A Solution Manual and Notes for: Statistical Methods for Forecasting by Boyas Abraham and Johannes Ledolter.

John L. Weatherwax*

July 27, 2009

Introduction

From amazon:

Presents the statistical methods and models used in the production of short-term forecasts. Deals with special topics such as transfer function analysis, Kalman filtering, state space models, Bayesian forecasting, and forecast evaluation. Explains their interconnections, and bridges the gap between theory and practice. Provides time series, autocorrelation, and partial autocorrelation plots. Includes examples and exercises using real data.

I would appreciate constructive feedback (sent to the email below) on any errors that are found in these notes. I hope you enjoy this book as much as I have and that these notes might help the further development of your skills in forecasting.

*wax@alum.mit.edu

Chapter 2: The Regression Model and its Applications in Forecasting

Notes On The Text

Derivations of some Properties of Least Square Estimators

The sum of squares total (SSTO) is defined as and can be simplified as

$$\begin{aligned}
\text{SSTO} &= \sum (y_t - \bar{y})^2 = \sum (y_t(y_t - \bar{y}) - \bar{y}(y_t - \bar{y})) \\
&= \sum y_t(y_t - \bar{y}) - \bar{y}\sum (y_t - \bar{y}) \\
&= \sum y_t(y_t - \bar{y}) \,,
\end{aligned}$$

since this second sum vanishes, that is $\sum (y_t - \bar{y}) = 0$. With this, the above becomes

$$\text{SSTO} = \sum y_t^2 - \bar{y}\sum y_t = \sum y_t^2 - n\bar{y}^2 \,. \tag{1}$$

Now the sum of squares due to the *regression* (SSR) is defined and can be simplified as

$$\begin{aligned}
\text{SSR} &= \sum (\hat{y}_t - \bar{y})^2 = \sum (\hat{y}_t^2 - 2\hat{y}_t\bar{y} + \bar{y}^2) \\
&= \sum \hat{y}_t^2 - 2\bar{y}\sum \hat{y}_t + n\bar{y}^2 \,.
\end{aligned}$$

Now the predicted response $\hat{y}_t$ is equal to the true response y_t plus some error residual e_t as $\hat{y}_t = y_t + e_t$, so that the second sum above becomes

$$\sum \hat{y}_t = \sum y_t + \sum e_t = n\bar{y} \,,$$

as $\sum e_t = 0$ if a constant (say β_0) is included in the regression. Finally we find for SSR given by

$$\text{SSR} = \sum \hat{y}_t^2 - n\bar{y}^2 \,. \tag{2}$$

Defining the sum of squares due to regression *error* (SSE) as $\sum (y_t - \hat{y}_t)^2$ we can decompose the total sum of squares as

$$\begin{aligned}
\text{SSTO} &= \sum (y_t - \bar{y})^2 = \sum (y_t - \hat{y}_t + \hat{y}_t - \bar{y})^2 \\
&= \sum (y_t - \hat{y}_t)^2 + 2(y_t - \hat{y}_t)(\hat{y}_t - \bar{y}) + (\hat{y}_t - \bar{y})^2 \\
&= \text{SSE} + 2\sum (y_t - \hat{y}_t)(\hat{y}_t - \bar{y}) + \text{SSR} \,.
\end{aligned}$$

The middle term of the above expression becomes

$$\sum (y_t - \hat{y}_t)(\hat{y}_t - \bar{y}) = \sum e_t(\hat{y}_t - \bar{y}) = \sum e_t\hat{y}_t - \bar{y}\sum e_t = \sum e_t\hat{y}_t \,.$$

This last expression (in vector notation) is $\hat{y}'e$. Now again in vector notation since the predicted values $\hat{y}$ are given by $\hat{y} = X\hat{\beta}$ the expression $\hat{y}'e$ becomes $\hat{\beta}'X'e$, which is zero since $X'e = 0$.

Prediction from regression models with *estimated* coefficients

In this section of the book we derive several properties of least squares estimates. One in particular is the variance of the forecast error. This can be derived as follows

$$
\begin{aligned}
V(y_k - \hat{y}_k^{\text{pred}}) &= V(\varepsilon_k + x_k'(\beta - \hat{\beta})) \\
&= \sigma^2 + V(x_k'(\beta - \hat{\beta})) \quad (3) \\
&= \sigma^2 + x_k' V(\beta - \hat{\beta}) x_k \quad (4) \\
&= \sigma^2 + x_k' V(\hat{\beta}) x_k \quad (5) \\
&= \sigma^2 + \sigma^2 x_k' (X'X)^{-1} x_k \, .
\end{aligned}
$$

Where in going from Equation 3 to Equation 4 we have used the fact that $V(a'x) = a'V(x)a$, when a is a constant vector and x is a random random vector. In going from Equation 4 to Equation 5 we have used the fact that adding a constant (here the true value of β) does not affect the value of a variance.

Some Discussion on the Examples

Example 2.1: Running Performance

When we add the additional variables X_2, X_3, X_4, X_5, representing height, weight, skin-fold sum and relative body fat, to test their *individual* significance as if they were the last variable added in the regression we compare the magnitude of their computed t statistic $t_{\hat{\beta}_i}$, against the value of $t_{\alpha/2}(n-p-1)$ as discussed in the section entitled: Hypothesis Tests for Individual Components. This example has $n = 14$ data points and when we have added all variables to the model we have $p = 5$ so that $t_{\alpha/2}(n-p-1) = t_{\alpha/2}(8)$. For a significance level for our test of $\alpha = 0.05$ this is the value of 2.306. The magnitude of all of the computed t statistics for the variables X_2, X_3, X_4, X_5 are smaller than this value and heuristically we can conclude that they are insignificant.

Example 2.2: Gas Mileage Data

For the gas mileage data set with $n = 38$ to determine if the estimate of the coefficient of the quadratic term β_2 is statistically significant we consider the partial F or partial t test. The partial t test could compare the magnitude of $t_{\hat{\beta}_2}$ against the value of

$$t_{\alpha/2}(n-p-1) = t_{0.025}(38-2-1) = t_{0.025}(35) = 2.032 \, .$$

Since $|t_{\hat{\beta}_2}|$ is less than this value we conclude the coefficient β_2 is actually insignificant.

The partial F test would consider the statistic $F^* = (t_{\hat{\beta}_2})^2 = (-1.50)^2 = 2.25$, which is to be compared against the value

$$F_\alpha(1, n-p-1) = F_{0.05}(1, 35) = 4.1213 \, .$$

Since $F^* < F_{0.05}(1, 35)$ we cannot reject H_0 at the level α. In other words there is *not* significant evidence that value of β_2 is different than zero, the same conclusion reached above. As discussed in the text, because of the identity

$$F_\alpha(1, n-p-1) = t^2_{\alpha/2}(n-p-1)\,, \tag{6}$$

the ordinary t test and the extra sum of squares test (the F test) are equivalent.

Exercise Solutions

Exercise 2.1 (an example with least squares)

Part (a): Forming the design matrix the rows of which are the augmented observations we find

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix},$$

where in this example $n = 10$. We desire estimations of the coefficients of our model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t\,. \tag{7}$$

The coefficients β_i are given by solving the normal equations

$$(X'X)\hat{\beta} = X'y\,. \tag{8}$$

The right-hand side of this expression is given by

$$X'y = \begin{bmatrix} \sum y_t \\ \sum x_{t1}y_t \\ \sum x_{t2}y_t \end{bmatrix} = \begin{bmatrix} 10 \\ 40 \\ 40 \end{bmatrix}.$$

Next we compute the normal matrix $X'X$ given by

$$\begin{aligned} X'X &= \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \\ &= \begin{bmatrix} n & \sum x_{t1} & \sum x_{t2} \\ \sum x_{t1} & \sum x_{t1}^2 & \sum x_{t1}x_{t2} \\ \sum x_{t2} & \sum x_{t1}x_{t2} & \sum x_{t2}^2 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 40 \end{bmatrix}. \end{aligned}$$

Thus $\hat{\beta}$ is given by $\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

Part (b): The concept of ANOVA is discussed in the text and the basic form for an ANOVA table given there looks like

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	$\text{SSR} = \sum(\hat{y}_t - \bar{y})^2 = \hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2$	p	$\text{MSR} = \frac{\text{SSR}}{p}$	$\frac{\text{MSR}}{\text{MSE}}$
Error	$\text{SSE} = e'e$	$n-p-1$	$\text{MSE} = \frac{\text{SSE}}{n-p-1}$	
Total (correlated for mean)	$\text{SSTO} = \sum(y_t - \bar{y})^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2$	$n-1$		

Now for this problem we have $\bar{y} = \frac{1}{10}\sum y_t = 1$ and using the above we find the product in SSR given by $\hat{\beta}'X'y = 130$ and we have enough information to calculate the value of SSR. Now to calculate SSE with the given information recall that

$$
\begin{aligned}
\text{SSE} &= e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2\hat{\beta}X'y + \hat{\beta}X'X\hat{\beta}\,.
\end{aligned}
$$

Note that from the given problem statement one can compute each of these terms. Doing so we find the SSE given by $e'e = 115$. Finally, the coefficient of determination

$$
R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSR} + \text{SSE}} = 0.774\,.
$$

Larger values of R^2 are "better" implying the ability to explain more of the variation in y around it mean. So for this problem we find that the specific ANOVA table is given by

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	$\text{SSR} = 120$	$p = 2$	$\text{MSR} = 60$	12
Error	$\text{SSE} = 35$	$n-p-1 = 7$	$\text{MSE} = 5$	
Total (correlated for mean)	$\text{SSTO} = 155$	$n-1 = 9$		

Part (c): Because the *estimated* covariance matrix of the least square estimator of $\hat{\beta}$ is given by

$$
\hat{V}(\hat{\beta}) = s^2(X'X)^{-1}\,, \tag{9}
$$

with s^2 is estimated by

$$
s^2 = \frac{1}{n-p-1}\text{SSE} = \frac{1}{n-p-1}\sum(y_t - \hat{y}_t)^2\,. \tag{10}
$$

The estimated standard error on an individual component β_i or $\hat{\beta}_i$ is given by $s_{\hat{\beta}_i} = s\sqrt{c_{ii}}$, with c_{ii} the diagonal element in $(X'X)^{-1}$. Defining C to be the inverse of $X'X$ we compute that for this problem

$$
C \equiv (X'X)^{-1} = \frac{1}{40}\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\,.
$$

The t statistic using in testing the null hypothesis $H_0 : \beta_i = \beta_{i0} = 0$ against the alternative $H_1 : \beta_1 \neq \beta_{i0} = 0$ is obtained by computing

$$t_i = \frac{\hat{\beta}_i - \beta_{i0}}{s\sqrt{c_{ii}}} \quad \text{for} \quad i = 0, 1, 2\,.$$

For the three regression coefficients given here we compute as a vector of t statistics (assuming $\beta_{i0} = 0$) given by $\begin{bmatrix} 1.4142 \\ 4.0000 \\ 2.8284 \end{bmatrix}$. To test significance of these values we select a value for α (say $\alpha = 0.05$) and then compute the corresponding $t_{\alpha/2}(n-p-1) = t_{0.025}(7) = 2.36$ threshold. If $|t_i| > t_{\alpha/2}(n-p-1)$ we reject H_0 in favor of H_1 at the significance of α. Since the value of the t statistics for β_i $i = 1, 2$ is larger than this value we conclude that the hypothesis H_0 can be rejected at the level α.

Part (d): For the joint test $H_0 : \beta_1 = \beta_2 = 0$ we need a simultaneous test of H_0 against the hypothesis H_1 : at least one of $\beta_i \neq 0$. If the null hypothesis is true then $F = \frac{\text{MSR}}{\text{MSE}}$, is given by an F distribution with p and $n-p-1$ degrees of freedom. Thus to asses the significance of the observed F value we should compare it with the value

$$F_\alpha(p, n-p-1) = F_{0.05}(2, 7) = 4.73\,.$$

Where if $F > F_\alpha(p, n-p-1) = 4.73$ we reject H_0 in favor of H_1 while if this is not true we cannot reject H_0 in favor of H_1. For this problem we find $F = 12$ considerably greater than $F_{0.05}(2, 7)$ indicating that we can reject H_0 at the significance of α. The simple numerical calculations performed in this problem are worked in the MATLAB script `prob_2_1.m`.

Exercise 2.2 (the normal equations for a linear trend model)

For this linear trend model $y_t = \beta_0 + \beta_1 t + \varepsilon_t$ for $t = 1, \cdots, n$, the normal equations are given by Equation 8 where in this case the design matrix $X'X$ is given by

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}.$$

So that with this expression $X'X$ then looks like

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n i \\ \sum_{i=1}^n i & \sum_{i=1}^n i^2 \end{bmatrix} = \begin{bmatrix} n & \frac{1}{2}n(n+1) \\ \frac{1}{2}n(n+1) & \frac{1}{6}n(n+1)(2n+1) \end{bmatrix}$$

Thus our least squares estimate of β is given by

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} n & \frac{1}{2}n(n+1) \\ \frac{1}{2}n(n+1) & \frac{1}{6}n(n+1)(2n+1) \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum_{t=1}^n t y_t \end{bmatrix}.$$

We can take the inverse of the above $X'X$ matrix to find

$$(X'X)^{-1} = \frac{1}{n(n^2-1)} \left[\begin{array}{cc} 2(1+n)(1+2n) & -6(1+n) \\ -6(1+n) & 12 \end{array} \right].$$

See the Mathematical file `prob_2_2.nb` where we perform these manipulations. Once we have this, our estimates of β can then be written by multiplying the above together as

$$\left[\begin{array}{c} \hat{\beta}_0 \\ \hat{\beta}_1 \end{array} \right] = \frac{1}{n(n^2-1)} \left[\begin{array}{c} 2(n+1)(2n+1)\sum y_t - 6(n+1)\sum t y_t \\ -6(n+1)\sum y_t + 12 \sum t y_t \end{array} \right].$$

With this we see that $\hat{\beta}_1$ is given by

$$\hat{\beta}_1 = \frac{12}{n(n^2-1)} \left(\sum_{t=1}^{n} \left(t - \frac{1}{2}(n+1) \right) y_t \right). \tag{11}$$

Since we can solve for $\sum t y_t$ in terms of $\hat{\beta}_1$ as

$$\frac{12}{n(n^2-1)} \sum t y_t = \hat{\beta}_1 + \frac{6}{n(n-1)} \sum y_t ,$$

we can put that in the equation for $\hat{\beta}_0$ to get that $\hat{\beta}_0$ in terms of $\hat{\beta}_1$ given by

$$\begin{aligned} \hat{\beta}_0 &= \frac{2(1+2n)}{n(n-1)} \sum y_t - 6(n+1) \left(\frac{1}{12} \hat{\beta}_1 + \frac{1}{2} \frac{1}{n(n-1)} \sum y_t \right) \\ &= \frac{1}{n} \sum y_t - \frac{1}{2}(n+1)\hat{\beta}_1 \\ &= \bar{y} - \frac{1}{2} \hat{\beta}_1 . \end{aligned}$$

These expressions will be used when we initialize the procedure used in performing double exponential smoothing in Chapter 3, see page 35.

Exercise 2.3 (an incomplete ANOVA table)

We are told to consider a linear model, $y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \varepsilon_t$, with explanatory variables X_1 per capital real income, X_2 relative price of beer, and a dependent variable Y of per-capita beer consumption.

Part (a): The requested test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at the $\alpha = 0.05$ level is a test of the individual coefficient and depends on the value of

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{s\sqrt{c_{ii}}} = \frac{1.14 - 0}{0.16} = 7.1250 .$$

This is to be compared with the value of $t_{\alpha/2}(n-p-1) = t_{0.025}(17-2-1) = t_{0.025}(14) = 2.144$. If $|t| > t_{0.025}(14)$ we reject H_0 in favor of H_1 while if $|t| < t_{0.025}(14)$ there is not enough evidence for rejecting H_0 at the significance level 0.05. In this case here t is significantly larger than $t_{0.025}(14)$ and we reject H_0 (therefore we "accept" the hypothesis H_1).

Part (b): Because SSTO = SSR + SSE and we are told that SSTO = 100 and SSE = 34 we know SSR = 66. Since $n = 17$ and $p = 2$ we find

$$\text{MSR} = \frac{\text{SSR}}{p} = \frac{66}{2} = 33\,,$$

and

$$\text{MSE} = \frac{\text{SSE}}{n-p-1} = \frac{34}{14} = 2.42\,.$$

Finally $F = \frac{\text{MSR}}{\text{MSE}} = 13.58$. Thus the ANOVA table in this case looks like

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	SSR = 66	$p = 2$	MSR = 33	13.58
Error	SSE = 34	$n - p - 1 = 14$	MSE = 2.42	
Total (correlated for mean)	SSTO = 100	$n - 1 = 16$		

and $R^2 = \frac{\text{SSR}}{\text{SSTO}} = \frac{66}{100} = 0.66$. To test the simultaneous hypothesis that $H_0 : \beta_1 = \beta_2 = 0$ against the hypothesis that H_1 : at least one $\beta_i \neq 0$, recall the F statistic above follows an F distribution with $p = 2$ and $n - p - 1 = 14$ degrees of freedom. Thus we need to compare the value of F obtained above with the value of $F_\alpha(p, n-p-1) = F_{0.05}(2, 14) = 3.7389$. If the value of the F ratio above is such that $F > F_{0.05}(p, n-p-1)$ then we can reject H_0 in favor of H_1 at the level α. Since this is true we can reject the hypothesis H_0. Some simple calculations for this problem are done in the MATLAB script `prob_2_3.m`.

Exercise 2.4 (trigonometrically spaced observations)

Part (a): For a problem of such small size we can write down the normal equations exactly. Our designs matrix in this case is given by

$$X = \left[\begin{array}{ccc} 1 & \cos(\theta) & \sin(\theta) \\ 1 & -\sin(\theta) & -\cos(\theta) \\ 1 & \sin(\theta) & \cos(\theta) \\ 1 & -\cos(\theta) & -\sin(\theta) \end{array}\right],$$

so that $X'X$ becomes

$$\begin{aligned} X'X &= \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \cos(\theta) & -\sin(\theta) & \sin(\theta) & -\cos(\theta) \\ \sin(\theta) & -\cos(\theta) & \cos(\theta) & -\sin(\theta) \end{array}\right] \left[\begin{array}{ccc} 1 & \cos(\theta) & \sin(\theta) \\ 1 & -\sin(\theta) & -\cos(\theta) \\ 1 & \sin(\theta) & \cos(\theta) \\ 1 & -\cos(\theta) & -\sin(\theta) \end{array}\right] \\ &= \left[\begin{array}{ccc} 4 & 0 & 0 \\ 0 & 2 & 4\cos(\theta)\sin(\theta) \\ 0 & 4\cos(\theta)\sin(\theta) & 2 \end{array}\right] = \left[\begin{array}{ccc} 4 & 0 & 0 \\ 0 & 2 & 2\sin(2\theta) \\ 0 & 2\sin(2\theta) & 2 \end{array}\right], \end{aligned}$$

so that our least squares estimate of β is given by $(X'X)^{-1}X'y$ or

$$\hat{\beta} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2\sin(2\theta) \\ 0 & 2\sin(2\theta) & 2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ y_1\cos(\theta) - y_2\sin(\theta) + y_3\sin(\theta) - y_4\cos(\theta) \\ y_1\sin(\theta) - y_2\cos(\theta) + y_3\cos(\theta) - y_4\sin(\theta) \end{bmatrix}.$$

Part (b): Because the variance of our estimate $\hat{\beta}$ is given by $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$ to evaluate this we thus need to compute $(X'X)^{-1}$. We find since this matrix is block diagonal that its inverse can be computed as

$$(X'X)^{-1} = \begin{bmatrix} 1/4 & 0 \quad 0 \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{bmatrix} 2 & 2\sin(2\theta) \\ 2\sin(2\theta) & 2 \end{bmatrix}^{-1} \end{bmatrix}.$$

Now

$$\begin{aligned} \begin{bmatrix} 2 & 2\sin(2\theta) \\ 2\sin(2\theta) & 2 \end{bmatrix}^{-1} &= \frac{1}{(4 - 4\sin^2(2\theta))} \begin{bmatrix} 2 & -2\sin(2\theta) \\ -2\sin(2\theta) & 2 \end{bmatrix} \\ &= \frac{1}{4\cos^2(2\theta)} \begin{bmatrix} 2 & -2\sin(2\theta) \\ -2\sin(2\theta) & 2 \end{bmatrix}. \end{aligned}$$

Thus

$$(X'X)^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\cos^2(2\theta)} & -2\frac{\sin(2\theta)}{\cos^2(2\theta)} \\ 0 & -\frac{2\sin(2\theta)}{\cos^2(2\theta)} & \frac{2}{\cos^2(2\theta)} \end{bmatrix}.$$

From which we see that when we multiply this by σ^2 and consider the second and third element of the diagonal we have

$$V(\hat{\beta}_1) = \frac{\sigma^2}{2\cos^2(2\theta)} = V(\hat{\beta}_2),$$

as we were to show.

Exercise 2.5 (some linear regression)

Part (a): For the given regression model we have $p = 2$, $n = 20$, $n - p - 1 = 17$. Since $\text{SSTO} = \text{SSR} + \text{SSE}$ we see that $\text{SSE} = \text{SSTO} - \text{SSR} = 200 - 60 = 134$. Then with these values the ANOVA table looks like

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	SSR = 66	$p = 2$	MSR = 33	4.1866
Error	SSE = 134	$n - p - 1 = 17$	MSE = 7.88	
Total (correlated for mean)	SSTO = 200	$n - 1 = 19$		

Part (b): Now we have $R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}} = 0.33$.

Part (c): To test the joint hypothesis that $\beta_1 = \beta_2 = 0$ (or that both coefficients are in fact zero) at the 0.05 significance level we compare the F ratio computed above with the value of $F_\alpha(p, n-p-1) = F_{0.05}(2, 17) = 3.59$. If $F > F_{0.05}(2, 17)$ we reject $H_0 : \beta_1 = \beta_2 = 0$ at the level α. In this case this is indeed true and we can reject the hypothesis H_0 at the level α.

Part (d): Under the simpler model SSR $= 50$ so the extra regression sum of squares

$$\text{SSR}(X_2|X_1) = \text{SSR}(X_1, X_2) - \text{SSR}(X_1) = 66 - 50 = 16\,.$$

To test the hypothesis that the addition of X_2 has significantly reduced SSR we compute the F^* statistic as

$$F^* = \frac{\text{SSR}(X_2|X_1)/(p-q)}{\text{SSE}(X_1, X_2)/(n-p-1)} = \frac{\text{SSR}(X_2|X_1)/(2-1)}{\text{SSE}(X_1, X_2)/(20-2-1)} = \frac{16}{(134/17)} = 2.03\,.$$

If $F^* > F_\alpha(p-q, n-p-1) = F_{0.05}(1, 17) = 4.45$ then we reject H_0 the hypothesis that all additional variables are indeed non-significant and should be taken as zero simultaneity. Since in this case we find that the value of F^* is *not* large enough we cannot reject the hypothesis H_0. This gives the indication that the coefficient β_2 does not provide enough reduction in the SSR when included in the model and should be dropped.

Part (e): If we consider the correlation of X_{2t} against the residuals, $e_t = y_t - \hat{y}_t$ in Part (d), if X_{2t} has predictive power this correlation should be non-zero.

Some simple calculations for this problem are performed in the MATLAB script `prob_2_5.m`.

Exercise 2.6 (the extra sum of squares)

Part (a): To test the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ at the $\alpha = 0.05$ level against the hypothesis H_1 : at least one $\beta_i \neq 0$ we considering a simultaneous test of all regression coefficients. In that case if the null hypothesis is true then the statistic F

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/p}{\text{SSE}/(n-p-1)}\,.$$

follows an F distribution with p and $n-p-1$ degrees of freedom. For this problem $p = 3$, $n = 10$, so $n-p-1 = 6$. To compute SSE recognized that since $R^2 = \frac{\text{SSR}}{\text{SSTO}}$ and SSTO $= 100$ we see that SSR $= 88$ so that SSE $=$ SSTO $-$ SSR $= 12$. Thus our F statistic for this joint test in this case becomes

$$F = \frac{88/3}{12/6} = 14.667\,.$$

While the value of $F_\alpha(p, n-p-1) = F_{0.05}(3, 6) = 4.75$. Since $F > F_{0.05}(3, 6)$ we conclude that we can reject H_0 at the significance level of α.

Part (b): Partial F tests for the variables X_1, X_2, X_3 involve deciding whether the inclusion of the variable X_i is helpful when considered after all the remaining variables X_j. Thus our partial F tests for the variables X_1, X_2, and X_3 involve computing F values given by

$$\begin{aligned} F_1 &= \frac{\text{SSR}(X_1|X_2, X_3)}{\text{SSE}(X_1, X_2, X_3)/(n-p-1)} = \frac{\text{SSR}(X_1|X_2, X_3)}{2} \\ F_2 &= \frac{\text{SSR}(X_2|X_1, X_3)}{\text{SSE}(X_1, X_2, X_3)/(n-p-1)} = \frac{\text{SSR}(X_2|X_1, X_3)}{2} = 1 \\ F_3 &= \frac{\text{SSR}(X_3|X_1, X_2)}{\text{SSE}(X_1, X_2, X_3)/(n-p-1)} = \frac{\text{SSR}(X_3|X_1, X_2)}{2}, \end{aligned}$$

which we now need to evaluate in terms of the given information. We begin with $\text{SSR}(X_1|X_2, X_3)$ since the others are similar. Now we find from its definition that

$$\text{SSR}(X_1|X_2, X_3) = \text{SSR}(X_1, X_2, X_3) - \text{SSR}(X_2, X_3).$$

We know the value of $\text{SSR}(X_1, X_2, X_3) = 88$ so to evaluate this we need to compute $\text{SSR}(X_2, X_3)$ from the given information. From the definition of $\text{SSR}(X_3|X_2)$ we find

$$\text{SSR}(X_3|X_2) = \text{SSR}(X_3, X_2) - \text{SSR}(X_2).$$

Therefore solving for $\text{SSR}(X_2, X_3)$ we find

$$\text{SSR}(X_2, X_3) = \text{SSR}(X_3|X_2) + \text{SSR}(X_2).$$

This finally gives for $\text{SSR}(X_1|X_2, X_3)$ the following expression

$$\text{SSR}(X_1|X_2, X_3) = \text{SSR}(X_1, X_2, X_3) - \text{SSR}(X_3|X_2) - \text{SSR}(X_2) = 88 - 1 - 40 = 47.$$

In the same way we have for the numerator in the expression for F_3 the following

$$\text{SSR}(X_3|X_1, X_2) = \text{SSR}(X_1, X_2, X_3) - \text{SSR}(X_1|X_2) - \text{SSR}(X_2) = 88 - 45 - 40 = 3.$$

With these expressions evaluated we can now compute F_i. We find

$$F_1 = 23.5, \quad F_2 = 1.0, \quad F_3 = 1.5.$$

These are to be compared to the value of

$$F_\alpha(p-q, n-p-1) = F_\alpha(1, n-p-1) = F_{0.05}(1, 6) = 5.98.$$

The only statistic that is larger than this value is F_1, implying that β_1 is the only statistically significant coefficient.

Part (c): To test the hypothesis $H_0 : \beta_2 = \beta_3 = 0$ against the alternative hypothesis that H_1 : at least one of β_2 or β_3 is nonzero, we recognize that this problem is comparing a larger model based on the three variables X_1, X_2, X_3 with the smaller model based only on X_1. Thus we compute the extra regression sum of squares attributed to the variables X_2 and X_3 as

$$\text{SSR}(X_2, X_3|X_1) = \text{SSR}(X_1, X_2, X_3) - \text{SSR}(X_1) = 88 - 82 = 6,$$

and since $\text{SSE}(X_1, X_2, X_3) = \text{SSTO} - \text{SSR}(X_1, X_2, X_3) = 100 - 88 = 12$. The statistics we then compute since $p = 3$ and $q = 1$ is

$$F^* = \frac{\text{SSR}(X_2, X_3|X_1)/(p-q)}{\text{SSE}(X_1, X_2, X_3)/(n-p-1)} = \frac{6/(3-1)}{18/(10-3-1)} = 1.5\,.$$

This is to be compared against the value of $F_\alpha(p-q, n-p-1) = F_{0.05}(2, 6) = 5.143$. Since the value of F^* is smaller than the value of $F_{0.05}(2, 6)$ we cannot reject H_0 at the $\alpha = 0.05$ significance level. This implies that we have more confidence that the coefficients β_2 and β_3 are in fact zero, the same conclusion reached in Part (b).

Some simple calculations for this problem are performed in the MATLAB script `prob_2_6.m`.

Exercise 2.7 (filling in an lack-of-fit ANOVA table)

Our regression model is given by $y_t = \beta_0 + \beta_1 x_t$ and from the description given we have $n = 3 + 3 + 3 + 3 + 5 = 17$, $p = 1$, and we have $k = 5$ level replications. For lack-of-fit ANOVA tables when we have multiple replications the sum of squares lack-of-fit SSLF is defined as

$$\text{SSLF} = \text{SSE} - \text{SSPE} = \text{SSE} - \sum_{i=1}^{k} \sum_{t=1}^{n_i} (y_t^{(i)} - \bar{y}^{(i)})^2 \qquad (12)$$

where SSPE is the sum of squares pure error. From the partial complete ANOVA table given we see that $\text{SSLF} = 30 - 9 = 21$. The remaining entries for various mean square expressions are simply the sum of squares expressions divided by the degrees of freedom. Thus the ANOVA table is given by

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	$\text{SSR} = 73$	$p = 1$	$\text{MSR} = 73$	$F = \frac{\text{MSR}}{\text{MSE}} = 36.5$
Error	$\text{SSE} = 30$	$n - p - 1 = 15$	$\text{MSE} = 2$	
Lack of fit	$\text{SSLF} = 21$	$k - p - 1 = 3$	$\text{MSLF} = 7$	$F_{\text{LF}} = \frac{\text{MSLF}}{\text{MSPE}} = 9.33$
Pure error	$\text{SSPE} = 9$	$n - k = 12$	$\text{MSPE} = 0.75$	
Total (correlated for mean)	$\text{SSTO} = 103$	$n - 1 = 16$		

Part (b): To asses the models adequacy is to recognize that under the hypothesis that there is no lack-of-fit i.e. that the given model is accurate, F_{LF}, follows a F distribution with $k - p - 1$ and $n - k$ degrees of freedom. Thus we want to compare our value of F_{LF} above with $F_\alpha(k - p - 1, n - k) = F_{0.05}(3, 12) = 3.49$. As the given ratio F_{LF} is larger than this value we can conclude that there is significant lack-of-fit present.

It is interesting to note that even with a model that is inadequate in terms of lack-of-fit to the data, we can still test the significance of the parameters we have included. In this case there is only one coefficient to be estimated in this model (besides the y's mean value β_0) we will test the hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$. Under hypothesis H_0 the F statistic above is given by a $F(p, n-p-1) = F(1, 15)$ distribution. At the level $\alpha = 0.05$ we determine the value $F_\alpha(p, n-p-1) F_{0.05}(1, 15) = 4.54$. Since $F > F_{0.05}(1, 15)$ we reject H_0 in favor of H_1 at level α. Thus our model maybe incorrect but in the incorrect model the coefficient is significant.

Part (c): If there was no lack of fit we could pool the variance estimates from MSLF and MSPE to derive one estimate. In the above discussion it was found that there was significant model miss-match and that this is not possible. In this case then we will take the value of MSE $= 2$ as an estimate of the variance σ^2.

Some simple calculations for this problem are performed in the MATLAB script `prob_2_7.m`.

Exercise 2.8 (a first order autoregressive model)

Note that in this exercise the model we propose for the time series z_t is given by $z_t - 100 = \beta(z_{t-1} - 100) + \varepsilon_t$. If we define the variable, y_t, as $y_t = z_t - 100$ our model of the process y_t is given by $y_t = \beta y_{t-1} + \varepsilon_t$, or a first order autoregressive sequence.

Part (a): To estimate the value of β we recognize that the above model is exactly the linear regression through the origin example where we know the ordinary least squares estimate of β is given by

$$\hat{\beta} = \frac{\sum_{t=2}^{10} y_{t-1} y_t}{\sum_{t=2}^{10} y_{t-1}^2} = \frac{y_1 y_2 + y_2 y_3 + \cdots + y_9 y_{10}}{y_1^2 + y_2^2 + \cdots + y_9^2} = \frac{160}{216 - 4^2} = 0.8\,.$$

Part (b): Given this model we would predict the value of z_t to be

$$\hat{z}_t = \hat{y}_t + 100 = \hat{\beta} y_{t-1} + 100\,,$$

using the value of $\hat{\beta}$ computed above. Thus

$$\hat{z}_{11} = 0.8 y_{10} + 100 = 0.8(4) + 100 = 103.2\,.$$

Now z_{12} would be estimated as $\hat{z}_{12} = 0.8 y_{11} + 100$. Since at the time we desire the prediction of z_{12} we have not observed y_{11} an estimate of y_{11} could be $\hat{y}_{11} = \hat{\beta} y_{10} = 0.8(4) = 3.2$ so that we would estimate z_{12} as

$$\hat{z}_{12} = 0.8(3.2) + 100 = 102.56\,.$$

Part (c): We are estimating the value of β ($p = 1$) from $n = 9$ samples, then from the discussion in the book a $100(1-\alpha)$ percent prediction interval for y_{11} would be given by

$$\hat{\beta} y_{10} \pm t_{\alpha/2}(n-2) \left(1 + \frac{y_{10}^2}{\sum_{t=1}^{9} y_t^2} \right)^{1/2} = 3.2 \pm 2.457 = [0.7426, 5.657]\,.$$

Here $t_{\alpha/2}(n-2)$ is the $100(1-\alpha/2)$ percentage point of a t distribution with $n - 2 = 7$ degrees of freedom. We find $t_{\alpha/2}(7) = 2.36$. Note that this is slightly different than the expression the book has in that they have $t_{\alpha/2}(n-1)$ in the case of simple linear regression through the origin. The books later expression does not match the general expression of $t_{\alpha/2}(n-p-1)$ and I believe it is incorrect. Using this the 95% prediction interval for z_{11} would be that for y_{11} but with the mean of 100 added back, so we find the 95% confidence interval for z_{11} given by

$$[100.74, 105.66] .$$

Some simple calculations for this problem are performed in the MATLAB script `prob_2_8.m`.

Exercise 2.9 (the adjusted coefficient of determination R_a^2)

The adjusted coefficient of determination R_a^2 is defined as

$$R_a^2 = 1 - \frac{\text{SSE}/(n-p-1)}{\text{SSTO}/(n-1)} = 1 - \frac{(n-1)\text{SSE}}{(n-p-1)\text{SSTO}} . \tag{13}$$

For this problem we are told that the regression performed resulted in a 70% reduction in the residual standard deviation over what was present before the regression. Thus I took this to mean that

$$\frac{\text{SSR}}{\text{SSTO}} = 0.7 .$$

With this we then have that $\frac{\text{SSE}}{\text{SSTO}} = 1 - \frac{\text{SSR}}{\text{SSTO}} = 0.3$ and the adjusted coefficient of determination R_a^2 using Equation 13 becomes

$$R_a^2 = 1 - \frac{3(n-1)}{10(n-p-1)} .$$

This reduction in residual standard deviation will be weighted by the number of variable p used to achieve it.

Exercise 2.10 (some manipulation of the linear regression model)

Part (a): The vector prediction of the linear model are given by $\hat{y} = X\hat{\beta}$ so that $\hat{y}'\hat{y} = \hat{\beta}'X'X\hat{\beta}$. The estimated $\hat{\beta}$ is obtained from $\hat{\beta} = (X'X)^{-1}(X'y)$. When we put that expression in for the right most value of $\hat{\beta}$ only we find that

$$\hat{y}'\hat{y} = \hat{\beta}'(X'y) ,$$

as we were to show.

Part (b): The F statistics is calculated as

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/p}{\text{SSE}/(n-p-1)} = \left(\frac{n-p-1}{p}\right)\frac{\text{SSR}}{\text{SSE}} .$$

While the coefficient of determination R^2 is given by

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = \frac{\text{SSR}}{\text{SSR}+\text{SSE}} = \frac{\text{SSR}/\text{SSE}}{\text{SSR}/\text{SSE}+1} .$$

Solving this later equation for $\frac{\text{SSR}}{\text{SSE}}$ we find $\frac{\text{SSR}}{\text{SSE}} = \frac{R^2}{1-R^2}$, thus our F statistic now becomes

$$F = \left(\frac{n-p-1}{p}\right)\left(\frac{R^2}{1-R^2}\right),$$

as we were to show.

Exercise 2.11 (the bias associated with an incorrect model)

The least squares estimate for β_0 is given by the sample mean $\hat{\beta}_0 = \frac{1}{n}\sum_{t=1}^n y_t$. If the true model for y_t is in fact given by $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, i.e. a linear function of x_t then the estimate for β_0 computed above is equal to

$$\frac{1}{n}\sum_{t=1}^n y_t = \beta_0 + \frac{\beta_1}{n}\sum_{t=1}^n x_t + \frac{1}{n}\sum_{t=1}^n \varepsilon_t \,.$$

The expectation of this estimator $\frac{1}{n}\sum_{t=1}^n y_t$ is then given by

$$\beta_0 + \frac{\beta_1}{n}\sum_{t=1}^n x_t \,,$$

which has a bias from the true value of β_0 of

$$\frac{\beta_1}{n}\sum_{t=1}^n x_t = \beta_1 \bar{x} \,,$$

where $\bar{x}$ is the mean of the x values.

Exercise 2.12 (more bias with an incorrect model)

Part (a): If we assume our process follows the simple linear model $y = X_1\beta_1 + \varepsilon$, when in fact it satisfies the more complicated linear model $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ then our least squares estimate for β_1 assuming the simple model is the standard expression $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$. The expectation of this is given by

$$\begin{aligned} E(\hat{\beta}_1) &= (X_1'X_1)^{-1}X_1'E(y) \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \,, \end{aligned}$$

since the process y actually satisfies $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ the expectation of y is given by $E(y) = X_1\beta_1 + X_2\beta_2$.

Part (b): Now $\hat{\beta}_1$ will be an unbiased estimate of β_1 if and only if $X_1'X_2\beta_2 = 0$. That is β_2 happens to be in the null space of the matrix $X_1'X_2$ or $X_1'X_2$ is the zero matrix. This second condition implies that the predictors in X_1 and X_2 are orthogonal to each other.

