Statistics 435/535/711 28 February 2022
P. Shaman

Distributed Lag Models

Let y be a dependent variable and x an independent variable. The model

$$y_t = \alpha + \sum_{k=0}^{\infty} \beta_k x_{t-k} + \varepsilon_t$$

is called a *distributed lag model*. The effect of the independent variable on the dependent variable is distributed over a number of lags (e.g., y is sales and x is advertising).

A special case (geometric decay) is given by

$$\beta_k = \beta\gamma^k, \qquad |\gamma| < 1.$$

For this case the model becomes

$$y_t = \alpha + \beta(x_t + \gamma x_{t-1} + \gamma^2 x_{t-2} + \gamma^3 x_{t-3} + \cdots) + \varepsilon_t.$$

The backward shift operator B is defined by

$$Bx_t = x_{t-1}, \quad B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}, \ldots,$$

so that $B^k x_t = x_{t-k}, k = 1, \ldots$. Then for geometric decay the model is

$$\begin{aligned} y_t &= \alpha + \beta(1 + \gamma B + \gamma^2 B^2 + \gamma^3 B^3 + \cdots) x_t + \varepsilon_t \\ &= \alpha + \beta(1 - \gamma B)^{-1} x_t + \varepsilon_t, \end{aligned}$$

by use of the geometric series summation. Now multiply on both sides by the operator $1 - \gamma B$. This yields

$$y_t = (1 - \gamma)\alpha + \gamma y_{t-1} + \beta x_t + \varepsilon_t - \gamma \varepsilon_{t-1},$$

or

$$y_t = \alpha' + \gamma y_{t-1} + \beta x_t + \varepsilon_t'.$$

Thus, for the special case of the distributed lag model defined by geometric decay, the lag one value of the dependent variable appears on the right-hand side. This is first-order autoregressive structure.

Annual Lydia Pinkham data

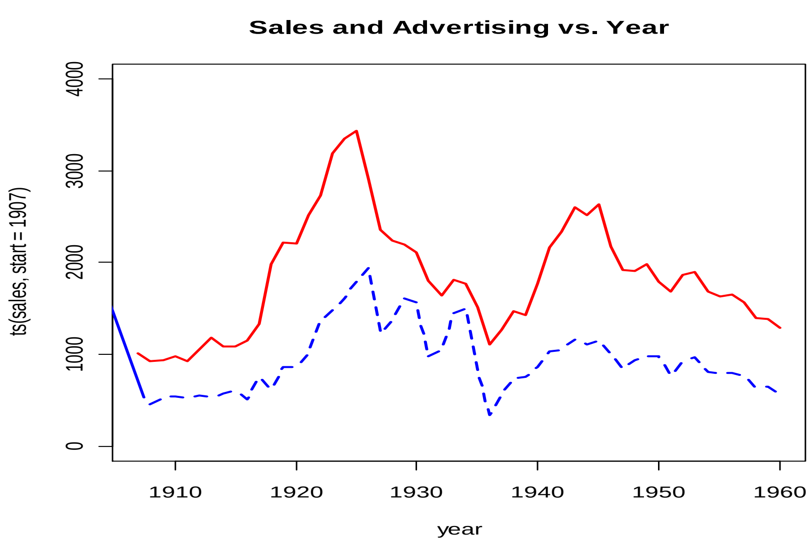
Annual sales and advertising expenditures by the Lydia E. Pinkham Medicine Company, in thousands of dollars, for the years 1907 to 1960, are in the file Lydia.txt. The purpose of the analysis presented here is to study the relationship between sales and advertising expenditures for this firm, and to discuss implications of the results. There are two major reasons why the Lydia Pinkham data are especially useful to study the connection between sales and advertising expenditures more generally: (i) The firm had no competition—no other companies marketed the same product. (ii) The firm employed no sales representatives—advertising, much of it print advertising in newspapers and magazines, was the only method of marketing it used.

A detailed discussion of the company, its history and operations, is in the file Palda posted on Canvas. The file contains pages copied from Kristian Palda's 1964 book.

```
> lydia<-read.csv("G:/Stat71122Spring/Lydia.txt",header=T)
> attach(lydia)
> head(lydia)
  year sales advrtsng advcopy Ind1 Ind2 Ind3 lagsales lag2sales lagadv
1 1907  1016      608       1    1    0    0       NA        NA     NA
2 1908   921      451       1    1    0    0     1016        NA    608
3 1909   934      529       1    1    0    0      921      1016    451
4 1910   976      543       1    1    0    0      934       921    529
5 1911   930      525       1    1    0    0      976       934    543
6 1912  1052      549       1    1    0    0      930       976    525
```

The following plot shows the sales and advertising time series. The sales series is in red and the advertising series is in blue.

```
> plot(ts(sales,start=1907),ylim=c(0,4000),xlab="year",main="Sales and
Advertising vs. Year",lty=1,lwd=2,col="red")
> lines(ts(advrtsng,start=c(1907,1)),lty=2,lwd=2,col="blue")
```



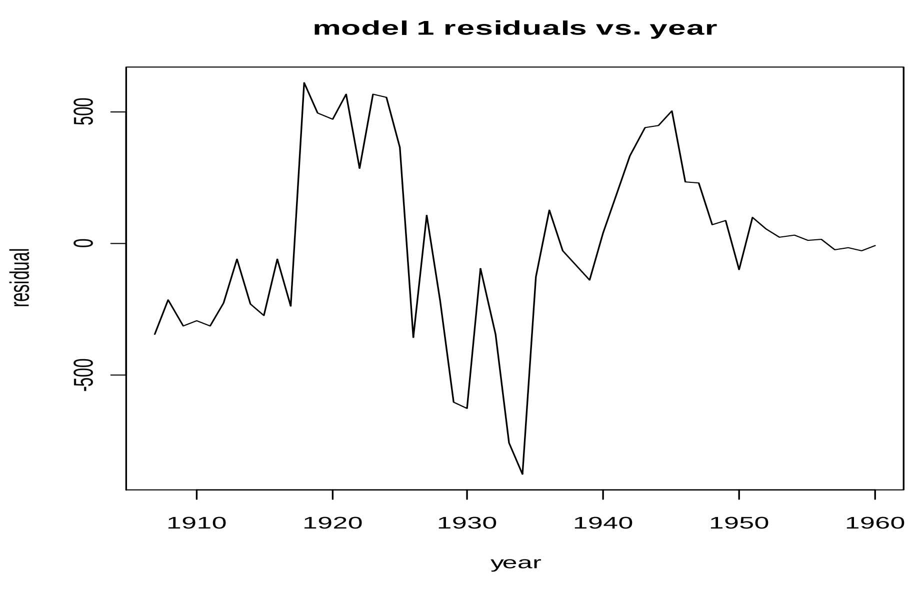
The sales and advertising series have similar traces. The advertising series appears to lag the sales series by perhaps a year in its pattern.

The model below is simply $sales_t = \alpha + \beta_1 advrtsng_t + \varepsilon_t$. While contemporaneous advertising is highly significant, the plot of residuals from the model shows remaining structure.

```
> model1<-lm(sales~advrtsng);summary(model1)

Call:
lm(formula = sales ~ advrtsng)

Residuals:
    Min      1Q  Median      3Q     Max 
-876.45 -222.95  -12.24  221.09  611.77 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 488.8327   127.4390   3.836  0.00034 ***
advrtsng      1.4346     0.1269  11.308 1.26e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 343.5 on 52 degrees of freedom
Multiple R-squared: 0.7109,     Adjusted R-squared: 0.7053 
F-statistic: 127.9 on 1 and 52 DF,  p-value: 1.259e-15
```

The residual plot follows.

```
> 
plot(ts(resid(model1),start=1907),xlab="year",ylab="residual",main="mo
del 1 residuals vs. year")
```



Palda defined four different time periods, corresponding to different regimes of advertising copy. I’ll discuss this in class. The time periods are the following:

Period	Years
1	1907–1914
2	1915–1925
3	1926–1940
4	1941–1960

Compare the residual plot above. Palda used 1908, not 1907, as the starting point of his analysis of the data.

