9.6　Lagged predictors

Sometimes, the impact of a predictor which is included in a regression model will not be simple and immediate. For example, an advertising campaign may impact sales for some time beyond the end of the campaign, and sales in one month will depend on the advertising expenditure in each of the past few months. Similarly, a change in a company's safety policy may reduce accidents immediately, but have a diminishing effect over time as employees take less care when they become familiar with the new working conditions.

In these situations, we need to allow for lagged effects of the predictor. Suppose that we have only one predictor in our model. Then a model which allows for lagged effects can be written as

$$y_t = \beta_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \cdots + \gamma_k x_{t-k} + \eta_t,$$

where η_t is an ARIMA process. The value of k can be selected using the AICc, along with the values of p and q for the ARIMA error.

Example: TV advertising and insurance quotations

A US insurance company advertises on national television in an attempt to increase the number of insurance quotations provided (and consequently the number of new policies). Figure 9.12 shows the number of quotations and the expenditure on television advertising for the company each month from January 2002 to April 2005.

```
autoplot(insurance, facets=TRUE) +
  xlab("Year") + ylab("") +
  ggtitle("Insurance advertising and quotations")
```
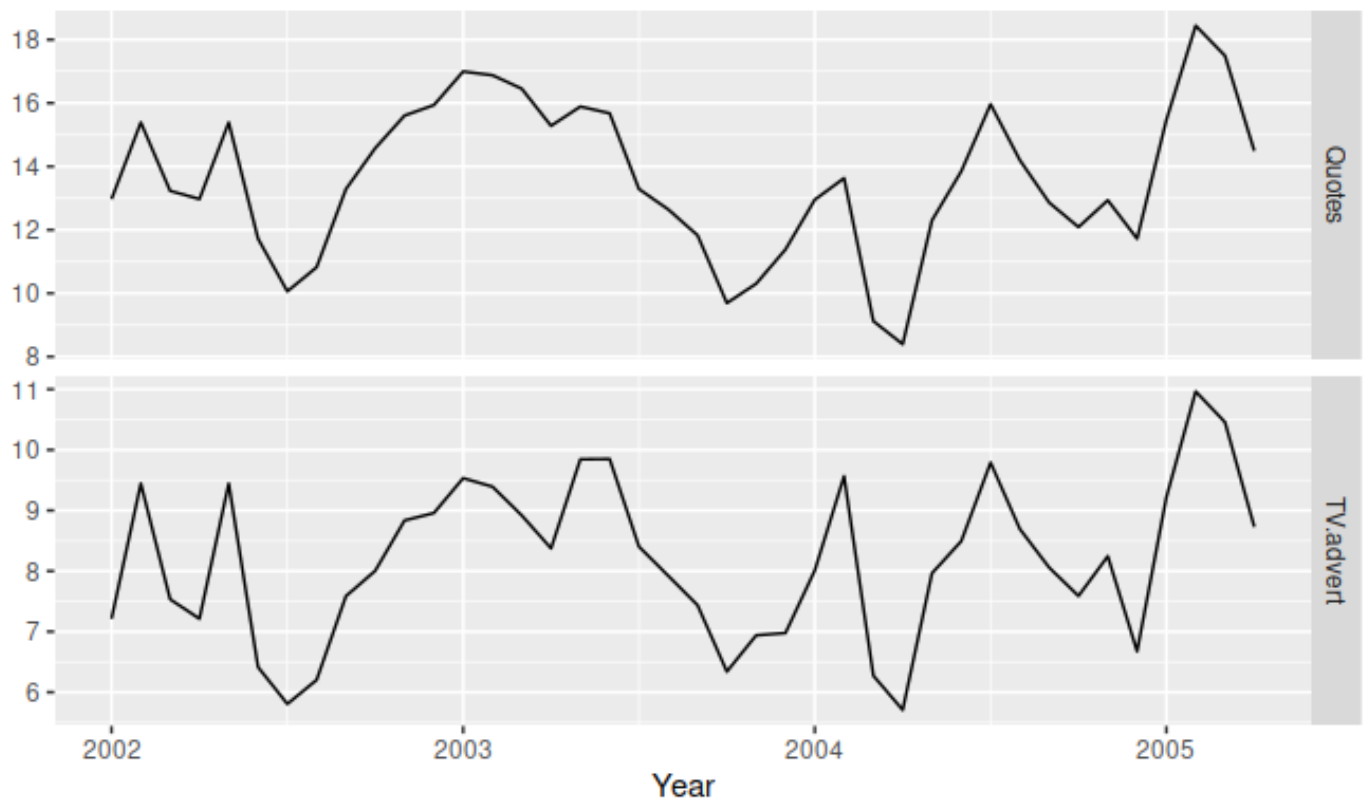
Insurance advertising and quotations

