

Rob J Hyndman  
George Athanasopoulos

# FORECASTING

## PRINCIPLES AND PRACTICE

A comprehensive introduction to the latest forecasting methods using R. Learn to improve your forecast accuracy using dozens of real data examples.



3RD EDITION

 **OTexts**  
Oxford Texts in Finance and Statistics

## 10. Dynamic regression models

### 10.6 Lagged predictors

[OTexts.org/fpp3/](http://OTexts.org/fpp3/)

# Lagged predictors

Sometimes a change in  $x_t$  does not affect  $y_t$  instantaneously

# Lagged predictors

Sometimes a change in  $x_t$  does not affect  $y_t$  instantaneously

- $y_t$  = sales,  $x_t$  = advertising.
- $y_t$  = stream flow,  $x_t$  = rainfall.
- $y_t$  = size of herd,  $x_t$  = breeding stock.

# Lagged predictors

Sometimes a change in  $x_t$  does not affect  $y_t$  instantaneously

- $y_t$  = sales,  $x_t$  = advertising.
  - $y_t$  = stream flow,  $x_t$  = rainfall.
  - $y_t$  = size of herd,  $x_t$  = breeding stock.
- 
- These are dynamic systems with input ( $x_t$ ) and output ( $y_t$ ).
  - $x_t$  is often a leading indicator.
  - There can be multiple predictors.

# Lagged predictors

The model include present and past values of predictor:

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

where  $\eta_t$  is an ARIMA process.

# Lagged predictors

The model include present and past values of predictor:

$$y_t = a + \gamma_0 x_t + \gamma_1 x_{t-1} + \dots + \gamma_k x_{t-k} + \eta_t$$

where  $\eta_t$  is an ARIMA process.

**Rewrite model as**

$$\begin{aligned} y_t &= a + (\gamma_0 + \gamma_1 B + \gamma_2 B^2 + \dots + \gamma_k B^k) x_t + \eta_t \\ &= a + \gamma(B) x_t + \eta_t. \end{aligned}$$

# Lagged predictors

The model include present and past values of predictor:

$$y_t = a + \gamma_0 x_t + \gamma_1 x_{t-1} + \dots + \gamma_k x_{t-k} + \eta_t$$

where  $\eta_t$  is an ARIMA process.

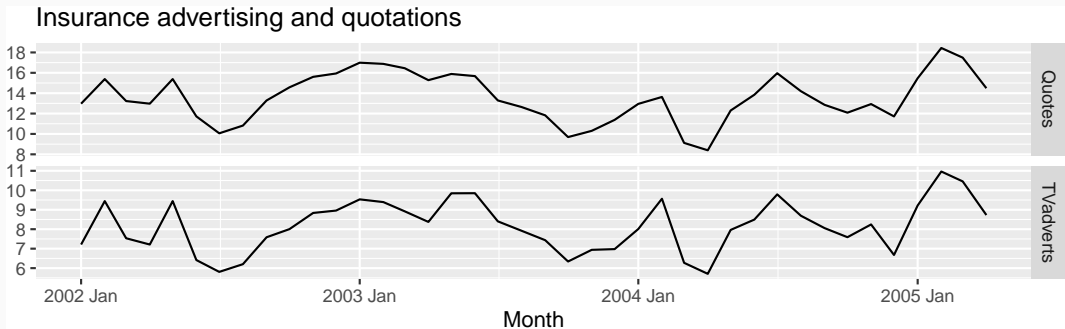
**Rewrite model as**

$$\begin{aligned} y_t &= a + (\gamma_0 + \gamma_1 B + \gamma_2 B^2 + \dots + \gamma_k B^k) x_t + \eta_t \\ &= a + \gamma(B) x_t + \eta_t. \end{aligned}$$

- $\gamma(B)$  is called a *transfer function* since it describes how change in  $x_t$  is transferred to  $y_t$ .
- $x$  can influence  $y$ , but  $y$  is not allowed to influence  $x$ .