The labels for these time periods are given by the categorical variable *advcopy*. Recall that, in defining a factor variable, R uses as the base for comparison the category for which the label is first in the dictionary ordering of the labels. This category does not receive a dummy variable. Each of the other categories is assigned a 1,0 dummy.

The following model adds a factor variable for *advcopy* to the previous analysis.

```
> fadvcopy<-as.factor(advcopy)
> model2<-lm(sales~advrtsng+fadvcopy);summary(model2)

Call:
lm(formula = sales ~ advrtsng + fadvcopy)

Residuals:
    Min      1Q  Median      3Q     Max 
-606.08 -115.16   10.76  177.56  389.25 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 271.8320   107.9132   2.519 0.015082 *  
advrtsng      1.3759     0.1098  12.533  < 2e-16 ***
fadvcopy2   583.3284   130.9357   4.455 4.87e-05 ***
fadvcopy3    16.8703   128.6327   0.131 0.896193    
fadvcopy4   400.5189   113.7859   3.520 0.000943 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 255.3 on 49 degrees of freedom
Multiple R-squared: 0.8494,     Adjusted R-squared: 0.8371 
F-statistic: 69.11 on 4 and 49 DF,  p-value: < 2.2e-16
```

This model gives four estimated straight lines relating sales to advertising. The lines all have the same slope, but different intercepts. The “Intercept” estimate provided by R is the intercept for time period 1, and the estimated coefficients labeled *fadvcopyj*, $j = 2, 3, 4$, are additive adjustments for the other time periods. In the following, S_t denotes sales in year t, and A_t advertising in year t. We can summarize the estimation for model 2 as follows:

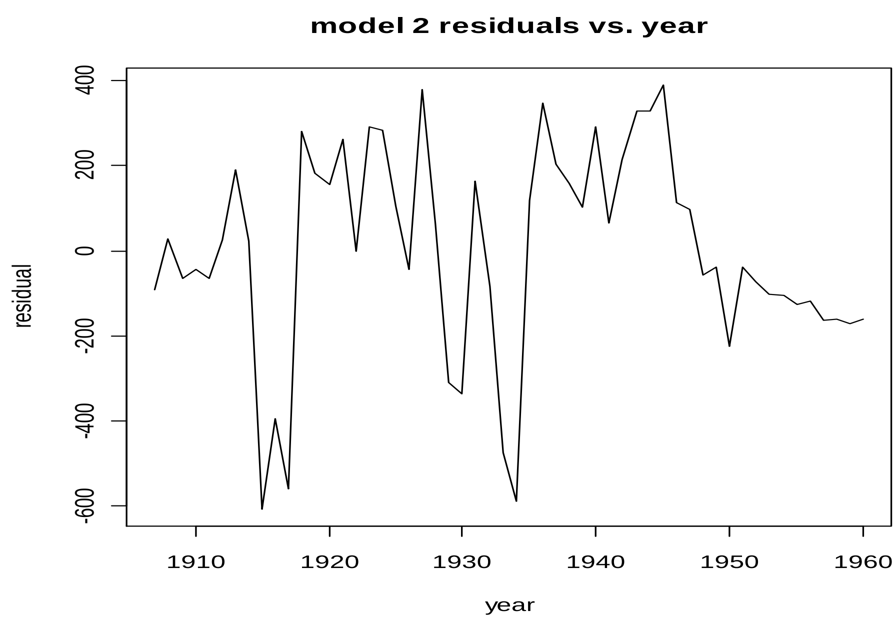
1907–1914 $S_t = 271.832 + 1.376\,A_t$

1915–1925 $S_t = 271.832 + 583.328 + 1.376\,A_t = 855.160 + 1.376\,A_t$

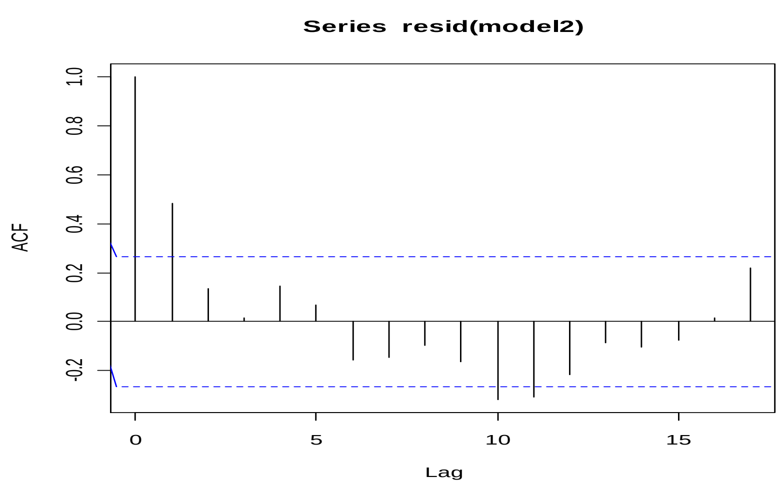
1926–1940 $S_t = 271.832 + 16.870 + 1.376\,A_t = 288.702 + 1.376\,A_t$

1941–1960 $S_t = 271.832 + 400.519 + 1.376\,A_t = 672.351 + 1.376\,A_t$

Here is the residual plot:



Let’s examine the residual autocorrelations.



There is remaining correlation structure. Although this model certainly represents improvement over model 1, it is less than satisfactory.

The same overall fit with a different set of dummy variables, self-defined, follows. The dummies are

$$
\begin{aligned}
Ind1 &= 1 \text{ for } 1907\text{–}1914 \\
&= 0 \text{ otherwise,} \\[1em]
Ind2 &= 1 \text{ for } 1915\text{–}1925 \\
&= 0 \text{ otherwise,} \\[1em]
Ind3 &= 1 \text{ for } 1926\text{–}1940 \\
&= 0 \text{ otherwise.}
\end{aligned}
$$

With this formulation, the base for comparison (it doesn't receive a dummy) is the period 1941–1960.

```
> model3<-lm(sales~advrtsng+Ind1+Ind2+Ind3);summary(model3)

Call:
lm(formula = sales ~ advrtsng + Ind1 + Ind2 + Ind3)

Residuals:
    Min      1Q  Median      3Q     Max 
-606.08 -115.16   10.76  177.56  389.25 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  672.3509   113.6994   5.913 3.18e-07 ***
advrtsng       1.3759     0.1098  12.533  < 2e-16 ***
Ind1        -400.5189   113.7859  -3.520 0.000943 ***
Ind2         182.8095    97.2061   1.881 0.065971 .  
Ind3        -383.6485    90.5743  -4.236 1.00e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 255.3 on 49 degrees of freedom
Multiple R-squared: 0.8494,     Adjusted R-squared: 0.8371 
F-statistic: 69.11 on 4 and 49 DF,  p-value: < 2.2e-16
```

Before trying to repair the fit of model 2 (the same overall as that of model 3), let's digress to consider the prevalence of advertising in the U.S. economy.

The table below shows annual U.S. advertising expenditures and gross national product, in billions of 1988 dollars. The advertising figures estimate all outlays at the national, local, and individual levels.