Exercise 2.14 (related linear models)

Part (a): Our models for the dependent variables y_t and z_t are given by

$$\begin{aligned} y_t &= \beta_0 + \beta_1 + \beta_2 x_t + v_t \\ z_t &= \beta_0 - \beta_1 + \beta_2 x_t + w_t \,. \end{aligned}$$

A regression model for the given specifications follow if we list the $2m$ observations of z at $x \pm c$ followed by the $n - 2m$ observations of y at $x = 0$ as the matrix set of equations

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \\ z_{m+1} \\ z_{m+2} \\ \vdots \\ z_{2m} \\ y_{2m+1} \\ y_{2m+2} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & -1 & -c \\ 1 & -1 & -c \\ \vdots & \vdots & \vdots \\ 1 & -1 & -c \\ 1 & -1 & +c \\ 1 & -1 & +c \\ \vdots & \vdots & \vdots \\ 1 & -1 & +c \\ 1 & +1 & 0 \\ 1 & +1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & +1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \\ w_{m+1} \\ w_{m+2} \\ \vdots \\ w_{2m} \\ v_{2m+1} \\ v_{2m+2} \\ \vdots \\ v_n \end{bmatrix} .$$

Part (b): As presented above determining β_0, β_1, and β_2 is a standard least squares problem and the least squares estimate of these β is given by

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1} X' y \,.$$

Now in this case the matrix $X'X$ is given by

$$\begin{aligned} X'X &= \begin{bmatrix} n & -2m + (n - 2m) & -cm + cm \\ n - 4m & n & mc - mc + 0 \\ 0 & 0 & mc^2 + mc^2 \end{bmatrix} \\ &= \begin{bmatrix} n & n - 4m & 0 \\ n - 4m & n & 0 \\ 0 & 0 & 2mc^2 \end{bmatrix} . \end{aligned}$$

The inverse of this matrix is given by (since it is block diagonal it can be easily inverted)

$$(X'X)^{-1} = \begin{bmatrix} \frac{n}{8m(n-2m)} & \frac{-n+4m}{8m(n-2m)} & 0 \\ \frac{-n+4m}{8m(n-2m)} & \frac{n}{8m(n-2m)} & 0 \\ 0 & 0 & \frac{1}{2mc^2} \end{bmatrix} .$$

Then the expression $(X'X)^{-1}X'$ is given by

$$
\begin{aligned}
(X'X)^{-1}X' &= \begin{bmatrix} \frac{n}{8m(n-2m)} & \frac{-n+4m}{8m(n-2m)} & 0 \\ \frac{-n+4m}{8m(n-2m)} & \frac{n}{8m(n-2m)} & 0 \\ 0 & & \frac{1}{2mc^2} \end{bmatrix} \begin{bmatrix} +1 & \cdots & +1 & +1 & \cdots & +1 & 1 & \cdots & 1 \\ -1 & \cdots & -1 & -1 & \cdots & -1 & 1 & \cdots & 1 \\ -c & \cdots & -c & +c & \cdots & +c & 0 & \cdots & 0 \end{bmatrix} \\
&= \begin{bmatrix} +\frac{1}{4m} & \cdots & +\frac{1}{4m} & +\frac{1}{4m} & \cdots & +\frac{1}{4m} & \frac{1}{2(n-2m)} & \cdots & \frac{1}{2(n-2m)} \\ -\frac{1}{4m} & \cdots & -\frac{1}{4m} & -\frac{1}{4m} & \cdots & -\frac{1}{4m} & \frac{1}{2(n-2m)} & \cdots & \frac{1}{2(n-2m)} \\ -\frac{1}{2mc} & \cdots & -\frac{1}{2mc} & +\frac{1}{2mc} & \cdots & +\frac{1}{2mc} & 0 & \cdots & 0 \end{bmatrix} .
\end{aligned}
$$

When we multiply this expression by the concatenated response vector of z_t and y_t we find our least squares estimates of β given by

$$
\begin{aligned}
\hat{\beta}_0 &= +\frac{1}{4m}\sum_{t=1}^{m} z_t + \frac{1}{4m}\sum_{t=m+1}^{2m} z_t + \frac{1}{2(n-2m)}\sum_{t=2m+1}^{n} y_t \\
\hat{\beta}_1 &= -\frac{1}{4m}\sum_{t=1}^{m} z_t - \frac{1}{4m}\sum_{t=m+1}^{2m} z_t + \frac{1}{2(n-2m)}\sum_{t=2m+1}^{n} y_t \\
\hat{\beta}_2 &= -\frac{1}{2mc}\sum_{t=1}^{m} z_t + \frac{1}{2mc}\sum_{t=m+1}^{2m} z_t \, .
\end{aligned}
$$

Now the covariance of this estimate is given by

$$
V(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{bmatrix} \frac{n}{8m(n-2m)} & \frac{-n+4m}{8m(n-2m)} & 0 \\ \frac{-n+4m}{8m(n-2m)} & \frac{n}{8m(n-2m)} & 0 \\ 0 & & \frac{1}{2mc^2} \end{bmatrix} .
$$

We know that our predicted model for y is given by $\hat{y}^{\text{pred}} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 x$ and that the variance of its estimate is given by

$$
\hat{V}(y - \hat{y}^{\text{pred}}) = s^2(1 + \mathbf{x}'(X'X)^{-1}\mathbf{x}) \, .
$$

When we want to evaluate y at $x = c$ our least squares augmented vector used in the above is given by $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ c \end{bmatrix}$, so that the inner product expression above becomes

$$
\begin{aligned}
\mathbf{x}'(X'X)^{-1}\mathbf{x} &= \begin{bmatrix} 1 & 1 & c \end{bmatrix} \begin{bmatrix} \frac{n}{8m(n-2m)} & \frac{-n+4m}{8m(n-2m)} & 0 \\ \frac{-n+4m}{8m(n-2m)} & \frac{n}{8m(n-2m)} & 0 \\ 0 & & \frac{1}{2mc^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ c \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & c \end{bmatrix} \begin{bmatrix} \frac{1}{2(n-2m)} \\ \frac{1}{2(n-2m)} \\ \frac{1}{2mc} \end{bmatrix} = \frac{1}{n-2m} + \frac{1}{2m} .
\end{aligned}
$$

With this expression we find

$$
\hat{V}(y - \hat{y}^{\text{pred}}) = 1 + \frac{1}{n-2m} + \frac{1}{2m} .
$$

When $n = 7$ this becomes

$$\hat{V}(y - \hat{y}^{\text{pred}}) = 1 + \frac{1}{7 - 2m} + \frac{1}{2m}.$$

To find the value of m that minimizes this variance we take the derivative of the above expression, set the result equal to zero and solve for m. We find $m = \frac{7}{4}$, with a value of $\hat{V}(y - \hat{y}^{\text{pred}}) = \frac{4}{7}$. The algebra for some of this problem can be found in the Mathematica file `prob_2_14.nb`.

Exercise 2.15 (regression for predicting housing prices)

The initial regression model (using all variables) for housing prices is taken to be

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t\,,$$

with the variables x_{ti} given as in the text. For this problem because we do not know a-priori which variables will be most informative we will initially estimate $\hat{\beta}_i$, the standard error $s_{\hat{\beta}_i}$, and the t statistics $t_{\hat{\beta}_i}$, for each variable. The values of the individual t statistics will signify which variable to keep in the regression. This problem is worked in the MATLAB script `prob_2_15.m` where we first load the Narula and Wellington data set using the MATLAB function `load_narula_wellington.m`. We then construct a four variable $p = 4$ regression. The estimated values of beta for this regression are given by

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 2.4393 \\ 2.2972 \\ -0.0546 \\ 13.9319 \\ -0.0412 \end{bmatrix}.$$

From these values, while not a scientific argument, the small magnitude of $\hat{\beta}_2$ and $\hat{\beta}_4$ makes it less likely that they will be significant variables in the regression unless the standard error on these variables is particularly small. Computing the standard errors of these variables we find that

$$\mathbf{s}_\beta = \begin{bmatrix} s_{\hat{\beta}_0} \\ s_{\hat{\beta}_1} \\ s_{\hat{\beta}_2} \\ s_{\hat{\beta}_3} \\ s_{\hat{\beta}_4} \end{bmatrix} = \begin{bmatrix} 4.1980 \\ 0.5490 \\ 0.4862 \\ 2.9961 \\ 0.0681 \end{bmatrix}.$$

Thus we see that the second and fourth standard errors *are* indeed smaller that then others, so the t statistics will fully determine the significance of X_2 and X_4. The t statistic for each of these estimates of β is given by

$$\mathbf{t}_{\hat{\beta}} = \begin{bmatrix} t_{\hat{\beta}_0} \\ t_{\hat{\beta}_1} \\ t_{\hat{\beta}_2} \\ t_{\hat{\beta}_3} \\ t_{\hat{\beta}_4} \end{bmatrix} = \begin{bmatrix} 0.5811 \\ 4.1843 \\ -0.1124 \\ 4.6501 \\ -0.6052 \end{bmatrix}.$$

For a given significance level α say 0.05 these numbers are to be compared with the value of

$$t_{\alpha/2}(n-p-1) = t_{0.025}(23) = 2.06\,,$$

from which only *two* coefficients $\hat{\beta}_1$ and $\hat{\beta}_3$ are seen to be significant. We are unable to simply drop all other variables (at once) because the t tests we are performing consider each variable as if it was the *last* variable added to the regression. We will instead drop the least significant variable (the one with the smallest t test value) and perform the regression again. Dropping the variable X_2 (lot size) we next compute a regression using X_1, X_3, and X_4 only. This give beta estimates, standard errors, and t statistics of

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 2.2329 \\ 2.2951 \\ 13.7855 \\ -0.0387 \end{bmatrix}, \quad \mathbf{s}_{\hat{\beta}} = \begin{bmatrix} 3.6965 \\ 0.5373 \\ 2.6415 \\ 0.0631 \end{bmatrix}, \quad \mathbf{t}_{\hat{\beta}} = \begin{bmatrix} 0.6041 \\ 4.2716 \\ 5.2188 \\ -0.6140 \end{bmatrix}.$$

The t statistics are to be compared with the value of $t_{0.025}(24) = 2.06$, from which only X_1 is significant. Deleting the insignificant variable X_4 (age of the house) we can perform a regression with only *two* variables X_1 and X_3. Performing this gives beta estimates, standard errors, and t statistics of

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0.4753 \\ 2.4235 \\ 13.4073 \end{bmatrix}, \quad \mathbf{s}_{\hat{\beta}} = \begin{bmatrix} 2.3096 \\ 0.4887 \\ 2.5365 \end{bmatrix}, \quad \mathbf{t}_{\hat{\beta}} = \begin{bmatrix} 0.2058 \\ 4.9589 \\ 5.2857 \end{bmatrix}.$$

The t statistics are to be compared with $t_{0.025}(25) = 2.06$, from which we conclude that both the variables X_2 and X_3 are significant. Computing the coefficient of determination R^2 for linear regression we find $R^2 = 0.9241$ a very good fit. To predict the sale price of the house specified we have

$$X_1 = 10.00\,, X_2 = 6.000\,, X_3 = 1.500\,, X_4 = 20\,,$$

of which we only need the values of X_1 and X_3. The augmented vector is then $\begin{bmatrix} 1 \\ X_1 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10.00 \\ 1.5 \end{bmatrix}$. Our predicted sale price is $\hat{y} = \begin{bmatrix} 1 \\ 10.00 \\ 1.5 \end{bmatrix}' \hat{\beta} = 44.82$. The $100(1-\alpha)$ percent prediction interval is given by

$$\hat{y}^{\text{pred}} \pm t_{\alpha/2}(n-p-1)s(1 + x'(X'X)^{-1}x)^{1/2}\,.$$

Using the number given above we find our prediction interval given by

$$(35.85, 53.79)\,.$$

See the MATLAB script `prob_2_15.m` were we perform these calculations.

Exercise 2.16 (various regression topics)

Part (a): False. Independence of the residual is a goal of a regression but it may not be achieved in practice. Reasons for it to not be true could result from an incorrect model functional form. If the residuals are not independent but have some significant correlation, a plot of x_{ti} (or $\hat{y}_t$) as the independent variable and e_t (the residual value) as the dependent variable may highlight the functional form that needs to be added to the model.

Part (b): True. The regression sum of squares $\sum_t (\hat{y}_t - \bar{y})^2$ is equivalent to $\hat{\beta}' X' y - n\bar{y}^2$ (See Equation 2 in these notes), but $\hat{\beta} = (X'X)^{-1} X y$ so the regression sum of squares becomes

$$y' X' (X'X)^{-1} X' y - n\bar{y}^2 \,,$$

the quoted expression.

Part (c): The answer to this is yes, by adding additional variables we will always decrease the MSE, since the numerator in its definition $\text{SSE} = \sum (y_t - \hat{y}_t)^2$ will decrease. The interesting problem to study is whether this decrease in MSE is statistically significant or not. This is described in the section of the book entitled: General Hypothesis Tests: The Extra Sum of Squares Principle, where one considers an extended model (of all variables say $X_1, X_2, \cdots, X_{q-1}, X_q, X_{q+1}, \cdots, X_p$) against the more simplistic model (say $X_1, X_2, \cdots, X_{q-1}, X_q$ with fewer variables) and computes

$$\begin{aligned}
\text{SSR}(X_{q+1}, \cdots, X_p | X_1, \cdots, X_q) &= \text{SSR}(X_1, \cdots, X_q, \cdots, X_p) - \text{SSR}(X_1, \cdots, X_q) \\
&= (\text{SSTO} - \text{SSE}(X_1, \cdots, X_q, \cdots, X_p)) \\
&- (\text{SSTO} - \text{SSE}(X_1, \cdots, X_q)) \\
&= \text{SSE}(X_1, \cdots, X_q) - \text{SSE}(X_1, \cdots, X_q, \cdots, X_p) \,,
\end{aligned}$$

which we can see measures how much error reduction there is when we include the additional variables $X_q, X_{q+1}, \cdots, X_p$. We then compute the statistic

$$F^* = \frac{\text{SSR}(X_{q+1}, \cdots, X_p | X_1, \cdots, X_q)/(p-q)}{\text{SSE}(X_1, \cdots, X_p)/(n-p-1)} \,,$$

which we can see measures the fractional decrease in error achieved when we include the new variables over the entire error obtainable when using all variables. The statistic is distributed as an F distribution with $p-q$ and $n-p-1$ degrees of freedom. If this F^* is "large enough" we can conclude that the addition of the variables $X_{q+1}, \cdots, X_p$ have statistically reduced the mean square error.

Part (d): The t statistic given by $t_{\hat{\beta}_i} = \hat{\beta}_i / s_{\hat{\beta}_i}$ represents the contribution of the predictor x_{ti} as if it were the *last* variable to enter the model. Thus one could find that while each variable and statistics is small (implying they are all insignificant) the entire regression is significant. Recall the argument when two variables X_1 and X_2 are highly correlated and both good predictors of Y. Then taken together in a regression model they each will have a very small t statistic but the overall regression model would still be a good one.

Part (e): I believe this is true since in the case of orthogonal columns the reduction in error i.e. SSE should be the same independent of the other variables we may have included in the regression. That is this error reduction is independent of the other variables

Part (f): False. The definition of R^2 is

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} . \tag{14}$$

Using the definition of SSR and SSTO. The expression in the book for this part however has a numerator of $\sum_t (\hat{y}_t - y_t)^2$.

Part (g): In the simple linear regression model where $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ the least squares estimate of β_1 is given by

$$\hat{\beta}_1 = \frac{n \sum x_t y_t - (\sum x_t)(\sum y_t)}{n \sum x_t^2 - (\sum x_t)^2} . \tag{15}$$

Dividing the top and bottom of this expression by n^2 gives

$$\hat{\beta}_1 = \frac{\frac{1}{n} \sum x_t y_t - (\frac{1}{n} \sum x_t)(\frac{1}{n} \sum y_t)}{\frac{1}{n} \sum x_t^2 - (\frac{1}{n} \sum x_t)^2} = \frac{\hat{r}}{s_x^2} ,$$

where

$$\hat{r} = E((x_t - \bar{x})(y_t - \bar{y})) = E(x_t y_t) - \bar{x}\bar{y} ,$$

is the sample covariance, and $s_x^2 = E((x_t - \bar{x})^2) = E(x_t^2) - \bar{x}^2$ is the sample variance. If we define r to be the sample correlation coefficient so $r = \frac{\hat{r}}{s_x s_y}$ (or $\hat{r} = s_x s_y r$) the above estimate for β_1 becomes

$$\hat{\beta}_1 = \frac{s_x s_y r}{s_x^2} = \frac{s_y r}{s_x} ,$$

as we were to show.

Part (h): The estimate of β_1 given in the simple linear regression model can also be written as in Equation 15 from which there seems to be no benefit in selecting values of x_t who's distances from $\bar{x}$ are large. If we however look at the variance in the error in our estimate of β_1 given by

$$V(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{\sum (x_t - \bar{x})^2} \right) ,$$

we see that our variance in the estimate of β_1 will decrease if we have many samples of x_t that are distributed from the mean $\bar{x}$.

Part (i): If the variance of ε_t in the simple linear regression model is proportional to x_t^2 then

$$V(\varepsilon_t) = C x_t^2 \sigma^2 \approx D(\beta_0 + \beta_1 x_t)^2 \sigma^2 = D' \eta_t^2 \sigma^2 = h(\eta_t)^2 \sigma^2 ,$$

for some constants C, D and D'. The "level" η_t defined as the value of $\beta_0 + \beta_1 x_t$, and the function h is defined as $h(\eta_t)^2 = D\eta_t^2$. This last equation implies that $h(\eta_t) \sim \eta_t$. From the discussion in the section on non-constant variance and variance stabilization techniques this corresponds to residuals whos standard deviation is proportional to the level and the transformation that should be applied to the dependent variable y_t to stabilize the variance is a logarithmic one.

Part (j): The described procedure would result in a model that is difficult to fit since there would be multiple ways to specify numerically the "level" of the dummy variable and still obtain the given outputs y_t. A better procedure is to introduce only *two* indicator functions IND_{t1} and IND_{t2} that would be based relative to a one of the factories (say denoted as factory 0) such that they are zero when the t-th item does not come from factory 1 or 2. Then the regression model we would attempt to fit would be given by (if we are only concerned about a constant output) value say

$$y_t = \beta_0 + \delta_1 \text{IND}_{t1} + \delta_2 \text{IND}_{t2} + \varepsilon_t \,.$$

Part (k): The model as given cannot be fit by least squares. A simple transformation of this model by taking the logarithm gives an alternative form for the same model of

$$\log(y) = \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \varepsilon \,,$$

which is linear in its unknown coefficients, β_i, and *can* be fit by ordinary least squares.

Part (l): Notice that the transformation of y_t given by

$$\log\left(\frac{1}{y_t} - 1\right) = \beta_0 + \beta_1 x_t + \log(\varepsilon_t) \,,$$

could be fit by ordinary least squares.

Part (m): The lack-of-fit tests can be applied to a regression analysis regardless of the dimensionality of the predictors X. The only requirement for the lack-of-fit tests is that for a *fixed* value of the predictor vector, X, we have some number of say $n > 1$ of measured *independent* responses y.

Part (n): In general, performing different variable selection criterion will lead to *different* sets of independent variables being selected. The effect is magnified the stronger the correlation present among the independent variables. For example, running forward selection and backwards selection to determine the optimal three variable regression will often result in different choices for the variables to include. This can be problematic if the purpose of the regression analysis is to hypothesis sources of causation (i.e. hypothesis of the sort: the value of Y is because of the value of X_i) but is usually not a problem if the purpose of the regression is for prediction.

Part (o): False. One should always remember correlation does *not* imply causation.

Part (p): If the F statistic is very large what one can be sure of is that the regression (as given) is statistically significant. It does not tell if the regression can be improved for example if it is found that there is correlation among the residuals. Another benefit obtained when we look at the residuals is that when plotting them against x_{ti} they can suggest possible terms to include in the model for example power of the dependent variable like x_{ti}^2. The residuals also give a clue if the assumption of constant variance is true. If not, the regression could be improved by variance-stabilization techniques or weighted least squares.

Part (q): A property of the least squares fit is that the fit of $\mathbf{y}$ produced (i.e. $\hat{\mathbf{y}}$) is such that $X'e = X'(y - \hat{y}) = 0$, and thus imposes certain linear restrictions among the n residuals. Now assuming that the design matrix includes a column of ones and the regression model therefore includes a constant term one of the these linear restrictions is $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} e = 0$, thus the residuals sum to zero. The regression given does not include a constant term and the students claim is false.

Exercise 2.17 (an example of a regression using indicator variables)

The given plots look like linear models so each will have the form $y_t = \beta_0 + \beta_1 x_{t1} + \varepsilon_t$, but modified to have an level shift via indicator function positioned at $x = 0$. To introduce this, define the function the "indicator" function IND_t as

$$\text{IND}_t = \begin{cases} 1 & x_t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

Then our regression model will be given by

$$y_t = \beta_0 + \beta_1 x_{t1} + \delta \, \text{IND}_t \,,$$

for some yet to be determined coefficient δ. The design matrix for this system from the table of x_{t1} and y_t is given by

$$X = \begin{bmatrix} 1 & -5 & 0 \\ 1 & -3 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 5 & 1 \end{bmatrix} .$$

With this expression for X, the ordinary least squares estimate of the coefficients $\hat{\beta}_i$ are given by

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1} X' y = \begin{bmatrix} 10.7083 \\ 1.1250 \\ 2.2500 \end{bmatrix} .$$

The standard errors on these coefficients are given by

$$\mathbf{s}_{\hat{\beta}} = \begin{bmatrix} 0.49257 \\ 0.13010 \\ 0.88878 \end{bmatrix} .$$

Using these two expressions together we compute the t statistic of these coefficients $t_{\hat{\beta}}$ of

$$\mathbf{t}_{\hat{\beta}} = \begin{bmatrix} 21.7399 \\ 8.6469 \\ 2.5316 \end{bmatrix} .$$

To determine if a given coefficient in the above regression is significant, these values are to be compared against the values of $t_{\alpha/2}(n-p-1)$, which for this problem (note that $p=1$ in this case because the indicator functions don't change the number of coefficients fitted at each level) and $\alpha = 0.05$ has a value of $t_{\alpha/2}(n-p-1) = 2.77$. All of the coefficients, but $\hat{\delta}$ in the above are larger than this value so for them we can reject the hypothesis H_0 that $\beta_i = 0$ at significance α. The third coefficient for $\hat{\delta}$ is slightly too small for the given threshold α. This indicates that perhaps there is *not* a level change as we cross the value of $x = 0$.

If we assume there is *not* a level change across zero (since at the given significance level, α, the resulting coefficient was not significant) then we would need to refit our regression model before computing the ANOVA table. When we do that we find for estimates of β, their standard errors, and their t statistics

$$\hat{\beta} = \begin{bmatrix} 11.83 \\ 1.41 \end{bmatrix}, \quad \mathbf{s}_{\hat{\beta}} = \begin{bmatrix} 0.3427 \\ 0.1003 \end{bmatrix}, \quad \mathbf{t}_{\hat{\beta}} = \begin{bmatrix} 34.5272 \\ 14.0950 \end{bmatrix}.$$

These t statistics should be compared to the value of $t_{\alpha/2}(4) = 2.77$, from which we see that each component is significant. For this final regression we can construct the ANOVA table

Source	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Regression	SSR = 141.01	$p = 1$	MSR = 140.01	198.66
Error	SSE = 2.82	$n-p-1=4$	MSE = 0.704	
Total (correlated for mean)	SSTO = 142.83	$n-1=5$		

The significance of the entire regression considering all regressands can be inferred from the F statistic calculated in the ANOVA table. We see from the above that its value is $F = 198.66$. This is to be compared against the threshold value of $F_{\alpha}(p, n-p-1) = F_{0.05}(6,4) = 7.7086$. The above F statistic is significantly larger than the this threshold value we can reject the hypothesis H_0 that all $\beta_i = 0$ at the significance level $\alpha = 0.05$.

The numerical calculations for this problem computing the above values can be found in the MATLAB script `prob_2_17.m`.

Exercise 2.18 (finding autocorrelated errors)

If the investigator finds that the error terms are autocorrelated i.e. $\varepsilon_t = \phi\varepsilon_{t-1} + a_t$, then as discussed in the book the initial regression model can be improved upon by instead regressing an autoregressive model of the form

$$y_t = \phi y_{t-1} + \beta_0(1-\phi) + \sum_{i=1}^{p} \beta_i(x_{ti} - \phi x_{t-1,i}) + a_i \,.$$

This will result in an improved model but requires a nonlinear least squares procedure. This in tern will result in *different* values for the coefficients β_i and we see that the investigators claims are in fact false.

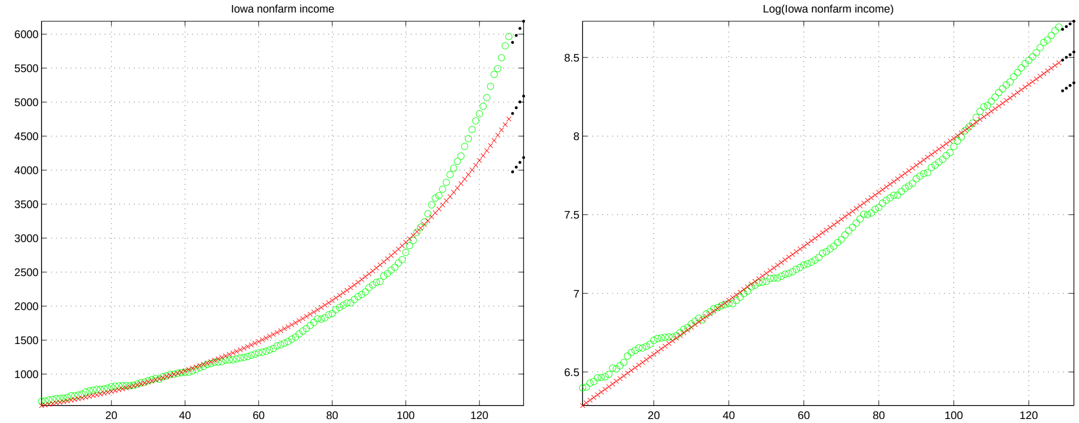


Figure 1: The quarterly Iowa nonfarm income example. **Left:** A plot of the raw data y_t (in green) vs. t for and the regression $e^{\ln(\hat{y}_t)}$ (in red). **Right:** A plot of the raw data for $\ln(y_t)$ (in green) vs. t for the Iowa nonfarm income and the regression $\ln(\hat{y}_t)$. The four-step look-ahead predictions (and their confidence bounds) are plotted as black dots on the right side of each plot. This regression is discussed in more detail below in the text that accompanies this problem.

Exercise 2.19 (quarterly Iowa nonfarm income)

See the Figure 1 (left) for the plot of y_t and Figure 1 (right) for a plot of $\ln(y_t)$ (both plots are in red). We see that the plot of $\ln(y_t)$ is very linear in appearance representing a constant growth rate. Because of this we fit a linear model to $\ln(y_t)$ using the MATLAB function developed for this chapter `wwx_regression.m`. When given a set of regressors X this function will compute the ordinary least squares estimates of a linear model using these variables. The user must preprend a column of ones if a constant function is desired in the regression. The resulting function call returns the least square estimates of β_i, their standard errors, their t statistics, and various ANOVA statistics. It is the responsibility of the user to then study the outputs of this function to decide with coefficient are actually statistically significant. This problem is implemented in the MATLAB script `prob_2_19.m`. When we run this script we first load the Iowa nonfarm data set using the MATLAB function `load_series_1.m`, constructs the augmented feature matrix, and compute estimates of the coefficients β_i for the model

$$\ln(y_t) = \beta_0 + \beta_1 t + \epsilon_t \,.$$

The routine `wwx_regression.m` gives t statistics for the estimates of β_i given by

$$\left[\begin{array}{c} t_{\hat{\beta}_0} \\ t_{\hat{\beta}_1} \end{array} \right] = \left[\begin{array}{c} 362.20 \\ 73.71 \end{array} \right] ,$$

which are to be compared with the threshold value of $t_{\alpha/2}(n-p-1) = t_{0.025}(126) = 1.97$. Since both values are significantly larger than $t_{0.025}(126)$ there is strong evidence to reject the hypothesis H_0 (that the coefficients are actually zero). To asses the completeness of the fit and the accuracy of the specified model, we next consider the *residuals* of the regression above. We plot these residuals as a function of time in Figure 2 (left). In that figure we see significant correlation. This indicates that the model above may *not* be adequate. To

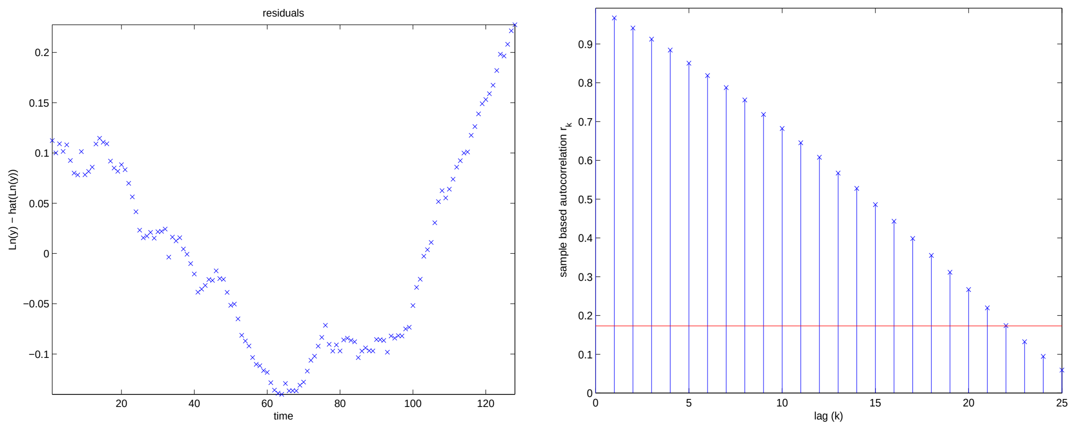


Figure 2: **Left:** A time series plot of the residuals found when predicting $\ln(y_t)$ in the Iowa nonfarm income. **Right:** A plot of the sample autocorrelation values, r_k, for the Iowa nonfarm income when fitting a linear regression to $\ln(y_t)$. The two sigma confidence bound on the hypothesis that $r_k \approx 0$ is drawn in red. Notice that there is significant error autocorrelation present in this example.

quantify this, we compute the sample autocorrelations, r_k, of the residuals $e_t = y_t - \hat{y}_t$. As discussed in the text, to determine whether these sample autocorrelations are significant we need to compare their magnitude to the value of $1.96n^{-1/2} = 0.1732$. If $|r_k| > 1.96n^{-1/2}$ we cannot reject the hypothesis H_0 (that they are insignificant) and to improve our regression must model their affect. A plot of the sample autocorrelations and the constant value of $1.96n^{-1/2}$ are plotted in Figure 2 (right). From that figure we see that many of the sample autocorrelations are in fact *larger* than this value and can conclude that the model, as given, in not sufficient and could be improved upon.

Even with this difficulty in our model specification, we can compute forecasts and prediction errors for the next four quarters using the given model. For a given input vector $\mathbf{x} = \begin{bmatrix} 1 \\ t \end{bmatrix}$, our predictions for $\ln(y_t)$ will be given by $\ln(\hat{y}_t^{\text{pred}}) = x'\hat{\beta}$ and confidence intervals on $\ln(y_t)$ given by

$$\ln(\hat{y}_t^{\text{pred}}) \pm t_{\alpha/2}(n-p-1)s(1 + x'(X'X)^{-1}x)^{1/2}\,.$$

Here s is an estimate of the noise variance and X is the design matrix. To derive estimates and confidence intervals for y_t directly we would need to exponentiate the above expressions. When these are implemented an executed they produce "log" confidence intervals (confidence intervals on $\log(y)$) and "direct" confidence intervals (confidence intervals on y) drawn in black dots in Figure 1. In both cases it does not appear that the confidence bounds are that good. A better method to forecast this data will be given in Chapter 3 where it will be forecast with simple exponential smoothing.

Exercise 2.20 (bias introduced by correlated errors)

For the linear regression model $y_t = \beta x_t + \varepsilon_t$ the standard least squares estimate of β is given by

$$\hat{\beta} = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2} . \tag{17}$$

Since we are assuming that x_t are *not* random and that we know the exact model followed by y_t the expectation of this expression is given by

$$E(\hat{\beta}) = \frac{1}{\sum_{t=1}^n x_t^2} \sum_{t=1}^n x_t E(y_t) = \frac{\sum_{t=1}^n \beta x_t^2}{\sum_{t=1}^n x_t^2} = \beta .$$

Since $E(y_t) = \beta x_t$, i.e. the expectation of ε_t is still zero. From this we see that $\hat{\beta}$ is an unbiased estimate of β. To compute the variance of our estimate $\hat{\beta}$ we will use the identity that $\text{Var}(\hat{\beta}) = E(\hat{\beta}^2) - E(\hat{\beta})^2$ and thus we need to compute $E(\hat{\beta}^2)$. We find

$$\begin{aligned} E(\hat{\beta}^2) &= \frac{1}{\left(\sum_{t=1}^n x_t^2\right)^2} E\left(\sum_{u=1}^n \sum_{v=1}^n x_u y_u x_v y_v \right) \\ &= \frac{1}{\left(\sum_{t=1}^n x_t^2\right)^2} \sum_{u=1}^n \sum_{v=1}^n x_u x_v E\left(y_u y_v\right) . \end{aligned}$$

Now the expectation of $E(y_u y_v)$ involves the correlation structure of the error terms ε and is computed by

$$\begin{aligned} E(y_u y_v) &= E((\beta x_u + \varepsilon_u)(\beta x_v + \varepsilon_v)) \\ &= \beta^2 x_u x_v + E(\varepsilon_u \varepsilon_v) \\ &= \beta^2 x_u x_v + \begin{cases} \sigma^2 & u = v \\ \sigma^2 \rho & |u - v| = 1 \\ 0 & \text{otherwise} \end{cases} . \end{aligned}$$

Using this expression the above sum in $E(\hat{\beta}^2)$ becomes

$$\begin{aligned} E(\hat{\beta}^2) &= \frac{\beta^2}{\left(\sum_{t=1}^n x_t^2\right)^2} \sum_{u=1}^n \sum_{v=1}^n x_u^2 x_v^2 + \frac{1}{\left(\sum_{t=1}^n x_t^2\right)^2} \sum_{u=1}^n x_u^2 \sigma^2 + \frac{1}{\left(\sum_{t=1}^n x_t^2\right)^2} \sum_{u=1}^{n-1} 2 x_u x_{u+1} \sigma^2 \rho \\ &= \beta^2 + \frac{\sigma^2}{\left(\sum_{t=1}^n x_t^2\right)^2} \left(\sum_{u=1}^n x_u^2 + 2\rho \sum_{u=1}^{n-1} x_u x_{u+1} \right) . \end{aligned}$$

Then using this expression we find

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{t=1}^n x_t^2} \left(1 + 2\rho \left(\frac{\sum x_u x_{u+1}}{\sum x_u^2} \right) \right) ,$$

which decomposes the result for $\text{Var}(\hat{\beta})$ into two terms. The first represented by $\frac{\sigma^2}{\sum_{t=1}^n x_t^2}$ is the term obtained when $\rho = 0$ and represents the standard least squares variance while the second term represents the change in variance due to the correlation in the residuals.

Part (b): The t statistic for the hypothesis $\beta = 0$ is determined by computing $\frac{\hat{\beta}}{\sqrt{\text{Var}(\hat{\beta})}}$. Since the values of x_t are fixed when the regression is performed the sign (positive or negative) of the expression $\sum x_u x_{u+1}$ is determined. We can assume without loss of generality that it is positive. In that case if $\rho > 0$ the t statistic will be smaller than it should be and we would be more likely to conclude that the coefficient β were zero when in fact it was not. In the other case when $\rho < 0$ the expression for the variance would be smaller than expected and the value of the t statistic would lead us to think that the estimated regression coefficient for β was significant when in fact it may not be.

Exercise 2.21 (residual variances that depend on $1/x_t^2$)

Part (a): As in exercise 2.20 the usual least-squares estimate of β for the model $y_t = \beta x_t + \varepsilon_t$ is given by Equation 17. As is shown in the previous problem this estimate is an unbiased estimator of β. Its variance can be calculated by recognizing that to have the given

$$V(y_t) = V(\varepsilon_t) = \frac{\sigma^2}{x_t^2}, \tag{18}$$

match the general form of

$$V(\varepsilon_t) = \frac{\sigma^2}{\omega_t}, \tag{19}$$

we need to have $\omega_t = x_t^2$. Thus the precision matrix Ω, in this case is given by

$$\Omega^{-1} = \begin{bmatrix} \omega_1^{-1} & & & \\ & \omega_2^{-1} & & \\ & & \ddots & \\ & & & \omega_n^{-1} \end{bmatrix} = \begin{bmatrix} 1/x_1^2 & & & \\ & 1/x_2^2 & & \\ & & \ddots & \\ & & & 1/x_n^2 \end{bmatrix}.$$

Now the variance of the ordinary least squares estimate of β when the individual measurements y_t have different variances given by Equation 19 is given by

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1} X' \Omega^{-1} X (X'X)^{-1}. \tag{20}$$

For the type of regression we are performing here (simple linear regression through the origin) the design matrix is

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

so that $X'X = \sum x_t^2$ and $X'\Omega^{-1}X = n$, thus the variance of our estimate of β when using ordinary least squares is given by

$$V(\hat{\beta}) = \frac{n\sigma^2}{(\sum_t x_t^2)^2}.$$

Part (b): The weighted least squares estimate of β is given by

$$\hat{\beta}_* = (X'\Omega X)^{-1} X'\Omega y\,. \tag{21}$$

For the given definition of Ω and ω_t we get

$$\begin{aligned} X'\Omega X &= \sum \omega_t x_t^2 = \sum x_t^4 \\ X'\Omega y &= \sum \omega_t x_t y_t = \sum x_t^3 y_t\,, \end{aligned}$$

so that using Equation 21 we get

$$\hat{\beta}_* = \frac{\sum \omega_t x_t y_t}{\sum \omega_t x_t^2} = \frac{\sum x_t^3 y_t}{\sum x_t^4}\,.$$

This is an unbiased estimate since using $E(y_t) = \beta x_t$, we can compute $E(\hat{\beta}_*) = \beta$. The variance of the weighted least squares estimate is given by

$$V(\hat{\beta}_*) = \sigma^2 (X'\Omega X)^{-1}\,, \tag{22}$$

which in this case becomes

$$V(\hat{\beta}_*) = \frac{\sigma^2}{\sum x_t^4}\,.$$

Part (d): If we measure that $V(y_t) = \eta_t^2$ then we can use variance stabilization techniques to transform the dependent variable y_t to variable that has constant variance as discussed in the book. Specifically, we would pick a transformation $g(\cdot)$ to apply such that

$$g'(\eta_t) = \frac{1}{\eta_t} \quad \text{or} \quad g(\eta_t) = \ln(\eta_t)\,,$$

so we would apply the logarithm to the dependent data y_t and attempt to fit a model instead to the terms $\ln(y_t)$. Because of this transformation the variance of $\ln(y_t)$ should now be constant.

Exercise 2.22 (deriving estimates from averages)

If we are given the model $y_{ij} = \beta_0 + \beta_1 x_j + \varepsilon_{ij}$ for $i = 1, 2, \cdots, n_j$ and for each of the c "classes" $j = 1, 2, \cdots, c$, then the standard least squares regression formulation would seek to find values for β_i such that (in the least squares sense) the following equality is satisfied

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \\ \vdots \\ y_{1c} \\ y_{2c} \\ \vdots \\ y_{n_c c} \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_1 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_1 \\ 1 & x_2 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_c \\ 1 & x_c \\ \vdots & \vdots \\ 1 & x_c \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} .$$

Here the matrix in front of the column vector $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is denoted as X and called the *design matrix*. To solve this least squares problem for β_i we need to compute the product $X'X$ which for the above design matrix becomes

$$X'X = \begin{bmatrix} n & \sum_{j=1}^{c} n_j x_j \\ \sum_{j=1}^{c} n_j x_j & \sum_{j=1}^{c} n_j x_j^2 \end{bmatrix} ,$$

and the product $X'y$ or

$$X'y = \begin{bmatrix} \sum_{j=1}^{c} \sum_{i=1}^{n_j} y_{ij} \\ \sum_{j=1}^{c} x_j \sum_{i=1}^{n_j} y_{ij} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{c} n_j \bar{y}_j \\ \sum_{j=1}^{c} x_j n_j \bar{y}_j \end{bmatrix} .$$

From these two expressions we can invert the matrix $X'X$ and apply that operator to $X'y$ to compute the solution to the normal equation $\hat{\beta} = (X'X)^{-1}(X'y)$. Note that these estimates of β_i are computed using only the information that we have.

