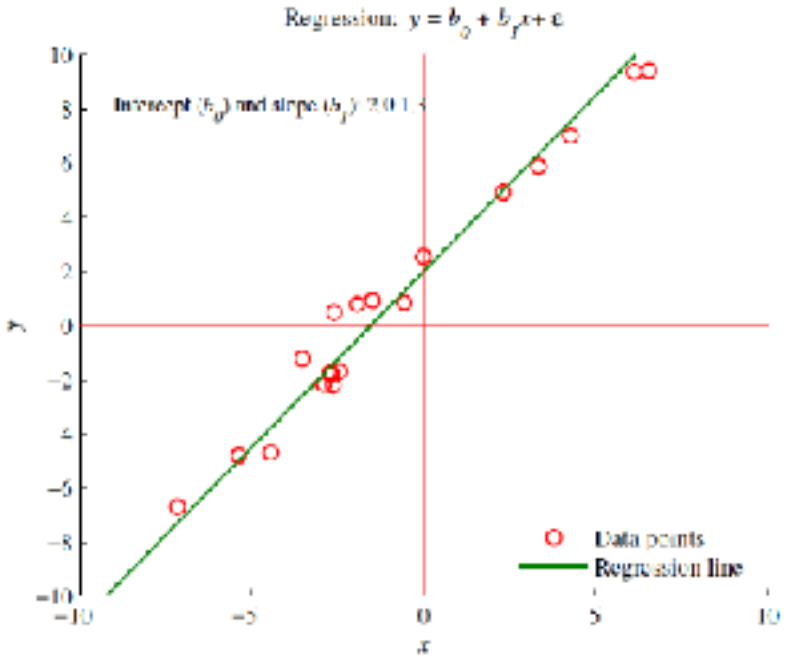
Least-Squares estimation

The LS estimates are the values of b_0 and b_1 (also denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$) that minimize the loss function:

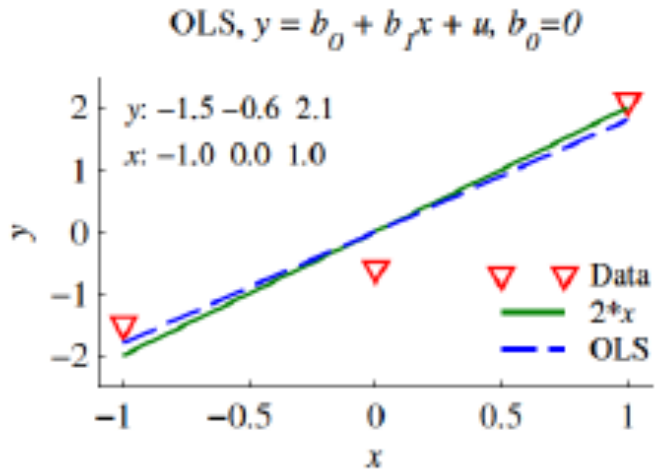
$$\sum_{t=1}^{T} \hat{\epsilon}_t^2 = \sum_{t=1}^{T} (Y_t - b_0 - b_1 X_t)^2$$

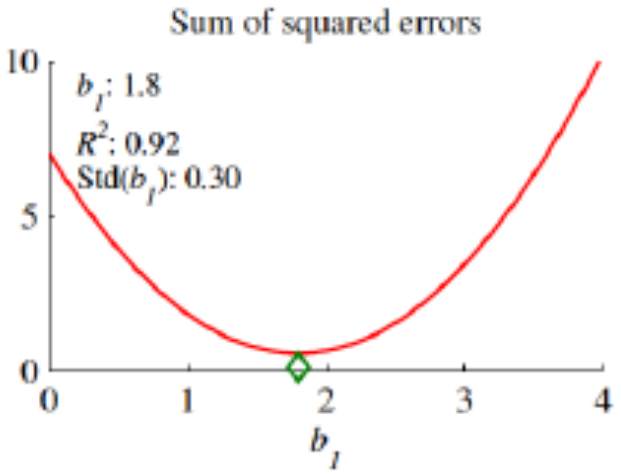
where $\hat{\epsilon}_t = \mathrm{Y_t} - \hat{Y}_t = Y_t - b_0 - b_1 X_t$ is the difference between the Y-values and the fitted values. It is called residuals.

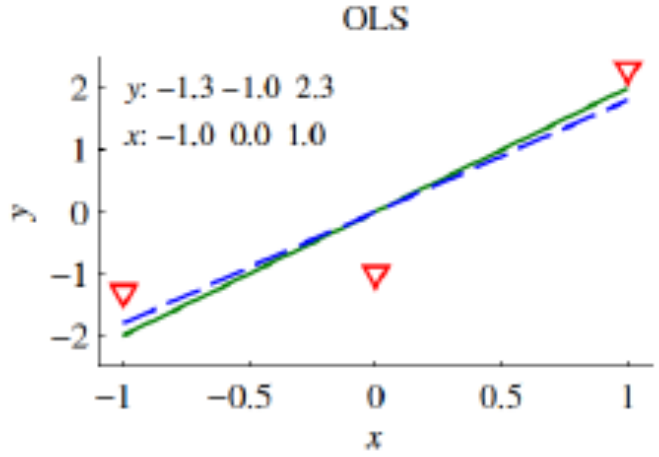
The objective is to pick values of b_0 and b_1 in order to make the model fit the data as closely as possible – where close is taken to be a small variance of the unexplained residual $\hat{\epsilon}_t$.

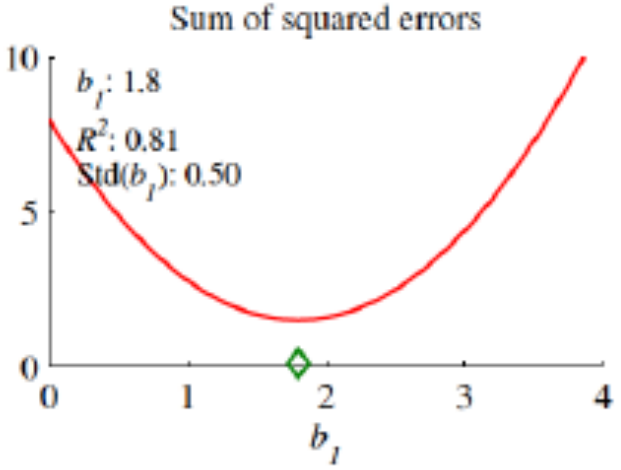


Example of OLS









First order conditions

First order conditions for minimizing a differentiable function $f(b)$: $\frac{\partial f(b)}{\partial b} = 0.$

The first order conditions are that the derivatives of the loss function with respect to b_0 and b_1 should be zero.

$$\frac{\partial}{\partial b_0} \sum \left(Y_t - b_0 - b_1 X_t\right)^2 = -2 \sum \left(Y_t - b_0 - b_1 X_t\right) 1 = 0$$

$$\frac{\partial}{\partial b_1} \sum \left(Y_t - b_0 - b_1 X_t\right)^2 = -2 \sum \left(Y_t - b_0 - b_1 X_t\right) X_t = 0$$

It leads to the OLS estimates:

$$b_1 = \hat{\beta}_1 = \frac{\frac{\sum_{t=1}^{T} X_t Y_t}{T} - \bar{X}\bar{Y}}{\frac{\sum_{t=1}^{T} X_t^2}{T} - \bar{X}^2} = \frac{cov(X,Y)}{var(X)} \qquad b_0 = \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

The slope estimator is the sample covariance of Y and X, divided by the sample variance of the regressor X. In a simpler case, when the means of Y and X are zero, then we can disregard the constant. It gives

$$b_1 = \hat{\beta}_1 = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$$

Example: Weekly interest rate vs. AAA yield

```
Call:
lm(formula = aaa_dif ~ cm10_dif)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.000109   0.002221   -0.05     0.96
cm10_dif     0.615762   0.012117   50.82   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.066 on 878 degrees of freedom
Multiple R-Squared: 0.746,      Adjusted R-squared: 0.746
F-statistic: 2.58e+03 on 1 and 878 DF,  p-value: <2e-16
```

From the output we see that the least-squares estimates of the intercept and slope are -0.000109 and 0.616. The Residual standard error is 0.066; this is what we call $\hat{\sigma}_\epsilon^2$ or s, the estimate of σ_ϵ^2.

Properties of estimators

Unbiasedness and Consistency

The bias of an estimator $\hat{\theta}$ is defined as: $Bias\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta.$

$\hat{\theta}$ is unbiased if $\mathrm{E}\left(\hat{\theta}\right) = \theta.$

An estimator $\hat{\theta}_n$ is said to be consistent if:

$$\lim_{n\to\infty} \Pr\left(\left|\hat{\theta}_n - \theta\right| < \epsilon\right) = 1 \quad \forall\ \epsilon > 0.$$

Asymptotic Normality

Provide a method to perform inference.
The primary tool in econometrics for inference is the central limit theorem.