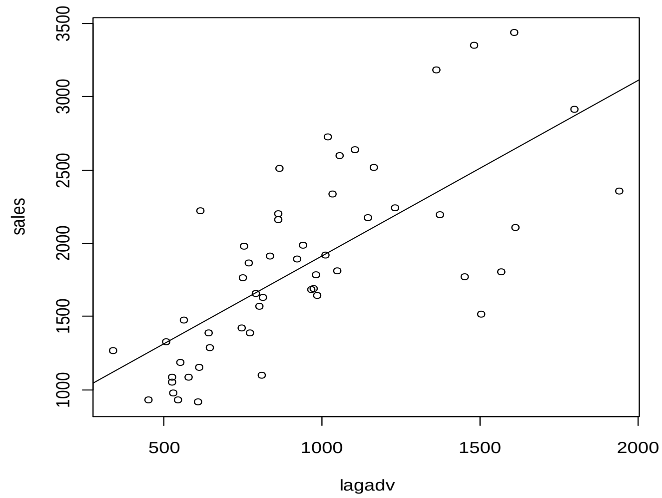
Year	Adv	GNP	Adv/GNP	Year	Adv	GNP	Adv/GNP
1940	2.11	100.4	0.0210	1980	53.55	2732.0	0.0196
1950	5.70	288.3	0.0198	1981	60.43	3052.6	0.0198
1960	11.96	515.3	0.0232	1982	66.58	3166.0	0.0210
1970	19.55	1015.5	0.0193	1983	75.85	3405.7	0.0223
1975	27.90	1598.4	0.0175	1984	87.82	3765.0	0.0233
1976	33.30	1782.8	0.0187	1985	94.75	3998.1	0.0237
1977	37.44	1990.5	0.0188	1986	102.14	4235.0	0.0241
1978	43.33	2249.7	0.0193	1987	109.65	4526.7	0.0242
1979	48.78	2508.2	0.0194	1988	118.05	4864.3	0.0243

Some studies of the relationship between sales and advertising have investigated whether current advertising creates future product demand. This issue has policy implications. Some have argued that for tax purposes advertising expenditures should be treated as long-term investments and therefore amortized over time, as opposed to being entirely expensed in the year in which they occur. U.S. tax regulations permit full current expensing, and those advocating amortization argue that the tax code favors advertising over other investment outlays. This argument indicates that estimation of the duration of time over which advertising affects sales is an important empirical question. We explore this question with the Lydia Pinkham data, first focusing on the annual data. This data set is especially well-suited to address the question.

Let's return to consideration of how to improve the model relating sales and advertising for the Lydia Pinkham data. Below is a rough initial attempt to explore the relationship between sales and lagged advertising. The results suggest that advertising from previous years does have an impact upon current sales. Note in particular the decay of the slope in the relationship as the lag of advertising increases. The decrease is steady.

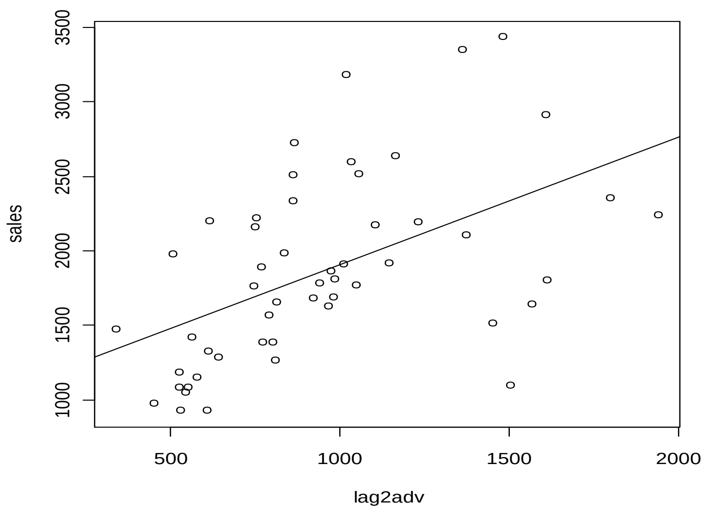
```
> lag2adv<-c(NA,NA,head(advrtsng,-2))
> lag3adv<-c(NA,NA,NA,head(advrtsng,-3))
> lag4adv<-c(NA,NA,NA,NA,head(advrtsng,-4))
> lag5adv<-c(NA,NA,NA,NA,NA,head(advrtsng,-5))
```

```
> plot(lagadv,sales)
> abline(lm(sales~lagadv))
```



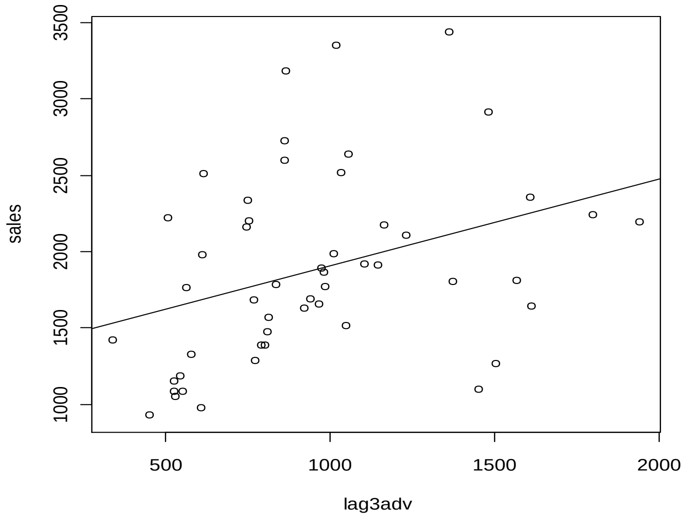

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 718.2953   169.0437   4.249 9.12e-05 ***
lagadv        1.1965     0.1672   7.156 3.09e-09 ***
```

```
> plot(lag2adv,sales)
> abline(lm(sales~lag2adv))
```



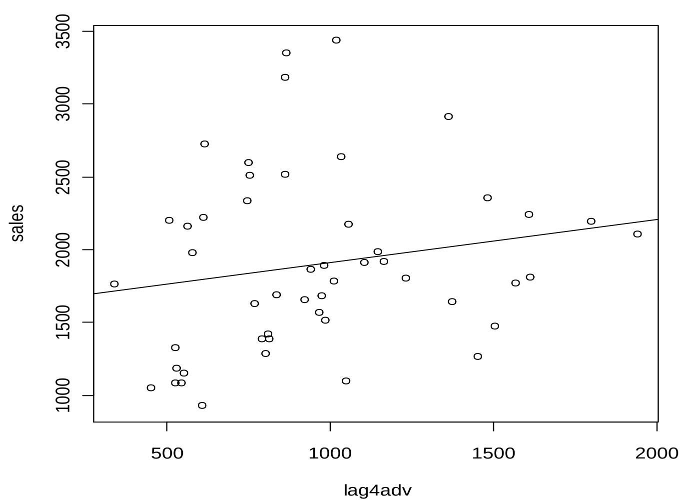

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1049.8999   205.1306   5.118 4.94e-06 ***
lag2adv        0.8580     0.2018   4.253 9.24e-05 ***
```

```
> plot(lag3adv,sales)
> abline(lm(sales~lag3adv))
```




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1341.8332   224.6665   5.973 2.58e-07 ***
lag3adv        0.5654     0.2197   2.574   0.0131 *
```

```
> plot(lag4adv,sales)
> abline(lm(sales~lag4adv))
```




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1613.4595   233.8801   6.899 1.05e-08 ***
lag4adv        0.2983     0.2277   1.310    0.196
```

Let's fit a distributed lag model with geometric decay,

$$S_t = \alpha + \beta(A_t + \gamma A_{t-1} + \gamma^2 A_{t-2} + \gamma^3 A_{t-3} + \cdots) + \varepsilon_t$$

$$= \alpha + \beta(1 - \gamma B)^{-1} A_t + \varepsilon_t,$$

where B is the backward shift operator. Recall that this formulation is equivalent to the autoregressive representation

$$S_t = (1-\gamma)\alpha + \gamma S_{t-1} + \beta A_t + \varepsilon_t - \gamma \varepsilon_{t-1}.$$

In this model the contemporaneous effect of advertising is β, the cumulative effect after one more time period is $\beta(1 + \gamma)$, and the cumulative effect after $m - 1$ more time periods is

$$\beta(1 + \gamma + \gamma^2 + \cdots + \gamma^{m-1}) = \beta(1 - \gamma^m)/(1 - \gamma).$$