Chapter 3: Regression and Exponential Smoothing Methods To Forecast Nonseasonal Time Series

Notes On The Text

Notes on the Constant Mean Model

For the constant mean model we assume that $z_n(l) = \beta$. To estimate β we can use ordinary least squares with a design matrix given by

$$X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} .$$

Then we see that $X'X = n$ and the ordinary least squares estimate of β is given by $\hat{\beta} = \frac{1}{n} \sum z_t = \bar{z}$. From the expression for the variance of the forecast error given in Chapter 2 of

$$\hat{V}(y_k - \hat{y}_k^{\text{pred}}) = \sigma^2 [1 + x_k'(X'X)^{-1} x_k] , \tag{23}$$

we find since in the constant mean model the augmented state vector is simply $x_k = 1$ (only the constant 1 i.e. there are no regressands) that the variance in our l step-ahead prediction is given by

$$V(z_{n+l} - \hat{z}_n(l)) = \sigma^2 \left(1 + \frac{1(1)}{n} \right) ,$$

which is equation 3.4 in the book. We will compute an estimate of σ above as

$$\sigma^2 \approx \frac{1}{(n-p-1)} \text{SSE} = \frac{1}{(n-p-1)} \sum_t (z_t - \hat{z}_t)^2 = \frac{1}{n-1} \sum_t (z_t - \bar{z})^2 .$$

Examples of discounted least squares: the locally constant model

We can use the general framework provided in the section entitled "Regression models with time as independent variable" to derive the corresponding formulas for the *constant mean model* introduced earlier. Equation 3.13 from the book is

$$f_{n+j} = f'(j)\beta + \varepsilon_{n+j} , \tag{24}$$

which for the constant mean model we take only one fitting function (the constant) given by

$$f_1(j) = 1 .$$

Then our general l step-ahead prediction is given by $\hat{z}_n(l) = f'(l)\hat{\beta}_n = \hat{\beta}_n$, with $\hat{\beta}_n$ given by $\hat{\beta}_n = F_n^{-1} h_n$ where

$$
\begin{aligned}
h_n &= \sum_{j=0}^{n-1} \omega^j f(-j) z_{n-j} = \sum_{j=0}^{n-1} \omega^j z_{n-j} \\
F_n &= \sum_{j=0}^{n-1} \omega^j f(-j) f'(-j) = \sum_{j=0}^{n-1} \omega^j = \frac{1-\omega^n}{1-\omega} .
\end{aligned}
$$

So that using these $\hat{\beta}_n$ is given by

$$
\hat{\beta}_n = \left(\frac{1-\omega}{1-\omega^n} \right) \sum_{j=0}^{n-1} \omega^j z_{n-j} , \tag{25}
$$

which is the equation 3.30 in the book. Taking the limit as $\omega \to 1$ we find

$$
\lim_{\omega \to 1} \left(\frac{(-1)}{-n\omega^{n-1}} \right) \sum_{j=0}^{n-1} z_{n-j} = \frac{1}{n} \sum_{j=0}^{n-1} z_{n-j} ,
$$

the expression for the average of the z_n values.

Examples of discounted least squares: the locally constant linear trend model

In general the discounted least squares procedure gives the l-step ahead estimates of $\hat{z}_n(l) = f'(l)\hat{\beta}_n$ where $\hat{\beta}_n = F_n^{-1} h_n$. We can specify this general framework to the locally constant linear trend model by taking $f(j) = \begin{bmatrix} 1 \\ j \end{bmatrix}$ so that

$$
\begin{aligned}
F_n &= \sum_{j=0}^{n-1} \omega^j f(-j) f'(-j) = \sum_{j=0}^{n-1} \omega^j \begin{bmatrix} 1 \\ -j \end{bmatrix} \begin{bmatrix} 1 & -j \end{bmatrix} = \sum_{j=0}^{n-1} \omega^j \begin{bmatrix} 1 & -j \\ -j & j^2 \end{bmatrix} . & (26) \\
h_n &= \sum_{j=0}^{n-1} \omega^j f(-j) z_{n-j} = \sum_{j=0}^{n-1} \omega^j \begin{bmatrix} 1 \\ -j \end{bmatrix} z_{n-j} . & (27)
\end{aligned}
$$

By using the above discounted least squares formulation to derive the locally constant linear trend model we have fitting functions $\mathbf{f}(t)$ with components $f_1(j) = 1$, $f_2(j) = j$, and a transition matrix $f(j+1) = Lf(j)$ given by $L = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. The large n estimates of $\hat{\beta}_n = (\hat{\beta}_{0,n}, \hat{\beta}_{1,n})'$ are given by solving for $\hat{\beta}_n = F_\infty^{-1} h_\infty$, which from Equation 26 and 27 we can compute

$$
\begin{aligned}
\hat{\beta}_{0,n} &= (1-\omega^2) \sum \omega^j z_{n-j} - (1-\omega)^2 \sum j\omega^j z_{n-j} \\
\hat{\beta}_{1,n} &= (1-\omega)^2 \sum \omega^j z_{n-j} - \frac{(1-\omega)^3}{\omega} \sum j\omega^j z_{n-j} . & (28)
\end{aligned}
$$

The lower limit of these sums is $j = 0$ while the upper limit is $j = n - 1$. These are the books equations 3.33 and can be used to compute the regression coefficients in this case.

The locally linear trend model viewed as double exponential smoothing

As the locally constant mean model can be interpreted as simple exponential smoothing the locally linear model can be interpreted as "double" exponential smoothing an extension of simple exponential smoothing. Using the output of a first order smooth $S_n^{[1]}$ we will compute a *second order* smooth $S_n^{[2]}$ using the equations

$$
\begin{aligned}
S_n^{[1]} &= (1-\omega)z_n + \omega S_{n-1}^{[1]} = (1-\omega)\sum_{j=0}^{n-1}\omega^j z_{n-j} && (29)\\
S_n^{[2]} &= (1-\omega)S_n^{[1]} + \omega S_{n-1}^{[2]} . && (30)
\end{aligned}
$$

Lets express $S_n^{[2]}$ directly in terms of the elements of our time series z_n. We will do this by writing $S_n^{[2]}$ in terms of $S_n^{[1]}$ and $S_0^{[2]}$, and then using the known summation expression for $S_n^{[1]}$ in terms of z_n expressed by Equation 29. We begin by writing the above expression for $S_n^{[2]}$ in Equation 30 for $n = 1$, we find

$$
S_1^{[2]} = (1-\omega)S_1^{[1]} + \omega S_0^{[2]} .
$$

Doing the same thing for $n = 2$ we obtain

$$
\begin{aligned}
S_2^{[2]} &= (1-\omega)S_2^{[1]} + \omega S_1^{[2]} \\
&= (1-\omega)S_2^{[1]} + \omega(1-\omega)S_1^{[1]} + \omega^2 S_0^{[2]} .
\end{aligned}
$$

Again for $n = 3$ we obtain

$$
\begin{aligned}
S_3^{[2]} &= (1-\omega)S_3^{[1]} + \omega S_2^{[2]} \\
&= (1-\omega)S_3^{[1]} + \omega(1-\omega)(S_2^{[1]} + \omega S_1^{[1]}) + \omega^3 S_0^{[2]} \\
&= (1-\omega)(S_3^{[1]} + \omega S_2^{[1]} + \omega^2 S_1^{[1]}) + \omega^3 S_0^{[2]} .
\end{aligned}
$$

This pattern continues and it looks like the general expression for $S_n^{[2]}$ in terms of $S_n^{[1]}$ is

$$
S_n^{[2]} = (1-\omega)\sum_{k=1}^{n}\omega^{n-k}S_k^{[1]} + \omega^n S_0^{[2]} . \tag{31}
$$

Since we have an expression for $S_k^{[1]}$ in terms of the original series z_n given by Equation 29 we can put this into Equation 31 to find

$$
\begin{aligned}
S_n^{[2]} &= (1-\omega)^2\sum_{k=1}^{n}\omega^{n-k}\sum_{j=0}^{k-1}\omega^j z_{k-j} + \omega^n S_0^{[2]} \\
&= (1-\omega)^2\sum_{k=1}^{n}\sum_{j=0}^{k-1}\omega^{n-k+j} z_{k-j} + \omega^n S_0^{[2]} .
\end{aligned}
$$

Now to further simplify this double summation lets write it out and look for a pattern. We find

$$\begin{aligned}
S_n^{[2]} &= (1-\omega)\left(S_n^{[1]} + \omega S_{n-1}^{[1]} + \omega^2 S_{n-2}^{[1]} + \cdots + \omega^{n-1} S_1^{[1]}\right) + \omega^n S_0^{[2]} \\
&= (1-\omega)^2 \left(\sum_{j=0}^{n-1} \omega^j z_{n-j} + \omega \sum_{j=0}^{n-2} \omega^j z_{n-1-j} + \omega^2 \sum_{j=0}^{n-3} \omega^j z_{n-2-j} + \cdots + \omega^{n-1} z_1\right) + \omega^n S_0^{[2]} \\
&= (1-\omega)^2 \left(z_n + \omega z_{n-1} + \omega^2 z_{n-2} + \cdots + \omega^{n-1} z_1\right. \\
&+ \qquad \omega z_{n-1} + \omega^2 z_{n-2} + \cdots + \omega^{n-1} z_1 \\
&+ \qquad \omega^2 z_{n-2} + \cdots + \omega^{n-1} z_1 \\
&\vdots \\
&+ \qquad \left.\omega^{n-1} z_1\right) + \omega^n S_0^{[2]} .
\end{aligned}$$

Summing "down the columns" of the above expression the pattern is now clear and we see that we have

$$\begin{aligned}
S_n^{[2]} &= (1-\omega)^2 \left(z_n + 2\omega z_{n-1} + 3\omega^2 z_{n-2} + 4\omega^3 z_{n-3} + \cdots + n\omega^{n-1} z_1\right) + \omega^n S_n^{[2]} \\
&= (1-\omega)^2 \sum_{j=0}^{n-1} (j+1)\omega^j z_{n-j} + \omega^n S_n^{[2]} ,
\end{aligned} \tag{32}$$

which when we take the limit $n \to \infty$ is the expression given in equation 3.35 in the book.

From the expressions 29 and 32 for $S_n^{[1]}$ and $S_n^{[2]}$ in terms of the individual series z_{n-j} we can write the discounted sums of z_n as $\sum_j \omega^j z_{n-j} = \frac{S_n^{[1]}}{1-\omega}$ and

$$\sum_j j\omega^j z_{n-j} = \frac{S_n^{[2]}}{(1-\omega)^2} - \sum_j \omega^j z_{n-j} = \frac{S_n^{[2]}}{(1-\omega)^2} - \frac{S_n^{[1]}}{1-\omega} .$$

Thus the expression for $\hat{\beta}_{0,n}$ and $\hat{\beta}_{1,n}$ derived in Equation 28 become in terms of $S_n^{[1]}$ and $S_n^{[2]}$

$$\hat{\beta}_{0,n} = (1+\omega)S_n^{[1]} - S_n^{[2]} + (1-\omega)S_n^{[1]} = 2S_n^{[1]} - S_n^{[2]} \tag{33}$$

$$\begin{aligned}
\hat{\beta}_{1,n} &= (1-\omega)S_n^{[1]} - \frac{1-\omega}{\omega}(S_n^{[2]} - (1-\omega)S_n^{[1]}) \\
&= \frac{1-\omega}{\omega}\left(\omega S_n^{[1]} - S_n^{[2]} + (1-\omega)S_n^{[1]}\right) \\
&= \frac{1-\omega}{\omega}\left(S_n^{[1]} - S_n^{[2]}\right) ,
\end{aligned} \tag{34}$$

which are the books equations 3.38.

The updating equations used to update the value of β_n given the next measurement z_{n+1} are given by the books equation 3.29 which in this case becomes

$$\begin{aligned}
\hat{\beta}_{\mathbf{n+1}} &= L'\hat{\beta}_{\mathbf{n}} + F^{-1} f(0)(z_{n+1} - \hat{z}_n(1)) \\
&= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{0,n} \\ \hat{\beta}_{1,n} \end{bmatrix} + \begin{bmatrix} 1-\omega^2 & (1-\omega)^2 \\ (1-\omega)^2 & \frac{(1-\omega)^3}{\omega} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(z_{n+1} - \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{0,n} \\ \hat{\beta}_{1,n} \end{bmatrix}\right) \\
&= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{0,n} \\ \hat{\beta}_{1,n} \end{bmatrix} + \begin{bmatrix} 1-\omega^2 \\ (1-\omega)^2 \end{bmatrix} (z_{n+1} - \hat{\beta}_{0,n} - \hat{\beta}_{1,n}) .
\end{aligned}$$

In component form this is

$$\begin{aligned}
\hat{\beta}_{0,n+1} &= \hat{\beta}_{0,n} + \hat{\beta}_{1,n} + (1-\omega^2)(z_{n+1} - \hat{\beta}_{0,n} - \hat{\beta}_{1,n}) \\
&= (1-\omega^2) z_{n+1} + \omega^2(\hat{\beta}_{0,n} + \hat{\beta}_{1,n}) \quad (35) \\
\hat{\beta}_{1,n+1} &= \hat{\beta}_{1,n} + (1-\omega)^2 (z_{n+1} - \hat{\beta}_{0,n} - \hat{\beta}_{1,n}) \,. \quad (36)
\end{aligned}$$

To simplify the second equation above, solve Equation 35 for z_{n+1} to get $z_{n+1} = \frac{\hat{\beta}_{0,n+1} - \omega^2(\hat{\beta}_{0,n} + \hat{\beta}_{1,n})}{(1-\omega^2)}$, and put this into Equation 36. This gives

$$\begin{aligned}
\hat{\beta}_{1,n+1} &= \hat{\beta}_{1,n} + \frac{(1-\omega)^2}{1-\omega^2}\left(\hat{\beta}_{0,n+1} - \omega^2(\hat{\beta}_{0,n} + \hat{\beta}_{1,n}) - (1-\omega^2)(\hat{\beta}_{0,n} + \hat{\beta}_{1,n})\right) \\
&= \hat{\beta}_{1,n} + \left(\frac{1-\omega}{1+\omega}\right)\left(\hat{\beta}_{0,n+1} - \hat{\beta}_{0,n} - \hat{\beta}_{1,n}\right) \,,
\end{aligned}$$

since $\frac{(1-\omega)^2}{1-\omega^2} = \frac{1-\omega}{1+\omega}$. Thus we find $\hat{\beta}_{1,n+1}$ given by

$$\hat{\beta}_{1,n+1} = \left(\frac{1-\omega}{1+\omega}\right)(\hat{\beta}_{0,n+1} - \hat{\beta}_{0,n}) + \frac{2\omega}{1+\omega}\hat{\beta}_{1,n} \,.$$

which is equation 3.40 in the book.

Holt's Interpretation of Double Exponential Smoothing

Holt's procedure is given by selecting *two* discount factors ω_1 and ω_2 are specified such that each satisfies $0.7 \le \omega_i \le 0.95$, we have update factors for the mean level $\hat{\mu}_n$ and the slope $\hat{\beta}$ once we have a new measurement z_{n+1} given by

$$\begin{aligned}
\hat{\mu}_{n+1} &= (1-\omega_1) z_{n+1} + \omega_1(\hat{\mu}_n + \hat{\beta}_n) \\
\hat{\beta}_{n+1} &= (1-\omega_2)(\hat{\mu}_{n+1} - \hat{\mu}_n) + \omega_2 \hat{\beta}_n \\
&= (1-\omega_1)(1-\omega_2) z_{n+1} - (1-\omega_1)(1-\omega_2)\hat{\mu}_n + (\omega_1 + \omega_2 - \omega_1\omega_2)\hat{\beta}_n \,.
\end{aligned}$$

The Actual Implementation of Double Exponential Smoothing

To actually implement double exponential smoothing we will recursively iterate Equations 29 and 30 to evaluate $S_n^{[1]}$ and $S_n^{[2]}$ but to do so we will need starting values of $S_n^{[1]}$ and $S_n^{[2]}$ when $n = 0$. These values can be determined by evaluating the expressions for $\hat{\beta}_{0,n}$ and $\hat{\beta}_{1,n}$ given by Equations 33 and 34 at $n = 0$ and associating the values of $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$ with a the least squares coefficients of a linear fit of some subset of the given data. Evaluating Equations 33 and 34 at $n = 0$ gives

$$\begin{aligned}
\hat{\beta}_{0,0} &= 2S_0^{[1]} - S_0^{[2]} \quad (37) \\
\hat{\beta}_{1,0} &= \left(\frac{1-\omega}{\omega}\right)(S_0^{[1]} - S_0^{[2]}) \quad \Leftrightarrow \quad \frac{2\omega}{1-\omega}\hat{\beta}_{1,0} = 2S_0^{[1]} - 2S_0^{[2]} \,. \quad (38)
\end{aligned}$$

Subtract the first equation from the second equation to get

$$\frac{2\omega}{1-\omega}\hat{\beta}_{1,0} - \hat{\beta}_{0,0} = S_0^{[2]} - 2S_0^{[2]} \,,$$

or solving for $S_0^{[2]}$ we find

$$S_0^{[2]} = \hat{\beta}_{0,0} - \frac{2\omega}{1-\omega}\hat{\beta}_{1,0} \,. \tag{39}$$

Putting this expression into Equation 37 to get

$$S_0^{[1]} = \frac{1}{2}\left(2\hat{\beta}_{0,0} - 2\frac{\omega}{1-\omega}\hat{\beta}_{1,0}\right) = \hat{\beta}_{0,0} - \frac{\omega}{1-\omega}\hat{\beta}_{1,0} \,. \tag{40}$$

These are the books equations 3.43. The initial values of $S_0^{[1]}$ and $S_0^{[2]}$ are specified once initial values are given by $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$. One can initialize the values of $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$ using the expression for ordinary least squares of a linear fit derived in Chapter 2, see page 6.

Example 3.4: University of Iowa Student Enrollment

We can verify that we have implemented double exponential smoothing correctly by duplicating the books Example 3.4 which deals with the university of Iowa student enrollment. Using the MATLAB script `example_3_4.m` we first load the data set using the function `load_student_enrollment.m`. To get an intuitive understanding of how double exponential smoothing performs, we plot the original data, pick an arbitrary value for the relaxation parameter ω, and plot the corresponding double exponential smooth overlayed on the original time series. The result of this is plotted in Figure 3 (left). In addition, we then perform a simulation over a range of relaxation values $0.01 < \omega < 0.9$ (chosen so that they would duplicate Table 3.12 in the book) and from all simulations find the value of ω that minimize the one-step ahead prediction error. When the above MATLAB script is run it also produces plots of the autocorrelation of the residuals. Their magnitude matches quite well with the values given in the book in Table 3.14. Because of the close agreement with many of the results here we can be more certain that the code for double exponential smoothing in the MATLAB script `double_exp_smoothing.m` has been implemented correctly.

Prediction intervals for future values

In this section the book shows that the variance of the l-step-ahead prediction error is given by

$$V(e_n(l)) = \sigma^2 + f'(l)V(\hat{\beta}_n)f(l) \,, \tag{41}$$

and that the variance of our estimate of β is given by

$$V(\hat{\beta}_n) = F_n^{-1}V(h_n)F_n^{-1} \,. \tag{42}$$

If we assume that the observations z_t are uncorrelated and have the same variance σ^2 then we can evaluate $V(h_n)$. We have from the expression for the variance in terms of expectations that

$$V(h_n) = E\left(h_n h_n'\right) - E(h_n)E(h_n)' \tag{43}$$

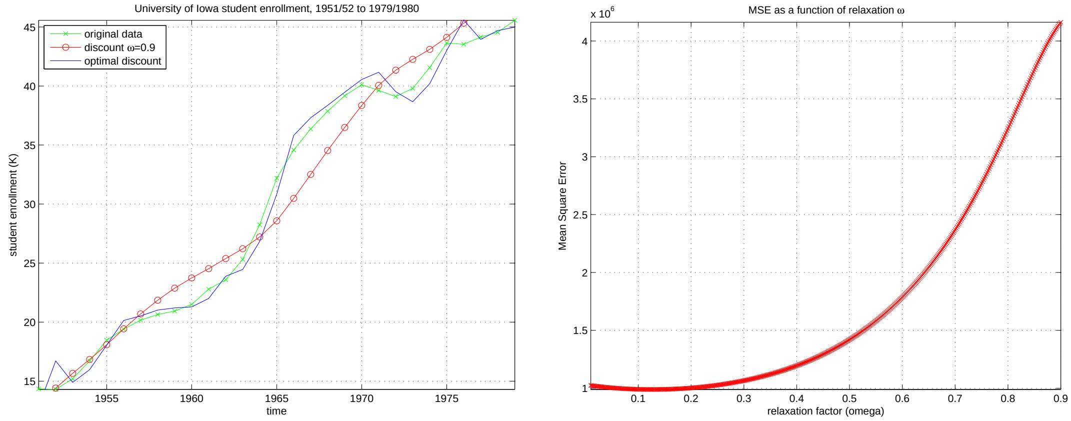


Figure 3: **Left:** The original university of Iowa student enrollment data (in green), the double exponential smooth of this data with a relaxation parameters $\omega = 0.9$ (in red), and the optimally selected smooth (in blue). **Right:** The SSE(ω) as a function of ω. The minimum was found to be $\omega_{\text{optimal}} = 0.125$ or $\alpha = 0.875$ and corresponds very closely with what the book reports.

Now from the definition of h_n we have the expectation in the second term above can be simplified as

$$E(h_n) = E\left(\sum_{j=0}^{n-1} \omega^j f(-j) z_{n-j}\right) = \sum_{j=0}^{n-1} \omega^j f(-j) E(z_{n-j}) = \sum_{j=0}^{n-1} \omega^j f(-j) f'(-j)\beta \,, \tag{44}$$

since we assume that $z_{n+j} = f'(j)\beta + \varepsilon_{n+j}$. We then square this expression to obtain

$$\begin{aligned} E(h_n)E(h_n)' &= \left(\sum_{j=0}^{n-1} \omega^j f(-j) f'(-j)\beta\right)\left(\sum_{k=0}^{n-1} \omega^k f(-k) f'(-k)\beta\right)' \\ &= \sum_{j=0}^{n-1}\sum_{k=0}^{n-1} \omega^j \omega^k f(-j) f'(-j)\beta\beta' f(-k) f'(-k) \,, \end{aligned}$$

as an expression for the second term in the expectation expansion of $V(h_n)$ in Equation 43. Note we can mentally check the dimensions of this product to determine that it indeed is a matrix as it should be. In addition, using the associativity property of matrix multiplications we can recognize that some of the products above evaluate to scalars and we can write

$$f(-j)f'(-j)\beta\beta' f(-k)f'(-k) = (f'(-j)\beta)\,(\beta' f(-k))\, f(-j)f'(-k) \,,$$

where we have factored out two scalar terms. To evaluate the first term in Equation 43 begin by expanding the summation as

$$\begin{aligned}
E(h_n h_n') &= E\left(\left(\sum_{j=0}^{n-1} \omega^j f(-j) z_{n-j}\right)\left(\sum_{k=0}^{n-1} \omega^j f(-k) z_{n-k}\right)'\right) \\
&= E\left(\sum_{j=0}^{n-1}\sum_{k=0}^{n-1} \omega^j \omega^k z_{n-j} z_{n-k} f(-j) f'(-k)\right) \\
&= \sum_{j=0}^{n-1}\sum_{k=0}^{n-1} \omega^j \omega^k E(z_{n-j} z_{n-k}) f(-j) f'(-k) .
\end{aligned}$$

Now to evaluate $E(z_{n-j}z_{n-k})$ recall our model for z_n to find

$$\begin{aligned}
E(z_{n-j}z_{n-k}) &= E\left((f'(-j)\beta + \varepsilon_{n-j})(f'(-k)\beta + \varepsilon_{n-k})\right) \\
&= f'(-j)\beta f'(-k)\beta + E(\varepsilon_{n-j}\varepsilon_{n-k}) \\
&= f'(-j)\beta f'(-k)\beta + \begin{cases} 0 & j \neq k \\ \sigma^2 & j = k \end{cases} .
\end{aligned}$$

Thus we can combined these two results to obtain

$$E(h_n h_n') = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1} \omega^{j+k} f(-j) f'(-k)(f'(-j)\beta f'(-k)\beta) + \sum_{j=0}^{n-1} \omega^{2j}\sigma^2 f(-j) f'(-j) .$$

Thus when we perform the subtraction needed in Equation 43 we find that the two double sums cancel and we are left with

$$V(h_n) = \sigma^2 \sum_{j=0}^{n-1} \omega^{2j} f(-j) f'(-j) , \tag{45}$$

the same as the expression given in the book.

Exercise Solutions

Exercise 3.1 (an example with exponential smoothing)

To update our forecast using the simple exponential smoothing (a locally constant mean model) where all look aheads result in the *same* prediction we have

$$\hat{z}_n(l) = \hat{z}_{n-1}(1) + (1-\omega)(z_n - \hat{z}_{n-1}(1)) . \tag{46}$$

In this problem we are told that $\hat{z}_{30}(2) = 102.5$ which also means $\hat{z}_{30}(1) = 102.5$ since the amount of look ahead does not matter. Thus the forecaster should predict using Equation 46 with $n = 31$ and $l = 1$ that

$$\begin{aligned}
\hat{z}_{31}(1) &= \hat{z}_{30}(1) + (1-\omega)(z_{30} - \hat{z}_{30}(1)) \\
&= 102.5 + 0.1(105 - 102.5) = 102.75 ,
\end{aligned}$$

for our new estimate of z_{32}.

Exercise 3.2 (predicting yearly sales)

Part (a): We will assume that the data provided to the simple exponential model is in the form of *yearly* sales so that the forecaster is not forecasting sales on a monthly time frame. Then we are told that the one-step-ahead yearly prediction is given by $\hat{z}_n(1) = 960$. If in January we have observed actual sales of 90 units, the forecasters revised forecast would be $960 - 90 = 870$ units to be sold in the next eleven months.

Part (b): A simple method of predicting the smoothing parameter α is to simulate the ones-step-ahead error

$$\text{SSE}(\alpha) = \sum_{t=1}^{n} e_{t-1}^2(1) = \sum_{t=1}^{n} (z_t - \hat{z}_{t-1}(1))^2 \,, \tag{47}$$

for many values of α and select the values of α that minimizes this expression.

Part (c): If the forecaster observes that a linear trend in the data z_t seems to be changing, then a better procedure to use would be double exponential smoothing which explicitly incorporates a linear component into the model.

Exercise 3.3 (the two-step-ahead forecast for simple exponential smoothing)

From the given definition of exponential smoothing

$$\hat{z}_n(1) = \alpha \left(z_n + (1-\alpha) z_{n-1} + (1-\alpha)^2 z_{n-2} + \cdots \right) \,, \tag{48}$$

if we consider $\hat{z}_n(2)$ defined in an analogous way as above but with the unknown time series value z_n replaced with its one-step-ahead estimate given by $\hat{z}_n(1)$. We find

$$\begin{aligned}
\hat{z}_n(2) &= \alpha \left(\hat{z}_n(1) + (1-\alpha) z_n + (1-\alpha)^2 z_{n-1} + \cdots \right) \\
&= \alpha \left[\alpha (z_n + (1-\alpha) z_{n-1} + (1-\alpha)^2 z_{n-2} + \cdots) \right. \\
&+ \left. (1-\alpha) z_n + (1-\alpha)^2 z_{n-1} + (1-\alpha)^3 z_{n-2} + \cdots \right] \\
&= \alpha \left[z_n + (1-\alpha)(\alpha + (1-\alpha)) z_{n-1} + (1-\alpha)^2 (\alpha + (1-\alpha)) z_{n-2} + \cdots \right. \\
&= \alpha \left[z_n + (1-\alpha) z_{n-1} + (1-\alpha)^2 z_{n-2} + \cdots \right. \\
&= \hat{z}_n(1) \,,
\end{aligned}$$

as we were to show.

Exercise 3.4 (simple exponential smoothing on the U.S. lumber data)

This problem is implemented in the MATLAB script `prob_3_4.m`. In that script we begin by loading the data set using the utility function `load_us_lumber.m`. We then specify an initial value to choose for the initial condition of the simple exponential smooth S_0. Some possible

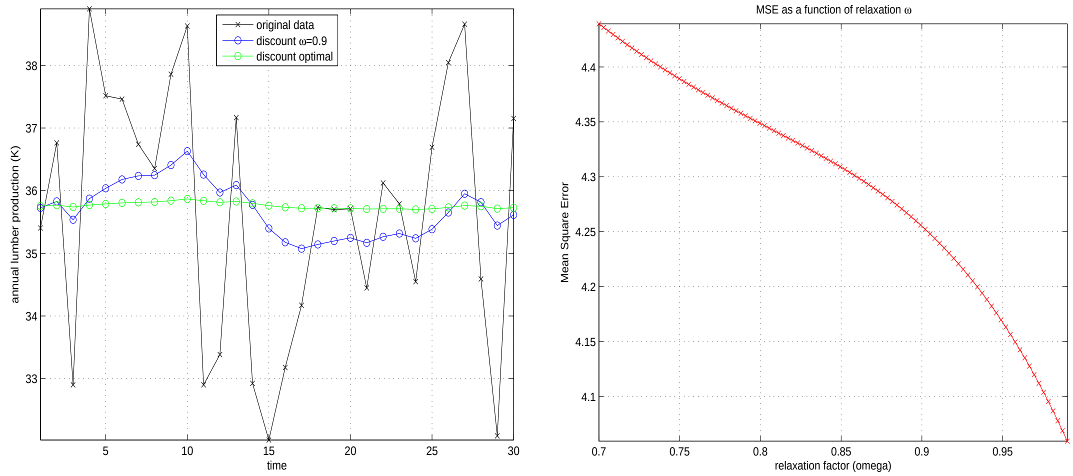


Figure 4: **Left:** The U.S. lumber data (in black), an example of simple exponential smoothing with a relaxation parameter $\omega = 0.9$ (in blue), and simple exponential smoothing with the optimal smoothing value $\omega = 0.99$ (in green). **Right:** A plot of the SSE(ω) given by Equation 47 as a function of $\omega = 1 - \alpha$. The minimum value of SSE(ω) is clearly located at the right end of the domain.

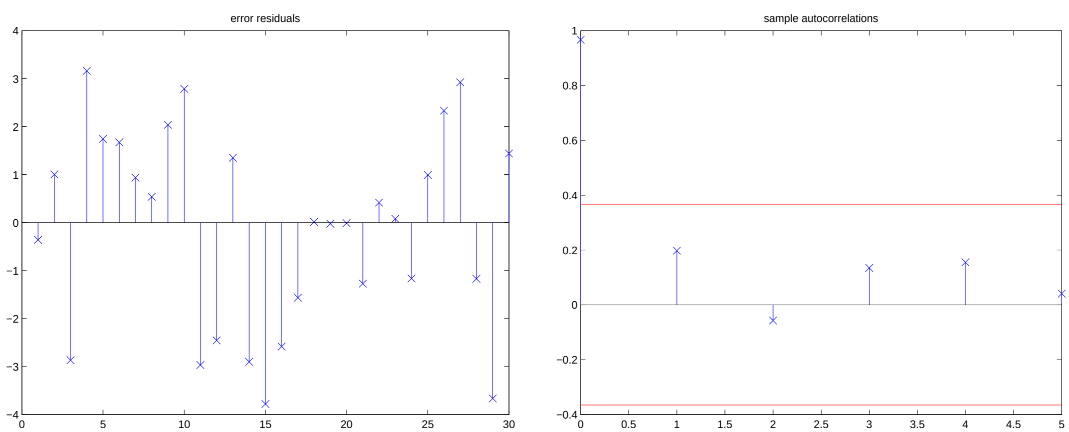


Figure 5: **Left:** The error residuals of the simple exponential smooth using the optimal value of ω. **Right:** The sample autocorrelation function of the error residuals.

simple choices are $S_0 = z_1$ the first measurement, $S_0 = \bar{z} = \frac{1}{n}\sum_t z_t$ the sample mean of the entire data set, the sample median, or the mean of some number of initial terms. Once an initial value for S_n is chosen, to visualize the output of simple exponential smoothing we pick a relaxation value randomly, say $\omega = 0.9$, and plot the resulting smoothed values S_n. The original data series and this initial smoothed value are shown in Figure 4 (left). We next seek to evaluate the *optimal* ω and determine how well we can forecast this time series with exponential smoothing. To do this we sample from a grid of ω values between $[0.7, 0.99]$ and for each, compute the one-step-ahead predictions using Equation 46 for $n = 0, 1, \cdots$. From these values of $\hat{z}_n(1)$ we compute the one-step-ahead prediction error using Equation 47 repeated here for convenience

$$\mathrm{SSE}(\omega) = \sum_{t=1}^{n} e_{t-1}^2(1) = \sum_{t=1}^{n} (z_t - \hat{z}_{t-1})^2 .$$

A plot of this expression for the ω values in the above grid is given in Figure 4 (right). From this plot we see that the optimal value of ω occurs at the right end of the domain in this case $\omega = 0.99$. This relatively large value of ω means that the mean level changes only slowly a fact that can be visually observed when we plot the values of $\hat{z}_n(1)$ for the optimal ω in Figure 4 (left) in green. Note that visually this has the appearance of a globally constant mean model. Since the optimal smooth is so similar to the globally constant mean model we could conclude that the significance of the model will be very similar to the results derived in Example 3.1. A plot of the error residuals in Figure 5 (left) appears to have mean zero and be uncorrelated. A plot of the sample autocorrelation function of the error residuals in Figure 5 (right) gives numerical values almost the same as that given in Example 3.1. In summary it appears that in finding the optimal $\omega \approx 1$ has reduced the more general exponential smoothing to the globally constant mean model.

Exercise 3.5 (simple exponential smoothing vs. a moving average)

In *simple exponential smoothing* the smoothed statistic S_n is given in terms of the time series by

$$S_n = S_n^{[1]} = (1-\omega)(z_n + \omega z_{n-1} + \omega^2 z_{n-2} + \cdots) . \tag{49}$$

Because we have an explicit model of our process z_t given by $z_t = \mu + \varepsilon_t$ we can study the variance of this statistic. We begin by computing the expectation of S_n. We find

$$\begin{aligned} E(S_n) &= (1-\omega)(E(z_n) + \omega E(z_{n-1}) + \omega^2 E(z_{n-2}) + \cdots) \\ &= (1-\omega)(\mu + \omega\mu + \omega^2\mu + \cdots) \\ &= \mu(1-\omega)(1 + \omega + \omega^2 + \cdots) = \mu . \end{aligned}$$

We next consider $E(S_n^2)$, which we find is given by (using $\alpha = 1-\omega$)

$$
\begin{aligned}
E(S_n^2) &= \alpha^2 E\left(\sum_{k,k'\geq 0} \omega^k \omega^{k'} z_{n-k} z_{n-k'}\right) \\
&= \alpha^2 E\left(\sum_{k,k'\geq 0} \omega^k \omega^{k'} (\mu + \varepsilon_{n-k})(\mu + \varepsilon_{n-k'})\right) \\
&= \alpha^2 E\left(\mu^2 \sum_{k,k'\geq 0} \omega^k \omega^{k'} + \mu \sum_{k,k'\geq 0} \omega^k \omega^{k'} \varepsilon_{n-k} + \mu \sum_{k,k'\geq 0} \omega^k \omega^{k'} \varepsilon_{n-k'} + \sum_{k,k'\geq 0} \omega^k \omega^{k'} \varepsilon_{n-k}\varepsilon_{n-k'}\right) \\
&= \alpha^2 \left(\frac{\mu^2}{(1-\omega)^2} + \sum_{k,k'\geq 0} \omega^k \omega^{k'} E(\varepsilon_{n-k}\varepsilon_{n-k'})\right) .
\end{aligned}
$$

Now the expectation in the second term can be evaluated since we assume that the ε_n are uncorrelated with variance σ^2, meaning that $E(\varepsilon_{n-k}\varepsilon_{n-k'}) = \sigma^2 \delta_{n-k,n-k'} = \sigma^2 \delta_{k,k'}$ so the above expression then becomes

$$
E(S_n^2) = \mu^2 + \alpha^2 \sigma^2 \sum_k \omega^{2k} = \mu^2 + \frac{\alpha^2\sigma^2}{1-\omega^2} = \mu^2 + \left(\frac{1-\omega}{1+\omega}\right)\sigma^2 .
$$

Thus $\text{Var}(S_n)$ can be computed using its expansion in terms of expectations and is given by

$$
\text{Var}(S_n) = E[S_n^2] - E[S_n]^2 = \left(\frac{1-\omega}{1+\omega}\right)\sigma^2 . \tag{50}
$$

For the *moving average* of the most recent N observations it definition is given by

$$
\bar{z}_t^{(N)} = \frac{1}{N}(z_t + z_{t-1} + \cdots + z_{t-N+1}) . \tag{51}
$$

From which we observe that $E(\bar{z}_t^{(N)}) = \mu$ and $(\bar{z}_t^{(N)})^2 = \frac{1}{N^2}\sum_{k,k'=0}^{N-1} z_{t-k} z_{t-k'}$ so that the expectation of $(\bar{z}_t^{(N)})^2$ is given by

$$
\begin{aligned}
E((\bar{z}_t^{(N)})^2) &= \frac{1}{N^2} \sum_{k,k'=0}^{N-1} E(z_{t-k} z_{t-k'}) \\
&= \frac{1}{N^2} \sum_{k,k'=0}^{N-1} \left(\mu^2 + \mu E(\varepsilon_{t-k}) + \mu E(\varepsilon_{t-k'}) + E(\varepsilon_{t-k}\varepsilon_{t-k'})\right) \\
&= \mu^2 + \frac{\sigma^2}{N^2} \sum_{k=0}^{N-1} 1 = \mu^2 + \frac{\sigma^2}{N} .
\end{aligned}
$$

Thus the variance of the moving average of the last N observations becomes

$$
\text{Var}(\bar{z}_t^{(N)}) = \mu^2 + \frac{\sigma^2}{N} - \mu^2 = \frac{\sigma^2}{N} . \tag{52}
$$

As suggested in the book lets consider the evaluation of $\text{Var}(S_n)$ using Equation 50 when $\alpha = \frac{2}{N+1}$. For that specific value of α we see that $\omega = 1 - \alpha = 1 - \frac{2}{N+1} = \frac{N-1}{N+1}$, and $1 + \omega = \frac{2N}{N+1}$, so that

$$\text{Var}(S_n) = \frac{2/(N+1)}{(2N/(N+1))}\sigma^2 = \frac{1}{N}\sigma^2 ,$$

the same as given in Equation 52 the variance of the moving average of the last N observations. In addition to equality we will have $V(S_n) \leq V(\bar{z}_t^{(N)})$ if

$$\frac{1-\omega}{1+\omega} \leq \frac{1}{N} ,$$

holds. On solving for ω we find that this requires

$$\omega \geq \frac{1 - \frac{1}{N}}{1 + \frac{1}{N}} = \frac{N-1}{N+1} ,$$

as the requirement on ω so that the variance of the simple exponential smooth is less than that of the N term simple moving average.

