

5.2 Least squares estimation

In practice, of course, we have a collection of observations but we do not know the values of the coefficients $\beta_0, \beta_1, \dots, \beta_k$. These need to be estimated from the data.

The least squares principle provides a way of choosing the coefficients effectively by minimising the sum of the squared errors. That is, we choose the values of $\beta_0, \beta_1, \dots, \beta_k$ that minimise

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_k x_{k,t})^2.$$

This is called **least squares** estimation because it gives the least value for the sum of squared errors. Finding the best estimates of the coefficients is often called “fitting” the model to the data, or sometimes “learning” or “training” the model. The line shown in Figure 5.3 was obtained in this way.

When we refer to the *estimated* coefficients, we will use the notation $\hat{\beta}_0, \dots, \hat{\beta}_k$. The equations for these will be given in Section 5.7.

The `tslm()` function fits a linear regression model to time series data. It is similar to the `lm()` function which is widely used for linear models, but `tslm()` provides additional facilities for handling time series.

Example: US consumption expenditure

A multiple linear regression model for US consumption is

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \varepsilon_t,$$

where y is the percentage change in real personal consumption expenditure, x_1 is the percentage change in real personal disposable income, x_2 is the percentage change in industrial production, x_3 is the percentage change in personal savings and x_4 is the change in the unemployment rate.

The following output provides information about the fitted model. The first column of `Coefficients` gives an estimate of each β coefficient and the second column gives its standard error (i.e., the standard deviation which would be obtained from repeatedly estimating the β coefficients on similar data sets). The standard error gives a measure of the uncertainty in the estimated β coefficient.

```
fit.consMR <- tslm(
  Consumption ~ Income + Production + Unemployment + Savings,
  data=uschange)
summary(fit.consMR)
#>
#> Call:
#> tslm(formula = Consumption ~ Income + Production + Unemployment +
#>       Savings, data = uschange)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -0.8830 -0.1764 -0.0368  0.1525  1.2055
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   0.26729    0.03721   7.18 1.7e-11 ***
#> Income        0.71448    0.04219  16.93 < 2e-16 ***
#> Production    0.04589    0.02588   1.77  0.078 .
#> Unemployment -0.20477    0.10550  -1.94  0.054 .
#> Savings       -0.04527    0.00278 -16.29 < 2e-16 ***
#> ---
#> Signif. codes:
#> 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.329 on 182 degrees of freedom
#> Multiple R-squared:  0.754, Adjusted R-squared:  0.749
#> F-statistic: 139 on 4 and 182 DF, p-value: <2e-16
```

For forecasting purposes, the final two columns are of limited interest. The “t value” is the ratio of an estimated β coefficient to its standard error and the last column gives the p-value: the probability of the estimated β coefficient being as large as it is if there

was no real relationship between consumption and the corresponding predictor. This is useful when studying the effect of each predictor, but is not particularly useful for forecasting.

Fitted values

Predictions of y can be obtained by using the estimated coefficients in the regression equation and setting the error term to zero. In general we write,

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t}. \quad (5.2)$$

Plugging in the values of $x_{1,t}, \dots, x_{k,t}$ for $t = 1, \dots, T$ returns predictions of y_t within the training-sample, referred to as *fitted values*. Note that these are predictions of the data used to estimate the model, not genuine forecasts of future values of y .

The following plots show the actual values compared to the fitted values for the percentage change in the US consumption expenditure series. The time plot in Figure 5.6 shows that the fitted values follow the actual data fairly closely. This is verified by the strong positive relationship shown by the scatterplot in Figure 5.7.

```
autoplot(uschange[, 'Consumption'], series="Data") +
  autolayer(fitted(fit.consMR), series="Fitted") +
  xlab("Year") + ylab("") +
  ggtitle("Percent change in US consumption expenditure") +
  guides(colour=guide_legend(title=" "))
```