As m increases, this converges to $\beta/(1 - \gamma)$ (because we assume $|\gamma| < 1$). Thus, the proportion of the long-run cumulative effect of advertising upon sales realized after m time periods, say p, is $1 - \gamma^m$. Given p, the solution for m is

$$m = \log_e(1 - p)/\log_e \gamma.$$

The model below is

$$S_t = \alpha + \beta_1 A_t + \beta_2 S_{t-1} + \beta_3 Ind1_t + \beta_4 Ind2_t + \beta_5 Ind3_t + \varepsilon_t.$$

```
> lagsales<-c(NA,head(sales,-1))
> lag2sales<-c(NA,NA,head(sales,-2))
> model4<-lm(sales~lagsales+advrtsng+Ind1+Ind2+Ind3);summary(model4)

Call:
lm(formula = sales ~ lagsales + advrtsng + Ind1 + Ind2 + Ind3)

Residuals:
    Min      1Q  Median      3Q     Max 
-371.39 -103.57  -30.79   88.24  389.88 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  254.63920   96.30805   2.644 0.011101 *  
lagsales       0.60734    0.08142   7.459 1.65e-09 ***
advrtsng       0.53450    0.13604   3.929 0.000278 ***
Ind1        -133.34673   88.95563  -1.499 0.140555    
Ind2         216.83964   67.21894   3.226 0.002288 ** 
Ind3        -202.50459   67.05600  -3.020 0.004079 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 176.2 on 47 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.929,      Adjusted R-squared: 0.9215
F-statistic:   123 on 5 and 47 DF,  p-value: < 2.2e-16
```

This fit is

$$S_t = 254.639 + 0.6073S_{t-1} + 0.5345A_t - 133.347Ind1_t + 216.840Ind2_t - 202.505Ind3_t.$$

It can be rewritten as

$$(1 - 0.6073B)\, S_t = 254.639 + 0.5345A_t - 133.347Ind1_t + 216.840Ind2_t - 202.505Ind3_t,$$

or

$$S_t = (1 - 0.6073B)^{-1}\, 254.639 + 0.5345(1 - 0.6073B)^{-1}A_t$$

$$+ (1 - 0.6073B)^{-1}(-133.347Ind1_t + 216.840Ind2_t - 202.505Ind3_t),$$

or

$$S_t = 648.431 + 0.5345(A_t + 0.6073A_{t-1} + 0.3689A_{t-2} + 0.2240A_{t-3} + 0.1361A_{t-4}$$

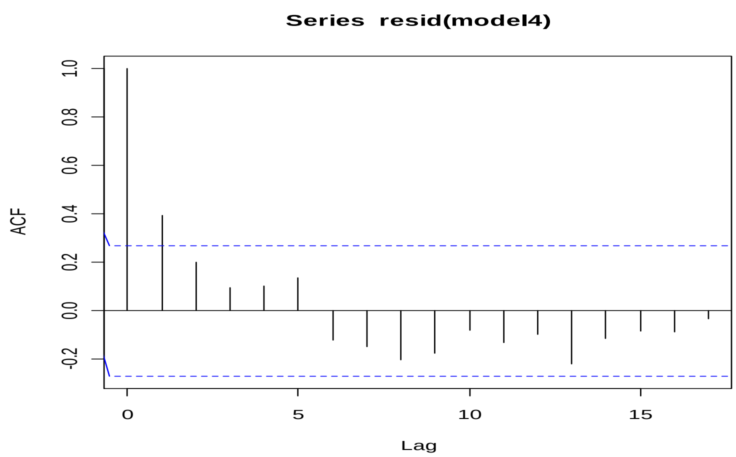
$$+ 0.0826A_{t-5} + \cdots)$$

$$+ (1 - 0.6073B)^{-1}(-133.347Ind1_t + 216.840Ind2_t - 202.505Ind3_t).$$

The implied 90 per cent duration interval is

$$\log(0.1)/\log(0.60734) = 4.62 \text{ years}.$$

```
> acf(resid(model4))
```



The residual lag one autocorrelation is significantly different from zero, and thus this model is not acceptable.

Let's add lag two of sales as an explanatory variable.

```
> model5<-
lm(sales~lagsales+lag2sales+advrtsng+Ind1+Ind2+Ind3);summary(model5)

Call:
lm(formula = sales ~ lagsales + lag2sales + advrtsng + Ind1 +
    Ind2 + Ind3)

Residuals:
    Min      1Q  Median      3Q     Max
-314.75  -97.10  -19.04   88.51  363.02

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  282.3863    95.4090   2.960   0.0049 **
lagsales       0.8561     0.1489   5.748 7.38e-07 ***
lag2sales     -0.2372     0.1207  -1.965   0.0556 .
advrtsng       0.4791     0.1362   3.519   0.0010 **
Ind1        -139.5593    91.8130  -1.520   0.1355
Ind2         176.1334    69.0602   2.550   0.0142 *
Ind3        -160.2427    69.0219  -2.322   0.0248 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 172.5 on 45 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.9319,     Adjusted R-squared: 0.9228
F-statistic: 102.7 on 6 and 45 DF,  p-value: < 2.2e-16
```

This fit is

$$S_t = 282.386 + 0.8561S_{t-1} - 0.2372S_{t-2} + 0.4791A_t$$

$$-139.559Ind1_t + 176.133Ind2_t - 160.243Ind3_t.$$

Rewrite as

$$(1 - 0.8561B + 0.2372B^2)\ S_t$$

$$= 282.386 + 0.4791A_t\ - 139.559Ind1_t + 176.133Ind2_t - 160.243Ind3_t.$$

Then

$$S_t = (1 - 0.8561B + 0.2372B^2)^{-1}\ 282.386$$

$$+ (1 - 0.8561B + 0.2372B^2)^{-1}\ 0.4791A_t$$

$$+ (1 - 0.8561B + 0.2372B^2)^{-1}\ (-139.559Ind1_t + 176.133Ind2_t - 160.243Ind3_t),$$

or

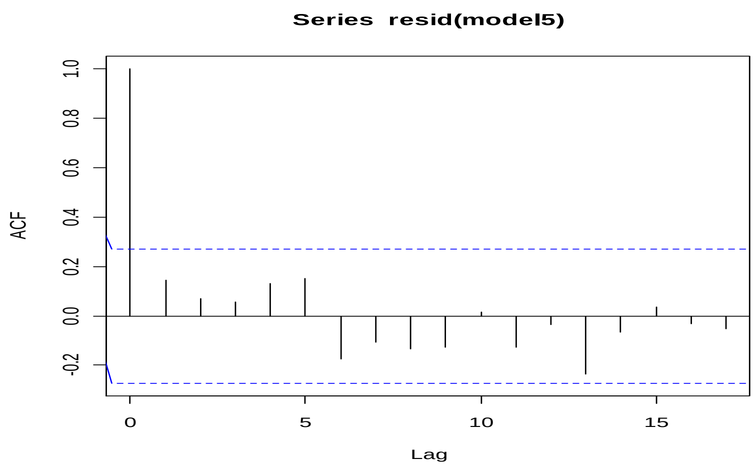
$$S_t = 740.945 + 0.4791(A_t + 0.8561A_{t-1} + 0.4957A_{t-2} + 0.2213A_{t-3} + 0.0719A_{t-4}$$

$$+ 0.0090A_{t-5} - 0.0093A_{t-6} - 0.0101A_{t-7} - 0.0064A_{t-8} + \cdots)$$

$$+ (1 - 0.8561B + 0.2372B^2)^{-1} (-139.559 Ind1_t + 176.133 Ind2_t - 160.243 Ind3_t).$$

A 90 per cent duration interval can be approximated from this representation (add the coefficients attached to advertising). The duration is roughly 3.1 years.

```
> acf(resid(model5))
```



None of the residual autocorrelations is significant.