Efficiency

Relative efficiency: If $\hat{\theta}_n$ and $\tilde{\theta}_n$ are consistent estimators of θ, then $\hat{\theta}_n$ is said to be relatively efficient to $\tilde{\theta}_n$ if

$AVar\left(\hat{\theta}_n\right) < AVar(\tilde{\theta}_n)$

Properties of $\hat{\beta}_1$ or b_1

Consider the simpler case of the coefficient estimator in the previous regression model:

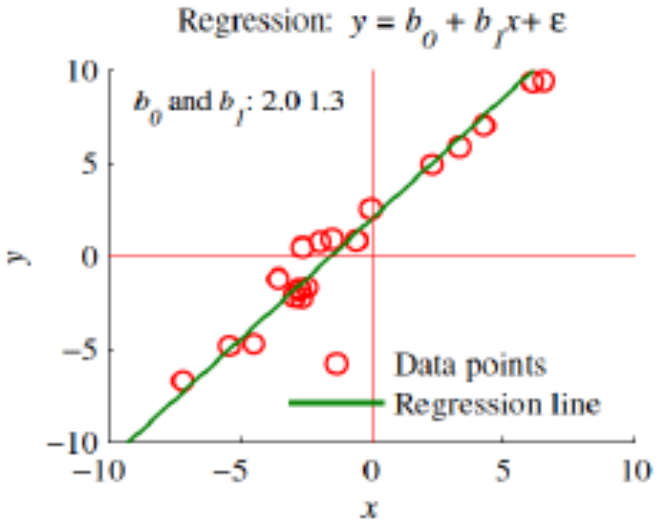
$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} X_t Y_t}{\sum_{i=1}^{n} X_i^2}$$

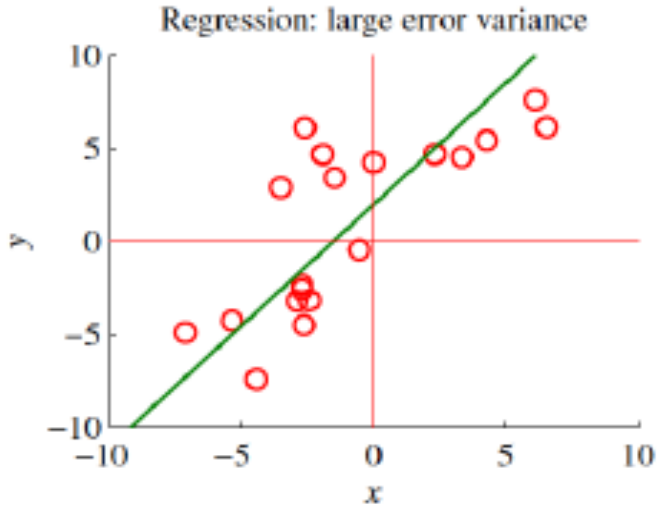
Use true parameter to substitute Y_t

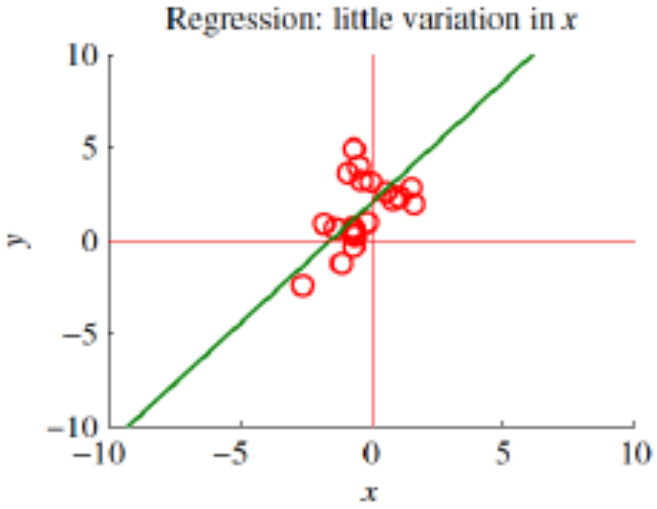
$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T} X_t(\beta_1 X_t + \epsilon_t)}{\sum_{t=1}^{T} X_t^2} \qquad = \beta_1 + \frac{(X_1\epsilon_1 + X_2\epsilon_2 + \ldots + X_T\epsilon_T)}{\sum_{t=1}^{T} X_t^2}$$

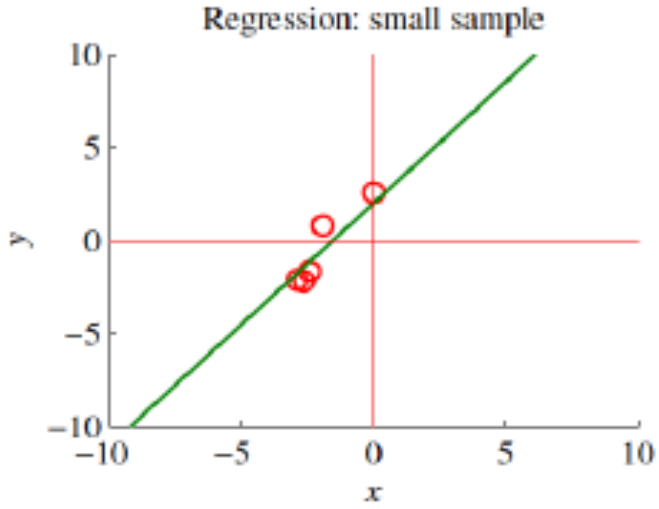
Assume fixed regressors, given the assumptions:
$E(\epsilon_t) = 0,$ and $Cov(X_t, \epsilon_t) = 0,$ we prove unbiasedness of slope: $E\left(\hat{\beta}_1\right) = \beta_1$.

Importance of error variance and variation of x









Variance of $\hat{\beta}_1$

It is useful to have a formula for the variance of an estimator to show how the estimator's precision depends on various aspects of the data such as the sample size and the values of the predictor variables.

If we assume IID noise, $E(\epsilon_t^2) = \sigma_\epsilon^2$, and $Cov(\epsilon_i,\ \epsilon_j) = 0$ if $i \neq j$

$$Var\left(\hat{\beta}_1\right) = \frac{1}{\left(\sum_{t=1}^{T} X_t^2\right)^2} Var(X_1\epsilon_1 + X_2\epsilon_2 + \ldots + X_T\epsilon_T)$$

$$= \frac{1}{\left(\sum_{t=1}^{T} X_t^2\right)^2} \left(X_1^2\sigma_\epsilon^2 + X_2^2\sigma_\epsilon^2 + \ldots + X_T^2\sigma_\epsilon^2\right)$$

$$= \frac{\sum_{t=1}^{T} X_t^2}{\left(\sum_{t=1}^{T} X_t^2\right)^2}\sigma_\epsilon^2 = \frac{\sigma_\epsilon^2}{\sum_{t=1}^{T} X_t^2}$$

Variance of $\hat{\beta}_1$

The variance of estimated coefficient $\hat{\beta}_1$ depends on the variance of error ϵ, sample size T, and the sample variance of X_t.

$$Var(\hat{\beta}_1) = \frac{\sigma_\epsilon^2 / T}{\sum_{t=1}^{T} (X_t)^2 / T} = \frac{\sigma_\epsilon^2}{T \hat{Var}(x_t)}$$

- The variance is increasing in σ_ϵ^2. More variability in the noise means more variable estimators.
- The variance is decreasing in both T and $\hat{Var}(X_t)$. Increasing $\hat{Var}(X_t)$ means that the X are spread farther apart, which makes the slope of the line easier to estimate.
- A significant practical question is whether one should use daily or weekly data, or perhaps even monthly or quarterly data. Does it matter which sampling frequency we use? Yes, the highest possible sampling frequency gives the most precise estimate of the slope, if data are stationary.

The error variance σ_ϵ^2 is not observable, usually estimated by residual variance

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{t=1}^{T} \hat{\epsilon}_t^2}{T - 2}.$$

Maximum likelihood estimation

Maximum-likelihood estimation (MLE) is to estimate the unknown parameters of a statistical model so that the particular parametric values make the observed data the most probable given the model.

➢ MLE needs an assumption of noise distribution

(1) Suppose the noises are zero-mean, normally distributed independent variables $\epsilon \sim N(0, \sigma_\epsilon^2 \boldsymbol{I})$, where σ_ϵ^2 is the common variance of the noises and $\boldsymbol{I}$ is the identity matrix.

(2) X is distributed independently of the errors ϵ.

➢ As the errors are independent draws from the same normal distribution, we can compute the log-likelihood function as follows:

$$logL = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_\epsilon^2) - \sum_{t=1}^{T}\left[\frac{(Y_t - \beta_0 - \beta_1 X_t)^2}{2\sigma_\epsilon^2}\right]$$

MLE

- The Maximum Likelihood (ML) principle requires maximization of the log-likelihood function. Maximizing the log-likelihood function entails solving the equations:

$$\frac{\partial logL}{\partial \beta_0} = 0, \frac{\partial logL}{\partial \beta_1} = 0, \frac{\partial logL}{\partial \sigma_\epsilon^2} = 0$$

- These equations can be explicitly written as follows:

$$\sum_{t=1}^{T} \left(Y_t - \beta_0 - \beta_1 X_t\right) = 0$$

$$\sum_{t=1}^{T} X_t \left(Y_t - \beta_0 - \beta_1 X_t\right) = 0$$

$$\mathrm{T}\sigma_\epsilon^2 - \sum_{t=1}^{T} \left[\left(Y_t - \beta_0 - \beta_1 X_t\right)^2\right] = 0$$

MLE

A little algebra shows that solving the first two equations yields:

$$\hat{\beta}_1 = \frac{\frac{\sum_{t=1}^{T} X_t Y_t}{T} - \bar{X}\bar{Y}}{\frac{\sum_{t=1}^{T} X_t^2}{T} - \bar{X}^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

where

$$\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t, \quad \bar{XY} = \frac{1}{T}\sum_{t=1}^{T} X_t Y_t$$

Substituting these expressions in the third equation:

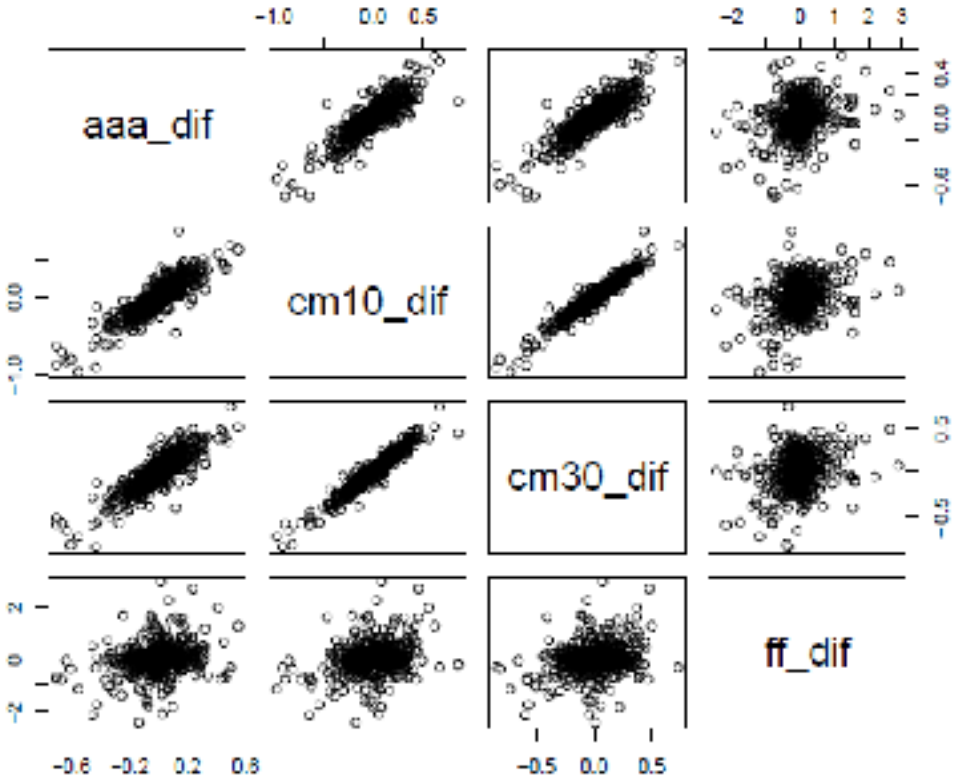
$$\frac{\partial log L}{\partial \sigma_\epsilon^2} = 0$$

yields the variance of the residuals: $\hat{\sigma}_\epsilon^2 = \frac{1}{\mathrm{T}}\sum_{t=1}^{T}\left[\left(Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t\right)^2\right]$