Each of the two techniques discussed have their own recursive update procedures to use when a new measurement z_{n+1} is available. For the technique of simple exponential smoothing, this update equation is given by Equation 46 (with $l = 1$) repeated here in the notation of a smooth and indexed from n rather than $t - 1$ as

$$S_{n+1} = \alpha z_{n+1} + (1 - \alpha)S_n . \tag{53}$$

For the moving average model $\bar{z}_t^{(N)}$ the recursive update equation can be computed as

$$\begin{aligned}
\bar{z}_t^{(N)} &= \frac{1}{N}(z_t + z_{t-1} + \cdots + z_{t-N+2} + z_{t-N+1}) \quad \text{so that} \\
\bar{z}_{t+1}^{(N)} &= \frac{1}{N}(z_{t+1} + z_t + z_{t-1} + \cdots + z_{t-N+2}) \\
&= \frac{1}{N}z_{t+1} + \frac{1}{N}(z_t + z_{t-1} + \cdots + z_{t-N+2} + z_{t-N+1}) - \frac{1}{N}z_{t-N+1} \\
&= \bar{z}_t^{(N)} + \frac{1}{N}(z_{t+1} - z_{t-N+1}) . \qquad (54)
\end{aligned}$$

Note that Equation 54 is more difficult to use than Equation 46 (equivalently Equation 53) since it requires the value of the data point z_{t-N+1} to update the forecast when the new datum z_{t+1} arrives. In the same way when the next point z_{t+2} arrives to update our forecast we will require the value of z_{t-N+2}. Thus all of the last N data points must be stored in memory as the algorithm processes incoming samples. This would appear to be a disadvantage over Equation 53 where only the *previous* value of the "state", S_n, is needed to be saved.

Exercise 3.6 (simple exponential smoothing of a series with an outlier)

As in Exercise 3.5 we assume our data generation process z_t is of the form $z_t = \mu + \varepsilon_t$. If at the single time t_0 one sample z_{t_0} is drawn from the level $\mu_1 = \mu + \delta$ then S_n when the number of samples is finite would look like for $n > t_0$ the following (this is similar to Equation 49)

$$S_n = c \sum_{k=1}^{n} \omega^{n-k} z_k = c \left(\sum_{k=1}^{t_0-1} \omega^{n-k} z_k + \omega^{n-t_0} z_{t_0} + \sum_{k=t_0+1}^{n} \omega^{n-k} z_k \right) ,$$

for $c = \frac{1-\omega}{1-\omega^n}$. If we are told that for this single time period t_0, the random sample we observe z_{t_0}, is given by $z_{t_0} = \mu + \delta + \varepsilon_{t_0}$ we find that S_n introduced above becomes

$$S_n = c \left(\sum_{k=1}^{t_0-1} \omega^{n-k} z_k + \omega^{n-t_0} (\mu + \delta + \varepsilon_{t_0}) + \sum_{k=t_0+1}^{n} \omega^{n-k} z_k \right) .$$

Taking the expectation of this expression to compute $E(S_n)$ we find

$$\begin{aligned}
E(S_n) &= c \sum_{k=1}^{n} \omega^{n-k} \mu + c\delta\omega^{n-t_0} \\
&= \mu c \left(\frac{1-\omega^n}{1-\omega} \right) + c\delta\omega^{n-t_0} \\
&= \mu + c\delta\omega^{n-t_0} .
\end{aligned}$$

As $n \to 0$ since $0 \leq \omega \leq 1$, the values of $\omega^n \to 0$ and so the disturbance $c\delta\omega^{n-t_0}$ in $E(S_n)$ produced by the sample z_{t_0} also decreases to zero. One could imagine a case where the measurement at z_{t_0} was the result of an *outlier* and was not truly representative of a large number of case. The above result shows that with simple exponential smoothing the perturbation introduced, eventually becomes negligible in its effect on the value of $E(S_n)$.

Exercise 3.7 (simple exponential smoothing of a series where the mean changes)

If the mean of our process z_t shifts to $\mu_1 = \mu + \delta$ from just μ for all times $t \geq t_0$ then we write S_n as in Exercise 3.6 as

$$S_n = c \left(\sum_{k=1}^{t_0-1} \omega^{n-k} z_k + \sum_{k=t_0}^{n} \omega^{n-k} z_k \right) .$$

Now the expectation of our measurements z_k when $k \geq t_0$ is given by $E(z_k) = \mu + \delta$ while when $t < t_0$ this expectation is $E(z_k) = \mu$ so the above when we take the expectation becomes

$$
\begin{aligned}
E(S_n) &= c\left(\sum_{k=1}^{t_0-1}\omega^{n-k}E(z_k)+\sum_{k=t_0}^{n}\omega^{n-k}E(z_k)\right)\\
&= c\left(\sum_{k=1}^{t_0-1}\omega^{n-k}\mu+\sum_{k=t_0}^{n}\omega^{n-k}(\mu+\delta)\right)\\
&= c\left(\mu\sum_{k=1}^{n}\omega^{n-k}+\delta\sum_{k=t_0}^{n}\omega^{n-k}\right)\\
&= \mu c\left(\frac{1-\omega^n}{1-\omega}\right)+c\delta\sum_{k=0}^{n-t_0}\omega^k\\
&= \mu+\delta\left(\frac{1-\omega^{n-t_0+1}}{1-\omega^n}\right),
\end{aligned}
$$

when $n \geq t_0$. Note that as $n \to \infty$ the value of $E(S_n)$ approaches $\mu+\delta=\mu_1$ the new mean. This gives an indication that if the underlying process changes, simple exponential smoothing will be able to track these changes.

Exercise 3.8 (the explicit representation of double exponential smoothing)

For double exponential smoothing the l step ahead predictors are given in terms of the first and second order smooths $S_n^{[1]}$ and $S_n^{[2]}$ and the smoothing constant α by

$$
\hat{z}_n(l) = \left(2+\frac{\alpha}{1-\alpha}l\right)S_n^{[1]} - \left(1+\frac{\alpha}{1-\alpha}l\right)S_n^{[2]}, \tag{55}
$$

where we know the expressions for $S_n^{[1]}$ and $S_n^{[2]}$ in terms of the series z_n from Equations 29 and 32. When we put these two expressions in we find $\hat{z}_n(l)$ becomes

$$
\begin{aligned}
\hat{z}_n(l) &= \alpha\left(2+\frac{\alpha}{1-\alpha}l\right)\sum_{j=0}^{n-1}\omega^j z_{n-j} - \alpha^2\left(1+\frac{\alpha}{1-\alpha}l\right)\sum_{j=0}^{n-1}(j+1)\omega^j z_{n-j}\\
&= \alpha\sum_{j=1}^{n}\left(2-\alpha j+\left(\frac{\alpha}{1-\alpha}\right)(1-\alpha j)l\right)\omega^{j-1}z_{n+1-j}.
\end{aligned}
$$

Thus in terms of the coefficients z_{n+1-j} we have $\pi_j^{(l)}$ given by

$$
\pi_j^{(l)} = \alpha\left(2-\alpha j+\left(\frac{\alpha}{1-\alpha}\right)(1-\alpha j)l\right)\omega^{j-1} \quad \text{for} \quad j \geq 1.
$$

When $\alpha = 0.1$ we have $\omega = 0.9$ and for $l = 1$ and $l = 2$ a plot of these coefficients is shown in Figure 6. This problem is worked in the MATLAB script `prob_3_8.m`.

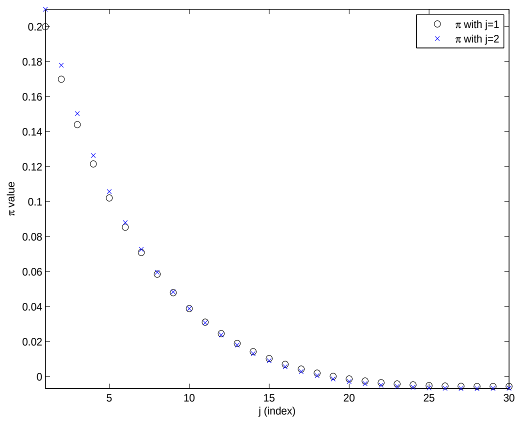


Figure 6: Plots of $\pi_j(l)$ for $l = 1$ and $l = 2$ as specified in Exercise 3.8.

Exercise 3.9 (Holt's method vs. double exponential smoothing)

Part (a): Holt's method iterates between performing a mean value update (updating the parameter β_0) and a slope value update (updating the parameter β_1) where different smoothing constants ω_i can be used for each update if desired. Specifically, we begin Holt's iterations by taking

$$\hat{\mu}_0 = 30 \quad \text{and} \quad \hat{\beta}_0 = 2\,,$$

and as each measurement $z_1, z_2, \cdots$ is observed we update these variables using the following

$$\begin{aligned} \hat{\mu}_{n+1} &= (1-\omega_1)z_{n+1} + \omega_1(\hat{\mu}_n + \hat{\beta}_n) && (56) \\ \hat{\beta}_{n+1} &= (1-\omega_2)(\hat{\mu}_{n+1} - \hat{\mu}_n) + \omega_2\hat{\beta}_n\,, && (57) \end{aligned}$$

for $n = 0, 1, 2, \cdots$ and for two discount factors $0 < \omega_i < 1$, for $i = 1, 2$. The prediction of the next measurement to be received from the point $n + 1$ onward is then given by

$$\hat{z}_{n+1}(l) = \hat{\mu}_{n+1} + \hat{\beta}_{n+1} l\,. \quad (58)$$

When we do this we obtain forecasts given by

```
33.120   35.040
34.574   36.451
36.765   38.671
```

Here $\hat{z}_{n+1}(1)$ is the first column and $\hat{z}_{n+1}(2)$ is the second column.

Part (b): For double exponential smoothing we need a way to calculate the initial smooth values $S_n^{[1]}$ and $S_n^{[2]}$ given the global fit from β_0 and β_1. Such a method is given by Equations 39 and 40. Once we have these variables translated we can use the MATLAB function `double_exp_smoothing.m` with the initial smoothing $S_0^{[1]}$ and $S_0^{[2]}$ specified to evaluate the required smooth. When we do this we find

```
32.0000   35.1600
33.2000   36.6580
34.7200   38.8648
```

for the predicted one and two step look aheads. Note to calculate $\hat{z}_{n+1}(l)$ when $l = 2$ we had to use Equation 55 and the computed smooths $S_n^{[1]}$ and $S_n^{[2]}$. Both parts of this problem are worked in the MATLAB script `prob_3_9.m`.

Exercise 3.10 (the variance of double exponential smoothing)

For this exercise it is helpful to note that if we *know* the true model is given by a functional form like $y_t = x_t'\beta + \varepsilon_t$, the deterministic vector x_t, and the value of β then $V(y_t) = V(\varepsilon_t) = \sigma^2$. If, however, we have to *estimate* β from data, then this estimation procedure introduces additional uncertainty in the variance of our *prediction* $\hat{y}_t = x_t'\hat{\beta}$ such that $V(\hat{y}_t) \neq V(\varepsilon)$. In terms of the a one-step-ahead error

$$e_t(1) = z_{t+1} - \hat{y}_t = z_{t+1} - x_t'\hat{\beta} = x_t'\beta + \varepsilon_{t+1} - x_t'\hat{\beta} = x_t'(\beta - \hat{\beta}) + \varepsilon_{t+1} \,.$$

So that the variance of $e_t(1)$ is given by

$$V(e_t(1)) = V(\varepsilon_{t+1}) + x_t' V(\hat{\beta}) x_t > V(\varepsilon_{t+1}) \,,$$

which is an intuitive argument why after some algebra the book derived the result that

$$V(e_n(l)) = \sigma^2 c_l^2 \,. \tag{59}$$

where $c_l^2 > 1$.

Now for double exponential smoothing, specifically, we recognized this as as the same as a locally constant linear trend model, so if we use the results in Equation 59 or the book's equation 3.62 for double exponential smoothing we have a value of c_l^2 given by

$$c_l^2 = 1 + \frac{1-\omega}{(1+\omega)^3}((1 + 4\omega + 5\omega^2) + 2l(1-\omega)(1+3\omega) + 2l^2(1-\omega^2)) \,. \tag{60}$$

To use Equation 59 requires an estimate of σ^2 the variance of the error terms ε_t. Given an estimate of the mean absolute deviance $\hat{\Delta}_e$ defined as

$$\hat{\Delta}_e = \frac{1}{n}\sum_{t=1}^{n} |e_{t-1}(1)| = \frac{1}{n}\sum_{t=1}^{n} |z_t - \hat{z}_{t-1}(1)| \,, \tag{61}$$

we can estimate σ^2. In Montgomery and Johnson (1976) specifically it is shown that $\hat{\sigma}_e = 1.25|\hat{\Delta}_e|$, from which we can estimate σ^2 as

$$\hat{\sigma}^2 \approx \frac{\hat{\sigma}_e^2}{c_1^2} \,.$$

Using this, the $100(1-\lambda)\%$ prediction interval for $\hat{z}_n(l)$ is given by

$$\begin{aligned}\hat{z}_n(l) \pm u_{\lambda/2}\hat{\sigma}c_l &= \hat{z}_n(l) \pm u_{\lambda/2}\left(\frac{\hat{\sigma}_e}{c_1}\right)c_l \\ &= \hat{z}_n(l) \pm 1.25 u_{\lambda/2}|\hat{\Delta}_e|\left(\frac{c_l}{c_1}\right),\end{aligned}$$

where $u_{\lambda/2}$ is the $100(1-\lambda)\%$ percentage point for the standard normal distribution. In MATLAB $u_{\lambda/2}$ this value can be calculated by using

```
norminv( 1 - 0.5 * lambda, 0, 1 )
```

if the variable `lambda` holds the percentage desired i.e. `lambda = 0.05` for a 95% confidence interval.

The prediction interval for the *sum* of the next four observations is given from equation 3.68 in the book with $K = 4$ or

$$\sum_{l=1}^{4}\hat{z}_n(l) \pm u_{\lambda/2}\frac{\hat{\sigma}_e}{c_1}\left[4+\left(\sum_{l=1}^{4}f'(l)\right)F^{-1}F_*F^{-1}\left(\sum_{l=1}^{4}f(l)\right)\right]^{1/2}.$$

For double exponential smoothing $f(l) = \begin{bmatrix}1\\ l\end{bmatrix}$, so $\sum_{l=1}^{4} f(l) = \begin{bmatrix}4\\ 1+2+3+4\end{bmatrix} = \begin{bmatrix}4\\ 10\end{bmatrix}$, and $F^{-1}F_*F^{-1}$ is given by the expression above equation 3.62 in the book.

The *average* of the next K observations will have a variance given by

$$\begin{aligned}V\left(\frac{1}{K}\sum_{l=1}^{K}z_{n+l} - \frac{1}{K}\sum_{l=1}^{k}\hat{z}_n(l)\right) &= \frac{1}{K^2}V\left(\sum_{l=1}^{K}z_{n+l} - \sum_{l=1}^{k}\hat{z}_n(l)\right) \\ &= \frac{\sigma^2}{K^2}\left[K+\left(\sum_{l=1}^{K}f'(l)\right)F^{-1}F_*F^{-1}\left(\sum_{l=1}^{K}f(l)\right)\right],\end{aligned}$$

using the result from the book. Note the factor of $1/K^2$ outside this result. The remaining steps for this problem are the same in that we are attempting to estimate σ^2 from say

$$\hat{\sigma}^2 \approx \frac{\hat{\sigma}_e^2}{c_1^2} = \frac{1.25^2|\hat{\Delta}_e|^2}{c_1^2}.$$

When we take a smoothing factor of $\alpha = 0.1$ our relaxation coefficient is $\omega = 1-\alpha = 0.9$, so for the values of $l = 1, 2, 3, 4$ the value of c_l can be calculated above. These values are calculated in the MATLAB script `prob_3_10.m`.

Exercise 3.11 (some simple geometric sums)

For this problem we desire to compute $\sum_{j\geq 0} j^k \omega^j$ for $k = 1, 2, 3, 4$. Following the hint given we find that when $k = 1$ the desired sum is given by

$$\sum_{j\geq 0} j\omega^j = \omega \frac{\partial}{\partial \omega} \sum_{j\geq 0} \omega^j = \omega \frac{\partial}{\partial \omega} \left(\frac{1}{1-\omega} \right) = \frac{\omega}{(1-\omega)^2} . \tag{62}$$

For $k = 2$ using the previous result we find

$$\sum_{j\geq 0} j^2\omega^j = \omega \frac{\partial}{\partial \omega} \sum_{j\geq 0} j\omega^j = \omega \frac{\partial}{\partial \omega} \left(\frac{\omega}{(1-\omega)^2} \right) = \frac{\omega(1+\omega)}{(1-\omega)^3} . \tag{63}$$

For $k = 3$ using the previous result we find

$$\sum_{j\geq 0} j^3\omega^j = \omega \frac{\partial}{\partial \omega} \sum_{j\geq 0} j^2\omega^j = \omega \frac{\partial}{\partial \omega} \left(\frac{\omega(1+\omega)}{(1-\omega)^3} \right) = \frac{\omega(1+4\omega+\omega^2)}{(1-\omega)^4} . \tag{64}$$

Finally, for $k = 4$ we have

$$\sum_{j\geq 0} j^4\omega^j = \omega \frac{\partial}{\partial \omega} \sum_{j\geq 0} j^3\omega^j = \omega \frac{\partial}{\partial \omega} \left(\frac{\omega(2+3\omega+\omega^2)}{(1-\omega)^4} \right) = \frac{\omega(1+11\omega+11\omega^2+\omega^3)}{(1-\omega)^5} . \tag{65}$$

Exercise 3.12 (the general expression for the k-th order smooth)

We desire to show that for general k the k-th order smooth statistics can be written as

$$S_n^{[k]} = \frac{(1-\omega)^k}{(k-1)!} \sum_{j=0}^{n-1} \left[\prod_{i=1}^{k-1} (j+i) \right] \omega^j z_{n-j} \tag{66}$$

We will prove this by mathematical induction. To do so we begin by showing it is true for a few initial cases. When $k = 1$ this expression gives

$$S_n^{[1]} = (1-\omega) \sum_{j=0}^{n-1} \omega^j z_{n-j} ,$$

and when $k = 2$ this expression gives

$$S_n^{[2]} = (1-\omega) \sum_{j=0}^{n-1} (j+1)\omega^j z_{n-j} .$$

Both of which we have shown to be true based on Equations 29 and 32 earlier. Thus this relationship holds true when $k = 1$ and $k = 2$, if we assume that it holds true for all $k \leq K$, we can attempt to apply a inductive argument to prove the above expression is valid for all k. From a similar derivation given on Page 34 we will express $S_n^{[k+1]}$ in terms of $S_n^{[k]}$ and then sum all terms with the same value of z_{n-j}. From the recursive relationship of the $k+1$ smooth we have

$$\begin{aligned}
S_n^{[k+1]} &= (1-\omega)S_n^{[k]} + \omega S_{n-1}^{[k+1]} \\
&= (1-\omega)S_n^{[k]} + \omega((1-\omega)S_{n-1}^{[k]} + \omega S_{n-2}^{[k+1]}) \\
&= (1-\omega)S_n^{[k]} + \omega(1-\omega)S_{n-1}^{[k]} + \omega^2 S_{n-2}^{[k+1]} \\
&= (1-\omega)S_n^{[k]} + \omega(1-\omega)S_{n-1}^{[k]} + \omega^2(1-\omega)S_{n-2}^{[k]} + \omega^3 S_{n-3}^{[k+1]} .
\end{aligned}$$

The pattern above repeats and we have

$$\begin{aligned}
S_n^{[k+1]} &= (1-\omega)S_n^{[k]} + \omega(1-\omega)S_{n-1}^{[k]} + \omega^2(1-\omega)S_{n-2}^{[k]} + \cdots \\
&+ \omega^{n-1}(1-\omega)S_1^{[k]} + \omega^n(1-\omega)S_0^{[k+1]} \\
&= (1-\omega)\sum_{l=0}^{n} \omega^l S_{n-l}^{[k]} + \omega^n S_0^{[k+1]} .
\end{aligned}$$

Using the induction hypothesis on all of the terms $S_{n-l}^{[k]}$ we have

$$\begin{aligned}
S_n^{[k+1]} &= (1-\omega)\sum_{l=0}^{n} \omega^l \frac{(1-\omega)^k}{(k-1)!} \sum_{j=0}^{n-l-1} \left[\prod_{i=1}^{k-1}(j+i)\right] \omega^j z_{n-l-j} + \omega^n S_0^{[k+1]} \\
&= \frac{(1-\omega)^{k+1}}{(k-1)!} \sum_{l=0}^{n} \sum_{j=0}^{n-l-1} \left[\prod_{i=1}^{k-1}(j+i)\right] \omega^{l+j} z_{n-l-j} + \omega^n S_0^{[k+1]} .
\end{aligned}$$

Lets now change the order of the summation in the above expression. In terms of the index variables l and j we want to sum them "diagonally" across the l and j plane. That is we want the sum above to be in the following order

$$\begin{aligned}
(l,j) &\sim (0,0) \\
&+ (1,0)\,,(0,1) \\
&+ (2,0)\,,(1,1)\,,(0,2) \\
&+ (3,0)\,,(2,1)\,,(1,2)\,,(0,3) \\
&+ \cdots
\end{aligned}$$

Thus to sum in this direction we will make the following change of variables of the indices in the double sum above. Rather than perform an outer sum over l with an inner sum over j, we introduce two new index variables (p,q) defined as

$$p = l + j \quad \text{and} \quad q = j$$

so that

$$l = p - q \quad \text{and} \quad j = q\,.$$

When we make this substitution we find the sum above becomes

$$\sum_{l=0}^{n} \sum_{j=0}^{n-l-1} \left[\prod_{i=1}^{k-1}(j+i)\right] \omega^{l+j} z_{n-l-j} = \sum_{p=0}^{n-1} \sum_{q=0}^{p} \left[\prod_{i=1}^{k-1}(q+i)\right] \omega^p z_{n-p} .$$

To further evaluate this recall the definition of the factorial function [1] for $r = 1, 2, 3, \cdots$ as

$$t^{(r)} = t(t-1)(t-2)\cdots(t-r+1)\,, \tag{67}$$

which has the property that sums of factorial functions are simple to compute

$$\sum_t t^{(r)} = \frac{t^{(r+1)}}{r+1}\,. \tag{68}$$

With this background the inner sum above can be expressed in terms of the factorial function in that

$$\begin{aligned}\prod_{i=1}^{k-1}(q+i) &= (q+k-1)(q+k-2)\cdots(q+3)(q+2)(q+1) \\ &= (q+k-1)^{(k-1)}\,.\end{aligned}$$

Thus

$$\begin{aligned}\sum_{q=0}^{p}\left[\prod_{i=1}^{k-1}(q+i)\right] &= \sum_{q=0}^{p}(q+k-1)^{(k-1)} = \frac{1}{k}\left.(q+k-1)^{(k)}\right|_{q=0}^{p+1} \\ &= \frac{1}{k}((p+k)^{(k)} - (k-1)^{(k)}) \\ &= \frac{1}{k}(p+k)^{(k)}\,,\end{aligned}$$

since the second term vanishes. Combining these expressions we have finally come to

$$\begin{aligned}S_n^{[k+1]} &= \frac{(1-\omega)^{k+1}}{k!}\sum_{p=0}^{n-1}(p+k)^{(k)}\omega^p z_{n-p} + \omega^n S_0^{[k+1]} \\ &= \frac{(1-\omega)^{k+1}}{k!}\sum_{p=0}^{n-1}\left[\prod_{i=1}^{k}(p+i)\right]\omega^p z_{n-p} + \omega^n S_0^{[k+1]}\,,\end{aligned}$$

which as $n \to \infty$ is the desired expression.

Exercise 3.13 (comments about exponential smoothing)

Part (a): In simple exponential smoothing the forecasts $\hat{z}_n(l)$ are given by

$$\hat{z}_n(l) = c\sum_{t=0}^{n-1}\omega^t z_{n-t} = c(z_n + \omega z_{n-1} + \cdots + \omega^{n-1} z_1)\,, \tag{69}$$

where $c = \frac{1-\omega}{1-\omega^n}$ which are *constant* for all values of l.

Part (b): In double exponential smoothing the prediction l-step ahead are given by

$$\hat{z}_n(l) = f'(l)\hat{\beta}_n\,,$$

with $f(l) = \begin{bmatrix} 1 \\ l \end{bmatrix}$. This equation reduces to Equation 55 which expresses $\hat{z}_n(l)$ in terms of the first and second order smooths $S_n^{[1]}$ and $S_n^{[2]}$. To see if the value of $\hat{z}_n(3)$ is on a straight line containing the points $(n+1, \hat{z}_n(1))$ and $(n+2, \hat{z}_n(2))$ we will explicitly construct the line containing these two points and then see if the point $(n+3, \hat{z}_n(3))$ is on it. The line (t, y) connecting the two points $(n+1, \hat{z}_n(1))$ and $(n+2, \hat{z}_n(2))$ is given by

$$\begin{aligned} y - \hat{z}_n(1) &= \left(\frac{\hat{z}_n(2) - \hat{z}_n(1)}{n + 2 - (n+1)} \right) (t - (n+1)) \\ &= (\hat{z}_n(2) - \hat{z}_n(1))(t - (n+1)) \,. \end{aligned}$$

Now we can check if the point $(n+3, \hat{z}_n(3))$ is on it by evaluating both sides of the above expression with the help of Equation 55 and seeing if they are equal. When $t = n+3$ the right-hand-side becomes

$$2(\hat{z}_n(2) - \hat{z}_n(1)) = 2\left(\frac{\alpha}{1-\alpha}\right) S_n^{[1]} - 2\left(\frac{\alpha}{1-\alpha}\right) S_n^{[2]} \,,$$

while the left-hand-side of the above is given by

$$\hat{z}_n(3) - \hat{z}_n(1) = 2\left(\frac{\alpha}{1-\alpha}\right) S_n^{[1]} - 2\left(\frac{\alpha}{1-\alpha}\right) S_n^{[2]} \,.$$

Since these two expressions are equal the point $(n+3, \hat{z}_n(3))$ *does* indeed lie on a straight line between the points $(n+1, \hat{z}_n(1))$, and $(n+2, \hat{z}_n(2))$.

Part (c): Since simple exponential smoothing is equivalent to the locally constant model we can use the results from the book. Recall that the $100(1-\lambda)\%$ prediction interval for the locally constant model $z_{n+j} = \beta + \varepsilon_{n+j}$ is given by

$$S_n \pm u_{\lambda/2} \sigma \sqrt{\frac{2\alpha}{1-\omega^2}} \,, \tag{70}$$

which is the *same* for all l. This is to be expected since if we assume that the series is a constant our prediction interval around this constant should not change.

Exercise 3.14 (forecasting sales of computer software)

This problem is implemented as a series of function calls in the MATLAB script `prob_3_14.m`. For this problem we implemented MATLAB functions to perform single, double, and triple exponential smoothing in the files: `simple_exp_smoothing.m`, `double_exp_smoothing.m`, and `triple_exp_smoothing.m` respectively. These routines perform the requested forecasting procedures for a given input time series z_t and relaxation coefficient ω.

In addition, it is often desired to compute the *optimal* relaxation coefficient ω i.e. the one that gives the smallest mean square error MSE when evaluated on a provided data set. Given a grid of ω values the MATLAB functions `simple_exp_smoothing_optimum.m`,

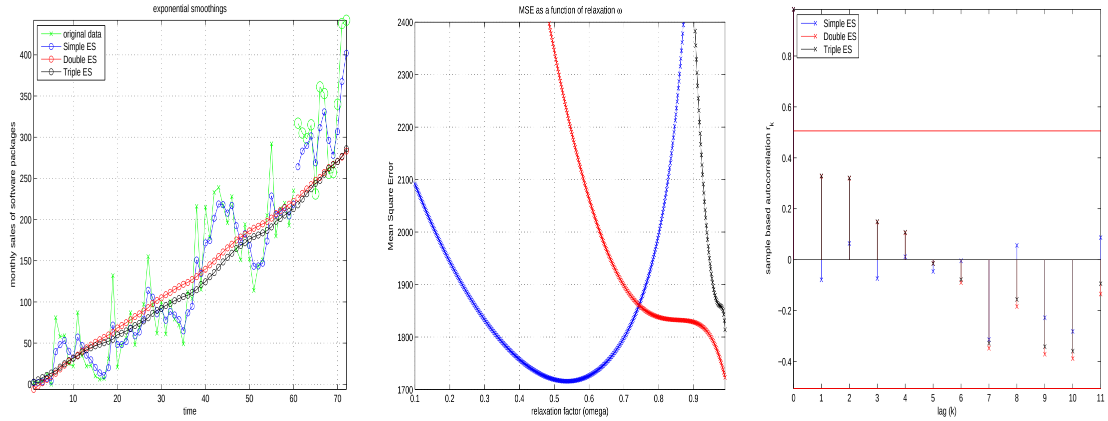


Figure 7: The sales of computer software example. **Left:** The raw data (in green), and the *optimal* single exponential smoothing (in blue), double exponential smoothing (in red) and triple exponential smoothing (in black). **Center:** Plots of the mean square error as a function of smoothing coefficient ω. **Right:** Plots of the sample autocorrelations as a function of the lag k.

`double_exp_smoothing_optimum.m`, and `triple_exp_smoothing_optimum.m` will apply their respective forecasting methods and return the value of ω that yields the smallest MSE. If a grid of value for ω is not specified, one is proposed, and results are returned relative to this grid.

Part (a): The original data set is plotted in Figure 7 (left) in green.

Part (b): As discussed in the book, since there are very few real world data sets that are best predicted with triple exponential smoothing we expect the data set to be characterized best by either simple exponential smoothing in which we have a slowly changing mean level or double exponential smoothing where we have a slowly changing mean and slope. In order to be complete, however, we will consider the use of triple exponential smoothing also.

As a summary, we present the steps involved in the implementation of simple, double, and triple exponential smoothing (assuming we are given an value for discount factor ω) here.

For **Simple Exponential Smoothing** we

- Estimate an initial value for the first order smooth $S_0 \equiv S_0^{[1]}$ (see below).
- Use the updating equation

  $$S_n = (1-\omega)z_n + \omega S_{n-1}$$

  to calculate the modified values of S_n as each new data point z_n is observed.
- Then the forecasts for l-step-ahead value $\hat{z}_n(l)$ given by

  $$\hat{z}_n(l) = S_n \,. \tag{71}$$

For **Double Exponential Smoothing** we

- Estimate initial values for the smoothed statistics $S_0^{[1]}$ and $S_0^{[2]}$ (see below).
- Use the updating equations

$$
\begin{aligned}
S_n^{[1]} &= (1-\omega)z_n + \omega S_{n-1}^{[1]} \\
S_n^{[2]} &= (1-\omega)S_n^{[1]} + \omega S_{n-1}^{[2]}
\end{aligned}
$$

  as each new data point z_n arrives.
- Then the forecasts for l-step-ahead value $\hat{z}_n(l)$ is by

$$
\hat{z}_n(l) = \left(2 + \frac{1-\omega}{\omega} l\right) S_n^{[1]} - \left(2 + \frac{1+\omega}{\omega} l\right) S_n^{[2]} \tag{72}
$$

For **Triple Exponential Smoothing** we have many of the same steps as above. We

- Estimate initial values for the smoothed statistics $S_0^{[1]}$, $S_0^{[2]}$, and $S_0^{[3]}$ (see below)
- Use the updating equations

$$
\begin{aligned}
S_n^{[1]} &= (1-\omega)z_n + \omega S_{n-1}^{[1]} \\
S_n^{[2]} &= (1-\omega)S_n^{[1]} + \omega S_{n-1}^{[2]} \\
S_n^{[3]} &= (1-\omega)S_n^{[2]} + \omega S_{n-1}^{[3]} .
\end{aligned}
$$

- Then the forecasts for l-step-ahead value $\hat{z}_n(l)$ is given by

$$
\hat{z}_n(l) = \left(2 + \frac{1-\omega}{\omega} l\right) S_n^{[1]} - \left(2 + \frac{1+\omega}{\omega} l\right) S_n^{[2]} \tag{73}
$$

As an example of how to estimate the values of $S_0^{[k]}$ for the various cases above, consider the double exponential smoothing. Methods to estimate the initial smoothing values like $S_0^{[1]}$ and $S_0^{[2]}$ involve estimating their equivalent local linear regression coefficients $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$ and then relating these values back to $S_0^{[1]}$ and $S_0^{[2]}$. To estimate $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$ we fit the constant linear model $z_t = \beta_0 + \beta_1 t + \varepsilon_t$ to some subset of the observations say the first half or first third or even the entire data set. Once the values of $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$ are estimated from them we derive estimates of $S_0^{[1]}$ and $S_0^{[1]}$ as

$$
\begin{aligned}
S_0^{[1]} &= \hat{\beta}_{0,0} - \left(\frac{\omega}{1-\omega}\right) \hat{\beta}_{1,0} \\
S_0^{[2]} &= \hat{\beta}_{0,0} - \left(\frac{2\omega}{1-\omega}\right) \hat{\beta}_{1,0} .
\end{aligned}
$$

Methods used to estimate $S_0^{[k]}$ for the other smoothing methods are similar.

When we implement these three smoothing methods in MATLAB and run them on the provided data set we obtain Figure 7 (left) which compares the actual data points with the three forecast methods. To determine the specific value for the relaxation parameter ω used for each smoothing method we selected a grid of values for ω, generated the simple, double, and exponential time series smooths for each of them and then selected the value of ω that gave the least mean square error (MSE) over the in-sample data. Plots of the mean square error as a function of ω that result are presented in Figure 7 (middle). The values of ω selected as optimal for the three smooths were found to be

$$0.53\,,\quad 0.99\,,\quad 0.99\,.$$

The value of $\omega \approx 1$ means that the double and triple exponential smoothing methods are concluding that the data appears best fit by a global linear (quadratic) function.

We next introduce the data points with the indices $61 - 72$. In Figure 7 (left) this data and the one-step-ahead value $\hat{z}_n(1)$ are also plotted. From this plot the predictions made by *simple* exponential smoothing appear best. If we compute the one-step-ahead prediction errors for the three models we find

```
   98.4267   90.8914   99.0962
   40.8526   73.2035   80.1411
   14.9368   60.8603   66.7064
   25.0207   72.7566   77.6324
  -70.5646  -16.5911  -12.8080
   92.1087  109.8409  113.4230
   41.4600   95.7459   97.8521
  -74.7371   -7.0782   -6.3045
  -39.1319   -9.8554   -9.3844
   61.9872   69.4237   69.6111
  131.2855  162.1182  161.2206
   74.4969  158.9518  156.0416
```

which have MSE expressions given by

$$33.012\,,\quad 71.689\,,\quad 74.436\,.$$

These show that indeed simple exponential smoothing produces the smallest value of the MSE. Finally, we present the sample autocorrelation functions, r_k, for the in-sample data an each of the methods in Figure 7 (right). In that plot it appears that for each method the sample value of r_k are well below their two sigma standard errors (drawn in red) indicating that they can be considered insignificant and we can rule out attempting to add an autoregressive error term to the model.

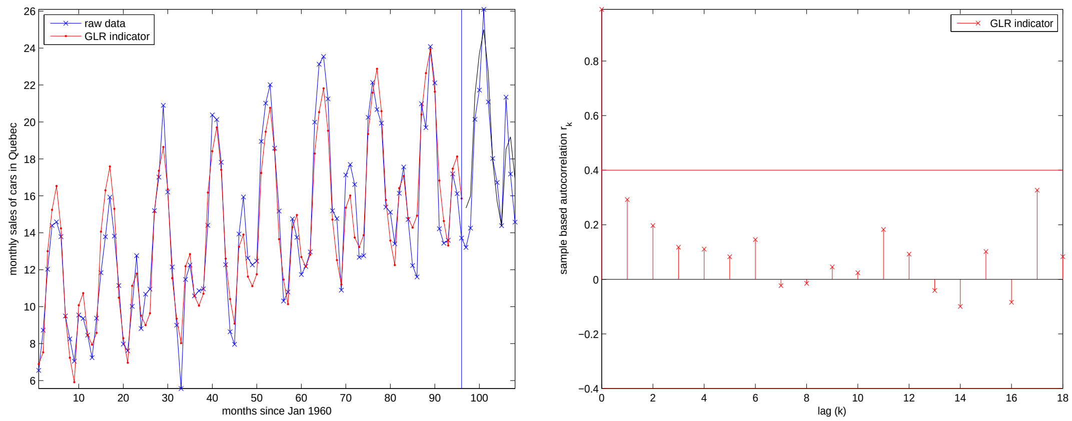


Figure 8: Duplication of Example 4.1: Quebec monthly car sales. **Left:** The raw data (in blue) and a global indicator variable model (in red and black). **Right:** A plot of the autocorrelation of the error residuals (with 2σ error bars).

Chapter 4: Regression and Exponential Smoothing Methods To Forecast Seasonal Time Series

Notes On The Text

Notes on Example 4.1: Quebec Monthly New Car Sales

In the MATLAB script `example_4_1` we attempt to duplicate the the results of this example in the book. When that script is run it loads the car data using the MATLAB function `load_monthly_car_sales`. We then generate a twelve indicator global linear model for this data of the form

$$z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{11} \delta_i \text{IND}_{ti} + \varepsilon_t \,.$$

When we run the above code we obtain values for the unknown coefficients given in Table 1.

These values agree with a similar table presented in the text. A plot of the original data and the fit obtained with the global linear model is shown in Figure 8 (left). The autocorrelations of the error residuals is plotted in Figure 8 (right).

Notes on Example 4.2: New Plant and Equipment Expenditures

In the MATLAB script `example_4_2` we attempt to duplicate the the results of this example in the book. When that script is run it loads the quarterly new plant equipment expenditure data set using the MATLAB function `load_quarterly_new_plant_equipment`. We then

Coefficient	Estimate	Standard Error	t Ratio
β_0	7.4018	0.5852	12.6481
β_1	0.0881	0.0053	16.4795
δ_1	-0.6051	0.7223	-0.8377
δ_2	-0.0502	0.7219	-0.0695
δ_3	5.3391	0.7216	7.3994
δ_4	7.4884	0.7212	10.3830
δ_5	8.6903	0.7209	12.0544
δ_6	6.3089	0.7207	8.7542
δ_7	1.4103	0.7204	1.9575
δ_8	-0.8702	0.7203	-1.2081
δ_9	-2.2845	0.7201	-3.1724
δ_{10}	1.7893	0.7200	2.4850
δ_{11}	2.3608	0.7200	3.2791

Table 1: Duplication of Example 4.1: Seasonal regression of Quebec monthly car sales using global linear model with seasonal *indicator* functions.

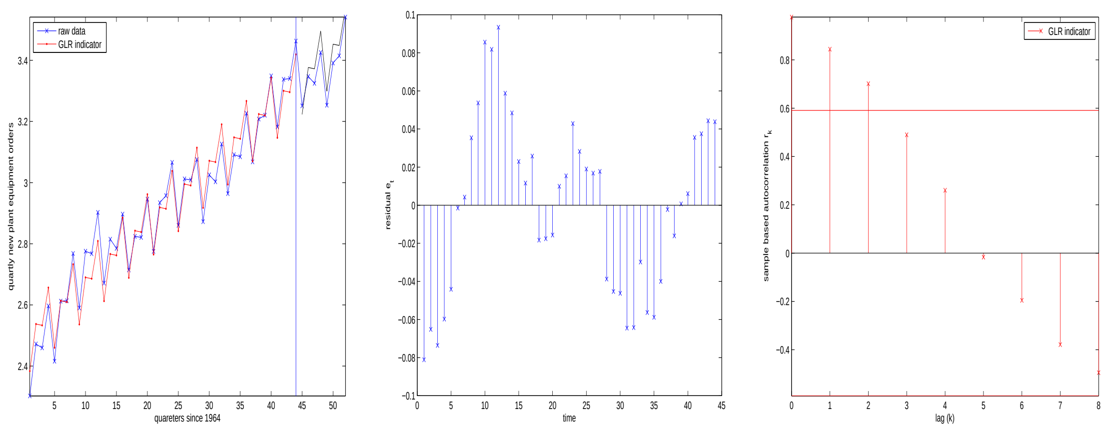


Figure 9: Duplication of Example 4.2: New plant and equipment expenditures modeled with a global linear trend with seasonal indicator functions. **Left:** The raw data (in blue) and a global indicator variable model (in red and black). **Middle:** The raw data and a global indicator variable model. **Right:** A plot of the sample autocorrelation of the error residuals (with error bars). Note that several values of r_k are significant indicating that further modeling improvements are possible, possibly with the inclusion of an AR model of the residuals.