Let's fit still another model. We add lag one advertising to the previous fit.

```
> options(digits=9)
> model6<-
lm(sales~lagsales+lag2sales+advrtsng+lagadv+Ind1+Ind2+Ind3);summary(mo
del6)

Call:
lm(formula = sales ~ lagsales + lag2sales + advrtsng + lagadv +
    Ind1 + Ind2 + Ind3)

Residuals:
      Min         1Q    Median         3Q        Max
-333.5635  -98.4101  -12.3733   85.5482  436.3308
```

```
Coefficients:
                Estimate  Std. Error  t value   Pr(>|t|)    
(Intercept)  259.860376   91.770977  2.83162 0.00695867 ** 
lagsales       0.965412    0.150265  6.42475 8.0049e-08 ***
lag2sales     -0.205023    0.116299 -1.76289 0.08486339 .  
advrtsng       0.549964    0.133858  4.10856 0.00017055 ***
lagadv        -0.344700    0.151048 -2.28206 0.02737726 *  
Ind1        -116.556442   88.376485 -1.31886 0.19403832    
Ind2         158.493912   66.492314  2.38364 0.02152282 *  
Ind3         -88.371460   73.133545 -1.20836 0.23336410    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 164.995 on 44 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.939119,   Adjusted R-squared: 0.929433 
F-statistic: 96.9603 on 7 and 44 DF,  p-value: < 2.22e-16
```

This fit can be written

$$S_t = 1084.509 + (1 - 0.965412B + 0.205023B^2)^{-1}(0.549964 - 0.3447B)A_t$$

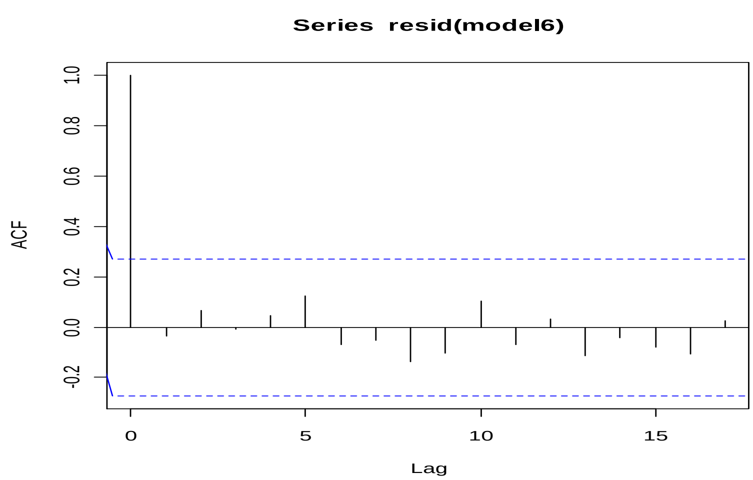
$$+ \text{ terms in } Indi_t \ (i\text{=1,2,3})$$

$$= 1084.509 + 0.5500(A_t + 0.3385A_{t-1} + 0.1219A_{t-2} + 0.0483A_{t-3} + \cdots)$$

$$+ \text{ terms in } Indi_t \ (i\text{=1,2,3}).$$

This suggests a 90 per cent duration interval of about 2.5 years.

```
> acf(resid(model6))
```

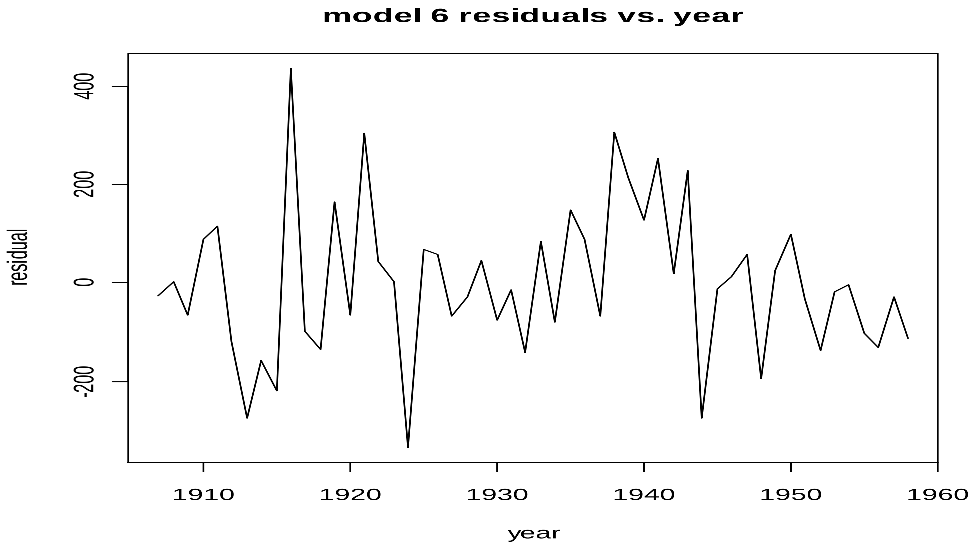


Thus, by this residual autocorrelation plot, model 6 is even better than model 5.

Note that model 4 incorporates geometric decay in the attenuation of the lagged advertising weights. Models 5 and 6 do not force this rate of decay, but the results are close to such decay for model 6.

The plot of the residuals vs. time for model 6 is given next.

```
>
plot(ts(resid(model6),start=c(1907,1)),xlab="year",ylab="residual",mai
n="model 6 residuals vs. year")
```



We have given alternative representations of the models above by inverting operators involving the backward shift operator. The calculations of the coefficients in the representations can be obtained via simple recursions. Let's consider a general framework. Suppose a model fit is

$$(1 + \alpha_1 B + \cdots + \alpha_p B^p) S_t = (\beta_0 + \beta_1 B + \cdots + \beta_r B^r)\, A_t + \cdots .$$

Rewrite as

$$S_t = (1 + \alpha_1 B + \cdots + \alpha_p B^p)^{-1} (\beta_0 + \beta_1 B + \cdots + \beta_r B^r)\, A_t + \cdots$$

$$= (\delta_0 + \delta_1 B + \delta_2 B^2 + \cdots)\, A_t .$$

Then we have

$$\beta_0 + \beta_1 B + \cdots + \beta_r B^r = (1 + \alpha_1 B + \cdots + \alpha_p B^p)(\, \delta_0 + \delta_1 B + \delta_2 B^2 + \cdots).$$

Equating coefficients of B^j on both sides, $j = 1, 2, ...,$ we can solve for the δ_j's. There are two cases to consider, $r \le p$ and $p < r$.

1. $r \leq p$

$$
\begin{aligned}
\beta_0 &= \delta_0, \\
\beta_j &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{j-1}\delta_1 + \alpha_j\delta_0, \quad j = 1,\ldots, r, \\
0 &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{j-1}\delta_1 + \alpha_j\delta_0, \quad j = r+1,\ldots, p, \\
0 &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{p-1}\delta_{j-p+1} + \alpha_p\delta_{j-p}, \quad j = p+1,\ p+2,\ldots .
\end{aligned}
\tag{1}
$$

The third line of (1) does not appear if $r = p$.

Then solve (1) for $\delta_0, \delta_1, \delta_2, \delta_3, \ldots$, in this order:

$$
\begin{aligned}
\delta_0 &= \beta_0, \\
\delta_j &= \beta_j - \alpha_1\delta_{j-1} - \cdots - \alpha_{j-1}\delta_1 - \alpha_j\delta_0, \quad j = 1,\ldots, r, \\
\delta_j &= -\alpha_1\delta_{j-1} - \cdots - \alpha_{j-1}\delta_1 - \alpha_j\delta_0, \quad j = r+1,\ldots, p, \\
\delta_j &= -\alpha_1\delta_{j-1} - \cdots - \alpha_{p-1}\delta_{j-p+1} - \alpha_p\delta_{j-p}, \quad j = p+1, p+2,\ldots .
\end{aligned}
$$

2. $p < r$

$$
\begin{aligned}
\beta_0 &= \delta_0, \\
\beta_j &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{j-1}\delta_1 + \alpha_j\delta_0, \quad j = 1,\ldots, p, \\
\beta_j &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{p-1}\delta_{j-p+1} + \alpha_p\delta_{j-p}, \quad j = p+1,\ldots, r, \\
0 &= \delta_j + \alpha_1\delta_{j-1} + \cdots + \alpha_{p-1}\delta_{j-p+1} + \alpha_p\delta_{j-p}, \quad j = r+1, r+2,\ldots .
\end{aligned}
\tag{2}
$$

Solve (2) for $\delta_0, \delta_1, \delta_2, \delta_3, \ldots$, in this order:

$$
\begin{aligned}
\delta_0 &= \beta_0, \\
\delta_j &= \beta_j - \alpha_1\delta_{j-1} - \cdots - \alpha_{j-1}\delta_1 - \alpha_j\delta_0, \quad j = 1,\ldots, p, \\
\delta_j &= \beta_j - \alpha_1\delta_{j-1} - \cdots - \alpha_{p-1}\delta_{j-p+1} - \alpha_p\delta_{j-p}, \quad j = p+1,\ldots, r, \\
0 &= -\alpha_1\delta_{j-1} - \cdots - \alpha_{p-1}\delta_{j-p+1} - \alpha_p\delta_{j--p}, \quad j = r+1, r+2,\ldots .
\end{aligned}
$$