Example:Multiple linear regression with interest rates

We continue the analysis of the weekly interest-rate data but now with changes in 30-year Treasury rate (cm30_dif) and changes in the Federal funds rate (ff_dif) as additional predictors. Thus p = 3. Here is a scatterplot matrix of the four time series. There is a strong linear relationship between all pairs of aaa_dif, cm10_dif, and cm30 dif, but ff_dif is not strongly related to the other series.

- Data: WeekInt.txt
- 2_MultipleLinearRegression.R



Multiple linear regression models

- If there are more than one predictor, we can write the following multiple linear regression equation:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \ldots + \beta_p X_{tp} + \epsilon_t, \qquad t = 1, \ldots, T.$$

which, with condition (1), implies that :

$$E\left[Y_t \,\middle|\, X_{t1}, \ldots, X_{tp}\right] = \beta_0 + \beta_1 X_{t1} + \ldots + \beta_p X_{tp}$$

The parameter β_0 is the intercept. The regression coefficients $\beta_1, \ldots, \beta_p$ are the slopes $\beta_j = \dfrac{\partial E\left[Y_t \,\middle|\, X_{t1}, \ldots, X_{tp}\right]}{\partial X_{tj}}$. Therefore, β_j is the change in the expected value of Y_t when X_{tj} changes one unit, keeping other predictors constant.

- The assumptions of linear regression model:

(1) linearity of the conditional expectation.

(2) independent and identically noise: $\epsilon_1, \ldots, \epsilon_T$ are independent with mean zero and constant variance $Var(\epsilon_t) = \sigma_\epsilon^2$ for all t.

Ordinary least squares method

Multiple regression: $Y_t = \beta_0 + \beta_1 X_{t1} + \ldots + \beta_p X_{tp} + \epsilon_t, \qquad t = 1, \ldots, T.$

Matrix form: $E(\boldsymbol{Y} \mid \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}, \qquad \boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)'$

- In the general case of a multivariate regression, the OLS method requires minimization of the sum of the squared errors. Consider the vector of errors:

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

- The ***sum of the squared residuals (SSR)*** $(= \epsilon_1^2 + \ \ldots \ + \epsilon_T^2)$ can be written as $SSR = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$. As $\boldsymbol{\epsilon} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$, we can also write $SSR = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$
- The OLS method requires that we minimize the SSR. To do so, we equate to zero the first derivatives of the SSR:

$$\frac{\partial (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{X} = \boldsymbol{0}$$

- This is a system of N equations. Solving this system, we obtain the estimators:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Maximum likelihood estimation

Multiple regression: $Y_t = \beta_0 + \beta_1 X_{t1} + \ldots + \beta_p X_{tp} + \epsilon_t, \qquad t = 1, \ldots, T.$

Matrix form: $E(\boldsymbol{Y} \mid \boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}, \qquad \boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)'$

- We make the same set of assumptions as we made in the case of a single regressor. Using the above notation, the log-likelihood function will have the form:

$$logL = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

- The maximum likelihood conditions are written as

$$\frac{\partial logL}{\partial \boldsymbol{\beta}} = 0, \frac{\partial logL}{\partial \sigma_\epsilon^2} = 0$$

- Solving the above system of equations gives the same form for the estimators as in the univariate case:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \qquad \hat{\sigma}_\epsilon^{\,2} = \frac{1}{T}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$$

Asymptotic behaviour of estimator

- Under the assumption that the errors are normally distributed, it can be demonstrated that the regression coefficients are jointly normally distributed as follows:

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta},\ \sigma_\epsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

  This expression is important because they allow to compute confidence intervals for the regression parameters.

- The variance estimator is not unbiased. It can be demonstrated that to obtain an unbiased estimator we have to apply a correction that takes into account the number of variables by replacing *T*:

$$\hat{\sigma}_\epsilon^{\,2} = \frac{1}{T-1-p}\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$$

- Remarks: The MLE method requires that we know the functional form of the distribution. If the distribution is known but not normal, we can still apply the MLE method but the estimators will be different.

Example:Multiple linear regression with interest rates

Call:

lm(formula = aaa_dif ~ cm10_dif + cm30_dif + ff_dif)

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.07e-05   2.18e-03   -0.04     0.97
cm10_dif     3.55e-01   4.51e-02    7.86  1.1e-14 ***
cm30_dif     3.00e-01   5.00e-02    6.00  2.9e-09 ***
ff_dif       4.12e-03   5.28e-03    0.78     0.44
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.0646 on 876 degrees of freedom
Multiple R-Squared: 0.756,      Adjusted R-squared: 0.755
F-statistic:  906 on 3 and 876 DF,  p-value: <2e-16
```

We see that $\widehat{\beta}_0 = -9.07 \times 10^{-05}$, $\widehat{\beta}_1 = 0.355$, $\widehat{\beta}_2 = 0.300$, and $\widehat{\beta}_3 = 0.00412$.

Standard errors, t-values and p-values

Each of the above coefficients has three other statistics associated with it.

- Standard error (SE), which is the estimated standard deviation of the least-squares estimator and tells us the precision of the estimator.
- t-value, which is the t-statistic for testing that the coefficient is 0. The t-value is the ratio of the estimate to its standard error. For example, for cm10_dif, the t-value is 7.86 = 0.355/0.0451.
- p-value (Pr > |t| in the lm output) for testing the null hypothesis that the coefficient is 0 versus the alternative that it is not 0. If a p-value for a slope parameter is small, as it is here for β_1, then this is evidence that the corresponding coefficient is not 0, which means that the predictor has a linear relationship with the response.

The *p*-value is large (0.44) for β_3 so we would not reject the null hypothesis that β_3 is zero. This result however should not be interpreted as stating that aaa_dif and ff_dif are unrelated, but only that ff_dif is not useful for predicting aaa_dif when cm10_dif and cm30 _dif are included in the regression model. *Since the Federal Funds rate is a short-term (overnight) rate, it is not surprising that ff_dif is less useful than changes in the 10- and 30-year Treasury rates for predicting aaa_dif.*

OLS & MLE

- We now establish the relationship between the MLE principle and the ordinary least squares (OLS) method.
- OLS is a general method to approximate a relationship between two or more variables. If we use the OLS method, the assumptions of linear regressions can be weakened. In particular, we need not assume that the errors are normally distributed but only assume that they are uncorrelated and have finite variance. The errors can therefore be regarded as a white noise sequence.
- The OLS estimators are the same estimators obtained with the MLE method; they have an optimality property. In fact, the Gauss-Markov theorem states that the above OLS estimators are the ***best linear unbiased estimators (BLUE)***.
- "Best" means that no other linear unbiased estimator has a lower variance. It should be noted explicitly that OLS and MLE are conceptually different methodologies: MLE seeks the optimal parameters of the distribution of the error terms, while OLS seeks to minimize the variance of error terms. The fact that the two estimators coincide was an important discovery.

Explanatory Power of a Regression

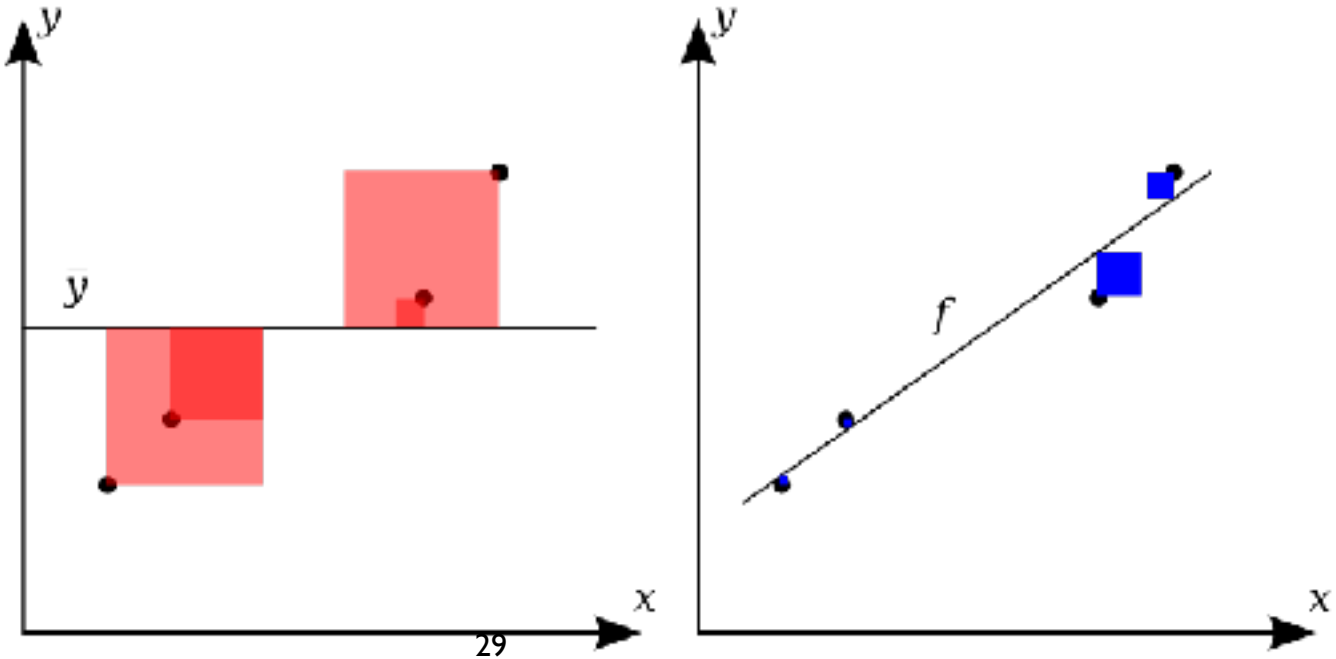
- The above computations to estimate regression parameters were carried out under the assumption that the data were generated by a linear regression function with uncorrelated and normally distributed noise. In general, we do not know if this is indeed the case.
- Though we can always estimate a linear regression model on any data sample by applying the estimators discussed above, we must now ask the question: When is a linear regression applicable and how can one establish the goodness (i.e., explanatory power) of a linear regression?
- Quite obviously, a linear regression model is applicable if the relationship between the variables is approximately linear. But
  - How can we check if this is indeed the case?
  - What happens if we fit a linear model to variables that have non-linear relationships, or if distributions are not normal?
- A number of tools have been devised to help answer these questions.

Coefficient of determination

A widely used measure of the quality and usefulness of a regression model is given by the coefficient of determination denoted by R^2 or R-squared.

- Total sum of squares (TSS): $S_Y^2 = \sum_{t=1}^{T} (Y_t - \bar{Y})^2$
- Explained sum of squares (ESS): $S_R^2 = \sum_{t=1}^{T} (\hat{Y}_t - \bar{Y})^2$
- Residual sum of squares (RSS): $S_\epsilon^2 = \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{T} \hat{\epsilon}_t^2$

TSS=ESS+ RSS



Coefficient of determination

- We can therefore define the coefficient of determination R^2 as:

$$R^2 = \frac{S_R^2}{S_Y^2}$$

and

$$1 - R^2 = \frac{S_\epsilon^2}{S_Y^2}$$

as the portion of the total fluctuation of the dependent variable, Y, explained by the regression relation.

- R^2 is a number between 0 and 1:
  - $R^2 = 0$ means that the regression has no explanatory power.
  - $R^2 = 1$ means that the regression has perfect explanatory power.
- It can be demonstrated that the coefficient of determination R^2 is distributed as the well-known Student *t* distribution. This fact allows one to determine intervals of confidence around a measure of the significance of a regression.

Adjusted R^2

- The quantity R^2 as a measure of the usefulness of a regression model suffers from the problem that a regression might fit data very well in sample but have no explanatory power out-of-sample. This occurs if the number of regressors is too high (overparametrization). Therefore an adjusted R^2 is sometimes used.
- The adjusted R^2 is defined as R^2 corrected by a penalty function that takes into account the number p of regressors in the model:

$$R^2_{Adjusted} = 1 - \left(1 - R^2\right)\frac{T-1}{T-p-1} = 1 - \frac{S^2_\epsilon/(T-p-1)}{S^2_Y/(T-1)}$$

The adjusted R^2 adjusts for the number of regressors in the model. It increases only if the new term improves the model more than would be explained by chance. It is always smaller than R^2, it can even be negative.

Mean Sums of Squares (MS) and *F*-Tests

- The ratio of the sum of squares to the degrees of freedom is the mean sum of squares:

$$mean\ sum\ of\ squares = \frac{sum\ of\ squares}{degree\ of\ freedom}$$

- Suppose we have two models, I and II, and the predictor variables in model I are a subset of those in model II, so that model I is a submodel of II. A common null hypothesis is that the data are generated by model I.
- To test this hypothesis, we use the excess regression sum of squares of model II relative to model I.

SS(II | I) = regression SS for model II - regression SS for model I

$$MS(II \mid I) = SS(II \mid I)/df_{II|I}, \quad where \qquad df_{II|I} = p_{II} - p_I$$

The *F*-statistic for testing the null hypothesis is

$$F = \frac{MS(II \mid I)}{\hat{\sigma}_\epsilon^2}$$

where $\hat{\sigma}_\epsilon^2$ is the mean residual sum of squares for model II. Under the null hypothesis, the *F*-statistic has an *F*-distribution with $df_{II|I}$ and $T - p_{II} - 1$ degrees of freedom. The null hypothesis is rejected if the *F*-statistic exceeds the α-upper quantile of this *F*-distribution.

Example: Weekly interest rates-Testing the one-predictor versus three-predictor model