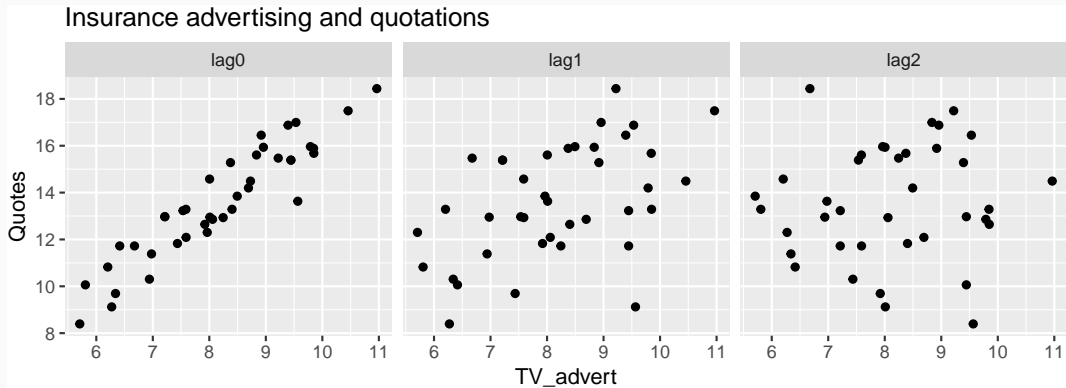
# Example: Insurance quotes and TV adverts

```
insurance |>
  pivot_longer(Quotes:TVadverts) |>
  ggplot(aes(x = Month, y = value)) +
  geom_line() +
  facet_grid(vars(name), scales = "free_y") +
  labs(y = NULL, title = "Insurance advertising and quotations")
```





# Example: Insurance quotes and TV adverts



# Example: Insurance quotes and TV adverts

```
fit <- insurance |>
  # Restrict data so models use same fitting period
  mutate(Quotes = c(NA, NA, NA, Quotes[4:40])) |>
  # Estimate models
  model(
    ARIMA(Quotes ~ pdq(d = 0) + TVadverts),
    ARIMA(Quotes ~ pdq(d = 0) + TVadverts + lag(TVadverts)),
    ARIMA(Quotes ~ pdq(d = 0) + TVadverts + lag(TVadverts) +
      lag(TVadverts, 2)),
    ARIMA(Quotes ~ pdq(d = 0) + TVadverts + lag(TVadverts) +
      lag(TVadverts, 2) + lag(TVadverts, 3))
  )
```

## Example: Insurance quotes and TV adverts

```
glance(fit)
```

Lag order	sigma2	log_lik	AIC	AICc	BIC
0	0.265	-28.3	66.6	68.3	75.0
1	0.209	-24.0	58.1	59.9	66.5
2	0.215	-24.0	60.0	62.6	70.2
3	0.206	-22.2	60.3	65.0	73.8

# Example: Insurance quotes and TV adverts

```
fit_best <- insurance |>  
  model(ARIMA(Quotes ~ pdq(d = 0) + TVadverts + lag(TVadverts)))  
report(fit_best)
```

```
## Series: Quotes
```

```
## Model: LM w/ ARIMA(1,0,2) errors
```

```
##
```

```
## Coefficients:
```

##	ar1	ma1	ma2	TVadverts	lag(TVadverts)	intercept
##	0.512	0.917	0.459	1.2527	0.1464	2.16
## s.e.	0.185	0.205	0.190	0.0588	0.0531	0.86

```
##
```

```
## sigma^2 estimated as 0.2166: log likelihood=-23.9
```

```
## AIC=61.9 AICc=65.4 BIC=73.7
```