It is easy to obtain numerical values for the δ_j's. We consider the model 6 fit as an example. The model involves $r = 1$, $p = 2$, the first case above. We obtain

$$S_t = (1 - 0.965412B + 0.205023B^2)^{-1}(0.549964 - 0.3447B)A_t + \cdots$$

$$= (\delta_0 + \delta_1 B + \delta_2 B^2 + \cdots)\, A_t.$$

Then

$$0.549964 - 0.3447B = (1 - 0.965412B + 0.205023B^2)(\,\delta_0 + \delta_1 B + \delta_2 B^2 + \cdots).$$

Equate coefficient of powers of B on both sides. Then

$$\delta_0 = 0.549964,$$

$$\delta_1 = -\,0.3447 + 0.965412(0.549964) = 0.186242,$$

$$\delta_j = 0.965412\ \delta_{j-1} - 0.205023\ \delta_{j-2}, \qquad j = 2, 3, \ldots .$$

The following R code allows us to calculate the 90 per cent duration interval. Initially we determine the first two deltas by hand calculation, as above.

```
> deltapartial<-delta<-c(rep(0,times=500))
> #deltapartial is the partial sum of the deltas
> delta[1]<-0.549964;delta[2]<-0.186242
> deltapartial[1]<-delta[1]
> deltapartial[2]<-deltapartial[1]+delta[2]
> for(j in 3:500){
+ j1<-j-1;j2<-j-2
+ delta[j]<-0.965412*delta[j1]-0.205023*delta[j2]
+ deltapartial[j]<-deltapartial[j1]+delta[j]
+ }

> deltapartial[500]*0.9
[1] 0.7709902

> deltapartial[1:20]
 [1] 0.5499640 0.7362060 0.8032510 0.8297931 0.8416715 0.8476972
0.8510792
 [8] 0.8531088 0.8543748 0.8551809 0.8556996 0.8560351 0.8562526
0.8563938
[15] 0.8564855 0.8565451 0.8565839 0.8566091 0.8566254 0.8566361
```

Thus, the 90 per cent duration interval is given by

$$2 + (0.77099 - 0.73621)/(0.80325 - 0.73621) = 2.52 \text{ years}.$$

Sales and Advertising---Some Analysis and Comments

Consider the bivariate setting of sales and advertising. Is there one-way causation, and if so, what is the direction? Or is there mutual causation? The questions are complicated by the fact that time series of both sales and advertising are highly positively autocorrelated. If we fail to account for the autocorrelation of one variable, say sales, we can be led to an erroneous interpretation of the nature of the effect of advertising upon sales.

One notion of causality is that advanced by Granger. Care should be taken in interpreting it. Let's explain Granger causality in the context of sales and advertising. Suppose we are given data on both sales and advertising up to time t. We say advertising causes sales in the Granger sense if sales can be better predicted by using past values of advertising than by not using these past values. Likewise, we can consider the prediction of Advertising_t and ask whether past values of sales improve the prediction. If causation proceeds in both directions we say there is feedback.

Here is one way we can proceed to investigate Granger causality. This method employs regression analysis. Construct the model

$$Adv_t = \alpha + \sum_{i=1}^{I} \beta_i \, Adv_{t-i} + \sum_{j=1}^{J} \gamma_j \, Sales_{t-j} + \varepsilon_t \, ,$$

where ε_t is a white noise error term and I and J are chosen large enough to accommodate a rich autocorrelation structure. Perform the partial F test of the null hypothesis that $\gamma_j = 0$, $j = 1,..., J$. If this null hypothesis is rejected, then sales is said to Granger-cause advertising. Otherwise there is no inference that sales Granger-causes advertising.

Next construct the model

$$Sales_t = \delta + \sum_{k=1}^{K} \mu_k \, Sales_{t-k} + \sum_{m=1}^{M} \pi_m \, Adv_{t-m} + \eta_t \, ,$$

where η_t is a white noise error term and K and M are chosen large enough to accommodate a rich autocorrelation structure. Perform the partial F test of the null hypothesis that $\pi_m = 0$, $m = 1,..., M$. If this null hypothesis is rejected, then advertising is said to Granger-cause sales. Otherwise there is no inference that advertising Granger-causes sales.

If we find that both sales Granger-causes advertising and advertising Granger-causes sales, then feedback is said to exist. As a variation of the above formulations, sometimes the contemporaneous terms with $j = 0$ and $m = 0$ are included in the regressions.

Here is a second way to investigate Granger causality. Fit an ARIMA model to the sales time series, and calculate the residuals. The residual time series will be approximately a white noise series. That is, the autocorrelation structure has been essentially removed from the sales time series. Next follow the same procedure for the advertising time series. Then calculate the cross correlations between the two *residual* series:

$$r_k = \mathrm{Corr}\left(e_{Adv,\,t}, e_{Sales,\,t-k}\right).$$

If r_k is significantly different from zero for any $k > 0$, then we have evidence that sales Granger-causes advertising. If a cross correlation is significant for $k < 0$, then there is evidence that advertising Granger-causes sales. There is feedback if cross correlations for both $k > 0$ and $k < 0$ are significant.

The two procedures described above have some drawbacks. The choices of *I*, *J*, *K*, and *M* are arbitrary, and the use of ARIMA modelling will affect the relationship between the original series. The use of cross spectral analysis will avoid some of these potential difficulties.

Let's explore Granger causality with the annual Lydia Pinkham data.

```
> lag3sales<-c(NA,NA,NA,head(sales,-3))
> lag4sales<-c(NA,NA,NA,NA,head(sales,-4))
```