```
Analysis of Variance Table

Model 1: aaa_dif ~ cm10_dif
Model 2: aaa_dif ~ cm10_dif + cm30_dif + ff_dif
  Res.Df  RSS  Df Sum of Sq    F  Pr(>F)
1    878 3.81
2    876 3.66   2      0.15 18.0 2.1e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The null hypothesis is that, in the three predictor model, the slopes of cm30_diff and ff_dif are zero. The small p value leads us to reject the null. The smaller model with only cm10_diff doesn't beat the alternative.

2_MultipleLinearRegression.R

Example: Weekly interest rates-Testing the two-predictor versus three-predictor model

```
Analysis of Variance Table

Model 1: aaa_dif ~ cm10_dif + cm30_dif
Model 2: aaa_dif ~ cm10_dif + cm30_dif + ff_dif
  Res.Df  RSS  Df Sum of Sq    F Pr(>F)
1    877 3.66
2    876 3.66   1    0.0025 0.61   0.44
```

The null hypothesis is that, in the three predictor model, the slope ff_dif is zero. We don't reject the null and concludes that the smaller model (with 2 predictors) is better.

2_MultipleLinearRegression.R

Model selection criteria

When there are many potential predictor variables, often we wish to find a subset of them that provides a parsimonious regression model.

- For linear regression models, AIC is

$$AIC = T\log\left(\hat{\sigma}_\epsilon^2\right) + 2(1+p)$$

where $1+p$ is the number of parameters in a model with p predictor variables.

$$(logL = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{Y}-\boldsymbol{X\beta}\right)'\left(\boldsymbol{Y}-\boldsymbol{X\beta}\right))$$

- BIC replaces $2(1+p)$ in AIC by $\log(T)(1+p)$. The first term, $T\log\left(\hat{\sigma}_\epsilon^2\right)$; is -2 times the log-likelihood evaluated at the MLE, assuming that the noise is Gaussian.
- Adjusted R square.
- Suppose there are M predictor variables. Let $\hat{\sigma}_{\epsilon,M}^2$ be the estimate of σ_ϵ^2 using all of them, and let $SSE(p)$ be the sum of squares for residual error for a model with some subset of only $p \leq M$ of the predictors.

$$C_p = \frac{SSE(p)}{\hat{\sigma}_{\epsilon,M}^2} - T + 2(p+1)$$

With C_p, AIC, and BIC, smaller values are better, but for adjusted R^2, larger values are better.

Example: Weekly interest rates-Model selection by AIC and BIC

"Best" means smallest for BIC and C_p and largest for adjusted R^2. There are three plots, one for each of BIC, C_p, and adjusted R^2. All three criteria are optimized by two predictor variables.

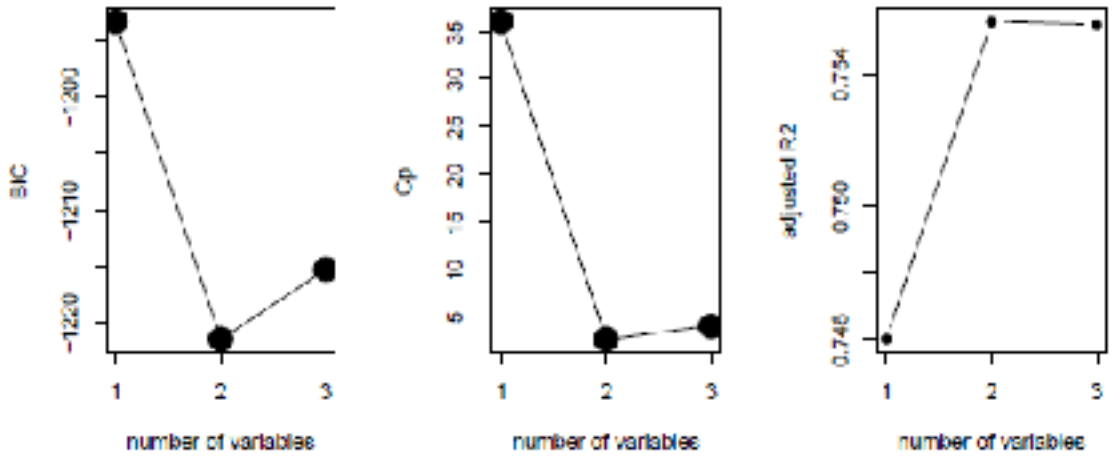


Fig. 12.5. *Changes in weekly interest rates. Plots for model selection.*

2_WeeklyInterestRatesModelSelection.R

Pitfalls of Regressions

It is important to understand when regressions are correctly applicable and when they are not. There are several situations where it would be inappropriate to use regressions. In particular, we analyze the following cases which represent possible pitfalls of regressions:

- Multicollinearlty
- Nonlinearity
- Nonnormality
- High leverage and outliers
- Autocorrelations
- Heteroscedasticity

Collinearity

Predictor variables exhibit *collinearity* when one of the predictors can be predicted well from the others.

Consequences of collinearity:
Coefficients in a multiple regression model can be surprising, taking on an unanticipated sign or being unexpectedly large or small.
The stronger the correlation between coefficients, the more the variance of their coefficients increases when both are included in the model. This can lead to a smaller t-statistic and correspondingly large P-value.

Collinearity

Housing Prices based on *Living Area* and *Bedrooms*.

Simple Regression Models

Variable	Coeff	SE(Coeff)	t-ratio	P-value
Intercept	13439.394	4992.353	2.692	0.00717
Living Area	113.123	2.682	42.173	<0.0001

An increase of $113.12 in price for each additional square foot of space.