Coefficient	Estimate	Standard Error	t Ratio
β_0	2.5777	0.0199	129.3897
β_1	0.0191	0.0006	33.4547
δ_1	-0.2138	0.0205	-10.4255
δ_2	-0.0792	0.0205	-3.8675
δ_3	-0.1026	0.0204	-5.0183

Table 2: Duplication of Example 4.2: A global indicator function model for the new plant and equipment expenditures data set. These numbers are slightly different than the ones presented in the text in that they were produced using the *entire* data set rather than a subset of size $n = 44$.

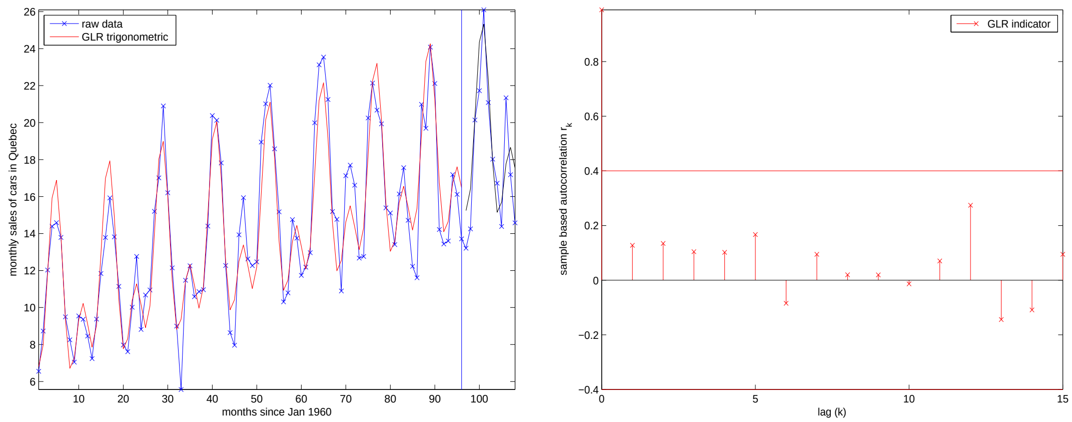


Figure 10: Duplication of Example 4.3: Quebec monthly car sales with a global linear *trigonometric* model. **Left:** The raw data (in blue) and a global indicator variable model (in red and black). **Right:** A plot of the autocorrelation of the error residuals (with error bars). Note the very large value of the $k = 12$ element of r_k.

generate a four ($s = 4$) indicator global linear model for the *log* of this data of the form

$$\log(z_t) = \beta_0 + \beta_1 t + \sum_{i=1}^{3} \delta_i \text{IND}_{ti} + \varepsilon_t \,.$$

When we run the above code we obtain values for the unknown coefficients given in the Table 2

Notes on Example 4.3: Another Look at Car Sales

In the MATLAB script `example_4_3` we attempt to duplicate the the results of this example in the book. When that script is run it loads the car data using the MATLAB function `load_monthly_car_sales`. We then generate a *two* frequency global linear model for this

Coefficient	Estimate	Standard Error	t Ratio
β_0	9.8750	0.3203	30.8338
β_1	0.0879	0.0057	15.3086
β_{11}	2.5704	0.2247	11.4408
β_{21}	-2.6624	0.2237	-11.9005
β_{12}	-2.9511	0.2239	-13.1826
β_{22}	0.8341	0.2237	3.7285

Table 3: Duplication of Example 4.3: A global linear model with trigonometric seasonal components for the Quebec monthly car sales data set. These results match quite well the ones given in Table 4.4 in the book.

data set of the form

$$z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{2} \left(\beta_{1i} \sin(\frac{2\pi i}{12} t) + \beta_{2i} \cos(\frac{2\pi i}{12} t) \right) + \varepsilon_t \,.$$

When we run the above code we obtain values for the unknown coefficients given in the Table 3. These values agree with a similar table presented in the text. A plot of the original data and the fit obtained with the global linear model can be seen in Figure 10 (left). The autocorrelations of the error residuals are plotted in Figure 10 (right). Note the relatively large value at the $k = 12$ lag. This indicates that the seasonal model can further be improved by adding terms that include higher order harmonics.

Notes on Section 4.3: Locally Constant Seasonal Models

In this section of these notes we attempt to duplicate several of the numerical studies presented in the section on *locally* constant seasonal models. We should note that when we consider locally constant seasonal models we initialize our estimate of β_n the local regression coefficients for $n = 0$ by evaluating a polynomial fit over the entire range or some subset of the data. Then the iterations used in updating the values of β_n as various samples of our time series z_n are given by

$$\hat{\beta}_{n+1} = L' \hat{\beta}_n + F^{-1} f(0)(z_{n+1} - \hat{z}_n(1)) \,, \tag{74}$$

where L is transition matrix specific to our fitting functions $f(l)$ in that $f(l) = Lf(l-1)$ and F is the steady state matrix $F = \sum_{j \geq 0} \omega^j f(-j) f'(-j)$. To begin, we let $n = 0$ in the above for which $z_{n+1} = z_1$ is our first observation and the expression $\hat{z}_n(1)$ above becomes

$$\hat{z}_n(1) = \hat{z}_0(1) = f'(1)\beta_0 \,,$$

or our initial guess at the initial value of z_1 based on the global fit that determined β_0.

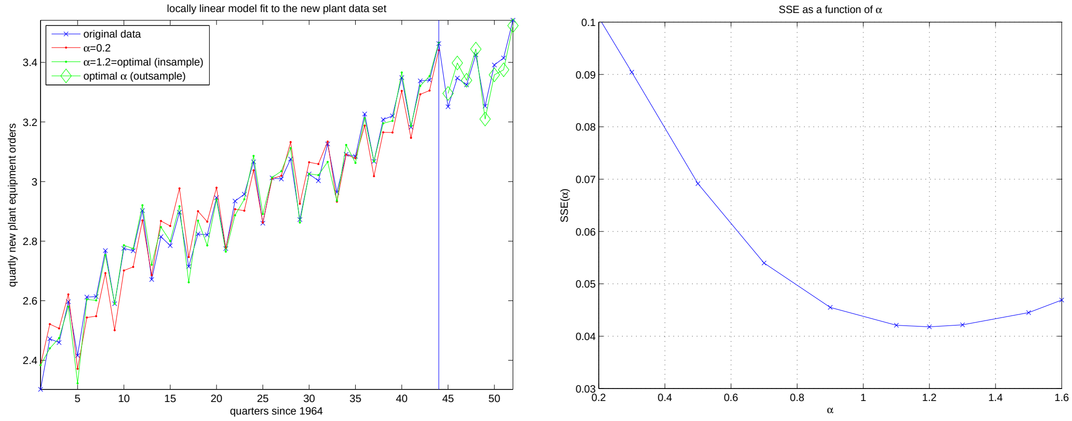


Figure 11: Duplication of a locally constant seasonal model of the new plant and equipment expenditure data set. **Left:** The raw data (in blue), a local model with an arbitrarily chosen value of the relaxation parameter $\alpha = 0.2$ (in red), and the fit produced from an optimally chosen $\alpha = 1.2$ (in green). The optimal fit is extended for $n \geq 44$ to show performance out of sample. **Right:** A plot of the sum of square one-step-ahead errors $\text{SSE}(\alpha) = \sum_n (\hat{z}_n(1) - z_{n+1})^2$ as a function of α, the relaxation parameter. The minimal value of SSE occurs at $\alpha = 1.2$.

Equipment Expenditure with Locally Constant Seasonal Indicators

In the MATLAB script `section_4_3_1` we attempt to duplicate some of the numerical results on the quarterly expenditures for new plant and equipment data set. For this example we will use a locally constant linear model with seasonal indicators. When that script is run it loads the quarterly new plant equipment expenditure data set using the MATLAB function `load_quarterly_new_plant_equipment`. We then model the log of the original time series as

$$z_{n+j} = \log(y_{n+j}) = \beta_0 + \beta_1 j + \sum_{i=1}^{3} \delta_i \text{IND}_{ji} + \varepsilon_{n+j}\,.$$

This model is fit in the MATLAB function `locally_constant_indicator_model` using Equation 74 to update the values for $\hat{\beta}_n$ as each sample z_n is presented. Then the MATLAB function `locally_constant_indicator_model_optimum` is used to search for an optimum value of the relaxation coefficient ω. When we run the above code we obtain values for the sum of square one-step-ahead errors given in the Table 4

These values agree quite well with table 4.6 presented in the text. A plot of the original data and the fit obtained with the local linear model can be seen in Figure 11 (left) a vertical line is drawn to represent the in-sample vs. out-sample domains. The plot of $\text{SSE}(\alpha)$ vs. α is plotted in Figure 11 (right). This later plot agrees with the one presented in the text.

α	SSE(α)
0.1000	0.1114
0.3000	0.0904
0.5000	0.0691
0.7000	0.0540
0.9000	0.0455
1.1000	0.0421
1.2000	0.0418
1.3000	0.0422
1.5000	0.0445
1.6000	0.0469

Table 4: A table of α vs. SSE(α) for the new plant and equipment expenditures example of regression with local seasonal indicator functions. This table duplicates the results from table 4.6 in the book.

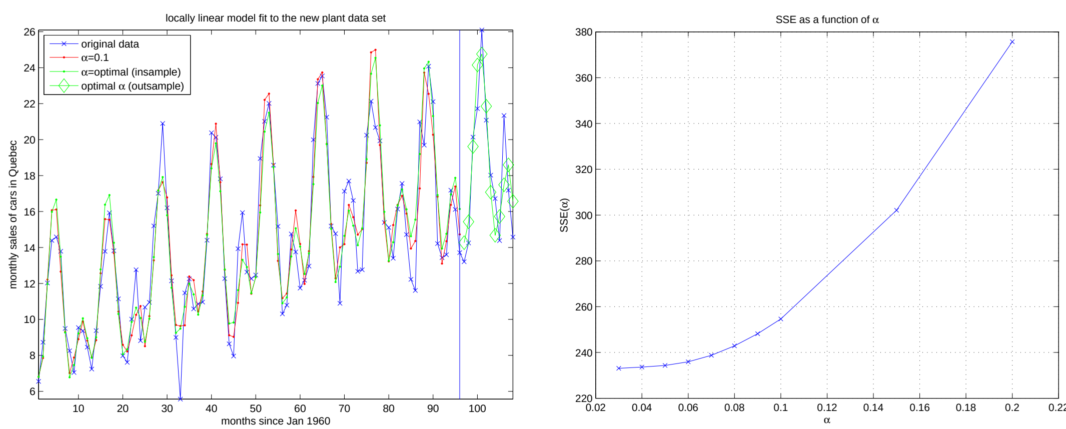


Figure 12: Duplication of a locally constant seasonal model using trigonometric functions for monthly Quebec cars sales data set.
Left: The raw data (in blue), a local model with an arbitrarily chosen value of the relaxation parameter $\alpha = 0.1$ (in red), and the fit produced from an optimally chosen $\alpha = 0.03$ (in green). The optimal fit is extended for $n \geq 96$ to show performance out of sample. **Right:** A plot of the sum of square one-step-ahead errors $\text{SSE}(\alpha) = \sum_n (\hat{z}_n(1) - z_{n+1})^2$ as a function of α, the relaxation parameter. The minimal value of SSE occurs at $\alpha = 0.03$.

Equipment Expenditure with Locally Constant Trigonometric Functions

In this subsection we duplicate the results from the book on locally constant seasonal methods using trigonometric functions applied specifically to the the monthly Quebec car sales data set. In the MATLAB script `section_4_3_2` we begin by loading this data set using the function `load_monthly_car_sales`. We then use the function `locally_constant_trigonometric_model` to construct local linear models with trigonometric seasonal indicators of the form

$$z_{n+j} = \beta_0 + \beta_1 j + \sum_{i=1}^{2} (\beta_{1i} \sin(f_i j) + \beta_{2i} \cos(f_i j)) + \varepsilon_t \,.$$

with $f_i = \frac{2\pi i}{12}$. We use the function `locally_constant_trigonometric_model_optimum` to find the optimal value of α. Results of this experiment are presented in Figure 12. which duplicate the results presented in the book quite nicely.

Exercise Solutions

Exercise 4.1 (specifying transition matrices L such that $f(j) = Lf(j-1)$)

We want to specify the matrix L such that

$$f(j) = Lf(j-1) \,, \tag{75}$$

for the given local fitting vectors $f(j)$. Note that if we expand $f(j-1)$ and write it as a linear combination of the elements of $f(j)$ as

$$f(j-1) = Mf(j) \quad \text{so} \quad f(j) = M^{-1} f(j-1) \,,$$

then the matrix L we desire is given by M^{-1}.

Part (a): Consider the expression $f(j-1)$ in this case. We find

$$\begin{aligned} f(j-1) &= \begin{bmatrix} 1 \\ \sin(\frac{\pi}{2}(j-1)) \\ \cos(\frac{\pi}{2}(j-1)) \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ \sin(\frac{\pi}{2}j)\cos(\frac{\pi}{2}) - \cos(\frac{\pi}{2}j)\sin(\frac{\pi}{2}) \\ \cos(\frac{\pi}{2}j)\cos(\frac{\pi}{2}) - \sin(\frac{\pi}{2}j)\sin(\frac{\pi}{2}) \end{bmatrix} = \begin{bmatrix} 1 \\ -\cos(\frac{\pi}{2}j) \\ \sin(\frac{\pi}{2}j) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \sin(\frac{\pi}{2}j) \\ \cos(\frac{\pi}{2}j) \end{bmatrix} \,. \end{aligned}$$

Since we have shown that $f(j-1) = Mf(j)$ for the matrix M defined implicitly above we have that $f(j) = M^{-1}f(j-1)$, thus computing this inverse we find

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \,.$$

Part (b): Consider the expression $f(j-1)$ in this case. Using some of the results from Part (a) we find that

$$f(j-1) = \begin{bmatrix} 1 \\ j-1 \\ -\cos(\frac{\pi}{2}j) \\ \sin(\frac{\pi}{2}j)) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ j \\ \sin(\frac{\pi}{2}(j-1)) \\ \cos(\frac{\pi}{2}(j-1)) \end{bmatrix} .$$

Taking the inverse of the matrix M (defined implicitly above) we find

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} .$$

Part (c): Consider $f(j-1)$ in this case. Using the results from Part (a) and Part (b) above we only need to evaluate the two new terms

$$j\sin(\frac{\pi}{2}j) \quad \text{and} \quad j\cos(\frac{\pi}{2}j) .$$

We find that

$$(j-1)\sin(\frac{\pi}{2}(j-1)) = (j-1)\left(-\cos(\frac{\pi}{2}j)\right) = -j\cos(\frac{\pi}{2}j) + \cos(\frac{\pi}{2}j) ,$$

and

$$(j-1)\cos(\frac{\pi}{2}(j-1)) = (j-1)\left(\sin(\frac{\pi}{2}j)\right) = j\sin(\frac{\pi}{2}j) - \sin(\frac{\pi}{2}j) ,$$

Thus $f'(j-1)$ becomes

$$\begin{aligned} f'(j-1) &= \begin{bmatrix} 1 \\ j-1 \\ -\cos(\frac{\pi}{2}j) \\ \sin(\frac{\pi}{2}j) \\ -j\cos(\frac{\pi}{2}j) + \cos(\frac{\pi}{2}j) \\ j\sin(\frac{\pi}{2}j) - \sin(\frac{\pi}{2}j) \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ j \\ \sin(\frac{\pi}{2}j) \\ \cos(\frac{\pi}{2}j) \\ j\sin(\frac{\pi}{2}j) \\ j\cos(\frac{\pi}{2}j) \end{bmatrix} . \end{aligned}$$

Thus taking the inverse of the matrix M defined implicitly above we find

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 & 0 \end{bmatrix} .$$

Part (d): Consider $f(j-1)$ in this case. Using the results from above we only need to evaluate the terms

$$\sin(\frac{2\pi}{12}j)\,, \quad \cos(\frac{2\pi}{12}j)\,, \quad \sin(\frac{4\pi}{12}j)\,, \quad \text{and} \quad \cos(\frac{4\pi}{12}j)\,,$$

evaluated at $j-1$. We find that

$$\begin{aligned}
\sin(\frac{2\pi}{12}(j-1)) &= \sin(\frac{2\pi}{12}j)\cos(\frac{2\pi}{12}) - \cos(\frac{2\pi}{12}j)\sin(\frac{2\pi}{12}) \\
&= \sin(\frac{2\pi}{12}j)\cos(\frac{\pi}{6}) - \cos(\frac{2\pi}{12}j)\sin(\frac{\pi}{6}) \\
&= \frac{\sqrt{3}}{2}\sin(\frac{2\pi}{12}j) - \frac{1}{2}\cos(\frac{2\pi}{12}j)\,,
\end{aligned}$$

and

$$\begin{aligned}
\cos(\frac{2\pi}{12}(j-1)) &= \cos(\frac{2\pi}{12}j)\cos(\frac{2\pi}{12}) + \sin(\frac{2\pi}{12}j)\sin(\frac{2\pi}{12}) \\
&= \frac{\sqrt{3}}{2}\cos(\frac{2\pi}{12}j) + \frac{1}{2}\sin(\frac{2\pi}{12}j)\,,
\end{aligned}$$

and

$$\begin{aligned}
\sin(\frac{4\pi}{12}(j-1)) &= \sin(\frac{4\pi}{12}j)\cos(\frac{\pi}{3}) - \cos(\frac{4\pi}{12}j)\sin(\frac{\pi}{3}) \\
&= \frac{1}{2}\sin(\frac{4\pi}{12}j) - \frac{\sqrt{3}}{2}\cos(\frac{4\pi}{12}j)\,,
\end{aligned}$$

with finally

$$\begin{aligned}
\cos(\frac{4\pi}{12}(j-1)) &= \cos(\frac{4\pi}{12}j)\cos(\frac{\pi}{3}) + \sin(\frac{4\pi}{12}j)\sin(\frac{\pi}{3}) \\
&= \frac{1}{2}\cos(\frac{4\pi}{12}j) + \frac{\sqrt{3}}{2}\sin(\frac{4\pi}{12}j)\,.
\end{aligned}$$

Thus $f(j-1)$ becomes

$$f(j-1) = \begin{bmatrix} 1 \\ j-1 \\ \frac{\sqrt{3}}{2}\sin(\frac{2\pi}{12}j) - \frac{1}{2}\cos(\frac{2\pi}{12}j) \\ \frac{\sqrt{3}}{2}\cos(\frac{2\pi}{12}j) + \frac{1}{2}\sin(\frac{2\pi}{12}j) \\ \frac{1}{2}\sin(\frac{4\pi}{12}j) - \frac{\sqrt{3}}{2}\cos(\frac{4\pi}{12}j) \\ \frac{1}{2}\cos(\frac{4\pi}{12}j) + \frac{\sqrt{3}}{2}\sin(\frac{4\pi}{12}j) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ j \\ \sin(\frac{2\pi}{12}j) \\ \cos(\frac{2\pi}{12}j) \\ \sin(\frac{4\pi}{12}j) \\ \cos(\frac{4\pi}{12}j) \end{bmatrix}.$$

Taking the inverse of the matrix M defined implicitly above we find

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Rather than computing the inverses of these matrices by hand one can use the MATLAB `inv` function with the format specifier `rat`. See the MATLAB script `prob_4_1` for an example of this.

Exercise 4.2 (general exponential smoothing of a trigonometric model)

Part (a): The transition matrix for this specification of fitting functions $f(j)$ is given in Exercise 4.1. Explicitly we have

$$
\begin{aligned}
f(j-1) &= \begin{bmatrix} 1 \\ \sin(\frac{2\pi}{12}(j-1)) \\ \cos(\frac{\pi}{2}(j-1)) \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{\sqrt{3}}{2}\sin(\frac{2\pi}{12}j) - \frac{1}{2}\cos(\frac{2\pi}{12}j) \\ \frac{\sqrt{3}}{2}\cos(\frac{2\pi}{12}j) + \frac{1}{2}\sin(\frac{2\pi}{12}j) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ \sin(\frac{2\pi}{12}j) \\ \cos(\frac{2\pi}{12}j) \end{bmatrix} .
\end{aligned}
$$

Since we have shown that $f(j-1) = Mf(j)$ for the matrix M defined implicitly above, taking its inverse on both sides we have that $f(j) = M^{-1}f(j-1)$. Thus computing this inverse we find the transition matrix L given by

$$
L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} .
$$

Part (b): The updating equations are derived in the section entitled: "Discounted Least Squares and general Exponential Smoothing", where one minimized over β as each sample z_n we observed the discounted least squares problem

$$
\sum_{j=0}^{n-1} \omega^j [z_{n-j} - f'(-j)\beta]^2 . \tag{76}
$$

This gave estimates $\hat{z}_n(l)$ for the l-step-ahead given by

$$
\hat{z}_n(l) = f'(l)\hat{\beta}_n , \tag{77}
$$

where the local expression for $\hat{\beta}$ is given by $\hat{\beta}_n = F_n^{-1}h_n$ where

$$
F_n = \sum_{j=0}^{n-1} \omega^j f(-j)f'(-j) \tag{78}
$$

$$
h_n = \sum_{j=0}^{n-1} \omega^j f(-j)z_{n-j} . \tag{79}
$$

Part (b): The results on variances of the l-step-ahead forecasts we show that

$$
V(e_n(l)) = \sigma^2 + f'(l)V(\hat{\beta}_n)f(l) ,
$$

with

$$
V(\hat{\beta}_n) = \sigma^2 F_n^{-1} \left(\sum_{j=0}^{n-1} \omega^{2j} f(-j)f'(-j) \right) F_n^{-1} .
$$

In the case given here with this specific fitting functions $f(j)$ we have the needed outer product

$$
\begin{aligned}
f(-j)f'(-j) &= \begin{bmatrix} 1 \\ -\sin(\frac{2\pi}{12}j) \\ \cos(\frac{2\pi}{12}j) \end{bmatrix} \begin{bmatrix} 1 & -\sin(\frac{2\pi}{12}j) & \cos(\frac{2\pi}{12}j) \end{bmatrix} \\
&= \begin{bmatrix} 1 & -\sin(\frac{2\pi}{12}j) & \cos(\frac{2\pi}{12}j) \\ -\sin(\frac{2\pi}{12}j) & \sin(\frac{2\pi}{12}j)^2 & -\sin(\frac{2\pi}{12}j)\cos(\frac{2\pi}{12}j) \\ \cos(\frac{2\pi}{12}j) & -\sin(\frac{2\pi}{12}j)\cos(\frac{2\pi}{12}j) & \cos(\frac{2\pi}{12}j)^2 \end{bmatrix} .
\end{aligned}
$$

If we multiply by ω^j or ω^{2j} depending on whether we are evaluating F_n or $V(\hat{\beta})$ and take the limit $n \to \infty$ we can use the results in Table 4.5 of the book which gives explicit representations of sums of trigonometric functions as above. Denoting this limit in n as F, i.e. $F = \sum_{j=0}^{\infty} \omega^j f(-j)f'(-j)$ we find the $(1,1)$ of F given by

$$F_{11} = \sum_{j=0}^{\infty} \omega^j = \frac{1}{1-\omega} .$$

The $(1,2)$ and $(2,1)$ components of F are given by

$$\sum_{j=0}^{\infty} \omega^j \sin(\frac{2\pi}{12}j) = \frac{\omega \sin(\frac{2\pi}{12})}{g_1} .$$

The $(1,3)$ and $(3,1)$ components of F are given by

$$\sum_{j=0}^{\infty} \omega^j \cos(\frac{2\pi}{12}j) = \frac{1-\omega \cos(\frac{2\pi}{12})}{g_1} .$$

The $(2,2)$ component of F is given by

$$\sum_{j=0}^{\infty} \omega^j \sin(\frac{2\pi}{12}j)^2 = \frac{1}{2}\left[\frac{1-\omega}{g_2} - \frac{1-\omega\cos(\frac{4\pi}{12})}{g_3}\right] .$$

The $(2,3)$ and $(3,2)$ components of F are given by

$$-\sum_{j=0}^{\infty} \omega^j \sin(\frac{2\pi}{12}j)\cos(\frac{2\pi}{12}j) = -\frac{1}{2}\left(\frac{\omega\sin(\frac{4\pi}{12})}{g_3}\right) .$$

Finally the $(3,3)$ component of F is given by

$$\sum_{j=0}^{\infty} \omega^j \cos(\frac{2\pi}{12}j)^2 = \frac{1}{2}\left[\frac{1-\omega}{g_2} + \frac{1-\omega\cos(\frac{4\pi}{12})}{g_3}\right] .$$

Where the expressions for g_1, g_2, and g_3 are given by

$$
\begin{aligned}
g_1 &= 1 - 2\omega\cos(\frac{2\pi}{12}) + \omega^2 \\
g_2 &= 1 - 2\omega + \omega^2 = (1-\omega)^2 \\
g_3 &= 1 - 2\omega\cos(\frac{4\pi}{12}) + \omega^2 .
\end{aligned}
$$

The steps needed to evaluate $\sum_{j=0}^{\infty} \omega^{2j} f(-j) f'(-j)$ are the same as to evaluate F and can be obtained from the manipulations just performed by replacing ω with ω^2. Once we have expressions for F and $\sum_{j=0}^{\infty} \omega^{2j} f(-j) f'(-j)$ we invert the former and multiply everything together to obtain $V(\hat{\beta})$ from which we can easily compute $V(e_n(l))$.

Exercise 4.3 (general exponential smoothing of a seasonal linear trend)

The local regression model we wish to consider is given by

$$z_{n+j} = \beta_0 + \beta_1 j + \delta \mathrm{IND}_j + \varepsilon_{n+j} .$$

Part (a): In this case the rows of the design matrix look like

$$f'(t) = (1, t, \mathrm{IND}_t) .$$

Then to find the transition matrix L such that $f(j) = L f(j-1)$, note that

$$\mathrm{IND}_j = \begin{cases} 1 & \text{if } j \text{ is odd} \\ 0 & \text{if } j \text{ is even} \end{cases} .$$

First assume that j is *even*. Then $j-1$ is odd and we have

$$f(j) = \begin{bmatrix} 1 \\ j \\ 0 \end{bmatrix} \quad \text{and} \quad f(j-1) = \begin{bmatrix} 1 \\ j-1 \\ \mathrm{IND}_{j-1} \end{bmatrix} = \begin{bmatrix} 1 \\ j-1 \\ 1 \end{bmatrix} ,$$

while if j is *odd* then we have

$$f(j) = \begin{bmatrix} 1 \\ j \\ 1 \end{bmatrix} \quad \text{and} \quad f(j-1) = \begin{bmatrix} 1 \\ j-1 \\ 0 \end{bmatrix} .$$

Both of these cases can be satisfied with a matrix M defined such that $f(j-1) = M f(j)$ as

$$M = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} ,$$

so that the desired matrix L is the inverse of this M and is given by

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix} .$$

Part (b): We begin with the initial value of our fitting functions $f(0)$ seen to be equal to

$$f(0) = \begin{bmatrix} 1 \\ 0 \\ \mathrm{IND}_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} .$$

Next using Equation 78 taking the limit as $n \to \infty$ to get F we need to evaluate

$$F = \sum_{j=0}^{\infty} \omega^j f(-j) f'(-j) \, ,$$

for the specified fitting functions $f(j)$. Since the outer product, $f(-j)f'(-j)$, is given by

$$\begin{aligned}
f(-j)f'(-j) &= \begin{bmatrix} 1 \\ -j \\ \text{IND}_{-j} \end{bmatrix} \begin{bmatrix} 1 & -j & \text{IND}_{-j} \end{bmatrix} \\
&= \begin{bmatrix} 1 & -j & \text{IND}_{-j} \\ -j & j^2 & -j\text{IND}_{-j} \\ \text{IND}_{-j} & -j\text{IND}_{-j} & \text{IND}_{-j}^2 \end{bmatrix} .
\end{aligned}$$

If j is even this matrix becomes

$$f(-j)f'(-j) = \begin{bmatrix} 1 & -j & 0 \\ -j & j^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} ,$$

while if j is odd then this matrix becomes

$$f(-j)f'(-j) = \begin{bmatrix} 1 & -j & 1 \\ -j & j^2 & -j \\ 1 & -j & 1 \end{bmatrix} .$$

When we multiply the matrix $f'(-j)f(-j)$ by ω^j and sum as required to calculate F some of the sums we will need to evaluate are given as

$$\sum_{j=0}^{\infty} \omega^j = \frac{1}{1-\omega} \, , \quad \sum_{j=0}^{\infty} j\omega^j = \frac{\omega}{(1-\omega)^2} \, , \quad \sum_{j=0}^{\infty} j^2\omega^j = \frac{\omega(1+\omega)}{(1-\omega)^3} \, .$$

These expressions can be used to directly evaluate the elements F_{11}, F_{12}, F_{21}, and F_{22} of F. To evaluate the other elements of F namely F_{13}, F_{23}, F_{31}, F_{32}, and F_{33} we need to evaluate

$$\begin{aligned}
F_{13} &= F_{31} = \sum_{j=0;j\text{odd}}^{\infty} \omega^j = \sum_{k=0}^{\infty} \omega^{2k+1} = \omega \sum_{k=0}^{\infty} (\omega^2)^k = \frac{\omega}{1-\omega^2} \\
F_{23} &= F_{32} = \sum_{j=0;j\text{odd}}^{\infty} -j\omega^j = -\sum_{k=0}^{\infty} (2k+1)\omega^{2k+1} \\
&= -\omega \left[2 \sum_{k=0}^{\infty} k\omega^{2k} + \sum_{k=0}^{\infty} \omega^{2k} \right] = -\omega \left[2 \frac{\omega^2}{(1-\omega^2)^2} + \frac{1}{1-\omega^2} \right] \\
&= -\omega \left[\frac{1+\omega^2}{(1-\omega^2)^2} \right] \\
F_{33} &= \sum_{j=0;j\text{odd}}^{\infty} \omega^j = \frac{\omega}{1-\omega^2} \, .
\end{aligned}$$

Then F as a matrix looks like

$$F = \begin{bmatrix} \frac{1}{1-\omega} & -\frac{\omega}{(1-\omega)^2} & \frac{\omega}{1-\omega^2} \\ -\frac{\omega}{(1-\omega)^2} & \frac{\omega(1+\omega)}{(1-\omega)^3} & -\frac{\omega(1+\omega^2)}{(1-\omega^2)^2} \\ \frac{\omega}{1-\omega^2} & -\frac{\omega(1+\omega^2)}{(1-\omega^2)^2} & \frac{\omega}{1-\omega^2} \end{bmatrix}.$$

If we invert this matrix we find that $F^{-1}f(0)$ is given by

$$F^{-1}f(0) = \begin{bmatrix} 1-\omega^2 \\ \frac{1}{2}(-1+\omega)^2(1+\omega) \\ \frac{1}{2}(-1+\omega)(1+\omega)^2 \end{bmatrix}.$$

So that in the limit of $\omega \to 0$ we find that

$$f_* = \lim_{\omega\to 0} F^{-1}f(0) = \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix},$$

as we were to show.

Part (c): Since from above we have shown that $f_* = (1, \frac{1}{2}, -\frac{1}{2})'$ the update equation for $\hat{\beta}_n$ is given by

$$\begin{aligned} \hat{\beta}_{n+1} &= L'\hat{\beta}_n + \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}(z_{n+1} - \hat{z}_n(1)) \\ &= L'\hat{\beta}_n + \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}(z_{n+1} - f'(1)\hat{\beta}_n) \\ &= \left(L' - \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} f'(1)\right)\hat{\beta}_n + \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} z_{n+1}. \end{aligned}$$

Now $f(1)$ is given by

$$f(1) = \begin{bmatrix} 1 \\ 1 \\ \mathrm{IND}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

so the matrix in-front of $\hat{\beta}_n$ is given by

$$L' - \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

With this the update equations for $\hat{\beta}_{n+1}$ becomes

$$\begin{bmatrix} \hat{\beta}_{0,n+1} \\ \hat{\beta}_{1,n+1} \\ \hat{\delta}_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} \hat{\beta}_{0,n} \\ \hat{\beta}_{1,n} \\ \hat{\delta}_n \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} z_{n+1}.$$

The first equation gives $\hat{\beta}_{0,n+1} = z_{n+1}$ or that $\hat{\beta}_{0,n} = z_n$ and the problem simplifies since we explicitly know $\hat{\beta}_{0,n}$. By considering only the last two equations we have

$$\begin{aligned}
\begin{bmatrix} \hat{\beta}_{1,n+1} \\ \hat{\delta}_{n+1} \end{bmatrix} &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{1,n} \\ \hat{\delta}_n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \hat{\beta}_{0,n} + \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} z_{n+1} \\
&= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{1,n} \\ \hat{\delta}_n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} (z_{n+1} - z_n) \,.
\end{aligned}$$

To solve this later equation we recognized that it is a forced linear difference equation and can be solved by methods presented in [1]. Solving it we find that $\hat{\beta}_{1,n}$ and $\hat{\delta}_n$ are given by the expressions presented in the MATHEMATICA file `prob_4_3.nb`. Using these expressions we can then evaluate

$$\begin{aligned}
\hat{z}_n(1) &= f'(1)\hat{\beta}_n = \hat{\beta}_{0,n} + \hat{\beta}_{1,n} + \hat{\delta}_n \\
&= z_n + (z_{n-1} - z_{n-2}) = z_n + z_{n-1} + z_{n-2} \quad \text{for} \quad n > 2 \,,
\end{aligned}$$

as expected. Note that in the analytic expressions for $\hat{\beta}_{1,n}$ and $\hat{\delta}_n$, the initial conditions are not important when we consider values of $n > 2$. In the same way we can compute

$$\begin{aligned}
\hat{z}_n(2) &= f'(2)\hat{\beta}_n = \hat{\beta}_{0,n} + 2\hat{\beta}_{1,n} \\
&= z_n + (-z_{n-2} + z_n) = 2z_n - z_{n-2} \,,
\end{aligned}$$

as expected. Some of the algebra for this problem is worked in the MATHEMATICA file `prob_4_3.nb`.

Exercise 4.4 (evaluating some simple geometric sums)

We will show one result from Table 4.5 since the other results can be computed in similar ways. The sum we will evaluate is $\sum_{j=0}^{\infty} \omega^j \sin(fj)$. Using Euler's fundamental relationship of $\sin(fj) = \frac{1}{2i}(e^{fji} - e^{-fji})$, where i is the imaginary unit we have that

$$\begin{aligned}
\sum_{j=0}^{\infty} \omega^j \sin(fj) &= \frac{1}{2i} \left(\sum_{j=0}^{\infty} \omega^j e^{fji} - \sum_{j=0}^{\infty} \omega^j e^{-fji} \right) \\
&= \frac{1}{2i} \left(\sum_{j=0}^{\infty} (\omega e^{fi})^j - \sum_{j=0}^{\infty} (\omega e^{-fi})^j \right) \\
&= \frac{1}{2i} \left(\frac{1}{1 - \omega e^{fi}} - \frac{1}{1 - \omega e^{-fi}} \right) \\
&= \frac{1}{2i} \left(\frac{\omega(e^{fi} - e^{-fi})}{(1 - \omega e^{fi})(1 - \omega e^{-fi})} \right) \\
&= \frac{\omega \sin(f)}{1 - 2\omega \cos(f) + \omega^2} \,.
\end{aligned}$$

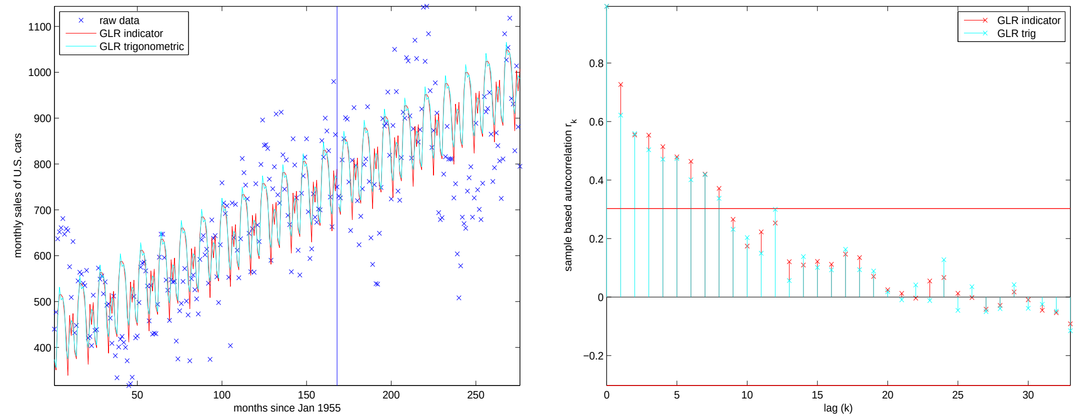


Figure 13: The sales of U.S. passenger cars fit with *global* seasonal models. The indicator function model is plotted in *red*, while the trigonometric model is plotted in cyan. **Left:** The raw data and various model predictions. **Right:** The sample autocorrelation of the one-step-ahead error residuals r_k under both models.

Exercise 4.5 (U.S. retail sales of passenger cars)

For this problem we load the suggested data into MATLAB using the script `load_monthly_us_retail_sales`. When we do this and then plot the given time series we observe the plot shown in Figure 13 (left). We will consider only the first $n = 168$ of these values and fit several different seasonal regression models.

Part (a): Since we are given monthly data we will take the number of seasons $s = 12$ (one for each month). For a globally constant linear trend model with seasonal indicators we will therefore take a model of the following form

$$z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{11} \delta_i \text{IND}_{ti} + \varepsilon_t \,. \tag{80}$$

A model of this form can be fit using ordinary least squares with the MATLAB routine `wwx_regression`, once we have specified the appropriate design matrix X. For example the first 11 rows of an appropriate design matrix X will be given by

$$X = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 6 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

Coefficient	Estimate	Standard Error	t Ratio
β_0	449.6181	27.3980	16.4106
β_1	2.0201	0.1440	14.0246
δ_1	-84.7787	34.1716	-2.4810
δ_2	-102.7274	34.1652	-3.0068
δ_3	21.2525	34.1595	0.6222
δ_4	57.0181	34.1543	1.6694
δ_5	55.3551	34.1497	1.6210
δ_6	47.0493	34.1458	1.3779
δ_7	9.6720	34.1425	0.2833
δ_8	-43.4195	34.1397	-1.2718
δ_9	-128.7968	34.1376	-3.7729
δ_10	0.1831	34.1361	0.0054
δ_11	-45.9085	34.1352	-1.3449

Table 5: Estimate of the coefficients in the global linear model with seasonal indicator functions for the U.S. sales of passenger cars data set.

and this pattern repeats for all the remaining rows of X. A matrix like this can be generated using the MATLAB function `gen_global_linear_seasonal_indicator_X` by providing it appropriate arguments. In this parametrization $t = 1$ corresponds to the month of January and the value of δ_i in the above model represents the seasonal effect of the i-th seasonal period relative to the s-th period (December). When we run the first part of the code in `prob_4_5` we obtain a table of coefficient estimate of results like that shown in Table 5. The resulting model predictions from these coefficients is overlayed onto the data points themselves and can be seen in Figure 13 (left) in red.