The following output shows that sales Granger-causes advertising.

```
> modelG1<-
lm(advrtsng~lagadv+lag2adv+lag3adv+lag4adv+lagsales+lag2sales+lag3sales
+lag4sales+Ind1+Ind2+Ind3);summary(modelG1)

Residuals:
    Min      1Q  Median      3Q     Max 
-319.76  -68.73   -9.47  104.03  305.05 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.802e+02  1.100e+02  -1.639 0.109441    
lagadv       3.737e-01  1.830e-01   2.042 0.048142 *  
lag2adv     -3.120e-01  1.821e-01  -1.713 0.094825 .  
lag3adv      1.146e-01  1.786e-01   0.642 0.524997    
lag4adv      1.580e-01  1.773e-01   0.891 0.378432    
lagsales     5.970e-01  1.465e-01   4.076 0.000225 ***
lag2sales   -4.588e-01  2.152e-01  -2.132 0.039526 *  
lag3sales    2.672e-01  2.217e-01   1.205 0.235543    
lag4sales   -5.637e-03  1.586e-01  -0.036 0.971832    
Ind1         1.313e+02  1.019e+02   1.288 0.205596    
Ind2         1.236e+02  6.987e+01   1.769 0.084942 .  
Ind3         1.051e+02  8.848e+01   1.188 0.242198    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 163.9 on 38 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared: 0.8459,     Adjusted R-squared: 0.8012 
F-statistic: 18.96 on 11 and 38 DF,  p-value: 3.609e-12

> modelG2<-
lm(advrtsng~lagadv+lag2adv+lag3adv+lag4adv+Ind1+Ind2+Ind3);summary(mode
lG2)

> anova(modelG2,modelG1)
Analysis of Variance Table

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)    
1     42 1679599                                  
2     38 1020341  4    659258 6.1381 0.0006524 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value for the partial F test is 0.00065. A similar result is obtained if we include the contemporaneous sales term.

However, for these annual data we find that advertising does not Granger-cause sales, unless we include the contemporaneous advertising term and include it in the test (the *p* value is then 0.01).

A study of sales and advertising by Ashley, Granger, and Schmalensee (*Econometrica*, 1980, 1149-1168) used quarterly U.S. data from 1956 to 1975. The authors found that aggregate national consumption (sales) Granger-causes advertising, but that advertising does not Granger-cause aggregate national consumption.

Recall the discussion of the 90 per cent duration interval in consideration of the Lydia Pinkham data. Suppose the model is

$$\text{Sales}_t = \alpha + \beta_0 \text{Adv}_t + \beta_0 \gamma \text{Adv}_{t-1} + \beta_0 \gamma^2 \text{Adv}_{t-2} + \cdots + u_t ,$$

where u_t is white noise. This can be rewritten as

$$\text{Sales}_t = \alpha(1-\gamma) + \gamma \text{Sales}_{t-1} + \beta_0 \text{Adv}_t + (u_t - \gamma u_{t-1}) .$$

The 90 per cent duration interval is then obtained as $\ln(0.1)/\ln\gamma$. Clarke (*Journal of Marketing Research*, 1976, 345–357) surveyed more than 70 studies that estimated duration levels, all using the above model with geometric decay. Estimates of the 90 per cent duration interval varied widely, ranging from 1.3 to 1368 months. Different products were involved in the studies, and thus some variation is to be expected. But Clarke found excessive variation. The studies involved data with varying data intervals (weekly, monthly, bimonthly, quarterly, and annual). The table below shows Clarke's findings, with "unreasonable" studies not included. Standard deviations are in parentheses.

Data interval	Mean estimate of γ	Mean estimate of 90% duration interval*	Number of studies
weekly	0.537 (.057)	0.9 (.2)	2
monthly	0.440 (.027)	3.0 (.2)	10
bimonthly	0.493 (.086)	9.0 (6.9)	10
quarterly	0.599 (.086)	25.1 (6.9)	10
annual	0.560 (.031)	56.5 (5.1)	27

*all figures are in months

Clarke concluded that the duration interval estimates from annual data are not likely to be accurate, and that estimates should be obtained from series with shorter data intervals. In particular, one may want to match the data interval to the average time interval between purchases of the product. Many authors believe that the 90 per cent duration interval for advertising for most established products is less than a year and that annual data should not be employed for estimation.

The geometric decay model is undoubtedly inappropriate for many of the studies summarized by Clarke. Hence one needs to be cautious in drawing conclusions from Clarke's study.

Another Model

Here is another fit for the annual Lydia Pinkham data. It adds interaction terms to the geometric decay model, the interactions between the time periods and contemporaneous advertising, and between the time periods and lagged sales. Partial *F* tests show that the interactions between time periods and contemporaneous advertising are statistically significant, and that the interactions between time periods and lagged sales are marginally significant ($p = 0.08$).

```
> modelintrctn1<-
lm(sales~advrtsng+lagsales+Ind1+Ind2+Ind3+advrtsng*Ind1+advrtsng*Ind2+a
dvrtsng*Ind3+lagsales*Ind1+lagsales*Ind2+lagsales*Ind3);summary(modelin
trctn1)

Call:
lm(formula = sales ~ advrtsng + lagsales + Ind1 + Ind2 + Ind3 +
    advrtsng * Ind1 + advrtsng * Ind2 + advrtsng * Ind3 + lagsales *
    Ind1 + lagsales * Ind2 + lagsales * Ind3)

Residuals:
    Min      1Q  Median      3Q     Max
-265.83  -67.05  -18.07   67.72  440.64

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -182.4488   180.7390  -1.009  0.31867
advrtsng         1.6753     0.3270   5.123 7.55e-06 ***
lagsales         0.3042     0.1506   2.020  0.04998 *
Ind1           247.2927   920.2953   0.269  0.78950
Ind2           403.9173   217.3517   1.858  0.07031 .
Ind3           629.2076   218.6483   2.878  0.00633 **
advrtsng:Ind1   -0.7603     1.5671  -0.485  0.63012
advrtsng:Ind2   -1.5474     0.4814  -3.214  0.00255 **
advrtsng:Ind3   -1.2114     0.3506  -3.455  0.00129 **
lagsales:Ind1    0.1588     0.6539   0.243  0.80937
lagsales:Ind2    0.6277     0.2428   2.585  0.01340 *
lagsales:Ind3    0.1403     0.1758   0.798  0.42925
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 141.9 on 41 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.9598,     Adjusted R-squared: 0.949
F-statistic: 89.03 on 11 and 41 DF,  p-value: < 2.2e-16
```

Partial *F* tests for the two sets of interactions follow.

```
> modelintrctn2<-
lm(sales~advrtsng+lagsales+Ind1+Ind2+Ind3+lagsales*Ind1+lagsales*Ind2+l
agsales*Ind3);summary(modelintrctn2)

> modelintrctn3<-
lm(sales~advrtsng+lagsales+Ind1+Ind2+Ind3+advrtsng*Ind1+advrtsng*Ind2+a
dvrtsng*Ind3);summary(modelintrctn3)
```

```
> anova(modelintrctn2,modelintrctn1)
Analysis of Variance Table

  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1     44 1102297
2     41  825567  3    276730 4.5811 0.007411 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(modelintrctn3,modelintrctn1)
Analysis of Variance Table

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     44 969425
2     41 825567  3    143858 2.3815 0.0834 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interactions between *advrtsng* and the time period dummies are significant collectively (p = 0.0074), and the interactions between *lagsales* and the time period dummies are marginally significant (p = 0.0834). Let's retain both sets of interactions.

This model with interactions fits a separate geometric decay model to each of the four time periods. Here are the fits, and, for each, estimates of the 90 per cent duration interval (the lines in italics show the fits without the interaction terms, as given previously by model 4):

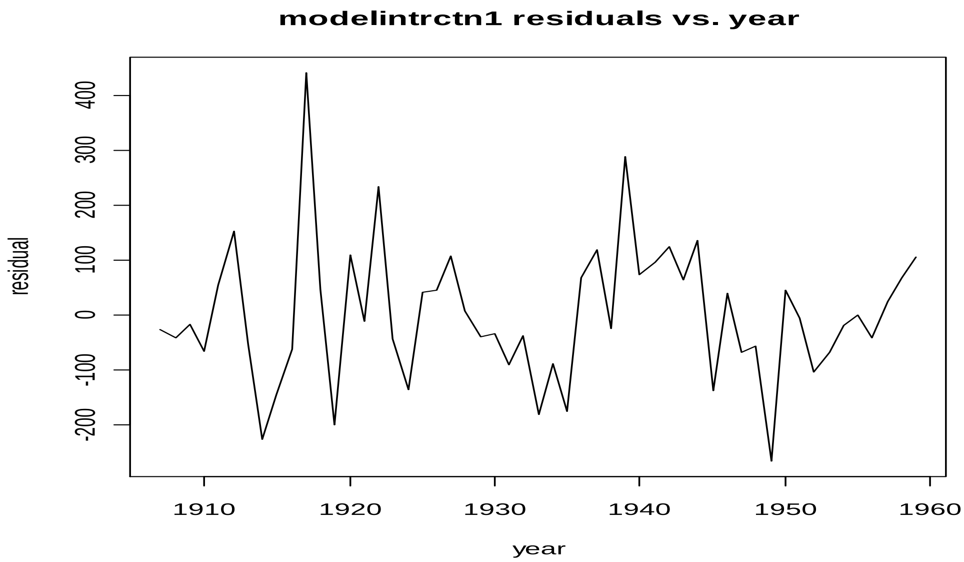
1907-1914: $S_t = 64.8439 + 0.9150A_t + 0.4630S_{t-1}$, 2.99 years
$S_t = 121.2925 + 0.5345A_t + 0.6073S_{t-1}$