Variable	Coeff	SE(Coeff)	t ratio	P value
Intercept	59063	9057	6.915	<0.0001
Bedrooms	48218	2656	18.151	<0.0001

An increase of $48,218 in price for each additional bedroom.

Multiple Regression Model

Variable	Coeff	SE(Coeff)	t ratio	P value
Intercept	[illegible]	[illegible]	[illegible]	[illegible]
Living Area	125.405	3.527	35.555	<0.0001
Bedrooms	−14196.769	[illegible]	[illegible]	[illegible]

The **coefficient** on *Bedrooms* seems *counterintuitive*.

39

Collinearity and variance inflation

- Note that $var\left(\hat{\beta}_j\right)$ is σ_ϵ^2 multiplied by the corresponding (j+1,j+1)-th element of the main diagonal of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. After operating, the variance of the j-th predictor can be expressed as

$$var\left(\hat{\beta}_j\right) = \frac{\sigma_\epsilon^2}{T\hat{Var}\left(X_j\right)} \frac{1}{1 - R_j^2}$$

where R_j^2 measures how well X_j can be predicted from the other Xs, i.e. regressing X_j on the $p - 1$ other predictors.

- Define ***variance inflation factor (VIF)*** of X_j as:

$$VIF_j = \frac{1}{1 - R_j^2} = \frac{TSS\left(X_j\right)}{ESS(X\_j \mid \boldsymbol{X}_{-j})}$$

The VIF of a variable tells us how much the variance of $\hat{\beta}_j$ of the j-th predictor variable X_j is increased by having the other predictor variables $X_1, \ldots X_{j-1}, X_{(j+1)}, \ldots X_P$ in the model.

For example, if a variable has a VIF of 4, then the variance of its $\hat{\beta}$ is four times larger than it would be if the other predictors were either deleted or were not correlated with it. The standard error is increased by a factor of 2.

- It is important to keep in mind that VIF_j tells us nothing about the relationship between the response and *j*th predictor. Rather, it tells us only how correlated the *j*th predictor is with the other predictors.

Partial residual plot

Partial residual plot is a graphical technique that attempts to show the relationship between a given independent variable X_j and the response variable Y given that other independent variables are also in the model. The partial residual for the jth predictor variable is

$$Y_t - \left(\hat{\beta}_0 + \sum_{j' \neq j} X_{j't} \hat{\beta}_{j'} \right) = \hat{Y}_t + \hat{\epsilon}_t - \left(\hat{\beta}_0 + \sum_{j' \neq j} X_{j't} \hat{\beta}_{j'} \right) = X_{jt} \hat{\beta}_j + \hat{\epsilon}_t$$

(The scatter plot of X_j and Y does not take into account the effect of the other independent variables).

Example: Partial residual plots for the weekly interest-rate example

- Partial residual plots for the weekly interest-rate example are shown in Figures 12.7(a) and (b). For comparison, scatterplots of cm10_dif and cm30_dif versus aaa_dif with the corresponding one-variable fitted lines are shown in panels (c) and (d).
- The main conclusion from examining the plots is that the slopes in (a) and (b) are nonzero, though they are shallower than the slopes in (c) and (d).

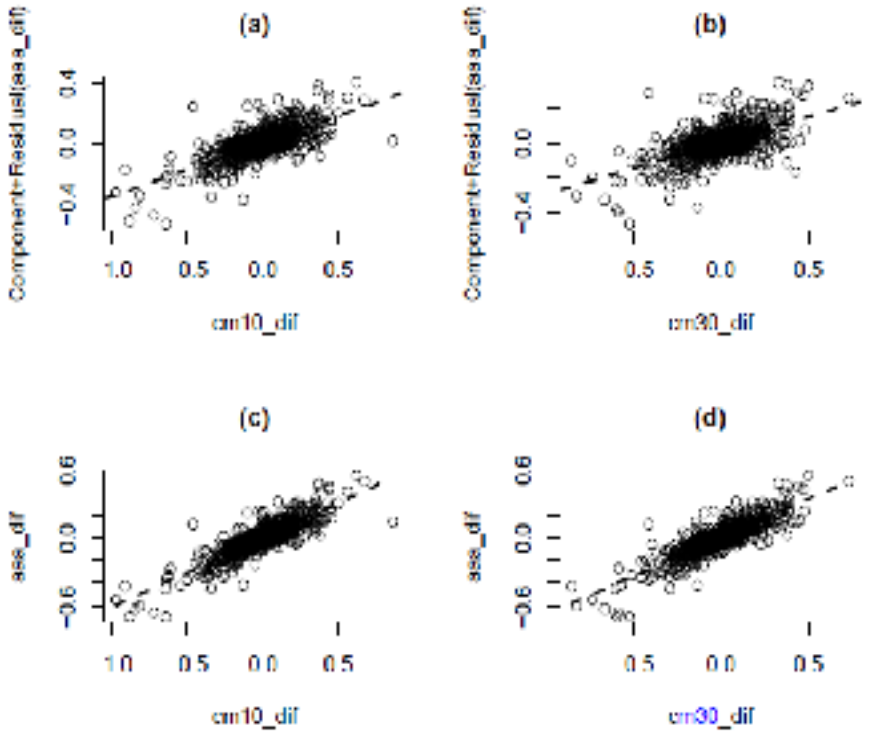


Fig. 12.7. Partial residual plots for the weekly interest rates [panels (a) and (b)] and scatterplots of the predictors and the response [panels (c) and (d)].

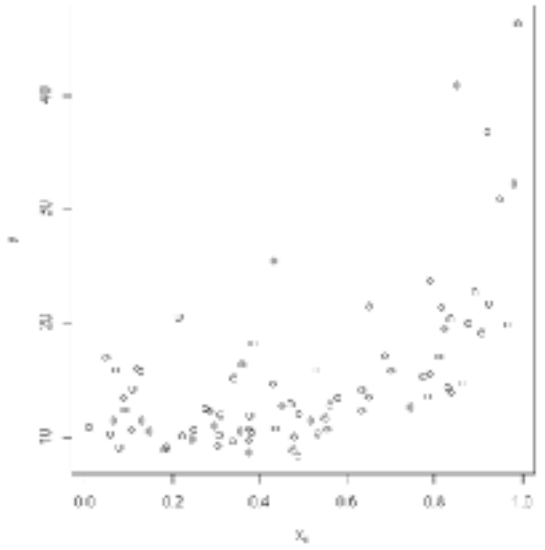
Due to collinearity, the effect of cm10_dif on aaa_dif when cm30_dif is in the model [panel (a)] is less than when cm30_dif is not in the model [panel (c)]

2_WeeklyInterestRatesPartialResidual.R

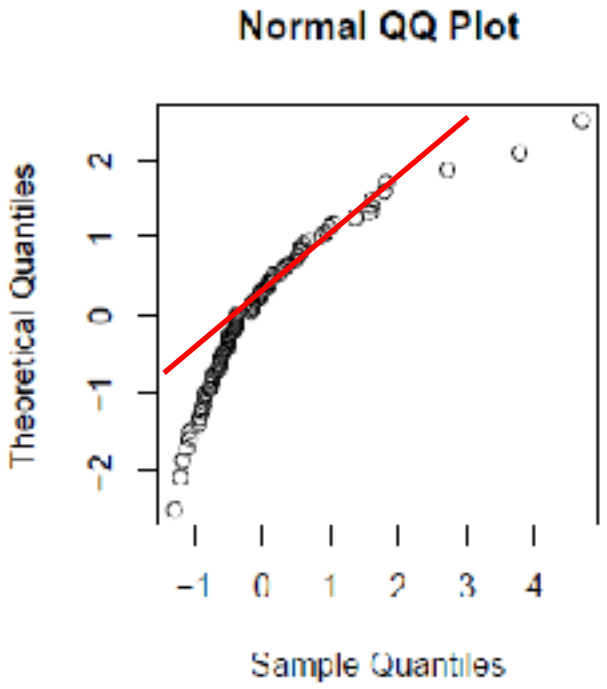
Detecting nonlinearity

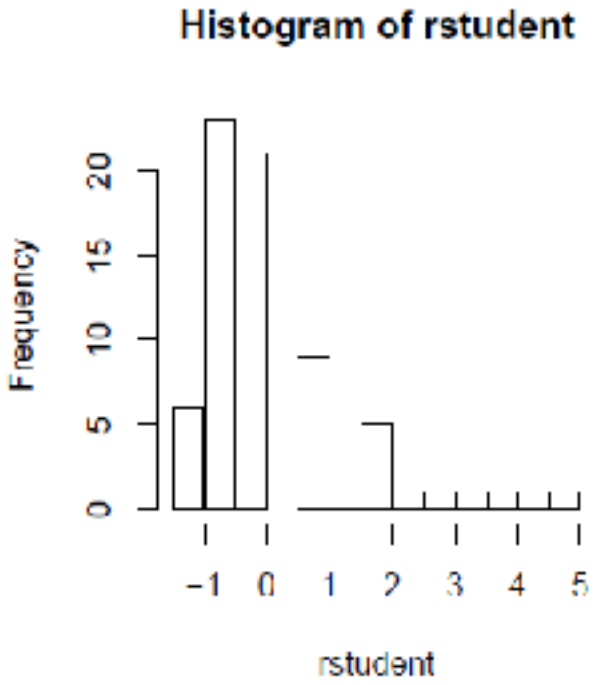
- Data were simulated to illustrate some of the techniques for diagnosing problems. In the example there are two predictor variables, X_1 and X_2. The assumed model is multiple linear regression, $Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i$.
- The scatterplot suggests non-linearity. One possibility is to take appropriate transformation of data, e.g. log transformation. Alternatively, one can adopt a non-linear model to reflect the empirical features of data.

❑ R: 2_DetectingNonlinearity.R



Detecting non-normality





R: 2_DetectingNonlinearity.R

Example: Residual plots for weekly interest changes

The normal plot in panel (a) shows heavy tails. A *t*-distribution was fit to the residuals, and the estimated degrees of freedom was 2.99, again indicating heavy tails. Panel (b) shows a QQ plot of the residuals and the quantiles of the fitted *t*-distribution with a 45o reference line. There is an excellent agreement between the data and the *t*-distribution.