We next plot the sample autocorrelation function (SACF) for the residual for this model we obtain the plot shown in Figure 13 (right) in red. One thing to note is that several of the autocorrelation are significant (greater than the horizontal red line). The significant non-zero values for the SACF indicate that this model could be improved by introducing an autoregressive model for the error residuals.

Part (b): For this part of the problem we apply a global constant linear trend model with sinusoidal harmonics. As discussed in the book we will have at most $\frac{s}{2} = \frac{12}{2} = 6$ harmonics for a monthly trend model. Thus the model we propose to fit is given by

$$z_t = \beta_0 + \beta_1 t + \sum_{i=1}^{6} (\beta_{1i} \sin(f_i t) + \beta_{2i} \cos(f_i t)) ,$$

with $f_i = \frac{2\pi}{s} i = \frac{2\pi}{12} i = \frac{\pi}{6} i$. In this case the design matrix is created by calling the MATLAB function `gen_global_linear_seasonal_trigonometric_X` and the resulting coefficient estimates obtained by calling `wwx_regression`. This is implemented in the second part of the problem `prob_4_5`. When that is run it produces the cyan plot in Figure 13 (left) and corresponding sample autocorrelations in cyan in Figure 13 (right). Again notice the strong autoregressive (AR) presence in the residuals of the one-step-ahead errors.

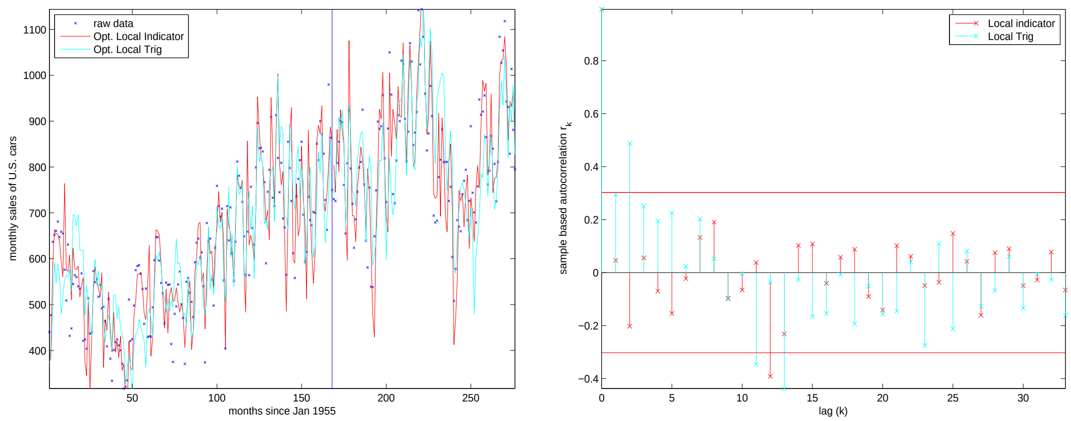


Figure 14: The sales of U.S. passenger cars fit with *local* seasonal models. The relaxation coefficient, ω, used in each case was chosen to minimize the mean square error of the in-sample data. The indicator function model is plotted in red, while the trigonometric model is plotted in cyan. **Left:** The raw data and various model predictions. **Right:** The sample autocorrelation of the one-step-ahead error residuals r_k under both models.

Part (c): In this part of the problem we implement a general exponential smoothing for the locally constant indicator version found in Part (a). This method is implemented in the the MATLAB function `locally_constant_indicator_model` which takes in an value for ω and produces the local seasonal smooth. The optimal value of ω can be found by using the MATLAB function `locally_constant_indicator_model_optimal`. Results from using these routines can be found in Figure 14 (left) and (right) in red.

Part (d): In this part of the problem we implement a general exponential smoothing for the locally constant trigonometric version of Part (b) above. This method is implemented in the the MATLAB function `locally_constant_trigonometric_model` and the optimal value for ω is found with the MATLAB function call `locally_constant_trigonometric_model_optimum`. Results from using these routines can be found in Figure 14 (left) and (right) in cyan.

If we compare the mean square error for each of the four methods we obtain the following numbers:

```
  1.9101     1.90205    0.7615    1.0797
```

from which it looks like the local indicator function performs best (have the smallest value) followed by the local trigonometric model followed by the global methods. This general trend is indicated visually in the performance plots above in that the two local methods seem to fit the data better than the two global methods.

This entire exercise is implemented in the MATLAB script `prob_4_5`.

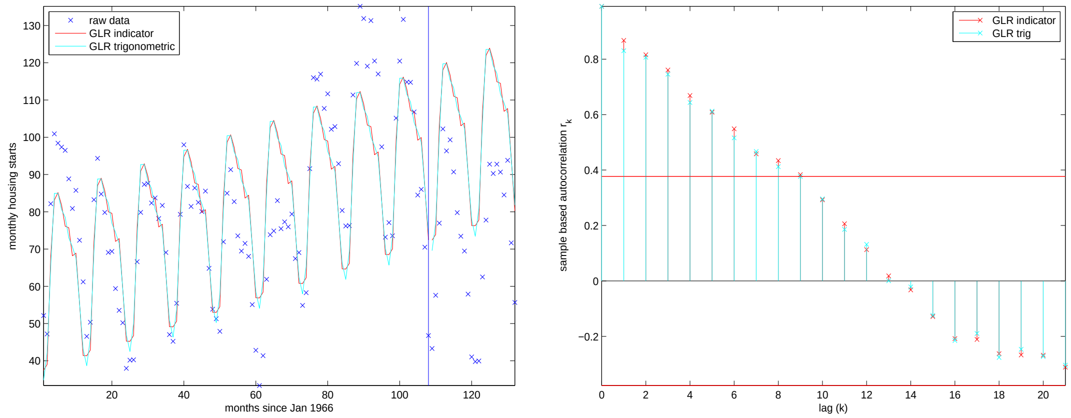


Figure 15: Data on housing starts on single family structures (in thousands) between January 1966 and December 1975 fit with *global* seasonal models. The indicator function model is plotted in red, while the trigonometric model is plotted in cyan. **Left:** The raw data and various model predictions. **Right:** The sample autocorrelation of the one-step-ahead error residuals r_k under both models.

Exercise 4.6 (predicting housing starts)

For this exercise we repeat exercise 4.5 but on the housing starts data set (series 6 in the data appendix). As in that exercise we fit both local and global seasonal models to the given data. The global models are presented in Figure 15 while the local models are presented in Figure 16. There we see that the local models do a much better job at fitting the data both in and out of sample. This can be verified when we consider the mean square error which for each of the four methods is given by

```
837.1728  830.7269   99.4340   93.3587
```

showing that indeed the local methods are superior to the global methods. It is interesting to note that the global indicator method and the global trigonometric method both produce model fits that are very close in appearance.

Exercise 4.7 (the orthogonality of global trigonometric regressions)

Part (a): The least squares estimate for the coefficients β will be given by computing $(X'X)^{-1}X'z$, where z is a vector of responses z_t where $t = 1, 2, \cdots, N$ and X is the design

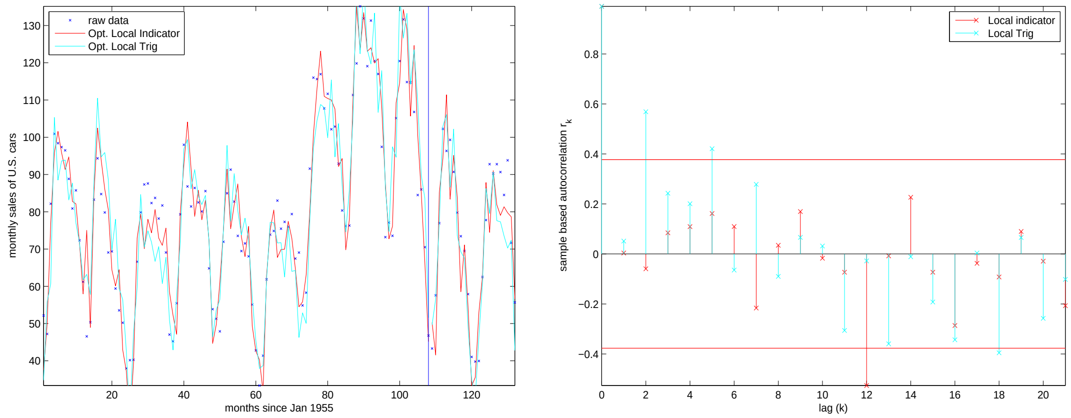


Figure 16: Data on housing starts on single family structures (in thousands) between January 1966 and December 1975 fit with *local* seasonal models. The relaxation coefficient, ω, used in each case was chosen to minimize the mean square error of the in-sample data. The indicator function model is plotted in *red*, while the trigonometric model is plotted in cyan. **Left:** The raw data and various model predictions. **Right:** The sample autocorrelation of the one-step-ahead error residuals r_k under both models. Note the strong seasonal autocorrelation at $k = 12$ in the local indicator method.

matrix which for these trigonometric fitting functions looks like

$$\left[\begin{array}{cccccccc} 1 & \sin(2\pi f_1 1) & \cos(2\pi f_1 1) & \sin(2\pi f_2 1) & \cos(2\pi f_2 1) & \cdots & \sin(2\pi f_q 1) & \cos(2\pi f_q 1) \\ 1 & \sin(2\pi f_1 2) & \cos(2\pi f_1 2) & \sin(2\pi f_2 2) & \cos(2\pi f_2 2) & \cdots & \sin(2\pi f_q 2) & \cos(2\pi f_q 2) \\ 1 & \sin(2\pi f_1 3) & \cos(2\pi f_1 3) & \sin(2\pi f_2 3) & \cos(2\pi f_2 3) & \cdots & \sin(2\pi f_q 3) & \cos(2\pi f_q 3) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \sin(2\pi f_1 N) & \cos(2\pi f_1 N) & \sin(2\pi f_2 N) & \cos(2\pi f_2 N) & \cdots & \sin(2\pi f_q N) & \cos(2\pi f_q N) \end{array}\right].$$

In the above matrix representation I have explicitly denoted the time value and not simplified the arguments of the trigonometric functions. Now the (i, j)th element of $X^T X$ is the product of the ith row of the matrix X' and jth column of the matrix X. Since row i of X' is the same thing as column i of X, we see that the (i, j)th element of $X^T X$ is the product of columns i and j of the matrix X. Thus the elements of the values in the product $X'X$ can then be determined. Computing a few elements by hand and looking for a pattern we find

$$\begin{aligned} (X'X)_{11} &= \sum_{t=1}^{N} 1 = N \\ (X'X)_{1j} &= \sum_{t=1}^{N} \sin(2\pi f_{\frac{j}{2}} t) \quad \text{for} \quad j \geq 2 \quad \text{and even} \\ (X'X)_{1j} &= \sum_{t=1}^{N} \cos(2\pi f_{\frac{j-1}{2}} t) \quad \text{for} \quad j > 2 \quad \text{and odd}\,. \end{aligned}$$

for the elements of the first row of $X'X$ (equivalently the first column of $X'X$, since $X'X$ is symmetric). For the second row of $X'X$ we then have

$$
\begin{aligned}
(X'X)_{22} &= \sum_{t=1}^{N} \sin(2\pi f_1 t)^2 = \frac{N}{2} \\
(X'X)_{2j} &= \sum_{t=1}^{N} \sin(2\pi f_1 t)\sin(2\pi f_{\frac{j}{2}} t) = 0 \quad \text{for} \quad j > 2 \quad \text{and even} \\
(X'X)_{2j} &= \sum_{t=1}^{N} \sin(2\pi f_1 t)\cos(2\pi f_{\frac{j-1}{2}} t) = 0 \quad \text{for} \quad j > 2 \quad \text{and odd}\,,
\end{aligned}
$$

for the elements of the second row of $X'X$. Continuing to the third row we have $(X'X)_{33} = \sum_{t=1}^{N} \cos(2\pi f_1 t)^2 = \frac{N}{2}$, and in the same way as before $(X'X)_{3j} = 0$ for all $j \neq 3$. If we generalized these results we see that the matrix $X'X$ is a diagonal matrix with N in the $(1,1)$ position and the value $\frac{N}{2}$ in all other diagonal positions.

Next the expression $X'z$ is a vector that has elements given by

$$
\begin{aligned}
(X'z)_1 &= \sum_{t=1}^{N} z_t \\
(X'z)_j &= \sum_{t=1}^{N} z_t \sin(2\pi f_{\frac{j}{2}} t) \quad j \quad \text{even} \\
(X'z)_j &= \sum_{t=1}^{N} z_t \cos(2\pi f_{\frac{j-1}{2}} t) \quad j \quad \text{odd}\,.
\end{aligned}
$$

Finally, inverting $X'X$ gives for the estimates of β

$$
\begin{aligned}
\hat{\beta}_0 &= \frac{1}{N}\sum_{t=1}^{N} z_t \\
\hat{\beta}_{1j} &= \frac{2}{N}\sum_{t=1}^{N} z_t \sin(2\pi f_j t) \\
\hat{\beta}_{2j} &= \frac{2}{N}\sum_{t=1}^{N} z_t \cos(2\pi f_j t)\,,
\end{aligned}
$$

as we were to show.

Chapter 5: Stochastic Time Series Models

Notes On The Text

Notes on second-order autoregressive process [AR(2)]

Using the operator notation discussed in the book, a second-order autoregressive process z_t can be written in terms of a_t as

$$z_t = (1 - \phi_1 B - \phi_2 B^2)^{-1} a_t \,,$$

here a_t is a sequence of uncorrelated random variables. If we also write this process using Wold's decomposition theorem which states that the process z_t can be written as

$$z_t = (1 + \psi_1 B + \psi_2 B^2 + \cdots) a_t = \psi(B) a_t \,,$$

for some coefficients ψ_t. For both expressions to hold true

$$(1 - \phi_1 B - \phi_2 B^2)^{-1} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \,,$$

or $1 - \phi_1 B - \phi_2 B^2$ and $\psi(B)$ must be multiplicative inverses. Thus computing this product

$$\begin{aligned}
(1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) &= 1 + \psi_1 B + \psi_2 B^2 + \cdots \\
&- \phi_1 B - \phi_1 \psi_1 B^2 - \phi_1 \psi_2 B^3 - \phi_2 \psi_2 B^4 + \cdots \\
&- \phi_2 B^2 - \phi_2 \psi_1 B^3 - \phi_2 \psi_2 B^4 + \cdots \\
&= 1 + (\psi_1 - \phi_1) B + (\psi_2 - \phi_1 \psi_1 - \phi_2) B^2 \\
&+ (\psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1) B^3 + \cdots
\end{aligned}$$

Setting the coefficients of B in this last expression equal to zero we find that the values ψ_t in the Wold decomposition in terms of the parameters of the AR(2) model are given by

$$\begin{aligned}
\psi_1 &= \phi_1 \\
\psi_2 &= \phi_1 \psi_1 + \phi_2 = \phi_1^2 + \phi_2 \\
\psi_3 &= \phi_1 \psi_2 + \phi_2 \psi_1 = \phi_1^3 + \phi_1 \phi_2 \,,
\end{aligned}$$

verifying the results in the book.

Notes on pth-order autoregressive process [AR(p)]

An AR(p) model has the following form

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_{p-1} z_{t-p+1} + \phi_p z_{t-p} + a_t \,. \tag{81}$$

Multiplying this expression by z_{t-k} on both sides and taking expectations we find

$$E(z_t z_{t-k}) = \phi_1 E(z_{k-1} z_{t-k}) + \phi_2 E(z_{t-2} z_{t-k}) + \cdots + E(z_{t-p} z_{t-k}) + E(a_t z_{t-k}) \,,$$

or recalling the definition the autocovariance (γ_k) this is equivalent to

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} + E(a_t z_{t-k}), \tag{82}$$

Recall that since z_{t-k} when written using Wold decomposition theorem only depends on $a_{t-k}, a_{t-k-1}, \cdots$ the random shocks a_t that comes later is independent of z_{t-k} if $k \neq 0$. Thus

$$E(a_t z_{t-k}) = \begin{cases} \sigma^2 & k = 0 \\ 0 & k = 1, 2, \cdots, \end{cases}$$

First when we then take $k = 0$ in Equation 82 we then get for γ_0 or the variance of our AR(p) process in terms of its parameters the following expression

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma^2 . \tag{83}$$

Dividing this by γ_0 we get recalling the definition of the autocorrelation function ρ_k of $\rho_k \equiv \frac{\gamma_k}{\gamma_0}$ that

$$1 = \phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p \rho_p + \frac{\sigma^2}{\gamma_0},$$

or

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2 - \cdots - \phi_p \rho_p}, \tag{84}$$

which is the books equation 5.26.

Next when $k > 0$ in Equation 82 the last term vanishes and we have

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \tag{85}$$

again dividing this by γ_0 we see that the autocorrelation functions ρ_k satisfy the same type of recursive difference equation as the process z_t. That is

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \tag{86}$$

When we explicitly enumerate Equations 86 for $k = 1, 2, \ldots, p$ we get (remembering $\rho_0 = 1$)

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \cdots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \cdots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \cdots + \phi_p . \end{aligned}$$

These equations are known as the **Yule-Walker** equations. This system can be written in matrix notation if we define a vector of autocorrelations $\rho = (\rho_1, \rho_2, \ldots, \rho_p)'$ and a vector of AR(p) parameters $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$, as

$$\rho = P\phi,$$

with the matrix P defined as

$$P = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}.$$

Then in this notation the parameters of the AR(p) model ϕ in terms of the autocorrelations ρ are given by $\phi = P^{-1}\rho$. With this general framework we can consider some specific cases. For an AR(2) model we get that the Yule-Walker equations give

$$\begin{aligned} \rho_1 &= \phi_1 + \rho_1\phi_2 \\ \rho_2 &= \phi_1\rho_1 + \phi_2 \,. \end{aligned}$$

In matrix form this is

$$\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} .$$

Solving for $\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$ we find

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \frac{1}{1-\rho_1^2} \begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \frac{1}{1-\rho_1^2} \begin{bmatrix} \rho_1 - \rho_1\rho_2 \\ -\rho_1^2 + \rho_2 \end{bmatrix} .$$

Thus

$$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \quad \text{and} \quad \phi_2 = \frac{-\rho_1^2+\rho_2}{1-\rho_1^2} \,,$$

in agreement with the results in the book.

Introduction of the partial autocorrelation function

Now if we express z_t as a AR(k) model as

$$z_t = \phi_{k1} z_{t-1} + \phi_{k2} z_{t-2} + \cdots + \phi_{kk} z_{t-k} + a_t \,, \tag{87}$$

we can use Crammer's rule to solve the Yule-Walker equations, $\phi = P^{-1}\rho$, for only the coefficient ϕ_{kk}. The general expression is presented in the book and specializes to the following simplifications for small k values

$$\begin{aligned} \phi_{11} &= \frac{|\rho_1|}{|1|} = \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} . \end{aligned}$$

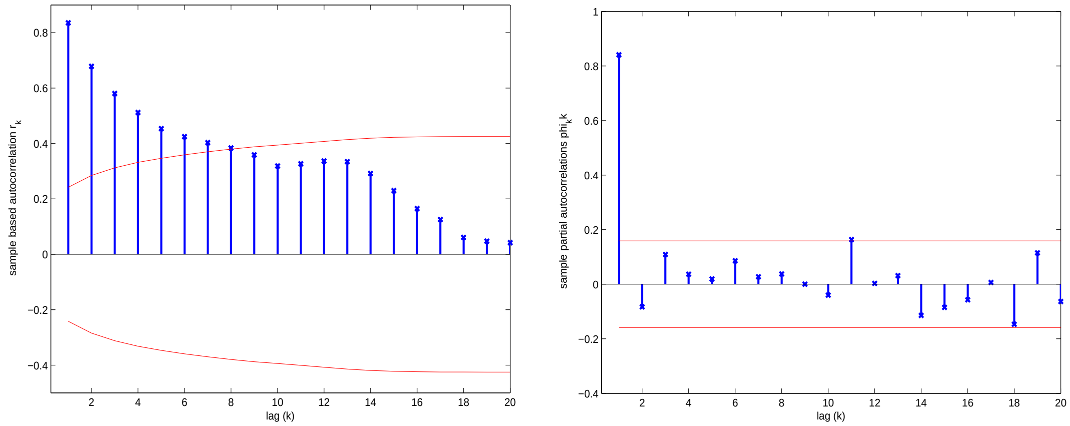


Figure 17: **Left:** The sample autocorrelation function, r_k (in blue) and confidence regions (in red) for the yield data time series. **Right:** The sample partial autocorrelation function ϕ_{kk} for the yield data time series. Note the very significant ϕ_{11} value indicating that this data set maybe fit with an AR(1) model.

The sample partial autocorrelation function for the yield data set

In the MATLAB script `section_5_2_2` we duplicate some of the results found in section 5.2.2 of the book. This script begins by presenting a plot of the sample autocorrelation function (SACF) r_k of the yield data set in Figure 17 (left). This plot agrees very well with the on presented in figure 5.2 of the book. Then using these values of r_k and the Levinson/Durbin algorithm implemented in the MATLAB function `spacf`, we compute the sample partial autocorrelation function (SPACF) of the yield data set in Figure 17 (right). These numbers presented here agree quite well with those presented in the book.

Notes on moving average processes of order q [MA(q)]

In this section we derive the autocovariance γ_k and partial autocorrelation ϕ_{kk} functions for another fundamental stochastic process; the MA(q) process. Which is defined as

$$z_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \,. \tag{88}$$

This representation is basically a truncated Wold decomposition in which we can write *any* stochastic process z_t as a series of uncorrelated random shocks a_t as

$$z_t - \mu = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \text{with} \quad \psi_0 = 1 \,. \tag{89}$$

Using the results from earlier in the book where we explicitly derived the autocovariances γ_k functions given the coefficients ψ_j in z_t Wold decomposition (duplicated here for convenience)

$$\begin{aligned}
\gamma_0 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\
\gamma_k &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \, ,
\end{aligned}$$

since $\psi_j = -\theta_j$ for $1 \le j \le q$ and $\psi_j = 0$ for $j > q$, we can easily calculate γ_k for MA(q) processes.

For example, for the MA(1) process $\psi_0 = 1$, $\psi_1 = -\theta_1$ and $\psi_j = 0$ for $j > 1$ and we find

$$\begin{aligned}
\gamma_0 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2(1+\theta_1^2) \\
\gamma_1 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2(\psi_0\psi_1 + \psi_1\psi_2 + \cdots) = -\theta_1\sigma^2 \\
\gamma_k &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = 0 \quad \text{for} \quad k > 1 \, .
\end{aligned}$$

For a MA(2) process defined as

$$z_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \, , \tag{90}$$

we see that $\psi_0 = 1$, $\psi_1 = -\theta_1$, $\psi_2 = -\theta_2$, and $\psi_j = 0$ for $j > 2$ so we find

$$\begin{aligned}
\gamma_0 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2(1+\theta_1^2+\theta_2^2) \\
\gamma_1 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2(\psi_0\psi_1 + \psi_1\psi_2) = \sigma^2(-\theta_1 + \theta_1\theta_2) \\
\gamma_2 &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+2} = \sigma^2(\psi_0\psi_2 + \psi_1\psi_3 + \cdots) = \sigma^2(-\theta_2) = -\theta_2\sigma^2 \\
\gamma_k &= 0 \quad \text{when} \quad k > 2 \, .
\end{aligned}$$

Examples forecasting with stochastic models

For developing forecasting models it is important to remember that the optimal l-step-ahead forecast is given by the following conditional expectation

$$z_n(l) = E(z_{n+l} | z_n, z_{n-1}, \ldots) \, , \tag{91}$$

and that the general expression for the forecast error $e_n(l)$ is given by

$$e_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + \psi_2 a_{n+l-2} + \cdots + \psi_{l-2} a_{n+2} + \psi_{l-1} a_{n+1} \,, \tag{92}$$

this later equation has a very simple expression for its variance. We have

$$V(e_n(l)) = \sigma^2 (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-2}^2 + \psi_{l-1}^2) \,. \tag{93}$$

With these relationships we can derive forecasts and error predictions for a number of common stochastic models.

The AR(1) Model: If we consider the AR(1) process

$$z_t - \mu = \phi(z_{t-1} - \mu) + a_t \,,$$

then to compute the two-step-ahead forecast, $z_n(2)$, we consider the conditional expectation

$$\begin{aligned} z_n(2) &= E(z_{n+2} | z_n, z_{n-1}, \dots) \\ &= E(\mu + \phi(z_{n+1} - \mu) + a_{t+2} | z_n, z_{n-1}, \dots) \\ &= \mu + \phi(z_n(1) - \mu) = \mu + \phi((\mu + \phi(z_n - \mu)) - \mu) \\ &= \mu + \phi^2 (z_n - \mu) \,. \end{aligned}$$

Next we compute the error in this prediction, $e_n(2)$, we find

$$\begin{aligned} e_n(2) &= z_{n+2} - z_n(2) \\ &= \mu + \phi(z_{n+1} - \mu) + a_{n+2} - (\mu + \phi^2 (z_n - \mu)) \\ &= a_{n+2} + \phi((z_{n+1} - \mu) - \phi(z_n - \mu)) \,, \end{aligned}$$

but $z_{n+1} - \mu = \phi(z_n - \mu) + a_{n+1}$ so the above becomes

$$e_n(2) = a_{n+2} + \phi a_{n+1} \,.$$

We can derive the general expression for $e_l(n)$ for an AR(1) model since in that case $\psi_j = \phi^j$ and Equation 92 becomes

$$e_n(l) = a_{n+l} + \phi a_{n+l-1} + \phi^2 a_{n+l-2} + \cdots + \phi^{l-2} a_{n+2} + \phi^{l-1} a_{n+1} \,,$$

so that the variance of the l-step-ahead error is thus seen to be

$$\begin{aligned} V(e_n(l)) &= \sigma^2 (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-2}^2 + \psi_{l-1}^2) \\ &= \sigma^2 \sum_{k=0}^{l-1} \phi^{2k} = \frac{\sigma^2 (1 - \phi^{2l})}{1 - \phi^2} \,, \end{aligned} \tag{94}$$

which is equation 5.68 in the book.

The ARIMA(0,1,1) Model: As another example consider an ARIMA(0,1,1) model written in its autoregressive operator form as $\pi(B) z_t = a_t$ or

$$\left(\frac{1 - B}{1 - \theta B} \right) z_t = a_t \,.$$

Expanding the function $\pi(B)$ in a Taylor series about B gives

$$\begin{aligned}
(1-B)\sum_{j=0}^{\infty}\theta^j B^j &= \sum_{j=0}^{\infty}\theta^j B^j - \sum_{j=0}^{\infty}\theta^j B^{j+1} \\
&= \sum_{j=0}^{\infty}\theta^j B^j - \sum_{j=1}^{\infty}\theta^{j-1} B^j \\
&= 1 + (1-\theta)\sum_{j=1}^{\infty}\theta^{j-1} B^j ,
\end{aligned}$$

so we have that z_t can be written

$$z_t = a_t + (1-\theta)\sum_{j=1}^{\infty}\theta^{j-1} z_{t-j} , \tag{95}$$

so that we recognize $\pi_j = (1-\theta)\theta^{j-1}$ for $j \geq 1$. Forecasts can be based off of this infinite series if needed.

The ARIMA(1,1,1) Model: As another forecasting example, next consider an ARIMA(1,1,1) process which has an general expression given by

$$(1-\phi B)(1-B)z_t = (1 - B - \phi B + \phi B^2)z_t = \theta_0 + (1-\theta B)a_t .$$

From which we see that z_t can be expressed as

$$z_t = (1+\phi)z_{t-1} - \phi z_{t-2} + \theta_0 + a_t - \theta a_{t-1} .$$

Using this the predictions are given by

$$\begin{aligned}
z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \cdots) \\
&= E((1+\phi)z_n - \phi z_{n-1} + \theta_0 + a_{n+1} - \theta a_n | z_n, z_{n-1}, \ldots) \\
&= \theta_0 + (1+\phi)z_n - \phi z_{n-1} - \theta a_n \\
z_n(2) &= E(z_{n+2}|z_n, z_{n-1}, \cdots) \\
&= E((1+\phi)z_{n+1} - \phi z_n + \theta_0 + a_{n+2} - \theta a_{n+1} | z_n, z_{n-1}, \ldots) \\
&= (1+\phi)z_n(1) - \phi z_n + \theta_0 ,
\end{aligned}$$

with $z_n(1)$ computed earlier. Note that an ARIMA(1,1,0) can be computed from this one by taking the value of $\theta = 0$.

The ARIMA(0,2,2) Model: As another forecasting example, consider an ARIMA(0,2,2) model given by

$$(1-B)^2 z_t = (1 - 2B + B^2)z_t = (1 - \theta_1 B - \theta_2 B^2)a_t .$$

From which we see that z_t can be written as

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} .$$

Then taking the condition expectation to derive the required forecasts we find

$$\begin{aligned} z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \dots) \\ &= E(2z_n - z_{n-1} + a_{n+1} - \theta_1 a_n - \theta_2 a_{n-1}|z_n, z_{n-1}, \dots) \\ &= 2z_n - z_{n-1} - \theta_1 a_n - \theta_2 a_{n-1}\,, \end{aligned}$$

we are keeping the two terms a_n and a_{n-1} since these expressions can be explicitly written in terms of z_t for $t \leq n$ when we consider the pure autoregressive representation $\pi(B)(z_t - \mu) = a_t$. The term a_{n+1} is independent of z_t for $t \leq n$ and can be discarded. For the two-steps-ahead forecast $z_n(2)$ we find

$$\begin{aligned} z_n(2) &= E(z_{n+2}|z_n, z_{n-1}, \dots) \\ &= E(2z_{n+1} - z_n + a_{n+2} - \theta_1 a_{n+1} - \theta_2 a_n|z_n, z_{n-1}, \dots) \\ &= 2z_n(1) - z_n - \theta_2 a_n\,. \end{aligned}$$

If we continue these derivations we begin to see that for general l the l-step-ahead prediction, $z_n(l)$ is related to earlier predictions as

$$z_n(l) = 2z_n(l-1) - z_n(l-2) \quad \text{for} \quad l \geq 3\,.$$

Notes on updating forecasts using stochastic models

We begin this section by recalling that the minimum mean square error (MMSE) l-step-ahead forecast can be specified in terms of ψ_j the coefficients in the linear filter representation of z_t as

$$z_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \dots\,, \tag{96}$$

this is equation 5.62 in the book. Now incrementing n in the above as $n \to n+1$ to indicate that we have observed one more sample z_n this becomes

$$z_{n+1}(l) = \psi_l a_{n+1} + (\psi_{l+1} a_n + \psi_{l+2} a_{n-1} + \cdots)\,,$$

where the term in parenthesis above we recognized as $z_n(l+1)$, or the $l+1$-step-ahead prediction from the nth observation. Thus

$$z_{n+1}(l) = \psi_l a_{n+1} + z_n(l+1)\,. \tag{97}$$

This can be expressed without reference to the value of a_{n+1}, the shock that arrives at the timestep $n+1$, by recalling the l-step-ahead forecast *error* or

$$e_n(l) = z_{n+l} - z_n(l) = a_{n+1} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1}\,. \tag{98}$$

When we take $l = 1$ in this expression we find

$$z_{n+1} - z_n(1) = a_{n+1}\,, \tag{99}$$

as an expression for a_{n+1} in terms of the error residual of the one-step-ahead error. Using this fact we arrive at a recursive forecast updating expressions entirely in terms of z_t as

$$z_{n+1}(l) = z_n(l+1) + \psi_l(z_{n+1} - z_n(1))\,, \tag{100}$$

or the books equation 5.85.

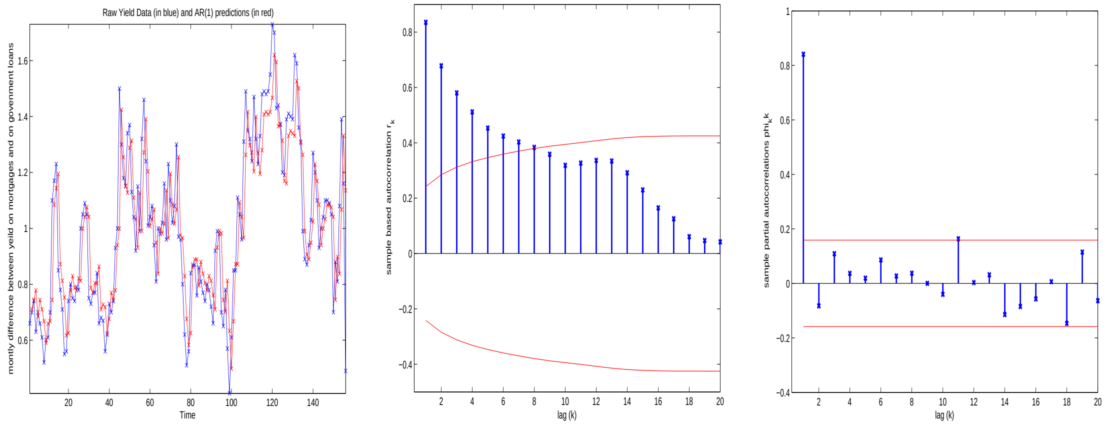


Figure 18: **Left:** The raw data (in blue) for the yield data set and the AR(1) model predictions (in red). **Middle:** The sample autocorrelation function, r_k (in blue) and confidence regions (in red) for the yield data time series. **Right:** The sample partial autocorrelation function ϕ_{kk} for the yield data time series. Note the very significant ϕ_{11} value indicating that this data set maybe fit with an AR(1) model.

Duplication of Section 5.8.1 ARIMA model of the yield data

In this section we duplicate the results on modeling the yield data set found in Section 5.8.1 from the book. See the MATLAB script `model_yield_section_5_8_1.m`. Since MATLAB does not have a free maximum likelihood estimator for ARIMA models we will compute estimate of the ARIMA(p,d,q) coefficients in the model using the R function `arima` which does exactly what is needed for these examples. We begin by presenting the yield data in Figure 18 (left), the sample autocorrelation function in Figure 18 (middle) and the sample partial autocorrelation function in Figure 18 (right). The basic exponential decay seen in the autocorrelation function in addition to the significant $k = 1$ component in the sample partial autocorrelation function indicate that an AR(1) model might be appropriate for this data set. Thus we choose to model this data with

$$z_t - \mu = \phi(z_{t-1} - \mu) + a_t \,.$$

In the Matlab script `model_yield_section_5_8_1.m` we estimate the mean of the data set z_t from the first $n = 159$ samples (leaving three samples to evaluate forecasts on). We find a value and a standard error given by

$$\hat{\mu} = 0.99(0.09)\,,$$

and thus the mean is significant. We then subtract this mean from the series and then use the R function `arima` in the R script `model_yield_section_5_8_1.R` to compute the maximum likelihood estimate of the coefficients ϕ and σ^2 (the variance of the shock a_t). When this is done we find estimates given by

$$\hat{\phi} = 0.8518(0.04) \quad \hat{\sigma}^2 = 0.024\,,$$

in close agreement with what the book computes. Using these values and the l-step-ahead predictions which for an AR(1) model like this are given by $z_n(l) = \hat{\mu} + \hat{\phi}^l(z_n - \hat{\mu})$, we

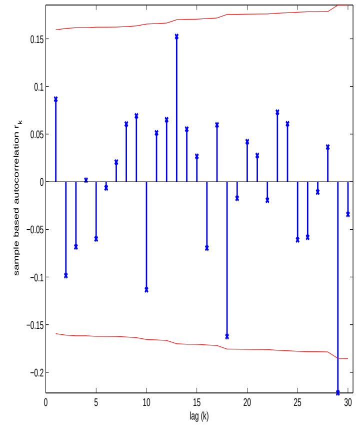


Figure 19: The sample autocorrelation function for the residuals of an AR(1) model fit to the yield data set. This plot is very similar to the one shown in the text.

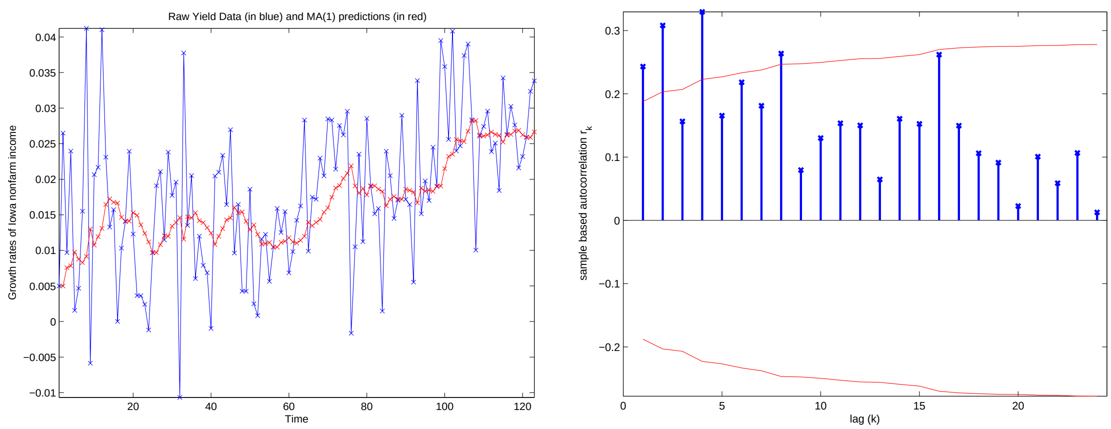


Figure 20: **Left:** The raw data (in blue) and predictions (in red) for the Iowa growth rates data set. Note that this series has a changing mean and is therefore not stationary. **Right:** The sample autocorrelation function, r_k (in blue) and confidence regions (in red) computed directly from the Iowa growth rates data time series.

can compute the prediction errors and their sample autocorrelation function. See Figure 19 where this is presented. Note that this matches very closely to a similar figure in the text.