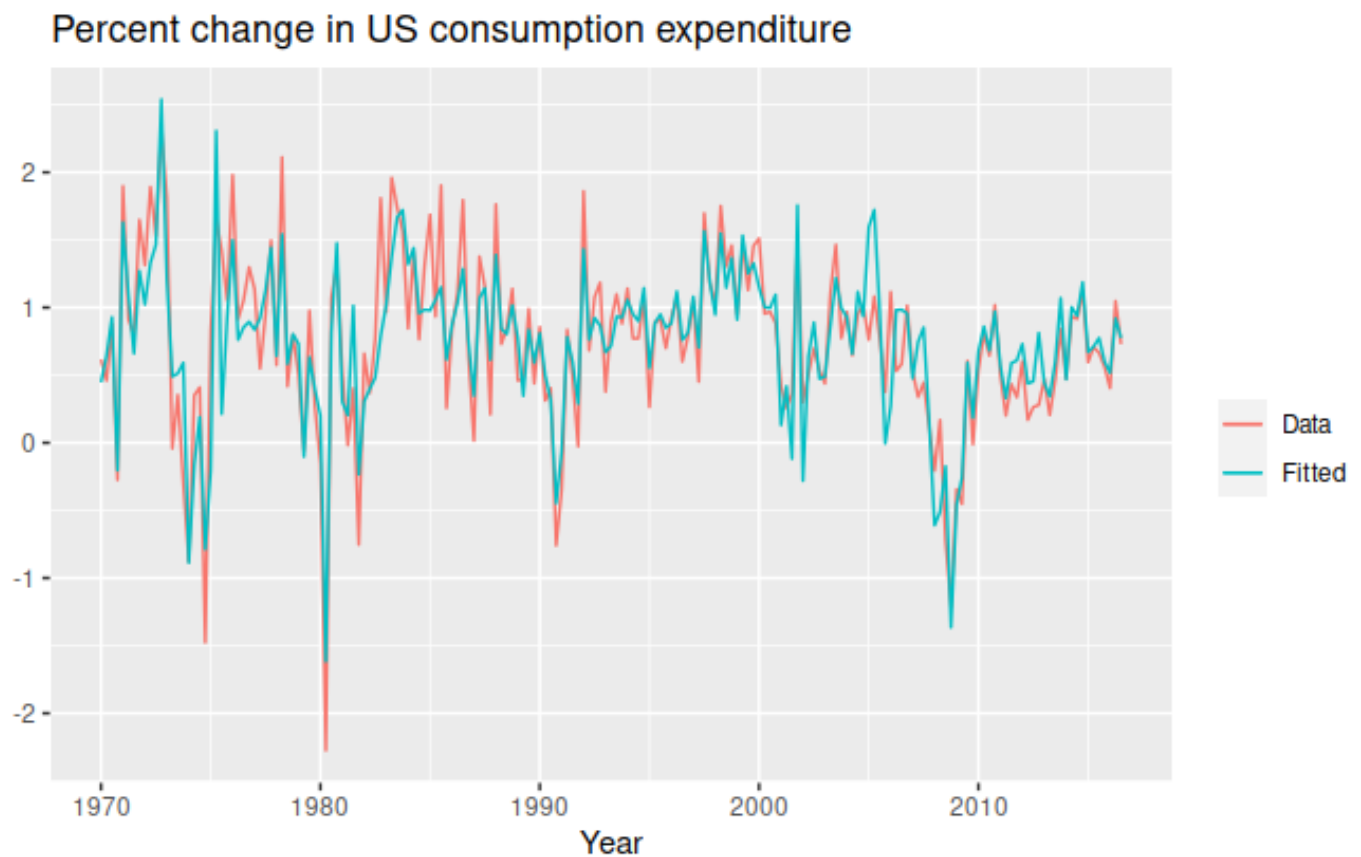


Figure 5.6: Time plot of actual US consumption expenditure and predicted US consumption expenditure.

```
cbind(Data = uschange[, "Consumption"],
      Fitted = fitted(fit.consMR)) %>%
as.data.frame() %>%
ggplot(aes(x=Data, y=Fitted)) +
  geom_point() +
  ylab("Fitted (predicted values)") +
  xlab("Data (actual values)") +
  ggtitle("Percent change in US consumption expenditure") +
  geom_abline(intercept=0, slope=1)
```