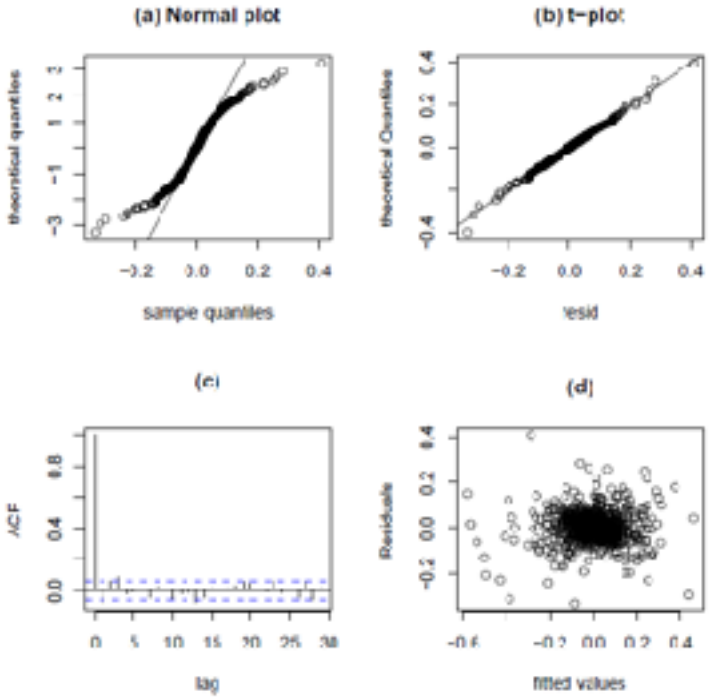
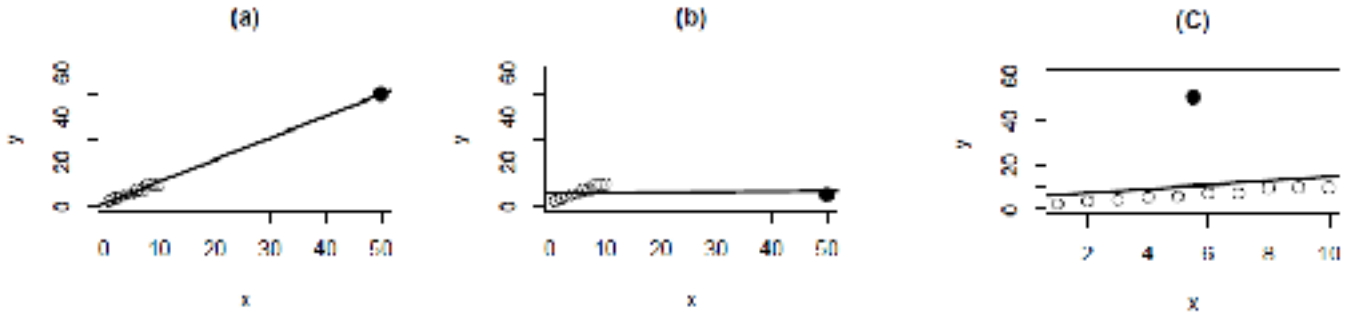


Fig. 13.10. *Residual plots for the regression of* aaa_dif *on* cm10_dif *and* cm30_dif.

- Data: 2_WeekInt.txt
- R: 2_ResidualPlotsForWeeklyInterestChanges.R

High-leverage points and residual outliers- Simulated data example

A high-leverage point is not necessarily a problem, only a potential problem.



In panel (a), Y is linearly related to X and the extreme X-value is, in fact, helpful as it increases the precision of the estimated slope.

In panel (b), the value of Y for the high-leverage point has been misrecorded as 5.254 rather than 50.254.This data point is called a residual outlier and has an extreme influence on the estimated slope.

In panel (c), X has been misrecorded for the high-leverage point as 5.5 instead of 50. Thus, this point is no longer high-leverage, but now it is a residual outlier. Its effect now is to bias the estimated intercept.

High-leverage points

Three important tools are often used for diagnosing problems of high-leverage points:

- leverages;
- externally studentized residuals; and
- **Cook's D**, which quantifies the overall influence of each observation on the fitted values.

Cook's D measures influence, and any case with a large Cook's D is called a high-influence case. Leverage and rstudent alone do not measure influence. Let $\hat{Y}_j(-i)$ be the j-th fitted value using estimates of the $\hat{\beta}$ s obtained with the i-th observation deleted. Then Cook's D for the i-th observation is

$$\frac{\sum_{j=1}^{n}\left\{\hat{Y}_j - \hat{Y}_j(-i)\right\}^2}{(p+1)\,\hat{\sigma}_\epsilon^2}$$

One way to use Cook's D is to plot the values of Cook's D against case number and look for unusually large values.

Example: Diagnostics

Residual plots and other diagnostics are shown in Figure 13.13 for a regression of *Y* on *X*. Describe any problems that you see and possible remedies.

There is a high-leverage point with a residual outlier. In panel (f) we see that this point is #58. One should investigate whether this observation is correct. Regardless of whether this observation is correct, it is so far detached from the other data that it is sensible to remove it and fit a model to the other data.

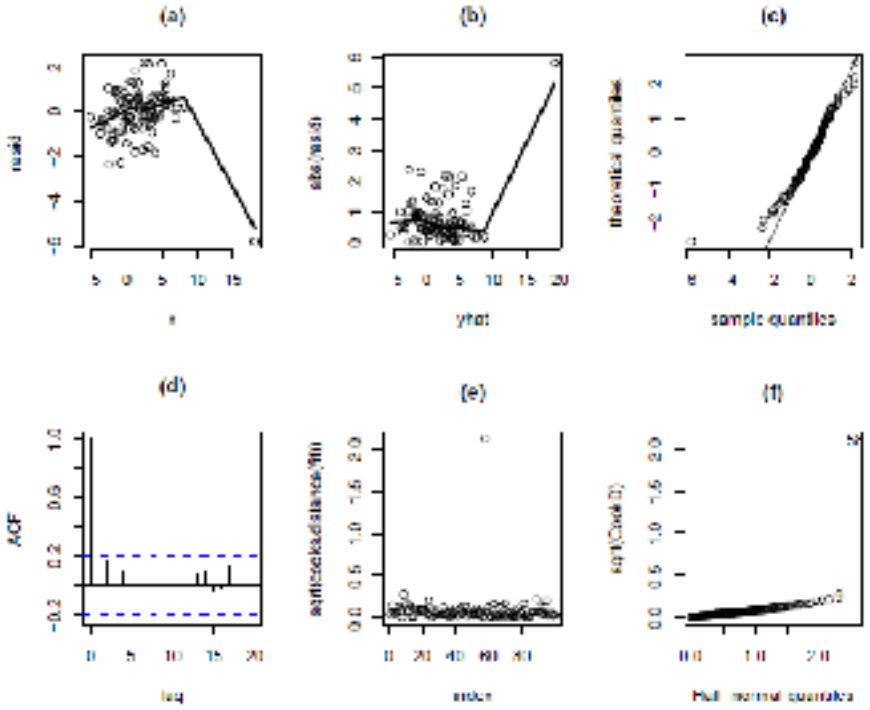
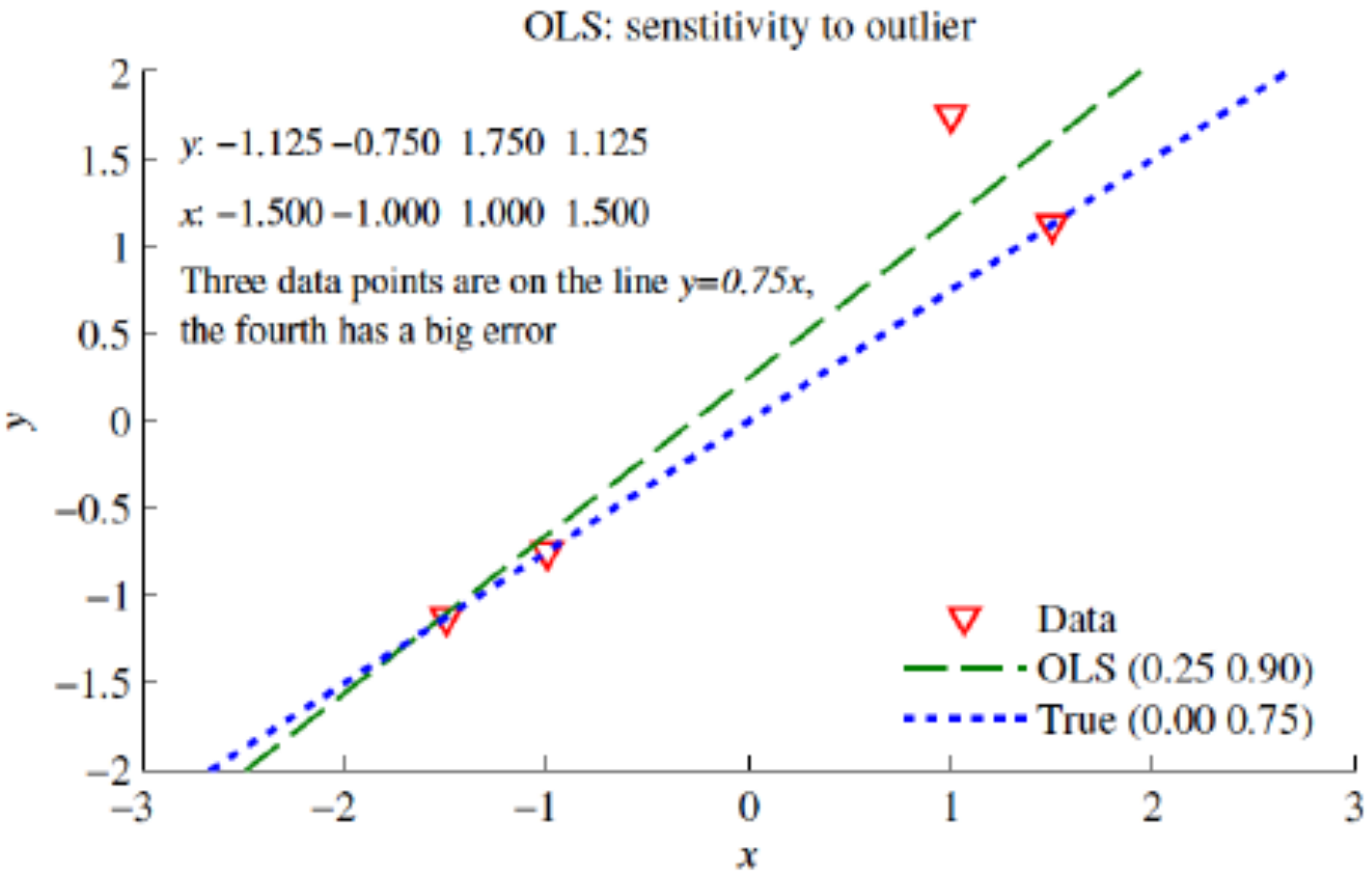