Duplication of Section 5.8.2 Quarterly Iowa Growth Rates

In this section we duplicate the results on modeling the growth rates data set found in Section 5.8.2 from the book. See the MATLAB script `model_yield_section_5_8_2.m`. We begin by presenting the Iowa growth rates (directly) data in Figure 20 (left). We note that this data set appears to have a changing mean and therefore cannot be stationary. To observe what we would obtain if we ignored this information in Figure 20 (right) we present a plot of the sample autocorrelation computed from the direct data series. In that figure we see that when viewed in autocorrelation space we see several significant autocorrelations that don't

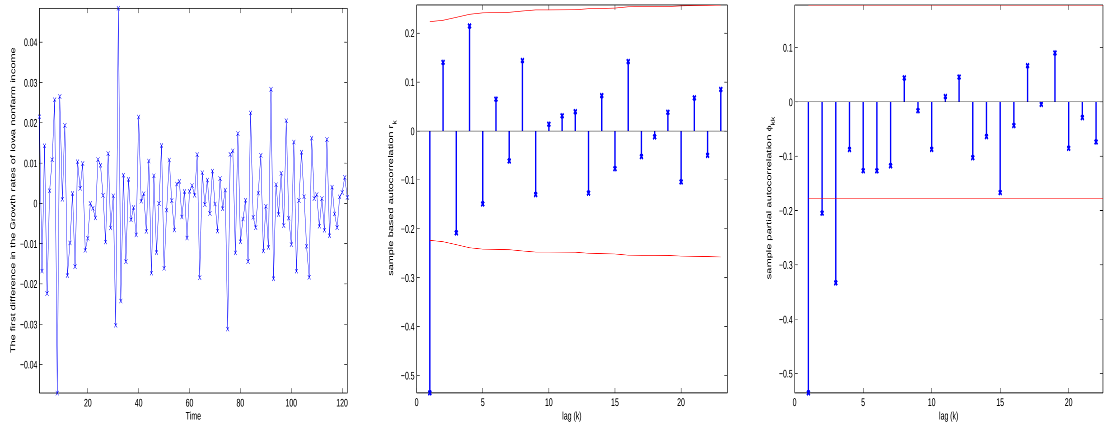


Figure 21: **Left:** A plot of the first difference of the raw Iowa growth rates data. **Middle:** The sample autocorrelation function, r_k (in blue) and confidence regions (in red) computed directly from the first difference above. **Right:** A plot of the sample partial autocorrelation function.

appear to decay with the index k. This behavior is indicative of a non-stationary sequence. To introduce stationarity we will consider differencing this series. We when we compute the sample standard deviations for the first three differences of this data we obtain

```
std(Y)=   0.010, std(diff(Y))=   0.013, std(diff(Y,2))=   0.023
```

The fact that the standard deviation increases as we take consecutive differences is indicative that we should take the first difference that gives the smallest value and so we further consider the first difference. We plot the first difference in Figure 21 (left) the sample autocorrelation function in Figure 21 (middle) and the sample partial autocorrelation function in Figure 21 (right). The basic exponential decay seen in the partial autocorrelation function in addition to the significant $k = 1$ component in the autocorrelation function indicate that an MA(1) model might be appropriate for this data set. Thus we choose to model the first difference of our original time series z_t with

$$\nabla z_t = \theta_0 + a_t - \theta a_{t-1} .$$

To compute the parameters for this model we begin by estimating the mean θ_0. In the Matlab script `model_yield_section_5_8_2.m` we estimate the mean of ∇z_t from the first $n = 122$ samples (leaving some additional samples to forecasts). We find a value and a standard error of $\hat{\theta}_0$ given by

$$\hat{\theta}_0 = 0.000226\,(0.00473)\,,$$

and thus the mean is insignificant. Next we estimate the value of the parameter θ and the variance of the shocks a_t using the R function `arima` in the R script `model_yield_section_5_8_2.R`. When this is done we find estimates given by

$$\hat{\theta} = 0.8883\,(0.04) \quad \hat{\sigma}^2 = 9.329\,10^{-5}\,,$$

which when $\hat{\sigma}^2$ is multiplied by 100^2 gives results which are in close agreement with what the book computes. Using these values and the one-step-ahead predictions for an ARIMA(0,1,1)

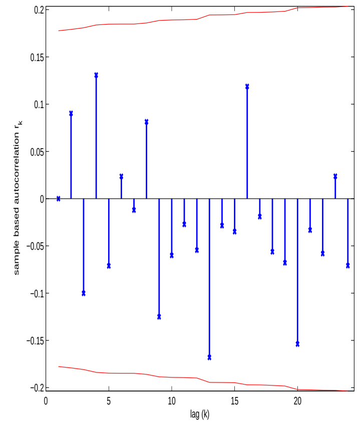


Figure 22: The sample autocorrelation function for the residuals of an ARIMA(0,1,1) model fit to the growth data set. This plot is very similar to the one shown in the text.

model like this are given by taking $\hat{a}_n = z_n - z_{n-1}(1)$ and $z_n(1) = z_n - \hat{\theta}\hat{a}_n$. We start by assuming $\hat{a}_1 = 0$ and then iterate

$$
\begin{aligned}
z_1(1) &= z_1 - \hat{\theta}_1 \hat{a}_1 = z_1 \\
\hat{a}_2 &= z_2 - z_1(1) \\
z_2(1) &= z_2 - \hat{\theta}\hat{a}_2 \\
\hat{a}_3 &= z_3 - z_2(1) \\
&\vdots \\
\hat{a}_{123} &= z_{123} - z_{122}(1)\,.
\end{aligned}
$$

In the MATLAB script for this section we use this procedure to compute estimates of the residuals $\hat{a}_n$ and their sample autocorrelation function. See Figure 22 where this is presented. Note that this matches very closely to a similar figure in the text.

Duplication of Section 5.8.3 Demand for Repair Parts

In this section we duplicate the results on modeling the repair for parts data set discussed in Section 5.8.3 from the book. See the MATLAB script `model_demand_section_5_8_3.m`. We begin by presenting the demand for parts (directly) data in Figure 23 (left). This data has a variance that is proportional (or approximately proportional) to the level z_t. Thus a logarithmic transformation will help stabilized the variance of this process. The log of the original data series is shown in Figure 23 (middle). We note that this data set appears to have a changing mean and therefore cannot be stationary. To observe what we would obtain if we ignored this information in Figure 23 (right) we present a plot of the sample autocorrelation computed from the direct data series. In that figure we see that when viewed in autocorrelation space we see several significant autocorrelations that decay quite slowly with the index k. This behavior is indicative of a non-stationary sequence. To introduce stationarity we will consider differencing this series. When we compute the sample standard deviations for the first three differences of this data we obtain

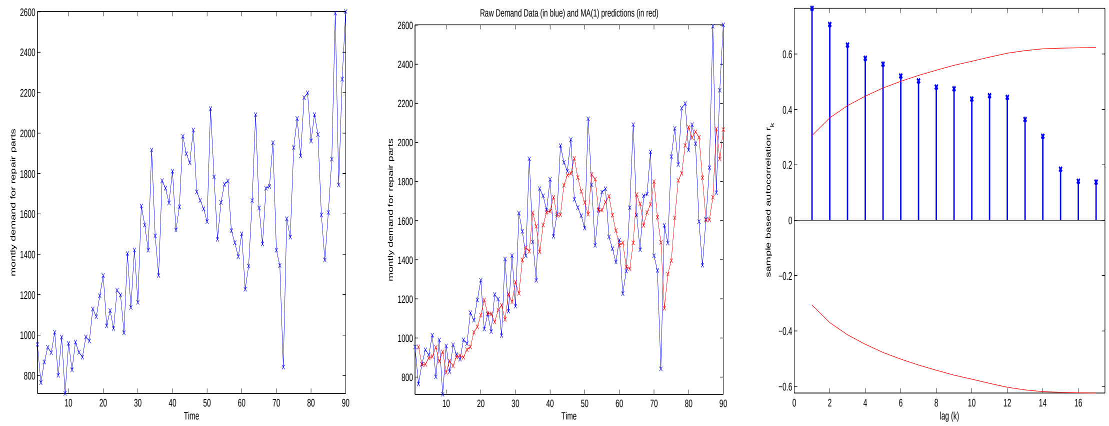


Figure 23: **Left:** The raw data (in blue) for the demand for parts data set. Note that this series has a changing mean and is therefore not stationary. In addition, it has a variance that is proportional (or approximately proportional) to the level z_t. Thus a logarithmic transformation will help stabilized the variance of this process. **Middle:** The log of the demand for parts data set (in blue) and predictions from the ARIMA model developed below. While not stationary the variance has been stabilized. **Left:** The sample autocorrelation function of this non-stationary series.

```
std(Y)= 0.30, std(diff(Y))= 0.18, std(diff(Y,2))= 0.31, std(diff(Y,3))= 0.58
```

The fact that the standard deviation increases as we take consecutive differences is indicative that we need only take the first difference to make the series stationary. We plot the first difference in Figure 24 (left) the sample autocorrelation function in Figure 24 (middle) and the sample partial autocorrelation function in Figure 24 (right). The two prominent spikes at lag $k = 1$ in both plots indicate that we may model this data well by assuming an ARMA(1,1) model for the difference in the log data. The book argues that only a MA(1) model is needed to fit this data set. For self study we attempt to fit an ARMA(1,1) model and then compare how well the ARMA(1,1) model performs relative to the MA(1) model the book uses. In the ARMA(1,1) case this means that we would model the series $y_t \equiv \log(z_t)$ as a ARIMA(1,1,1) model or

$$(1 - B)(1 - \phi B)(y_t - \mu) = (1 - \theta)a_t \,,$$

or

$$y_t - \mu = (1 + \phi)(y_{t-1} - \mu) - \phi(y_{t-2} - \mu) + a_t - \theta a_{t-1} \,.$$

To compute the parameters for this model we use the R function `arima`. When we explicitly compute the first difference of $\log(z_t)$ and use the option which requests that a mean value be included in the maximum likelihood estimates we find

$$\hat{\mu} = 0.0084 \, (0.0051)$$

and thus the mean is insignificant. In `model_demand_section_5_8_3.R` we then rerun the `arima` function this time not requesting that the mean estimated from the first difference. We estimate the value of the parameters ϕ and θ and the variance of the shocks a_t and find

$$\hat{\phi} = 0.0991 \, (0.2815) \quad \hat{\theta} = 0.64 \, (0.24) \quad \hat{\sigma}^2 = 0.025 \,,$$

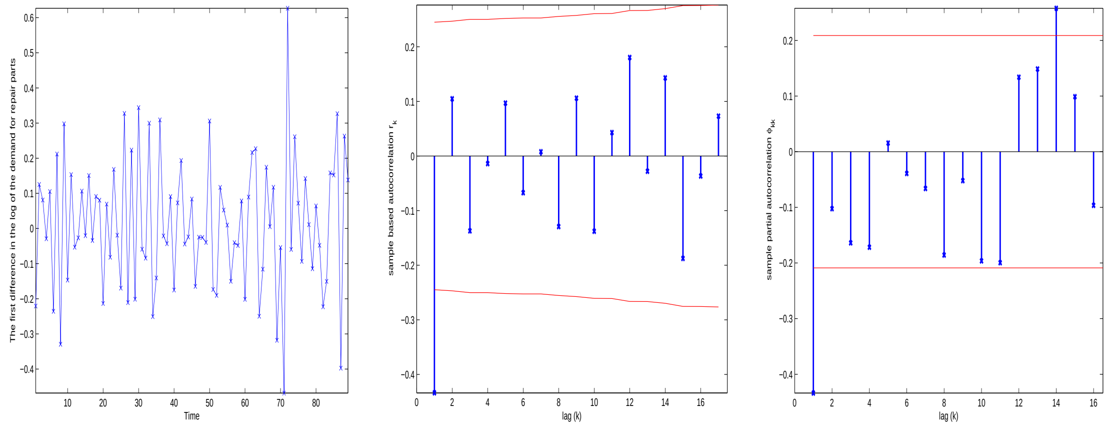


Figure 24: **Left:** A plot of the first difference of the log of the raw demand for parts data set. **Middle:** The sample autocorrelation function, r_k (in blue) and confidence regions (in red) computed directly from this data. **Right:** A plot of the sample partial autocorrelation function for this data set.

Thus we note that in fact the coefficient for ϕ is found to be insignificant and the books initial estimate of a MA(1) model for y_t was correct. Removing this parameter from the model and refitting we find that

$$\hat{\theta} = 0.55\,(0.11) \quad \hat{\sigma}^2 = 0.025\,,$$

in close agreement to the numbers the book found. For a MA(1) model the method of prediction was discussed above when we considered the previous data set. When we compute the prediction errors and their sample autocorrelation function. See Figure 25 where this is presented. Note that this matches very closely to a similar figure in the text.

Exercise Solutions

Exercise 5.1 (some sample autocorrelation functions)

Part (a): The sample autocorrelation of z_t plotted in Figure 26 (left) indicates that z_t has a very slowly decaying autocorrelation (every value of r_k is statistically different from zero) and therefore the sequence z_t is not stationary. Because of this we can consider taking the first difference of z_t and observing if the resulting series is stationary. The SACF of $(1-B)z_t$ appears to be that of an AR(1) model with $\phi < 0$ since the terms oscillate above and below zero.

To determine is this data further supports an AR(1) model we can construct from the sample autocorrelation of $(1-B)z_t$ the sample partial autocorrelation function (SPACF) $\hat{\phi}_{kk}$. Note that in the notation ϕ_{kj} the k index denotes the *total* order of the autoregressive model while the index j denotes which component of this total model we are considering i.e. $1 \leq j \leq k$,

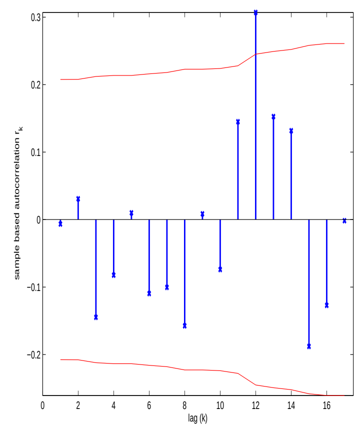


Figure 25: The sample autocorrelation function for the residuals of an MA(1) model fit to the demand for parts data set. This plot is very similar to the one shown in the text. Notice the significant component at the lag $k = 12$.

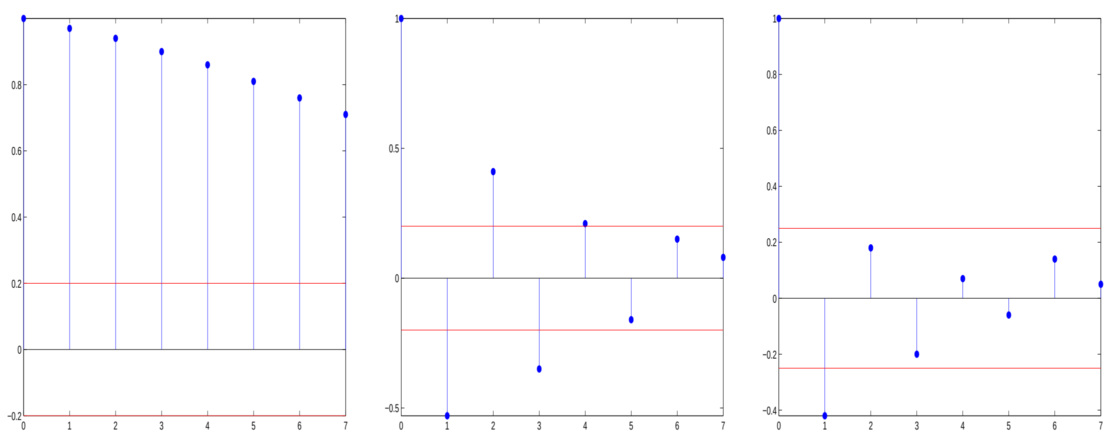


Figure 26: **Left:** The sample autocorrelation function, r_k, for the time series z_t in Part (a) of Exercise 5.1. **Middle:** The sample autocorrelation function, r_k, for the first difference of the time series z_t i.e. $z_t - z_{t-1}$ in Part (a) of Exercise 5.1. **Right:** The sample autocorrelation for the time series z_t in Part (b) of Exercise 5.1. Note standard two-sigma confidence regions are drawn as horizontal lines (at the values $2/\sqrt{n}$) in red.

so that z_t will have a representation given by

$$z_t = \phi_{k1} z_{t-1} + \phi_{k2} z_{t-2} + \cdots + \phi_{kk} z_{t-k} + a_t \, .$$

Then using the Levinson and Durbin algorithm discussed in the text and coded in the MATLAB function `spacf`, we see that the values we observe for this data set are given by (two two decimals)

```
-0.53    0.18   -0.11   -0.07   -0.00    0.04
```

We then look at which values of $\hat{\phi}_{kk}$ are significant by comparing the above magnitudes to the value of $2/\sqrt{n} = 0.2$. The first value is clearly significant and indicated that an approximate value to consider for ϕ would be near ≈ -0.5. Thus in summary for this data set one could consider an ARIMA(1,1,0) model for z_t.

Part (b): From Figure 26 (right) the only significant sample autocorrelation value r_k is r_1 which signifies we might be looking at data from a MA(1) model.

Exercise 5.2 (a stationary linear model)

To be stationary it is necessary that $E(z_t) = \mu$ where μ is a constant independent of time t. For the model given, $z_t = \beta_0 + \beta_1 t + a_t$, we have for this expectation $E(z_t) = \beta_0 + \beta_1 t$ which is not independent of t and hence z_t is *not* stationary. Now z_t is not stationary but its first difference

$$\begin{aligned} y_t &\equiv (1-B)z_t = \beta_0 + \beta_1 t + a_t - \beta_0 - \beta_1(t-1) - a_{t-1} \\ &= a_t - a_{t-1} - \beta_1 \, , \end{aligned}$$

would be stationary and could be fit with a MA(1) model.

Exercise 5.3 (a model for monthly sales)

Part (a): Since this is an ARMA(p,q) model for an ARMA(p,q) model to be stationary the roots of the autoregressive polynomial, $\phi(B)$, in the representation

$$\phi(B)(z_t - \mu) = \theta(B) a_t \, ,$$

where

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \, , \end{aligned}$$

must have roots *outside* the unit circle. Since in the case considered here the polynomial is $\phi(B) = (1-B)^2$, has a double root at $B = 1$ which is not outside the unit circle this model for z_t is not stationary.

Part (b): To be invertible the polynomial $\theta(B)$ must have roots outside the unit circle. In this case $\theta(B) = 1 - \theta B$, which has roots $B = \frac{1}{\theta} > 1$, so this model *is* invertible.

Part (c): The given model explicitly calculates the second differences of z_t as the expression

$$\begin{aligned}
(1-B)^2 z_t &= (1-B)(z_t - z_{t-1}) \\
&= z_t - z_{t-1} - z_{t-1} + z_{t+2} \\
&= z_t - 2z_{t-1} + z_{t+2}\,,
\end{aligned}$$

is the second difference. Since this equals $(1 - \theta B)a_t$ a MA(1) model and will have the simplest type of autocorrelation function. In general, for any stochastic model when $z_t - \mu$ is expressed in terms of its linear filter representation as

$$\begin{aligned}
z_t - \mu &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \\
&= \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \text{with} \quad \psi_0 = 1 \\
&= \psi(B) a_t\,,
\end{aligned}$$

that the autocorrelation function ρ_k for z_t is given by

$$\rho_k = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}\,. \tag{101}$$

In the case above the linear filter representation of $(1-B)^2 z_t$ is $\psi(B) = 1 - \theta B$, so $\psi_1 = -\theta$ and $\psi_j = 0$ for $j > 1$. Thus

$$\rho_1 = -\frac{\theta}{1+\theta^2} \quad \text{and} \quad \rho_k = 0 \quad \text{for} \quad k > 1\,.$$

Part (d): The π weights are the expression of this process in terms of an infinite autoregressive representation. That is

$$\pi(B)(z_t - \mu) = a_t\,,$$

so

$$\pi(B) = \frac{(1-B)^2}{(1-\theta B)} \tag{102}$$

so equating the coefficients in Equation 102 by multiplying both sides by $1 - \theta B$ we see that π_j must satisfy

$$(1-\theta B)(1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots) = (1-B)^2 = 1 - 2B + B^2\,,$$

or expanding and grouping the left hand side we find

$$\begin{aligned}
1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots - \theta B + \theta\pi_1 B^2 + \theta\pi_2 B^3 + \cdots &= 1 - 2B + B^2 \\
1 + (-\pi_1 - \theta)B + (-\pi_2 + \theta\pi_1)B^2 + (-\pi_3 B^3 + \theta\pi_2) + \cdots &= 1 - 2B + B^2\,.
\end{aligned}$$

So equating coefficients of B on both sides of this expression we see that

$$\begin{aligned}
-\pi_1 - \theta = -2 &\;\rightarrow\; \pi_1 = 2 - \theta \\
-\pi_2 + \theta\pi_1 = 1 &\;\rightarrow\; \pi_2 = \theta\pi_1 - 1 = -(1-\theta)^2 \\
-\pi_k + \theta\pi_{k-1} = 0 &\;\rightarrow\; \pi_k = \theta\pi_{k-1}\,,
\end{aligned}$$

for $k \geq 3$. Thus when we solve for this last expression for π_k we find

$$\pi_1 = 2 - \theta \quad \text{and} \quad \pi_k = -(1-\theta)^2\theta^{k-2} \quad \text{for} \quad k \geq 2\,.$$

The ψ weights are the coefficients in the linear filter representation

$$z_t - \mu = \psi(B)a_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots\,.$$

For this process we have

$$z_t = \frac{(1-\theta B)}{(1-B)^2} a_t = \psi(B)a_t\,,$$

so we multiply by $(1-B)^2$ on both sides the coefficients ψ_j must satisfy

$$\begin{aligned}
1 - \theta B &= (1 - 2B + B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) \\
&= 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \psi_4 B^4 + \cdots \\
&+ \quad -2B - 2\psi_1 B^2 - 2\psi_2 B^3 - 2\psi_3 B^4 - \cdots \\
&+ \quad B^2 + \psi_1 B^3 + \psi_2 B^4 + \cdots\,,
\end{aligned}$$

or

$$\begin{aligned}
1 - \theta B &= 1 + (\psi_1 - 2)B + (\psi_2 - 2\psi_1 + 1)B^2 + (\psi_3 - 2\psi_2 + \psi_1)B^3 \\
&+ (\psi_4 - 2\psi_3 + \psi_2)B^4 + \cdots + (\psi_k - 2\psi_{k-1} + \psi_{k-2})B^k + \cdots\,.
\end{aligned}$$

So that

$$\begin{aligned}
-\theta = \psi_1 - 2 &\rightarrow \psi_1 = 2 - \theta \\
\psi_2 = 2\psi_1 - 1 &\rightarrow \psi_2 = 3 - 2\theta \\
\psi_3 = 2\psi_2 - \psi_1 &\rightarrow \psi_3 = 4 - 3\theta \\
&\vdots \\
\psi_k = 2\psi_{k-1} - \psi_{k-2} &\rightarrow \psi_k = (k+1) - k\theta = (\theta - 1)k + 1\,,
\end{aligned}$$

for $k \geq 1$.

Part (e): The one-step-ahead forecast for an ARIMA(0,2,1) model such as this one can be derived as in section 5.4 Forecasting from the book. There it was shown that if we have the values of ψ_j for $j \geq 1$ in the linear system representation

$$z_t - \mu = \psi(B)a_t \quad \text{with} \quad \psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots\,,$$

then the minimum mean square error (MMSE) forecast and its variance is given by

$$z_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots \tag{103}$$

$$V(e_n(l)) = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\,. \tag{104}$$

Equation 103 is often more easily implemented by converting the model into its autoregressive formulation. Following the example in that section, we write our process as

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta a_{t-1}\,,$$

so that the one-step-ahead forecast prediction is given by

$$\begin{aligned} z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \ldots) \\ &= E(2z_n - z_{n-1} + a_{n+1} - \theta a_n | z_n, z_{n-1}, \ldots) \\ &= 2z_n - z_{n-1} - \theta a_n \,. \end{aligned}$$

As discussed in the text, this forecast involves the unknown values of θ and a_n. Once the model is estimated and an estimate of θ, say $\hat{\theta}$, is obtained we can calculate the residuals $\hat{a}_t$ using Equation 99 as

$$\hat{a}_t = z_t - 2z_{t-1} + z_{t-2} + \hat{\theta}\hat{a}_{t-1} \,,$$

and use the that value in the forecast. Of course an initial value of $\hat{a}_1$ will need to be specified. To get this process started. The value of $\hat{a}_1$ could be taken as $\hat{a}_1 = z_1 - z_0(1)$ or the initial residual. The variance of the one-step-ahead forecast is given by

$$V(e_n(1)) = \sigma^2 \,,$$

since $l = 1$ in the sum in Equation 104.

Exercise 5.4 (a model for an additive process)

For this exercise we will assume that our process y_t can be expressed as

$$y_t = x_t + z_t \,,$$

with x_t and AR(1) process of the form $(1 - \phi B)x_t = a_t$ and z_t a white-noise process.

Part (a): We want to evaluate the autocorrelation functions ρ_k for the process y_t. Multiply y_t above by y_{t-k} and take the expectation. We find since y_t is a zero mean process that

$$\begin{aligned} E(y_t y_{t-k}) &= E((x_t + z_t)(x_{t-k} + z_{t-k})) \\ &= E(x_t x_{t-k}) + E(z_t z_{t-k}) \,. \end{aligned}$$

Now since z_t is white-noise with variance σ_z^2 the expression $E(z_t z_{t-k}) = \sigma_z^2 \delta_{k,0}$, and since x_t is an AR(1) model we have derived (by considering the linear filter representation) that

$$\begin{aligned} E(x_t^2) &= \frac{\sigma^2}{1 - \phi^2} \\ E(x_t x_{t-k}) &= \frac{\sigma^2 \phi^k}{1 - \phi^2} \,. \end{aligned}$$

Here σ^2 is the variance of the random shocks a_t in the AR(1) model. Thus when $k = 0$ we have

$$E(y_t^2) = \frac{\sigma^2}{1 - \phi^2} + \sigma_z^2 \,,$$

and when $k > 0$ we have

$$E(y_t y_{t-k}) = \frac{\sigma^2 \phi^k}{1 - \phi^2} \,.$$

Thus ρ_k is given by

$$\rho_k = \frac{\frac{\sigma^2\phi^k}{1-\phi^2}}{\frac{\sigma^2}{1-\phi^2} + \sigma_z^2} = \frac{\sigma^2\phi^k}{\sigma^2 + (1-\phi^2)\sigma_z^2},$$

for $k \geq 1$.

Part (b): Since we have an exponential decay in the autocorrelation function ρ_k for y_t, which is the same type of phenomena as observed in an AR(1) model, an AR(1) model would be appropriate for y_t.

Exercise 5.5 (the k-th order partial autocorrelation ϕ_{kk})

Part (a): The definition of the k-th order partial autocorrelation of a process z_t or ϕ_{kk} is the last coefficient in an autoregressive model of k-th order i.e. we regress z_t on the k-previous values of z i.e.

$$z_t = \phi_{k1}z_{t-1} + \phi_{k2}z_{t-2} + \cdots + \phi_{kk}z_{t-k} + a_t,$$

then the value of ϕ_{kk} or the coefficient of the term z_{t-k} is defined as the partial autocorrelation of order k.

Part (b):

i: For the model

$$(1 - 1.2B + 0.8B^2)z_t = a_t,$$

since this is an AR(2) model with coefficients $\phi_1 = 1.2$ and $\phi_2 = -0.8$, we have that

$$\phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{2}{3},$$

and

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\frac{\phi_1^2}{1-\phi_2} + \phi_2 - \frac{\phi_1^2}{(1-\phi_1)^2}}{1 - \frac{\phi_1^2}{(1-\phi_1)^2}} = -\frac{4}{5},$$

and $\phi_{kk} = 0$ for $k > 2$. To derive these we have used the expression derived in the book and these notes on the autocorrelation function ρ_k for an AR(2) process. Note that an AR(2) process has only two non-zero values for ϕ_{kk}. These simple calculations are done in the MATLAB script `prob_5_5.m`.

ii: Since this is an AR(1) model $\phi_{11} = \rho_1 = \phi = 0.7$ and $\phi_{kk} = 0$ for $k > 1$. Note that an AR(1) process has only one non-zero value for ϕ_{kk}.

iii: Since the model

$$(1 - 0.7B)z_t = (1 - 0.5B)a_t,$$

is an ARMA(1,1) model the kth-order partial autocorrelation (PACF) will be a mix between the the PACF of an AR(1) and a MA(1) model. Specifically the PACF of this ARMA(1,1) model will have a single initial value $\phi_{11} = \rho_1 = 0.7$ and from that point onward it will decay like the PACF of a MA(1) process having exponential decay.

iv: Since the given model is a MA(1) model it will have a partial autocorrelation function that is infinite in extent and is dominated by contribution of damped exponentials. Specifically we can show that ϕ_{kk} for a MA(1) model looks like

$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} = -\frac{(0.5)^k(1-0.5^2)}{1-(0.5)^{2(k+1)}} \quad \text{for} \quad k > 0\,.$$

Exercise 5.6 (the variance of the mean of $z_t = a_t - a_{t-1}$)

Consider the sum $\frac{1}{n}\sum_{t=1}^{n} z_t$. Since we know the explicit model for z_t we can write this sum as

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}(a_t - a_{t-1}) &= \frac{1}{n}\sum_{t=1}^{n} a_t - \frac{1}{n}\sum_{t=1}^{n} a_{t-1} = \frac{1}{n}\sum_{t=1}^{n} a_t - \frac{1}{n}\sum_{t=0}^{n-1} a_t \\
&= \frac{1}{n}a_n - \frac{1}{n}a_0\,.
\end{aligned}$$

As a_0 and a_n are uncorrelated, the variance of this expression is then given by

$$V\left(\frac{1}{n}\sum_{t=1}^{n} z_t\right) = \frac{1}{n^2}(\sigma^2 + \sigma^2) = \frac{2\sigma^2}{n^2}\,,$$

as was to be shown.

Exercise 5.7 (H. Working's example)

Recall the definitions of

$${}_t\Delta_m = z_t - z_{t-m}\,,$$

and

$${}_t\Delta_m^* = \frac{1}{m}\left[z_t + z_{t+1} + z_{t+2} + \cdots + z_{t+m-1}\right] - \frac{1}{m}\left[z_{t-m} + z_{t-m+1} + z_{t-m+2} + \cdots + z_{t-1}\right]\,.$$

Then to calculate the variances of these expressions it is helpful to derive the linear filter representation of the time series z_t. From the assumed model for z_t of $z_t - z_{t-1} = a_t$ we have $(1-B)z_t = a_t$ or

$$z_t = (1-B)^{-1}a_t = \sum_{k=0}^{\infty} B^k a_t = \sum_{k=0}^{\infty} a_{t-k}\,.$$

Using this the linear filter representation of the first difference m-steps apart ${}_t\Delta_m$ is given by

$${}_t\Delta_m = \sum_{k=0}^{\infty} a_{t-k} - \sum_{k=0}^{\infty} a_{t-m-k} = \sum_{k=0}^{m-1} a_{t-k}\,,$$

this has zero expectation and so its variance is given by

$$V({}_t\Delta_m) = \sum_{k=0}^{m-1} V(a_{t-k}) = \sigma^2 m\,.$$

For the second expression ${}_t\Delta^*_m$, its linear filter representation can be written as (after multiplying both sides by m)

$$\begin{aligned}
m\,{}_t\Delta^*_m &= (z_t - z_{t-m}) + (z_{t+1} - z_{t-m+1}) + \cdots + (z_{t+m-1} - z_{t-1}) \\
&= \sum_{k=0}^{m-1} a_{t-k} + \sum_{k=0}^{m-1} a_{t+1-k} + \cdots + \sum_{k=0}^{m-1} a_{t+m-1-k} \\
&= a_{t-m+1} + a_{t-m+2} + a_{t-m+3} + \cdots + a_{t-3} + a_{t-2} + a_{t-1} + a_t \\
&+ \quad a_{t-m+2} + a_{t-m+3} + \cdots + a_{t-3} + a_{t-2} + a_{t-1} + a_t + a_{t+1} \\
&+ \quad a_{t-m+3} + \cdots + a_{t-3} + a_{t-2} + a_{t-1} + a_t + a_{t+1} + a_{t+2} \\
&\vdots \\
&+ \quad a_t + \cdots + a_{t+m-2} + a_{t+m-1} \\
&= a_{t-m+1} + 2a_{t-m+2} + \cdots + ma_t + (m-1)a_{t+1} + (m-2)a_{t+2} + \cdots + 2a_{t+m-2} + a_{t+m-1} \\
&= \sum_{k=1}^{m} k a_{t-m+k} + \sum_{k=1}^{m-1} k a_{t+m-k}\,.
\end{aligned}$$

This expression has zero mean so (remembering to divide back by m) we find its variance given by

$$\begin{aligned}
V({}_t\Delta^*_m) &= \frac{1}{m^2} E\left(\left(\sum_{k=1}^{m} k a_{t-m+k} + \sum_{k=1}^{m-1} k a_{t+m-k}\right)^2\right) \\
&= \frac{1}{m^2} E\left(\sum_{k=1}^{m} k^2 a^2_{t-m+k} + \sum_{k=1}^{m-1} k^2 a^2_{t+m-k}\right),
\end{aligned}$$

since the cross terms all vanish. Thus

$$\begin{aligned}
V({}_t\Delta^*_m) &= \frac{\sigma^2}{m^2} \sum_{k=1}^{m} k^2 + \frac{\sigma^2}{m^2} \sum_{k=1}^{m} k^2 \\
&= \frac{2\sigma^2}{m^2} \sum_{k=1}^{m} k^2 = \frac{\sigma^2}{3}\left(\frac{(m+1)(2m+1)}{m}\right).
\end{aligned}$$

Thus the ratio desired is given by

$$\frac{V({}_t\Delta^*_m)}{V({}_t\Delta m)} = \frac{1}{3}\left(1 + \frac{1}{m}\right)\left(2 + \frac{1}{m}\right) \to \frac{2}{3}\,,$$

as $m \to \infty$ as we were to show.

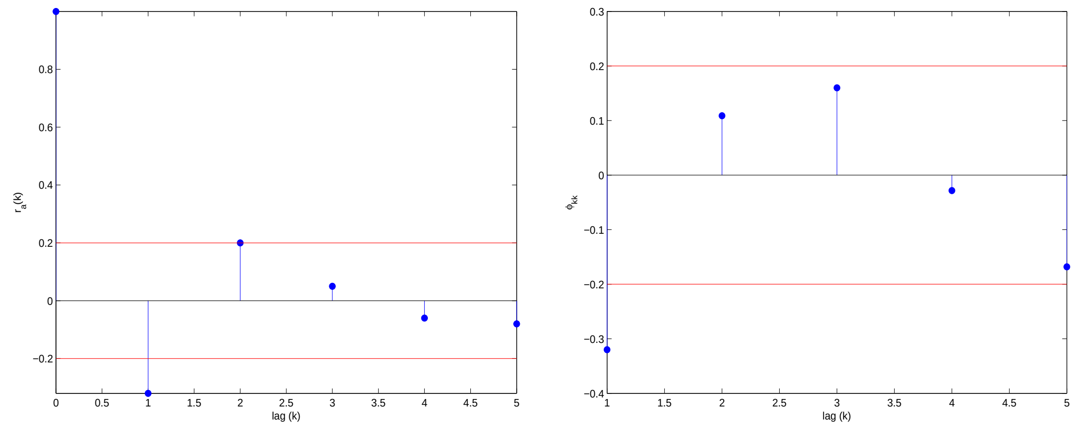


Figure 27: **Left:** A stem plot of the autocorrelation of the residuals $r_{\hat{a}}(k)$ for Exercise 5.9. **Right:** A stem plot of the sample partial autocorrelation function for the residuals time series in Exercise 5.9.

Exercise 5.8 (appropriate ARMA models for sums)

We are told that each z_{ti} is given by a MA(1) model such that $z_{ti} = (1 - \theta_i B)a_{ti}$ with known variances for the random shocks a_{ti}, given by $V(a_{ti}) = \sigma_i^2$. Then the total annual cost, C_t, is the sum of 100 MA(1) models

$$C_t = \sum_{i=1}^{100} z_{ti} \, ,$$

Part (a): Now the total cost is simply a sum of 100 uncorrelated random variables z_{ti} the variance of these individual random variables is given by $\text{Var}(z_{ti}) = \sigma_i^2(1 + \theta_i^2)$, so the total variance of C_t is the sum of these individual variances $\sum_{i=1}^{100} \sigma_i^2(1 + \theta_i^2)$. Then in this case C_t is simply a zero-mean random variable with known variance.

Exercise 5.9 (considering an ARIMA model)

Consider the stem plot of the sample autocorrelation of the residuals from an ARIMA(0,1,1) model shown in Figure 27 (left). On this plot we also plot the associated two sigma standard errors

$$s[r_{\hat{a}}(k)] \approx n^{-1/2} = 1/10 = 0.1 \, ,$$

we see that the lag one $k = 1$ (and possibly the lag two autocorrelation) is significant. In addition, we plot the sample partial autocorrelation function (SPACF) in Figure 27 (right). There we see a significant value for ϕ_{11}. These two plots taken together imply that the residuals a_t might satisfy an AR(1) model with $\phi < 0$. If a_t satisfies an AR(1) model the $(1 - \phi B)a_t = b_t$ with b_t white noise. This implies that the model for z_t, given as $(1 - B)z_t = (1 - \theta B)a_t$, should perhaps be modified to be

$$(1 - B)z_t = (1 - \theta B)a_t = (1 - \theta B)\left(\frac{1}{(1 - \phi B)} b_t\right) \, ,$$

or

$$(1 - \phi B)(1 - B)z_t = (1 - \theta B)b_t \,,$$

showing that z_t satisfies an ARIMA(1,1,1) model. See the MATLAB script `prob_5_9.m` for the code used to generate these plots.

Exercise 5.10 (some ARIMA models)

Model 1: An easy way to generate the forecast equations is to write the model explicitly in terms of z_t i.e. in its autoregressive form. For example, the first model

$$(1 - \phi B)(z_t - \mu) = a_t \,,$$

is equivalent to

$$z_t - \mu = \phi(z_{k-1} - \mu) + a_t \,,$$

which is an AR(1) process and is discussed in the book. Thus we find that

$$z_n(l) = \mu + \phi^l(z_n - \mu) \,, \tag{105}$$

is the l-step-ahead forecast and that

$$V(e_n(l)) = \frac{\sigma^2(1 - \phi^{2l})}{1 - \phi^2} \,, \tag{106}$$

is the variance of the forecast.

Model 2: For this ARMA(2,0) model we have

$$z_t = \mu + \phi_1(z_{t-1} - \mu) + \phi_2(z_{t-2} - \mu) + a_t \,,$$

so taking the conditional expectation of both sides to compute $z_n(1)$ we find

$$\begin{aligned}
z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \ldots) \\
&= E(\mu + \phi_1(z_n - \mu) + \phi_2(z_{n-1} - \mu) + a_{n+1}|z_n, z_{n-1}, \ldots) \\
&= \mu + \phi_1(z_n - \mu) + \phi_2(z_{n-1} - \mu) \,.
\end{aligned}$$

In general as shown in the book we have

$$\begin{aligned}
z_n(l) &= E(z_{n+l}|z_n, z_{n-1}, \ldots) \\
&= E(\mu + \phi_1(z_{n+l-1} - \mu) + \phi_2(z_{n+l-2} - \mu) + a_{n+l}|z_n, z_{n-1}, \ldots) \\
&= \mu + \phi_1(z_n(l-1) - \mu) + \phi_2(z_n(l-2) - \mu) \,, \qquad (107)
\end{aligned}$$

and the forecast variance is given by

$$v(e_n(l)) = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2) \,,$$

which relies on the given process' linear filter representations coefficients ψ_j in

$$z_n = a_n + \psi_1 a_{n-1} + \psi_2 a_{n-2} + \cdots \,.$$

For an AR(2) model as show on Page 77 of these notes the ψ_j-linear filter representation is given by

$$\psi_1 = \phi_1 \,, \quad \psi_2 = \phi_1^2 + \phi_2 \,, \quad \psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} \quad \text{for} \quad j \geq 3 \,.$$

With all of this information we can compute the desired quantities for three look-aheads $l = 1, 2, 3$. We specifically find

$$\begin{aligned} z_n(1) &= \mu + \phi_1(z_n - \mu) + \phi_2(z_{n-1} - \mu) \quad \text{with} \\ V(e_n(1)) &= \sigma^2 \,, \end{aligned}$$

for $l = 1$ and

$$\begin{aligned} z_n(2) &= \mu + \phi_1(z_n(1) - \mu) + \phi_2(z_n - \mu) \\ &= \mu + \phi_1(\phi_1(z_n - \mu) + \phi_2(z_{n-1} - \mu)) + \phi_2(z_n - \mu) \\ &= \mu + (\phi_1^2 + \phi_2)(z_n - \mu) + \phi_1\phi_2(z_{n-1} - \mu) \quad \text{with} \\ V(e_n(2)) &= \sigma^2(1 + \phi_1^2) \,, \end{aligned}$$

for $l = 2$ and finally

$$\begin{aligned} z_n(3) &= \mu + \phi_1(z_n(2) - \mu) + \phi_2(z_n(1) - \mu) \\ &= \mu + \phi_1((\phi_1^2 + \phi_2)(z_n - \mu) + \phi_1\phi_2(z_{n-1} - \mu)) \\ &+ \phi_2(\phi_1(z_n - \mu) + \phi_2(z_{n-1} - \mu)) \\ &= \mu + (\phi_1^3 + 2\phi_1\phi_2)(z_n - \mu) + (\phi_1^2\phi_2 + \phi_2^2)(z_{n-1} - \mu) \quad \text{with} \\ V(e_n(3)) &= \sigma^2(1 + \phi_1^2 + (\phi_1^2 + \phi_1)^2) \,. \end{aligned}$$

for $l = 3$.