Figure 9.12: Numbers of insurance quotations provided per month and the expenditure on advertising per month.

We will consider including advertising expenditure for up to four months; that is, the model may include advertising expenditure in the current month, and the three months before that. When comparing models, it is important that they all use the same training set. In the following code, we exclude the first three months in order to make fair comparisons.

```r
# Lagged predictors. Test 0, 1, 2 or 3 lags.
Advert <- cbind(
    AdLag0 = insurance[,"TV.advert"],
    AdLag1 = stats::lag(insurance[,"TV.advert"],-1),
    AdLag2 = stats::lag(insurance[,"TV.advert"],-2),
    AdLag3 = stats::lag(insurance[,"TV.advert"],-3)) %>%
  head(NROW(insurance))

# Restrict data so models use same fitting period
fit1 <- auto.arima(insurance[4:40,1], xreg=Advert[4:40,1],
  stationary=TRUE)
fit2 <- auto.arima(insurance[4:40,1], xreg=Advert[4:40,1:2],
  stationary=TRUE)
fit3 <- auto.arima(insurance[4:40,1], xreg=Advert[4:40,1:3],
  stationary=TRUE)
fit4 <- auto.arima(insurance[4:40,1], xreg=Advert[4:40,1:4],
  stationary=TRUE)
```

Next we choose the optimal lag length for advertising based on the AICc.

```r
c(fit1[["aicc"]],fit2[["aicc"]],fit3[["aicc"]],fit4[["aicc"]])
#> [1] 68.500 60.024 62.833 65.457
```

The best model (with the smallest AICc value) has two lagged predictors; that is, it includes advertising only in the current month and the previous month. So we now re-estimate that model, but using all the available data.

```
(fit <- auto.arima(insurance[,1], xreg=Advert[,1:2],
  stationary=TRUE))
#> Series: insurance[, 1]
#> Regression with ARIMA(3,0,0) errors
#>
#> Coefficients:
#>          ar1     ar2    ar3  intercept  AdLag0  AdLag1
#>        1.412  -0.932  0.359      2.039   1.256   0.162
#> s.e.   0.170   0.255  0.159      0.993   0.067   0.059
#>
#> sigma^2 estimated as 0.217:  log likelihood=-23.89
#> AIC=61.78   AICc=65.4   BIC=73.43
```

The chosen model has AR(3) errors. The model can be written as

$$y_t = 2.039 + 1.256x_t + 0.162x_{t-1} + \eta_t,$$

where y_t is the number of quotations provided in month t, x_t is the advertising expenditure in month t,

$$\eta_t = 1.412\eta_{t-1} - 0.932\eta_{t-2} + 0.359\eta_{t-3} + \varepsilon_t,$$

and ε_t is white noise.

We can calculate forecasts using this model if we assume future values for the advertising variable. If we set the future monthly advertising to 8 units, we get the forecasts in Figure 9.13.

```
fc8 <- forecast(fit, h=20,
  xreg=cbind(AdLag0 = rep(8,20),
             AdLag1 = c(Advert[40,1], rep(8,19))))
autoplot(fc8) + ylab("Quotes") +
  ggtitle("Forecast quotes with future advertising set to 8")
```
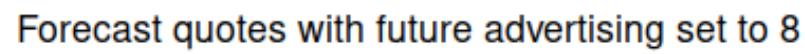
Figure 9.13: Forecasts of monthly insurance quotes, assuming that the future advertising expenditure is 8 units in each future month.