Fig. 13.13. *Residual plots and diagnostics for regression of Y on X in Problem 2. The residuals are studentized values. (a) Plot of residual versus x. (b) Plot of absolute residuals versus fitted values. (c) Normal QQ plot of residuals. (d) ACF plot of residuals. (e) Plot of the square root of Cook's D versus index (= observation number). (f) Half-normal plot of square root of Cook's D.*

Outliers

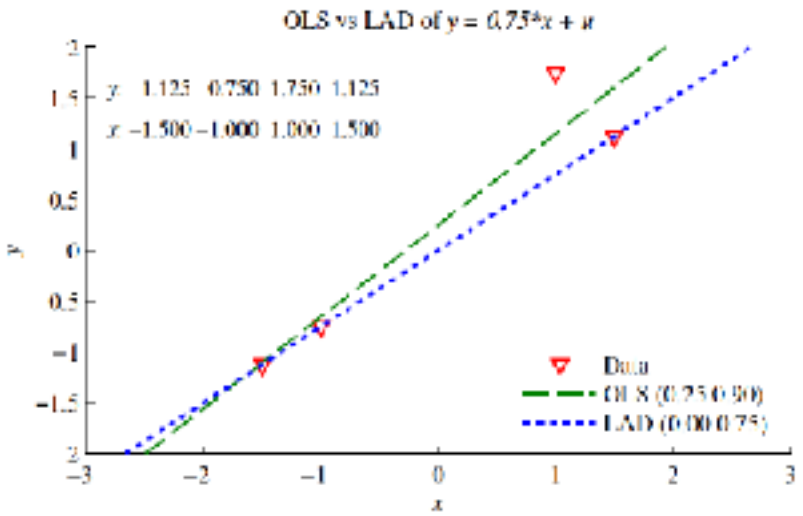

OLS: senstitivity to outlier
y: −1.125 −0.750 1.750 1.125
x: −1.500 −1.000 1.000 1.500
Three data points are on the line y=0.75x,
the fourth has a big error
Data
OLS (0.25 0.90)
True (0.00 0.75)
y
x
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−3
−2
−1
0
1
2
3

Robust regression estimation method

The OLS estimator is sensitive to outliers.

The least absolute derivation (LAD) estimator is less sensitive to outliers.

$$\hat{\beta}_{LAD} = argmin\left(\sum_{t=1}^{T} |y_t - b_0 - b_1 x_t|\right)$$



Example: Diagnostics

Residual plots and other diagnostics are shown in Figure 13.12 for a regression of *Y* on *X*. Describe any problems that you see and possible remedies.

Panel (a) shows clearly that the effect of X on Y is nonlinear. Because of the strong bias caused by the nonlinearity, the residuals are biased estimates of the noise and examination of the remaining plots is not useful. The remedy to this problem is to fit a nonlinear effect. A model that is quadratic in X seems like a good choice. After this model has been fit, the other diagnostic plots should be examined.

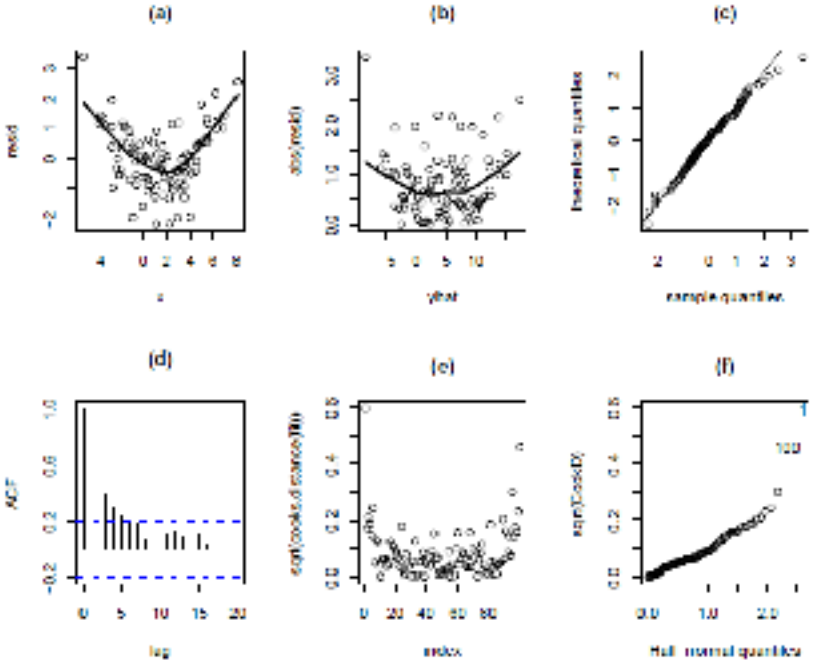


Fig. 13.12. *Residual plots and diagnostics for regression of Y on X in Problem 1. The residuals are studentized values. (a) Plot of residuals versus x. (b) Plot of absolute residuals versus fitted values. (c) Normal QQ plot of residuals. (d) ACF plot of residuals. (e) Plot of the square root of Cook's D versus index (= observation number). (f) Half-normal plot of square root of Cook's D.*

R lab

This section uses the data set USMacroG in R's AER package. This data set contains quarterly times series on 12 U.S. macroeconomic variables for the period 1950-2000. We will use the variables:

consumption = real consumption expenditures,
dpi = real disposable personal income,
cpi = consumer price index (inflation rate)
government = real government expenditures, and
unemp = unemployment rate.

Our goal is to predict changes in consumption from changes in the other variables.

- Data: Rlab2_1_USMacroG.txt
- R: Rlab2_1.R

R lab

Run the following R code to load the data, difference the data (since we wish to work with changes in these variables), and create a scatterplot matrix.

```
library(AER)
data("USMacroG")
MacroDiff= apply(USMacroG,2,diff)
pairs(cbind(consumption,dpi,cpi,government,unemp))
```

Problem 1. Describe any interesting features, such as, outliers, seen in the scatterplot matrix. Keep in mind that the goal is to predict changes in consumption. Which variables seem best suited for that purpose? Do you think there will be collinearity problems?

R lab

Next, run the code below to fit a multiple linear regression model to consumption using the other four variables as predictors.

```
fitLm1 = lm(consumption~dpi+cpi+government+unemp)
summary(fitLm1)
confint(fitLm1)
```

Problem 2 From the summary, which variables seem useful for predicting changes in consumption?

R lab

Next, print an AOV table.

```
anova(fitLm1)
```

Problem 3 For the purpose of variable selection, does the AOV table provide any useful information not already in the summary?

Upon examination of the p-values, we might be tempted to drop several variables from the regression model, but we will not do that since variables should be removed from a model one at a time. The reason is that, due to correlation between the predictors, when one is removed then the significance of the others changes. To remove variables sequentially, we will use the function stepAIC in the MASS package.

```
library(MASS)
fitLm2 = stepAIC(fitLm1)
summary(fitLm2)
```

R lab

Problem 4 Which variables are removed from the model, and in what order?

Now compare the initial and final models by AIC.

```
AIC(fitLm1)
AIC(fitLm2)
AIC(fitLm1)-AIC(fitLm2)
```

Problem 5 How much of an improvement in AIC was achieved by removing variables? Was the improvement huge? Is so, can you suggest why? If not, why not?

The function vif in the car package will compute variance inflation factors. A similar function with the same name is in the faraway package. Run

```
library(car)
vif(fitLm1)
vif(fitLm2)
```

Problem 6 Was there much collinearity in the original four-variable model? Was the collinearity reduced much by dropping two variables?

R lab

Partial residual plots, which are also called component plus residual or cr plots, can be constructed using the function cr.plot in the car package. Run

```
par(mfrow=c(2,2))
sp = 0.8
cr.plot(fitLm1,dpi,span=sp,col="black")
cr.plot(fitLm1,cpi,span=sp,col="black")
cr.plot(fitLm1,government,span=sp,col="black")
cr.plot(fitLm1,unemp,span=sp,col="black")
```