# Example: Insurance quotes and TV adverts

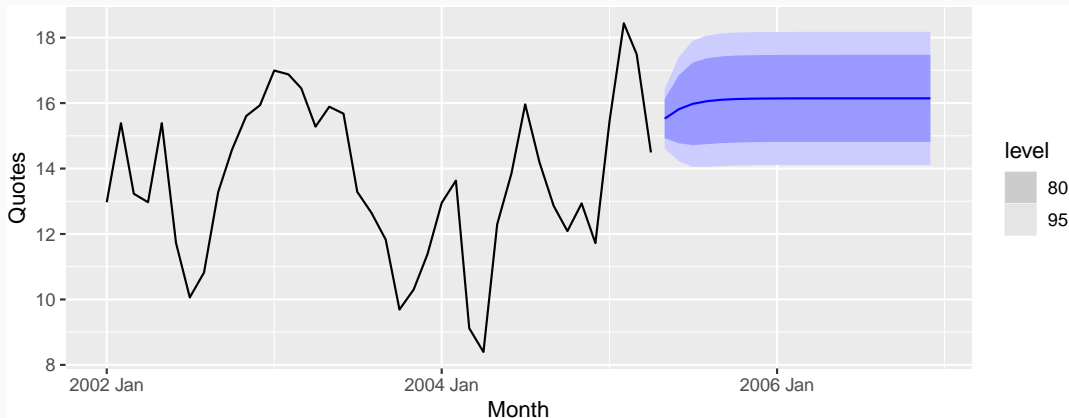
```
fit_best <- insurance |>  
  model(ARIMA(Quotes ~ pdq(d = 0) + TVadverts + lag(TVadverts)))  
report(fit_best)
```

```
## Series: Quotes  
## Model: LM w/ ARIMA(1,0,2) errors  
##  
## Coefficients:  
##          ar1      ma1      ma2  TVadverts  lag(TVadverts)  intercept  
##          0.512  0.917  0.459      1.2527           0.1464          2.16  
## s.e.    0.185  0.205  0.190      0.0588           0.0531          0.86  
##  
## sigma^2 estimated as 0.2166:  log likelihood=-23.9  
## AIC=61.9   AICc=65.4   BIC=73.7
```

$$y_t = 2.155 + 1.253x_t + 0.146x_{t-1} + \eta_t,$$
$$\eta_t = 0.512\eta_{t-1} + \varepsilon_t + 0.917\varepsilon_{t-1} + 0.459\varepsilon_{t-2},$$

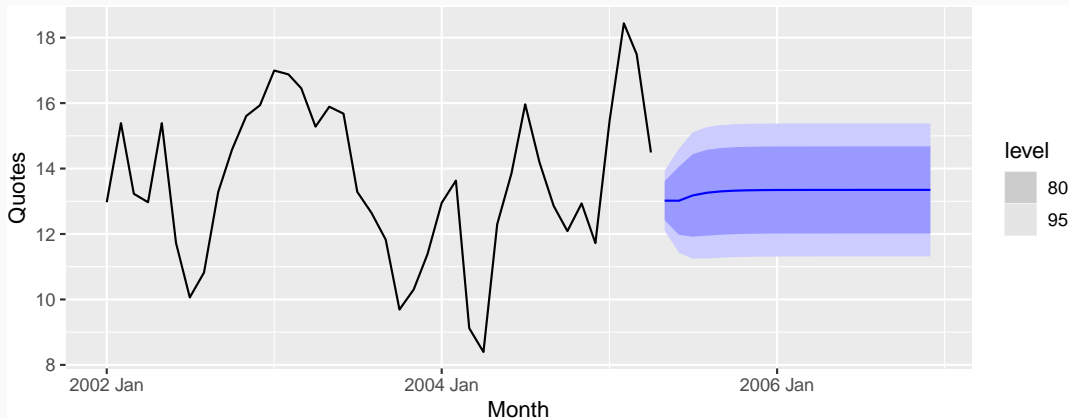
# Example: Insurance quotes and TV adverts

```
advert_a <- new_data(insurance, 20) |>  
  mutate(TVadverts = 10)  
forecast(fit_best, advert_a) |> autoplot(insurance)
```



# Example: Insurance quotes and TV adverts

```
advert_b <- new_data(insurance, 20) |>  
  mutate(TVadverts = 8)  
forecast(fit_best, advert_b) |> autoplot(insurance)
```



# Example: Insurance quotes and TV adverts

```
advert_c <- new_data(insurance, 20) |>  
  mutate(TVadverts = 6)  
forecast(fit_best, advert_c) |> autoplot(insurance)
```