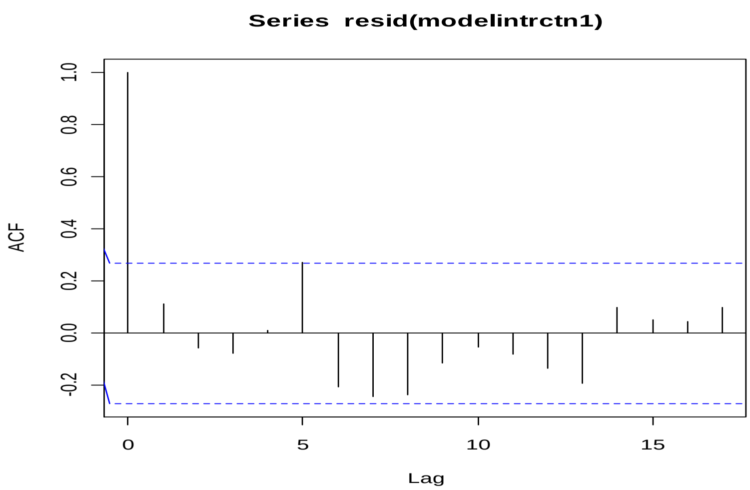
1915-1925: $S_t = 221.4685 + 0.1279A_t + 0.9319S_{t-1}$, 32.65 years
$S_t = 471.4788 + 0.5345A_t + 0.6073S_{t-1}$

1926-1940: $S_t = 446.7588 + 0.4638A_t + 0.4445S_{t-1}$, 2.84 years
$S_t = 52.1346 + 0.5345A_t + 0.6073S_{t-1}$

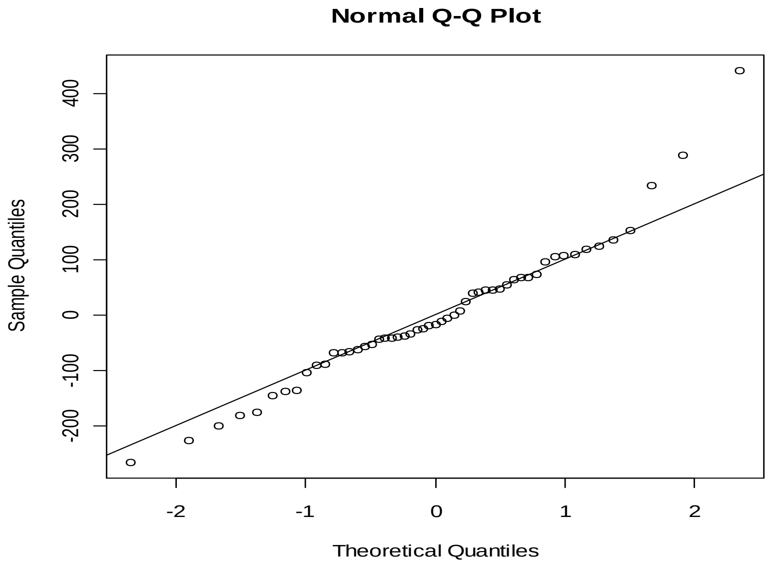
1941-1960: $S_t = -182.4488 + 1.6753A_t + 0.3042S_{t-1}$, 1.93 years
$S_t = 254.6392 + 0.5345A_t + 0.6073S_{t-1}$

This fit with interactions does give residuals which indicate decent reduction to white noise (see below). The fit does suggest some structural differences between the time periods. However, the time periods are rather short, especially the first two. And the result for the second time period does not lead to a sensible estimate of a 90 per cent duration interval.





The normal quantile plot of the residuals shows underprediction by the model of three relatively large annual sales figures.



More on use of dummy variables with R. Let's perform some more calculations with R for the Lydia Pinkham data.

```
> levels(fadvcopy)
[1] "1" "2" "3" "4"
```

The following fit is equivalent to the overall fit of model 6 above. Here the dummies defined by *fadavcopy* are used, rather than *Ind1*, *Ind2*, and *Ind3*.

```
> model71<-
lm(sales~fadvcopy+advrtsng+lagsales+lag2sales+lagadv);summary(model71)

Call:
lm(formula = sales ~ fadvcopy + advrtsng + lagsales + lag2sales +
    lagadv)

Residuals:
    Min      1Q  Median      3Q     Max
-333.56  -98.41  -12.37   85.55  436.33

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 143.3039    80.0307   1.791 0.080237 .
fadvcopy2   275.0504    97.9004   2.809 0.007376 **
fadvcopy3    28.1850    96.9147   0.291 0.772553
fadvcopy4   116.5564    88.3765   1.319 0.194038
advrtsng      0.5500     0.1339   4.109 0.000171 ***
lagsales      0.9654     0.1503   6.425    8e-08 ***
lag2sales    -0.2050     0.1163  -1.763 0.084863 .
lagadv       -0.3447     0.1510  -2.282 0.027377 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 165 on 44 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.9391,     Adjusted R-squared: 0.9294
F-statistic: 96.96 on 7 and 44 DF,  p-value: < 2.2e-16
```

Next we change the base for comparison to level 2. The refit follows.

```
> fadvcopy2<-factor(fadvcopy,levels=c(2,1,3,4))
> model72<-
lm(sales~fadvcopy2+advrtsng+lagsales+lag2sales+lagadv);summary(model72)

Call:
lm(formula = sales ~ fadvcopy2 + advrtsng + lagsales + lag2sales +
    lagadv)

Residuals:
    Min      1Q  Median      3Q     Max
-333.56  -98.41  -12.37   85.55  436.33

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  418.3543   102.9898   4.062 0.000197 ***
fadvcopy21  -275.0504    97.9004  -2.809 0.007376 **
fadvcopy23  -246.8654    88.3803  -2.793 0.007698 **
fadvcopy24  -158.4939    66.4923  -2.384 0.021523 *
advrtsng       0.5500     0.1339   4.109 0.000171 ***
lagsales       0.9654     0.1503   6.425    8e-08 ***
lag2sales     -0.2050     0.1163  -1.763 0.084863 .
lagadv        -0.3447     0.1510  -2.282 0.027377 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 165 on 44 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared: 0.9391,     Adjusted R-squared: 0.9294
F-statistic: 96.96 on 7 and 44 DF,  p-value: < 2.2e-16
```

Each of the two models above includes a factor variable with four levels among the explanatory variables. In the fitting of each, R creates three 0–1 dummy variables. And, as noted, the level which is not assigned a dummy, becoming the base for comparison, is the level whose label is first in the dictionary ordering or in the ordering which has been specified.

The two model fits give the same results. Each has a different assignment of dummy variables to deal with the four regimes of advertising copy. In each model there are an intercept and three adjustments to the intercept. When properly combined, one obtains a different overall intercept for each level of the categorical variables. The calculations are illustrated next.

model71 The base for comparison is level 1.

```
                          level   adjusted intercept
(Intercept) 143.3039        1     143.3039
fadvcopy2   275.0504        2     143.3039 + 275.0504 = 418.3543
fadvcopy3    28.1850        3     143.3039 + 28.1850 = 171.4889
fadvcopy4   116.5564        4     143.3039 + 116.5564 = 259.8603
```

model72 The base for comparison is level 2.

```
                          level   adjusted intercept

(Intercept) 418.3543        2     418.3543
fadvcopy21 -275.0504        1     418.3543 - 275.0504 = 143.3039
fadvcopy23 -246.8654        3     418.3543 – 246.8654 = 171.4889
fadvcopy24 -158.4939        4     418.3543 – 158.4939 = 259.8604
```

Summary and additional remarks

1. A distributed lag regression model contains as independent variables contemporaneous and time lagged values of one or more explanatory variables. In essence, such a model relates a response variable to current and past values of one or more explanatory variables.

2. In a special case of a distributed lag model, there is one explanatory variable, the response is related to the current and lagged values of this explanatory variable, and the coefficients of the lagged values of the explanatory variable decay in geometric fashion as the lag number increases. When this is the case, the model can be alternatively expressed as one with autoregressive lag one structure and just the contemporaneous version of the explanatory variable.

3. The Lydia Pinkham data set contains annual sales and advertising figures for the Lydia E. Pinkham Medicine Company. The company had no competitors, and its only vehicle for marketing the single product it sold was advertising, primarily in print. This makes the company’s data very well-suited to study the relationship between sales and advertising.

4. The distributed lag models allow one to estimate what is termed the 90 percent duration interval. When the response variable represents sales, and the distributed lag explanatory variable represents advertising, this interval estimates the length of time into the future that current advertising continues to have an impact upon sales. Such a calculation is useful in considering policy for the deduction of advertising expenses in tax policy.

5. Using the Lydia Pinkham data, these notes illustrate the use of several different methods of defining dummy variables.