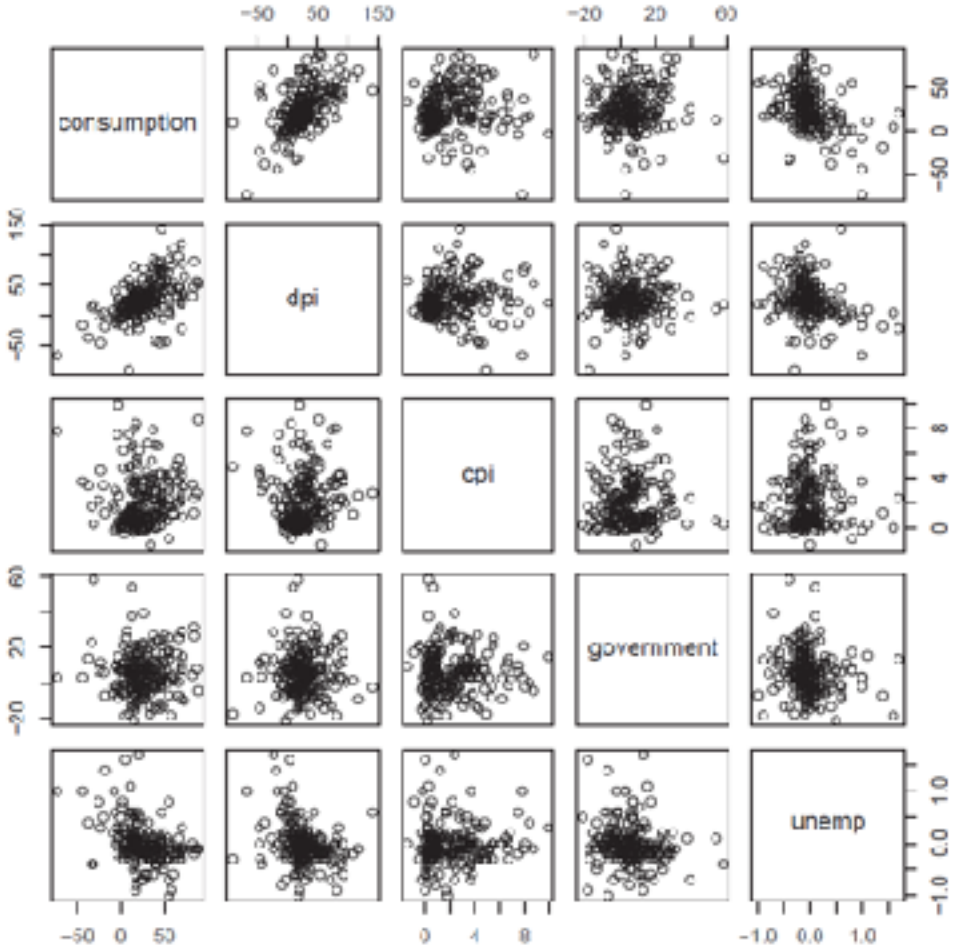
Besides dashed least-squares lines, the partial residual plots have solid lowess smooths through them unless this feature is turned off by specifying smooth=F. Lowess is an earlier version of loess. The smoothness of the lowess curves is determined by the parameter span, with larger values of span giving smoother plots. The default is span = 0.5. In the code above, span is 0.8 but can be changed for all four plots by changing the variable sp. A substantial deviation of the lowess curve from the least-squares line is an indication that the effect of the predictor is nonlinear. The default color of the cr.plot figure is red, but this can be changed as in the code above.

Problem 7 What conclusions can you draw from the partial residual plots?

R lab – Results & Discussions

Problem 1. Describe any interesting features, such as, outliers, seen in the scatterplot matrix. Keep in mind that the goal is to predict changes in consumption. Which variables seem best suited for that purpose? Do you think there will be collinearity problems?

No outliers are seen in the scatterplots. Changes in consumption show a positive relationship with changes in dpi and a negative relationship with changes in unemp, so these two variables should be most useful for predicting changes in consumption. The correlations between the predictors (changes in the variables other than consumption) are weak and collinearity will not be a serious problem.



R lab – Results & Discussions

Problem 2 From the summary, which variables seem useful for predicting changes in consumption?

```
Call:
lm(formula = consumption ~ dpi + cpi + government + unemp)

Residuals:
    Min      1Q  Median      3Q     Max
-60.626 -12.203  -2.678   9.862  59.283

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.752317   2.520168   5.854 1.97e-08 ***
dpi           0.353044   0.047982   7.358 4.87e-12 ***
cpi           0.726576   0.678754   1.070    0.286
government   -0.002158   0.118142  -0.018    0.985
unemp       -16.304368   3.855214  -4.229 3.58e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 20.39 on 198 degrees of freedom
Multiple R-squared: 0.3385,     Adjusted R-squared: 0.3252
F-statistic: 25.33 on 4 and 198 DF,  p-value: < 2.2e-16
```

Changes in dpi and unemp are highly significant and so are useful for prediction. Changes in cpi and government have large p-values and do not seem useful.

R lab – Results & Discussions

Problem 3 For the purpose of variable selection, does the AOV table provide any useful information not already in the summary?

```
Response: consumption
             Df Sum Sq Mean Sq F value    Pr(>F)
dpi           1  34258   34258 82.4294 < 2.2e-16 ***
cpi           1    253     253  0.6089    0.4361
government    1    171     171  0.4110    0.5222
unemp         1   7434    7434 17.8859 3.582e-05 ***
Residuals   198  82290     416
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The AOV table contains sums of squares and mean squares that are not in the summary. These are however not needed for variable selection.

R lab – Results & Discussions

Problem 4 Which variables are removed from the model, and in what order?

```
> fitLm2 = stepAIC(fitLm1)
Start:  AIC=1228.98
consumption ~ dpi + cpi + government + unemp

             Df Sum of Sq    RSS    AIC
- government  1    0.1387  82291   1227
- cpi         1       476  82767   1228
<none>                     82290   1229
- unemp       1      7434  89724   1245
- dpi         1     22500 104790   1276

Step:  AIC=1226.98
consumption ~ dpi + cpi + unemp

          Df Sum of Sq    RSS    AIC
- cpi      1       476  82767   1226
<none>                  82291   1227
- unemp    1      7604  89895   1243
- dpi      1     22542 104833   1274

Step:  AIC=1226.15
consumption ~ dpi + unemp

          Df Sum of Sq    RSS    AIC
<none>                  82767   1226
- unemp    1      7381  90148   1241
- dpi      1     22932 105699   1274
```

First changes in government is removed and then changes in cpi.

```
Call:
lm(formula = consumption ~ dpi + unemp)

Residuals:
    Min      1Q  Median      3Q     Max
-60.892 -12.660  -3.085   9.737  59.374

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.28476    1.91084   8.522 3.79e-15 ***
dpi          0.35567    0.04778   7.444 2.84e-12 ***
unemp      -16.01489    3.79216  -4.223 3.66e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.34 on 200 degrees of freedom
Multiple R-squared: 0.3347,     Adjusted R-squared: 0.3281
F-statistic: 50.31 on 2 and 200 DF,  p-value: < 2.2e-16
```

R lab – Results & Discussions

Problem 5 How much of an improvement in AIC was achieved by removing variables? Was the improvement huge? Is so, can you suggest why? If not, why not?

AIC decreased by 2.83 which is not a huge improvement. Dropping variables decreases the log-likelihood (which increases AIC) and decreases the number of variables (which decreases AIC). The decrease due to dropping variables is limited; it is twice the number of deleted variables. In this case, the maximum possible decrease in AIC from dropping variables is 4 and is achieved only if dropping the variables does not change the log-likelihood, so we should not have expected a huge decrease. Of course, when there are many variables then a huge decrease in AIC is possible if a very large number of variables can be dropped.

```
> AIC(fitLm1)
[1] 1807.064
> AIC(fitLm2)
[1] 1804.237
> AIC(fitLm1)-AIC(fitLm2)
[1] 2.827648
```

R lab – Results & Discussions

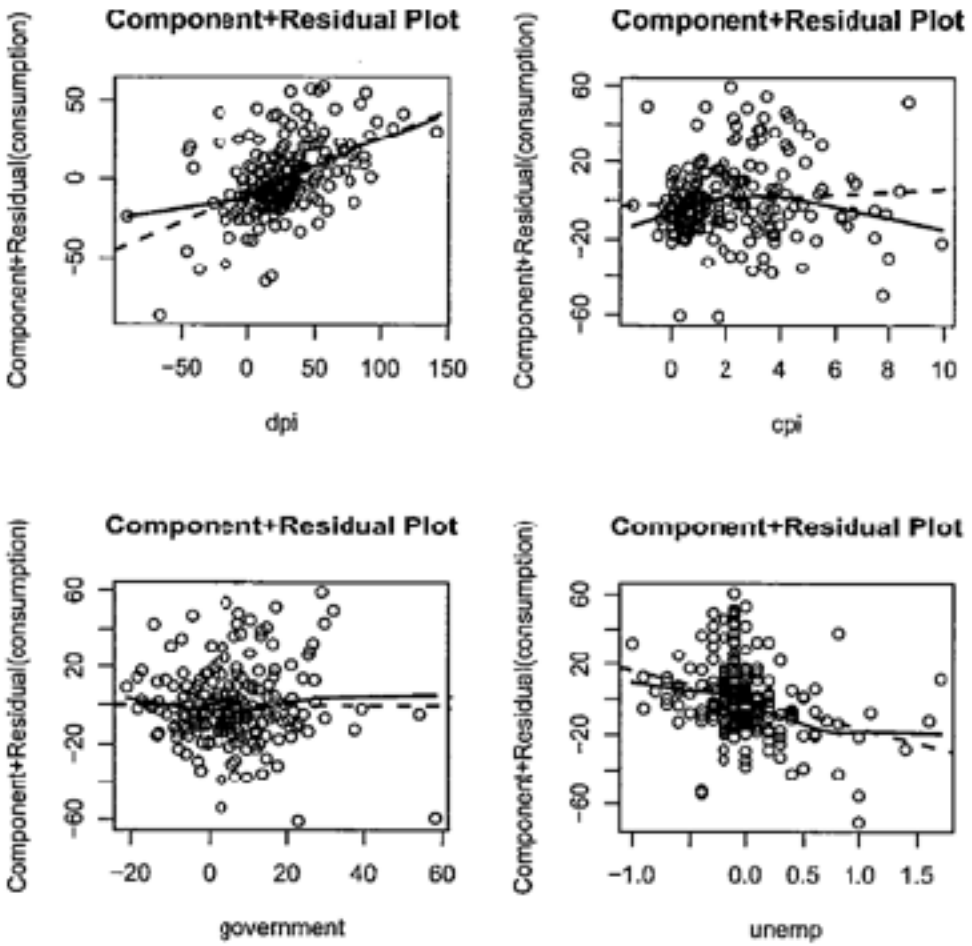
Problem 6 Was there much collinearity in the original four-variable model? Was the collinearity reduced much by dropping two variables?

```
> vif(fitLm1)
      dpi        cpi government      unemp
 1.100321   1.005814   1.024822   1.127610
> vif(fitLm2)
     dpi    unemp
1.095699 1.095699
```

There was little collinearity in the original model, since all four VIFs are near their lower bound of 1. Since there was little collinearity to begin with, it could not be much reduced.

R lab – Results & Discussions

Problem 7 What conclusions can you draw from the partial residual plots?



The least-squares lines for government and cpi are nearly horizontal, which agrees with the earlier result that these variables can be dropped. The lowess curves are close to the least-squares lines, at least relative to the random variation in the partial residuals, and this indicates that the effects of dpi and unemp on consumption are linear.