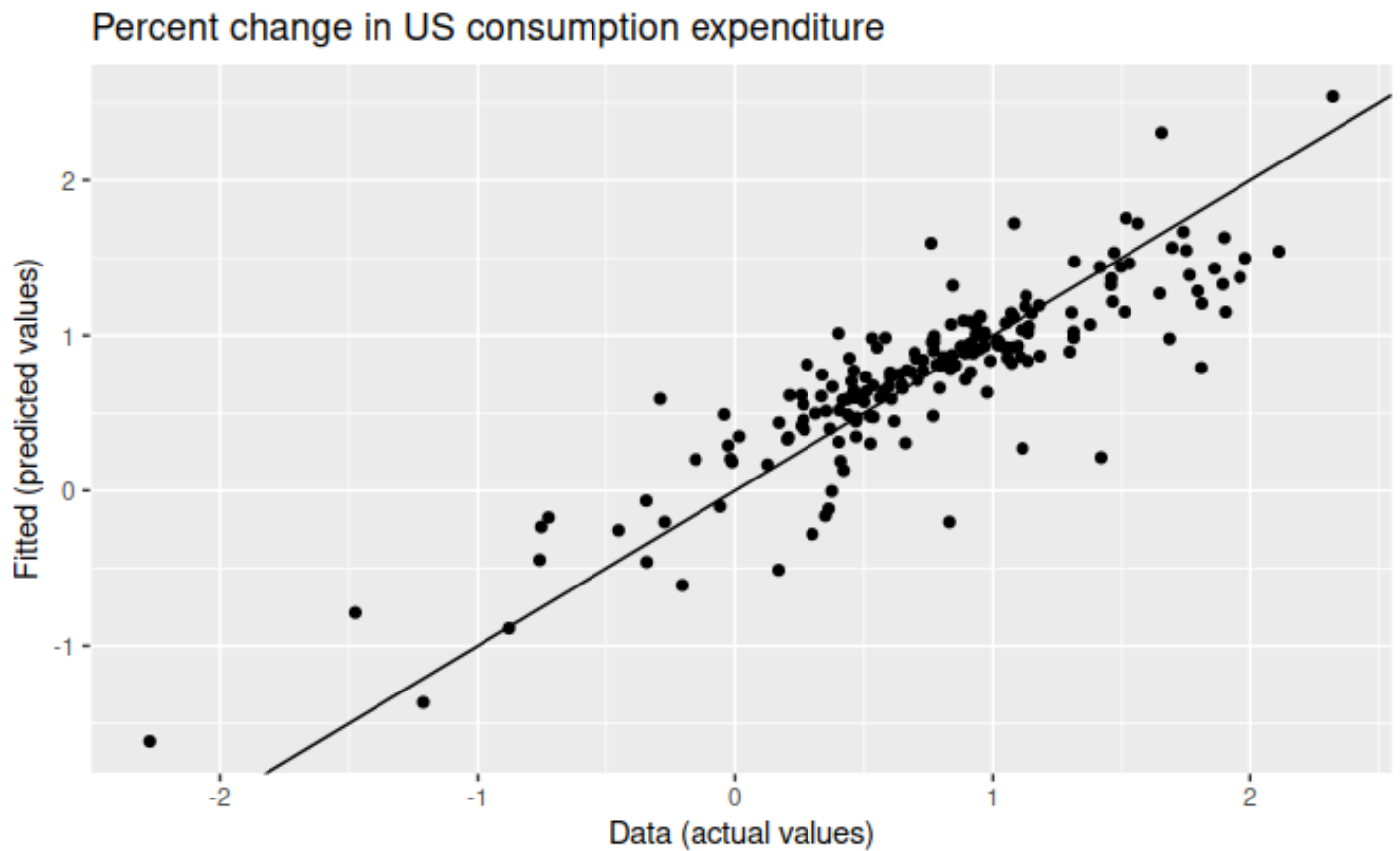


Figure 5.7: Actual US consumption expenditure plotted against predicted US consumption expenditure.

Goodness-of-fit

A common way to summarise how well a linear regression model fits the data is via the coefficient of determination, or R^2 . This can be calculated as the square of the correlation between the observed y values and the predicted \hat{y} values. Alternatively, it can also be calculated as,

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2},$$

where the summations are over all observations. Thus, it reflects the proportion of variation in the forecast variable that is accounted for (or explained) by the regression model.

In simple linear regression, the value of R^2 is also equal to the square of the correlation between y and x (provided an intercept has been included).

If the predictions are close to the actual values, we would expect R^2 to be close to 1. On the other hand, if the predictions are unrelated to the actual values, then $R^2 = 0$ (again, assuming there is an intercept). In all cases, R^2 lies between 0 and 1.

The R^2 value is used frequently, though often incorrectly, in forecasting. The value of R^2 will never decrease when adding an extra predictor to the model and this can lead to over-fitting. There are no set rules for what is a good R^2 value, and typical values of R^2 depend on the type of data used. Validating a model's forecasting performance on the test data is much better than measuring the R^2 value on the training data.

Example: US consumption expenditure

Figure 5.7 plots the actual consumption expenditure values versus the fitted values. The correlation between these variables is $r = 0.868$ hence $R^2 = 0.754$ (shown in the output above). In this case model does an excellent job as it explains 75.4% of the variation in the consumption data. Compare that to the R^2 value of 0.16 obtained from the simple regression with the same data set in Section 5.1. Adding the three extra predictors has allowed a lot more of the variation in the consumption data to be explained.

Standard error of the regression

Another measure of how well the model has fitted the data is the standard deviation of the residuals, which is often known as the “residual standard error.” This is shown in the above output with the value 0.329.

It is calculated using

$$\hat{\sigma}_e = \sqrt{\frac{1}{T - k - 1} \sum_{t=1}^T e_t^2}, \quad (5.3)$$

where k is the number of predictors in the model. Notice that we divide by $T - k - 1$ because we have estimated $k + 1$ parameters (the intercept and a coefficient for each predictor variable) in computing the residuals.

The standard error is related to the size of the average error that the model produces. We can compare this error to the sample mean of y or with the standard deviation of y to gain some perspective on the accuracy of the model.

The standard error will be used when generating prediction intervals, discussed in Section 5.6.