Model 3: For the model $(1 - \phi B)(1 - B)z_t = \theta_0 + (1 - \theta B)a_t$, to generate forecasts we will write this in terms of its autoregressive formulation as in the two cases above. Alternatively we could note that this model is exactly the ARIMA(1,1,1) model given in the example in the text and use the results there. Then we write the model as

$$z_t = \theta_0 + (1 + \phi)z_{t-1} - \phi z_{t-2} + a_t - \theta a_{t-1} \,,$$

which has forecasts given as in the book i.e.

$$\begin{aligned} z_n(1) &= E(z_{n+1} | z_n, z_{n-1}, \ldots) \\ &= E(\theta_0 + (1 + \phi)z_n - \phi z_{n-1} + a_{n+1} - \theta a_n | z_n, z_{n-1}, \ldots) \\ &= \theta_0 + (1 + \phi)z_n - \phi z_{n-1} - \theta a_n \,. \end{aligned}$$

To actually *use* this last forecast equation one would have to estimated a_n (along with θ_0, θ, and ϕ). How to do this is indicated by the general forecast error expansion Equation 98 evaluated at $l = 1$ which is

$$e_n(1) = z_{n+1} - z_n(1) = a_{n+1} \,.$$

Thus as an estimate of a_n we use the residual error in the forecast of z_n i.e.

$$a_n = z_n - z_{n-1}(1) \,.$$

To compute $z_n(2)$ we have

$$\begin{aligned} z_n(2) &= E(z_{n+2}|z_n, z_{n-1}, \ldots) \\ &= E(\theta_0 + (1+\phi)z_{n+1} - \phi z_n + a_{n+2} - \theta a_{n+1}|z_n, z_{n-1}, \ldots) \\ &= \theta_0 + (1+\phi)z_n(1) - \phi z_n \\ &= \theta_0 + (1+\phi)(\theta_0 + (1+\phi)z_n - \phi z_{n-1} - \theta a_n) - \phi z_n \\ &= \theta_0 + (1+\phi)\theta_0 + ((1+\phi)^2 - \phi)z_n - \phi(1+\phi)z_{n-1} - \theta(1+\phi)a_n \,. \end{aligned}$$

An expression for $z_n(3)$ is computed in the same way.

The variance of the l-step-ahead forecast error is best expressed in terms of the coefficients ψ_j in the Wold decomposition for the given model. This requires we evaluate

$$\begin{aligned} z_t &= \frac{1}{(1-\phi B)(1-B)}(\theta_0 + (1-\theta B)a_t) \\ &= \frac{1}{(1-\phi B)(1-B)}\theta_0 + \frac{1}{(1-\phi B)(1-B)}(1-\theta B)a_t \\ &= \theta_0 + \left(\frac{1-\theta B}{(1-\phi B)(1-B)}\right) a_t \,, \end{aligned}$$

where the last step is since the operator $\frac{1}{(1-\phi B)(1-B)}$ applied to a constant (here θ_0) is that constant again. Expanding $\frac{1-\theta B}{(1-\phi B)(1-B)}$ in a Taylor series in B gives

$$\begin{aligned} \psi_1 &= 1 \\ \psi_2 &= 1 + \phi + \theta \\ \psi_3 &= 1 + \phi + \phi^2 - (1+\phi)\theta \\ &\vdots \\ \psi_j &= \sum_{k=0}^{j} \phi^k - \left(\sum_{k=0}^{j-1} \phi^k\right)\theta = \left(\frac{\phi^{j+1}-1}{\phi-1}\right) - \left(\frac{\phi^j - 1}{\phi - 1}\right) \\ &= \left(\frac{1}{\phi-1}\right)(\phi^j(\phi-\theta) + (\theta - 1)) \,. \end{aligned}$$

These expressions are used in Equation 93 to compute the desired variances.

Model 4: For the ARIMA(0,2,3) model

$$(1-B)^2 z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t \,.$$

This can be expanded to give

$$z_t = 2z_{t-1} - z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} \,.$$

From this we compute

$$\begin{aligned} z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \ldots) \\ &= E(2z_n - z_{n-1} + a_{n+1} - \theta_1 a_n - \theta_2 a_{n-1} - \theta_3 a_{n-2}|z_n, z_{n-1}, \ldots) \\ &= 2z_n - z_{n-1} - \theta_1 a_n - \theta_2 a_{n-1} - \theta_3 a_{n-2} \,, \end{aligned}$$

where to actually make predictions using this formula the values of a_n would need to be estimated (see the previous model for a discussion of how to do this). To compute $z_n(2)$ we have

$$\begin{aligned}
z_n(2) &= E(z_{n+2}|z_n, z_{n-1}, \dots) \\
&= E(2z_{n+1} - z_n + a_{n+2} - \theta_1 a_{n+1} - \theta_2 a_n - \theta_3 a_{n-1}|z_n, z_{n-1}, \dots) \\
&= 2z_n(1) - z_n - \theta_2 a_n - \theta_3 a_{n-1} \\
&= 3z_n - 2z_{n-1} + (2\theta_1 - \theta_2)a_n + (-\theta_2 - \theta_3)a_{n-1} - \theta_3 a_{n-2} \,.
\end{aligned}$$

An expression for $z_n(3)$ is computed in the same way.

To determine the variances of the l-step-ahead forecasts we need to determine the values of ψ_j in the Wold decomposition. Thus we need to compute the jth Taylor coefficient of B in the expression

$$\frac{1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3}{(1-B)^2} \,,$$

Using Mathematica we find

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= 2 - \theta_1 \\
\psi_2 &= 3 - 2\theta_1 - \theta_2 \\
\psi_3 &= 4 - 3\theta_1 - 2\theta_2 - \theta_3 \\
&\vdots \\
\psi_j &= (j+1) - j\theta_1 - (j-1)\theta_2 - (j-2)\theta_3 \\
&= j(1 - \theta_1 - \theta_2 - \theta_3) + (1 + \theta_2 + 2\theta_3) \,.
\end{aligned}$$

These expressions are used in Equation 93 to compute the desired variances.

See the Mathematica file `prob_5_10.nb` where we derive these Taylor series expansions.

Exercise 5.11 (the ARMA(1,1) model)

Part (a): The autoregressive form of this model is given by finding the coefficients π_j in the polynomial $\pi(B)$ such that

$$\pi(B)(z_t - \mu) = a_t \,.$$

In the ARMA(1,1) model given here the specific functional form for the function $\pi(B)$ is $\frac{1-\phi B}{1-\theta B}$, so we need to compute the values of π_j such that

$$\pi(B) = \frac{1 - \phi B}{1 - \theta B} = 1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots \,.$$

These values were derived in the book for ARMA(1,1) models and found to be $\pi_j = (\phi - \theta)\theta^{j-1}$ for $j > 0$. Thus

$$\begin{aligned} z_t - \mu &= a_t + \sum_{j=1}^{\infty} \pi_j B^j z_t = a_t + \sum_{j=1}^{\infty} \pi_j z_{t-j} \\ &= a_t + \sum_{j=1}^{\infty} (\phi - \theta)\theta^{j-1} z_{t-j} \,. \end{aligned}$$

The MMSE forecast using the autoregressive formulation are then given by

$$\begin{aligned} z_n(1) &= E(z_{n+1} | z_n, z_{n-1}, \ldots) \\ &= E(\mu + a_{n+1} + \sum_{j=1}^{\infty} (\phi - \theta)\theta^{j-1} z_{n+1-j} | z_n, z_{n-1}, \ldots) \\ &= \mu + \sum_{j=1}^{\infty} (\phi - \theta)\theta^{j-1} z_{n+1-j} \,. \end{aligned}$$

Note that due to the MA(1) component the required number of autoregressive coefficients is infinite and because of this the given formulation may not be the best way to compute the MMSE predictor. For $z_n(2)$ we have a similar expression given by

$$\begin{aligned} z_n(2) &= E(\mu + a_{n+2} + \sum_{j=1}^{\infty} (\phi - \theta)\theta^{j-1} z_{n+2-j} | z_n, z_{n-1}, \ldots) \\ &= \mu + (\phi - \theta) z_n(1) + \sum_{j=2}^{\infty} (\phi - \theta)\theta^{j-1} z_{n+2-j} \,, \end{aligned}$$

with $z_n(1)$ given in the expression above.

Part (b): To calculate the l-step-ahead forecast error, we need the linear filter representation of

$$\psi(B) = \frac{1 - \theta B}{1 - \phi B} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \,.$$

We find in the book that an ARMA(1,1) model has these coefficients given by $\psi_j = (\phi - \theta)\phi^{j-1}$ for $j > 0$ so that the variance of our forecasts is given by

$$\begin{aligned} V(e_n(l)) &= \sigma^2 \left(1 + (\phi - \theta)^2 \sum_{j=1}^{l-1} \phi^{j-1} \right) \\ &= \sigma^2 \left(1 + \frac{(\phi - \theta)^2}{\phi} \left(\left. \frac{\phi^j}{\phi - 1} \right|_1^l \right) \right) \\ &= \sigma^2 \left(1 + \frac{(\phi - \theta)^2}{\phi} \left(\frac{\phi^l - \phi}{\phi - 1} \right) \right) \\ &= \sigma^2 \left(1 + (\phi - \theta)^2 \left(\frac{\phi^{l-1} - 1}{\phi - 1} \right) \right) \,. \end{aligned}$$

Part (c): As $\phi \to 1$ our prediction $z_n(1)$ becomes

$$
\begin{aligned}
z_n(1) &= \mu + \sum_{j=1}^{\infty} (1-\theta)\theta^{j-1} z_{n+1-j} \\
&= \mu + (1-\theta)(z_n + \theta z_{n-1} + \theta^2 z_{n-2} + \cdots)\,,
\end{aligned}
$$

the same as equation 5.76 for an ARIMA(0,1,1). To evaluate $V(e_n(1))$ as $\phi \to 1$ we need to evaluate the limit of

$$
\frac{\phi^{l-1} - 1}{\phi - 1} \to \frac{(l-1)\phi^{l-1}}{1} \to l - 1\,.
$$

Thus $V(e_n(l))$ goes to

$$
V(e_n(l)) = \sigma^2(1 + (l-1)(1-\theta)^2)\,,
$$

which is correct since when $\phi \to 1$ the model becomes an ARIMA(0,1,1) model this result matches the result derived in the text for the ARIMA(0,1,1) model.

Exercise 5.12 (an example forecasting with an AR(1) model)

We are told that estimates of the parameters θ and ϕ are given by $\hat{\theta}_0 = 50$ and $\hat{\phi}_0 = 0.6$ and in addition that $z_{100} = 115$. The forecast for an AR(1) model like this are given by

$$
\begin{aligned}
z_n(1) &= E(z_{n+1} | z_n, z_{n-1}, \dots) \\
&= E(\theta_0 + \phi z_n + a_{n+1} | z_n, z_{n-1}, \dots) \\
&= \hat{\theta}_0 + \hat{\phi} z_n\,.
\end{aligned}
$$

For $z_n(2)$ we have

$$
\begin{aligned}
z_n(2) &= E(\theta_0 + \phi z_{n+1} + a_{n+2} | z_n, z_{n-1}, \dots) \\
&= \hat{\theta}_0 + \hat{\phi} z_n(1)\,.
\end{aligned}
$$

Finally for $z_n(3)$ we have

$$
\begin{aligned}
z_n(3) &= E(\theta_0 + \phi z_{n+2} + a_{n+3} | z_n, z_{n-1}, \dots) \\
&= \hat{\theta}_0 + \hat{\phi} z_n(2)\,.
\end{aligned}
$$

Using the estimated values in the above we find forecasts for the periods 101, 102, and 103 of z_n are given by

$$
\begin{aligned}
\hat{z}_{101} &= z_{100}(1) = 119.0 \\
\hat{z}_{102} &= z_{100}(2) = 121.4 \\
\hat{z}_{103} &= z_{100}(3) = 122.84\,.
\end{aligned}
$$

See the MATLAB file `prob_5_12.m` where we perform these calculations.

Exercise 5.13 (updating the forecasts in an ARIMA(0,1,1) model)

To update our forecasts we use

$$z_{n+1}(l) = z_n(l+1) + \psi_l(z_{n+1} - z_n(1)),$$

but evaluated at $l-1$ or

$$z_{n+1}(l-1) = z_n(l) + \psi_{l-1}(z_{n+1} - z_n(1)).$$

Now for an ARIMA(0,1,1) model $\psi_j = 1-\theta$ for all j so the prediction update equation above specifically becomes

$$z_{n+1}(l-1) = z_n(l) + (1-\theta)(z_{n+1} - z_n(1)). \tag{108}$$

In addition for this type of model the eventual prediction equation is

$$z_n(l) = z_n(l-1) \quad \text{for} \quad l > 1, \tag{109}$$

or a constant i.e. all predictions are the same. Thus the knowledge of $z_{100}(10) = 26$ means we also know $z_{100}(l)$ for $1 \le l \le 10$ and they all equal the value of 26.

When we receive the measurement $z_{101} = 24$ our prior prediction for this value was $\hat{z}_{100}(1) = 26$, so using the update Equation 108 with $n = 100$ and $l = 10$ we find

$$\begin{aligned} z_{101}(9) &= z_{100}(10) + (1-\hat{\theta})(z_{101} - z_{100}(1)) \\ &= 26 + (1-0.6)(24-26) = 26 - 0.4(2) = 25.2\,. \end{aligned}$$

From Equation 109 this is the value of $z_{101}(l)$ for $1 \le l \le 9$ as well.

To compute a 95% confidence interval recall that in general it is given by

$$\hat{z}_n(l) \pm u_{\lambda/2}\{\hat{V}(e_n(l))\}^{1/2}. \tag{110}$$

Since an ARIMA(0,1,1) model has a forecast variance given by

$$V(e_n(l)) = \sigma^2(1 + (l-1)(1-\theta)^2), \tag{111}$$

we have confidence intervals specifically given by

$$\hat{z}_n(l) \pm u_{\lambda/2}\hat{\sigma}(1 + (l-1)(1-\theta)^2)^{1/2},$$

so our confidence intervals for the samples we are interested in $z_{102}, z_{103}, \cdots, z_{110}$ are given by (when $\hat{\sigma}^2 = 1$)

$$\begin{aligned} z_{101}(1) &\pm u_{\lambda/2} \\ z_{101}(2) &\pm u_{\lambda/2}(1 + (1-\hat{\theta})^2)^{1/2} \\ z_{101}(3) &\pm u_{\lambda/2}(1 + 2(1-\hat{\theta})^2)^{1/2} \\ &\vdots \\ z_{101}(9) &\pm u_{\lambda/2}(1 + 8(1-\hat{\theta})^2)^{1/2}. \end{aligned}$$

When we compute these nine confidence intervals for the values of $1 \le l \le 9$ we find

1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00
24.04	23.88	23.74	23.61	23.49	23.37	23.25	23.14	23.04
27.96	28.11	28.25	28.38	28.51	28.62	28.74	28.85	28.95

Each column contains the value of l, and the lower and upper bound of the given confidence interval. See the MATLAB file `prob_5_13.m` where we perform these calculations.

Exercise 5.14 (forecasting with an ARIMA(1,1,0) model)

Part (a): Note that an ARIMA(1,1,0) model $(1-\phi B)(1-B)z_t = a_t$ or

$$z_t = (1+\phi)z_{t-1} - \phi z_{t-2} + a_t \,,$$

can be obtained from the ARIMA(1,1,1) model discussed in the book by taking the deterministic trend parameter $\theta_0 = 0$ and the lag-one moving average coefficient $\theta = 0$. In the book we derive for those parameter settings

$$\begin{aligned}
z_n(1) &= (1+\phi)z_n - \phi z_{n-1} \\
z_n(2) &= (1+\phi)z_n(1) - \phi z_n \\
&\vdots \\
z_n(l) &= (1+\phi)z_n(l-1) - \phi z_n(l-2) \,.
\end{aligned}$$

Since $z_{49} = 33.4$ and $z_{50} = 33.9$ we can compute $z_{50}(1), z_{50}(2), \ldots z_{50}(5)$, using the above forecasting equations. We find

$$\begin{aligned}
z_n(1) &= 34.100 \\
z_n(2) &= 34.180 \\
z_n(3) &= 34.212 \\
z_n(4) &= 34.225 \\
z_n(5) &= 34.230 \,.
\end{aligned}$$

for the first five predictions.

To compute the prediction intervals about these forecasts we will use Equation 110 with $V(e_n(l))$ given by Equation 104 for $l = 1, \ldots, 5$. This requires ψ_l for an ARIMA(1,1,0) process. To get these values we will write the model as

$$z_t = \left(\frac{1}{1 - B - \phi B + \phi B^2} \right) a_t \,,$$

and compute the first five coefficients of the Taylor series of $\frac{1}{1-B-\phi B+\phi B^2}$. We find

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= 1+\phi \\
\psi_2 &= 1+\phi+\phi^2 \\
\psi_3 &= 1+\phi+\phi^2+\phi^3 \\
\psi_4 &= 1+\phi+\phi^2+\phi^3+\phi^4 \,,
\end{aligned}$$

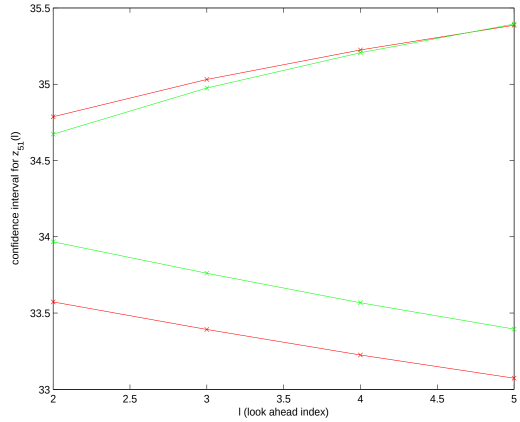


Figure 28: Plots of the confidence intervals for $z_{52}, z_{53}, \cdots, z_{55}$, before (in red) and after (in green) the measurement z_{51} is observed.

so that when we use

$$V(e_n(l)) = \sigma^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\,,$$

the confidence intervals we get are given for $l = 1, 2, \cdots, 5$ as

```
 1.0000    2.0000    3.0000    4.0000    5.0000
33.7472   33.5730   33.3927   33.2250   33.0731
34.4528   34.7870   35.0313   35.2246   35.3867
```

Part (b): To update these forecasts when the new measurement $z_{51} = 34.2$ is observed we will use Equation 100, Equation 110, and $V(e_n(l))$ given by Equation 104, this time evaluated at $n = 51$ and $l = 1, \ldots, 4$. When we do this we obtain the following new prediction intervals from z_{51} and $l \geq 1$ onwards

```
 1.0000    2.0000    3.0000    4.0000
33.9672   33.7610   33.5679   33.3951
34.6728   34.9750   35.2065   35.3947
```

In Figure 28 we show the confidence intervals for z_{52} through z_{55} before (in red) and after (in green) the measurement at z_{51} has be observed. Note that the addition of the measurement decreases our uncertainty in the location of the remaining points. See the MATLAB file `prob_5_14.m` and the Mathematica file `prob_5_14.nb` where we perform these calculations.

Exercise 5.15 (forecasts from an ARIMA(0,2,2) model)

The forecasts for an ARIMA(0,2,2) model like this were derived in the text. There it was noted that with $\theta_1 = 0.81$ and $\theta_2 = -0.38$ that

$$
\begin{aligned}
z_n(1) &= 2z_n - z_{n-1} - \theta_1 a_n - \theta_2 a_{n-1} \\
z_n(2) &= 2z_n(1) - z_{n-1} - \theta_2 a_n \\
z_n(3) &= 2z_n(2) - z_n(1) \\
&\vdots \\
z_n(l) &= 2z_n(l-1) - z_n(l-2) \quad \text{for} \quad l \geq 3\,.
\end{aligned}
$$

As stated in the exercise these predictions depend on estimates of a_{100} and a_{99} when predictive forward from $n = 100$. The derive these estimates recall that $a_n = z_n - z_n(1)$ so if we set a_{91} and a_{92} both equal to zero we can compute $z_n(1)$ for $n = 92, \cdots, 100$ and from these values compute a_n for $n = 92, \cdots, 100$. Specifically we compute

$$
\begin{aligned}
z_{92}(1) &= 2z_{92} - z_{91} - \hat{\theta}_1 a_{92} - \hat{\theta}_2 a_{91} = 2z_{92} - z_{91} \\
a_{93} &= z_{93} - z_{92}(1) \\
z_{93}(1) &= 2z_{93} - z_{92} - \hat{\theta}_1 a_{93} - \hat{\theta}_2 a_{92} = 2z_{92} - z_{91} - \hat{\theta}_1 a_{93} \\
a_{94} &= z_{94} - z_{93}(1) \\
&\vdots \\
z_{98}(1) &= 2z_{98} - z_{97} - \hat{\theta}_1 a_{98} - \hat{\theta}_2 a_{97} \\
a_{99} &= z_{99} - z_{98}(1) \\
z_{99}(1) &= 2z_{99} - z_{98} - \hat{\theta}_1 a_{99} - \hat{\theta}_2 a_{98} \\
a_{100} &= z_{100} - z_{99}(1)\,.
\end{aligned}
$$

With the two values a_{99} and a_{100} we can compute the forecasts $z_{100}(l)$ for $l = 1, 2, 3, \cdots, 10$ as requested. For the values of $a_{91}, a_{92}, \cdots, a_{100}$ we find

```
  0.00    0.00   -0.60   -1.28    0.58    0.86    0.47   -0.74   -0.18    0.63
```

and using these we find

```
18.5    19.0    19.6    20.1    20.7    21.3    21.8    22.4    22.9    23.5
```

for the values of $z_{100}(l)$ for $1 \leq l \leq 10$ expressed to one decimal. See the MATLAB script `prob_5_15.m` where we perform the numerics needed to obtain these results.

Exercise 5.16 (the autoregressive formulation of an ARIMA(0,1,1) model)

Part (a): From the given ARMA(1,1) model $(1-B)z_t = (1-0.8B)a_t$ which we can write as

$$\left(\frac{1-B}{1-0.8B}\right) z_t = a_t \,,$$

or expanding the fraction with the denominator $1-0.8B$ in its Taylor series we have

$$(1-B)\left(\sum_{j=0}^{\infty}(0.8B)^j\right) z_t = a_t \,.$$

Simplifying this summation on the left-hand-side we find

$$z_t = a_t + 0.2\sum_{j=1}^{\infty}(0.8)^{j-1} z_{t-j} \,.$$

From this expression we define $\bar{z}_{t-1} \equiv 0.2\sum_{j=1}^{\infty}(0.8)^{j-1}z_{t-j}$ and the coefficients π_j are given by

$$\pi_j = 0.2(0.8)^{j-1} \,,$$

for $j = 1, 2, \ldots$. Now consider the requested sum

$$\sum_{j=1}^{\infty}\pi_j = 0.2\sum_{j=0}^{\infty}(0.8)^j = 0.2\frac{1}{1-0.8} = 1 \,.$$

Part (b): The one-step-ahead forecast is given by

$$\begin{aligned}
z_t(1) &= E(z_{t+1}|z_t, z_{t-1}, \cdots) \\
&= E(\bar{z}_t + a_{t+1}|z_t, z_{t-1}, \cdots) \\
&= E(0.2\sum_{j=1}^{\infty} 0.8^{j-1} z_{t+1-j}|z_t, z_{t-1}, \cdots) \\
&= 0.2\sum_{j=1}^{\infty} 0.8^{j-1} z_{t+1-j} = \bar{z}_t \,.
\end{aligned}$$

The two step-ahead forecasts are of the form

$$\begin{aligned}
z_t(2) &= E(z_{t+2}|z_t, z_{t-1}, \cdots) \\
&= E(\bar{z}_{t+1} + a_{t+2}|z_t, z_{t-1}, \cdots) \\
&= E(0.2\sum_{j=1}^{\infty} 0.8^{j-1} z_{t+2-j}|z_t, z_{t-1}, \cdots) \\
&= 0.2\left(0.2\sum_{j=1}^{\infty} 0.8^{j} z_{t+1-j}\right) + 0.2\sum_{j=1}^{\infty} 0.8^{j} z_{t+1-j} \\
&= 0.2\left[\sum_{j=1}^{\infty}(0.2+0.8)0.8^{j-1} z_{t+1-j}\right] \\
&= 0.2\sum_{j=1}^{\infty} 0.8^{j-1} z_{t+1-j} = \bar{z}_t \,,
\end{aligned}$$

the same expression for $z_t(1)$.

Part (c): From the Exercise 5.18 (proved below) we have

$$\mathrm{Cov}(e_t(1), e_t(2)) = \sigma^2 \sum_{i=0}^{0} \psi_i \psi_{i+1} = \sigma^2 \psi_0 \psi_1 \,.$$

An ARIMA(0,1,1) model has $\psi_j = 1 - \theta$ for all $j \geq 1$ so $\psi_0 = 1$ and $\psi_1 = 1 - \theta$ and the above becomes

$$\mathrm{Cov}(e_t(1), e_t(2)) = \sigma^2 (1 - \theta) \,.$$

Exercise 5.17 (the covariance between forecast errors from different origins)

From the section on forecasting we have

$$e_t(l) = z_{t+l} - z_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-1} a_{t+1} = \sum_{i=0}^{l-1} \psi_i a_{t+l-i} \,,$$

with $\psi_0 = 1$. Since this expression has a zero mean and the a_t's are uncorrelated we have

$$\mathrm{Cov}(e_t(l), e_{t-j}(l)) = E(e_t(l) e_{t-j}(l)) \,.$$

Using the above summation notation but now for $e_{t-j}(l)$ (by shifting t j units to the left) we see that this expression is given by

$$e_{t-j}(l) = \sum_{i'=0}^{l-1} \psi_{i'} a_{t-j+l-i'} \,,$$

so that the product $e_t(l) e_{t-j}(l)$ is given by

$$e_t(l) e_{t-j}(l) = \sum_{i=0}^{l-1} \sum_{i'=0}^{l-1} \psi_i \psi_{i'} a_{t+l-i} a_{t-j+l-i'} \,.$$

The expectation passes through the summation and the above becomes

$$E(e_t(l) e_{t-j}(l)) = \sum_{i=0}^{l-1} \sum_{i'=0}^{l-1} \psi_i \psi_{i'} E(a_{t+l-i} a_{t-j+l-i'}) \,.$$

The individual expectations in the sum are zero unless $t + l - i = t - j + l - i'$ or $i = i' + j$ and the expectation is σ^2 when this equality is true. Since the range of the index i is $0 \leq i \leq l - 1$ and $i = i' + j$ when the expectation is non-zero we have that i' must satisfy $0 \leq i' + j \leq l - 1$ or $-j \leq i' \leq l - j - 1$. In addition, since i' is constrained to be $0 \leq i' \leq l - 1$ the overlapping region for non-zero expectation is given by

$$0 \leq i' \leq l - j - 1 \,.$$

Thus the sum above becomes

$$E(e_t(l)e_{t-j}(l)) = \sigma^2 \sum_{i'=0}^{l-j-1} \psi_{i'+j}\psi_{i'} \,.$$

On adding j to the summation limits this becomes

$$E(e_t(l)e_{t-j}(l)) = \sigma^2 \sum_{i'=j}^{l-1} \psi_{i'}\psi_{i'+j} \,,$$

the requested expression.

Exercise 5.18 (the covariance between forecast errors with different lead times)

As in exercise 5.17 consider

$$E(e_t(l)e_t(l+j)) = \sum_{i=0}^{l-1} \sum_{i'=0}^{l+j-1} \psi_i\psi_{i'} E(a_{t+l-i}a_{t+l+j-i'}) \,.$$

As before these expectations are zero unless $t+l-i = t+l+j-i'$ or $i = i'-j$ equivalently $i' = i+j$ depending on which index one wants to consider. Now the index i must be between $0 \leq i \leq l-1$ so that adding j to both sides of this inequality we have that $j \leq i+j \leq j+l-1$ or since $i' = i+j$ that i' is constrained as $j \leq i' \leq j+l-1$. But since i' itself is restricted to the range $0 \leq i' \leq l+j-1$ we see that the intersection of these two regions gives that

$$j \leq i' \leq l+j-1 \,.$$

All other combinations of indices give zero for the expectation. Using this in the expression above gives

$$E(e_t(l)e_t(l+j)) = \sigma^2 \sum_{i'=j}^{l+j-1} \psi_{i'}\psi_{i'-j} = \sigma^2 \sum_{i'=0}^{l-1} \psi_{i'+j}\psi_{i'} \,,$$

the requested expression.

Exercise 5.19 (the variance of the MMSE forecast of the sum)

Using the results from Exercise 5.18 we can evaluate this. We have for the variance of the forecast error the following

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{l=1}^{s} z_{t+l} - \sum_{l=1}^{s} z_t(l)\right) &= \operatorname{Cov}\left(\sum_{l=1}^{s} z_{t+l} - \sum_{l=1}^{s} z_t(l), \sum_{l=1}^{s} z_{t+l} - \sum_{l=1}^{s} z_t(l)\right) \\
&= \operatorname{Cov}\left(\sum_{l=1}^{s} e_t(l), \sum_{l=1}^{s} e_t(l)\right) \\
&= \sum_{l=1}^{s}\sum_{l'=1}^{s} \operatorname{Cov}(e_t(l), e_t(l')) \\
&= \sum_{l=1}^{s} \operatorname{Cov}(e_t(l), e_t(l)) + 2\sum_{l=1}^{s}\sum_{l'=l+1}^{s} \operatorname{Cov}(e_t(l), e_t(l')) \\
&= \sigma^2 \sum_{l=1}^{s}\left(\sum_{i=0}^{l-1} \psi_i^2\right) + 2\sum_{l=1}^{s}\sum_{l'=1}^{s-l} \operatorname{Cov}(e_t(l), e_t(l'+l)) \\
&= \sigma^2 s \sum_{i=0}^{l-1} \psi_i^2 + 2\sigma^2 \sum_{l=1}^{s}\sum_{l'=1}^{s-l}\left(\sum_{i=0}^{l-1} \psi_i \psi_{i+l'}\right).
\end{aligned}
$$

Now if z_t follows an ARIMA(0,1,1) model then $\psi_j = 1-\theta$ for all j and we have that

$$
\sum_{i=0}^{l-1} \psi_i^2 = (1-\theta)^2 l = \sum_{i=0}^{l-1} \psi_i \psi_{i+l'} \,,
$$

so the above expression becomes

$$
\sigma^2 s l (1-\theta)^2 + 2\sigma^2 l (1-\theta)^2 \sum_{l=1}^{s}\sum_{l'=1}^{s-l} 1 \,,
$$

Now the second sum above is given by

$$
\sum_{l=1}^{s}\sum_{l'=1}^{s-l} 1 = \sum_{l=1}^{s}(s-l) = (s-1) + (s-2) + \cdots + 2 + 1 = \sum_{l=1}^{s-1} l = \frac{1}{2}(s-1)s \,.
$$

Using this we finally arrive at

$$
\operatorname{Var}\left(\sum_{l=1}^{s} z_{t+l} - \sum_{l=1}^{s} z_t(l)\right) = \sigma^2 l (1-\theta)^2 s^2 \,.
$$

Exercise 5.20 (the eventual forecast functions)

For an ARIMA(p,d,q) model when $l > q$ the eventual forecast functions satisfy

$$
\phi(B)(1-B)^d z_n(l) = 0 \,, \tag{112}
$$

where the backwards shift operator, B, is now taken to operate on l. The solution to this difference equation is called the *eventual forecast function.*

Part (a): For $(1-B)z_t = (1-\theta_1 B - \theta_2 B^2)a_t$ or an ARIMA(0,1,2) model, for $l > 2$ the eventual forecast function satisfies

$$(1-B)z_t(l) = 0 \quad \text{or} \quad z_t(l) = z_t(l-1)\,.$$

Thus the eventual forecast for this model is a constant with an initial condition determined by the value of $z_n(1)$. If $\theta_2 = 0$ then we have the ARIMA(0,1,1) model

$$(1-B)z_t = (1-\theta_1 B)a_t\,,$$

or

$$z_t = z_{t-1} + a_t - \theta_1 a_{t-1}\,.$$

Thus our one-step-ahead predictions is

$$\begin{aligned} z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \dots) \\ &= E(z_n + a_{n+1} - \theta_1 a_n | z_n, z_{n-1}, \dots) \\ &= z_n - \theta_1 a_n\,. \end{aligned}$$

Since $z_n(l) = z_n(l-1)$ for $l > 1$ i.e. that all eventual forecasts are the same. That is

$$z_n(l) = z_n - \theta_1 a_n\,.$$

Part (b): For the ARIMA(1,3,1) model

$$(1-\phi B)(1-B)^3 z_t = (1-\theta B)a_t\,.$$

the eventual forecasts must satisfy (when $l > 1$) the following equation

$$(1-\phi B)(1-B)^3 z_n(l) = 0\,.$$

Exercise 5.21 (a linear trend with a stochastic intercept)

The linear trend model considered in this exercise is given by $z_t = \mu_t + \beta t + a_t$ with $\mu_t = \mu_{t-1} + \varepsilon_t$ where the sequences $\{a_t\}$ and $\{\varepsilon_t\}$ uncorrelated white-noise with variances given by $V(a_t) = \sigma^2$ and $V(\varepsilon_t) = \sigma_\varepsilon^2$.

Part (a): Note z_t is not stationary since $E(z_t) = E(\mu_t) + \beta t = \beta t$ is not independent of time, which is a necessary requirement to be stationary. Note that the first difference of z_t or

$$\begin{aligned} w_t &= z_t - z_{t-1} = \mu_t - \mu_{t-1} + \beta(t - t + 1) + a_t - a_{t-1} \\ &= \beta + \varepsilon_t + a_t - a_{t-1}\,, \end{aligned}$$

is stationary with a mean $w_t = z_t - z_{t-1}$ given by β.

Part (b): Lets compute the $1,2,3$ autocorrelations γ_k for w_t when $k = 1,2,3$. From the above since the mean of w_t is β we find

$$\begin{aligned}
\gamma_1 &= E((w_t - \beta)(w_{t-1} - \beta)) \\
&= E((\varepsilon_t + a_t - a_{t-1})(\varepsilon_{t-1} + a_{t-1} - a_{t-2})) \\
&= E(\varepsilon_t \varepsilon_{t-1}) + E(a_t a_{t-1}) - E(a_t a_{t-2}) - E(a_{t-1} a_{t-1}) + E(a_{t-1} a_{t-2}) \\
&= -\sigma^2 \,.
\end{aligned}$$

In the same way γ_2 is given by

$$\begin{aligned}
\gamma_2 &= E((w_t - \beta)(w_{t-2} - \beta)) \\
&= E(\varepsilon_t + a_t - a_{t-1})(\varepsilon_{t-2} + a_{t-2} - a_{t-3}) = 0 \,.
\end{aligned}$$

Finally γ_3 is given by

$$\begin{aligned}
\gamma_3 &= E((w_t - \beta)(w_{t-3} - \beta)) \\
&= E(\varepsilon_t + a_t - a_{t-1})(\varepsilon_{t-3} + a_{t-3} - a_{t-4}) = 0 \,.
\end{aligned}$$

Part (c): Since $\gamma_1 = -\sigma^2$ and $\gamma_k = 0$ for $k \geq 2$ this implies an ARIMA(0,1,1) model for z_t.

Part (d): If $\beta = 0$, then $z_t = \mu_t + a_t = \mu_{t-1} + \varepsilon_t + a_t$, which is a locally constant mean model. From the discussion in earlier chapters exponential smoothing is optimal in this case.

Exercise 5.22 (local mean model with stochastic mean)

Part (a): From Exercise 5.21 an approximation for z_t is given by an ARIMA(0,1,1) model. Note that the given model formulation leads to

$$z_t - z_{t-1} = \varepsilon_t + a_t - a_{t-1} \,,$$

Then the forecasts are predicted as

$$\begin{aligned}
z_n(l) &= E(z_{n+l} | z_n, z_{n-1}, \ldots) \\
&= E(z_{n+l-1} + \varepsilon_{n+l} + a_{n+l} - a_{n+l-1} | z_n, z_{n-1}, \ldots) \,.
\end{aligned}$$

To evaluate this expectation consider the situation when the term a_{n+l-1} needs to be retained in the forecast. This will happen if $n + l - 1 = n$ or $l = 1$. If $n + l - 1 > n$ then all of the shocks from ε_{n+l}, a_{n+l}, and a_{n+l-1} are unobservable and thus don't contributed to the forecasts. Thus we get

$$z_n(l) = \begin{cases} z_n(l-1) - a_n & l = 1 \\ z_n(l-1) & l > 1 \end{cases} \,.$$

To forecast using this model we need to be able to estimate a_n, which we can estimate using $z_n - z_{n-1}(1)$. Using the above we find

$$\begin{aligned}
z_n(1) &= z_n - a_n = 100.5 - 1 = 99.5 \\
z_n(2) &= z_n(1) = 99.5 \\
z_n(3) &= z_n(2) = 99.5 \,.
\end{aligned}$$

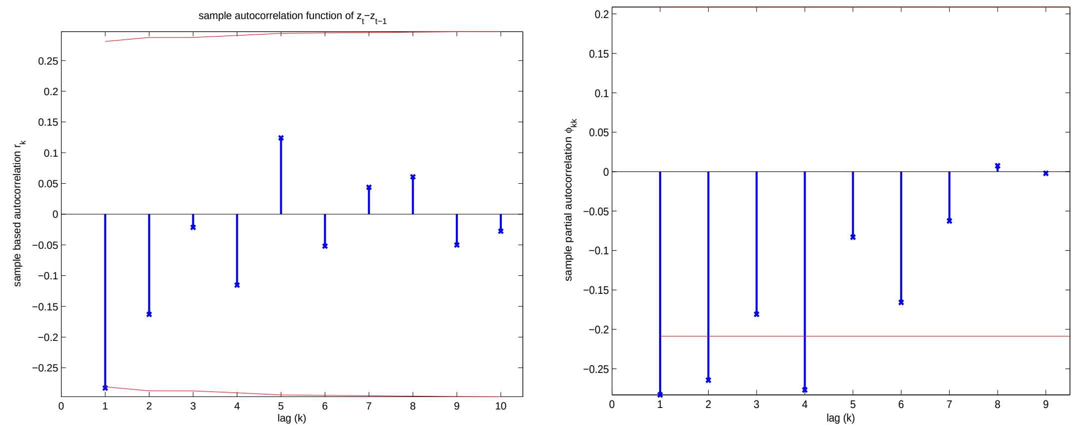


Figure 29: **Left:** The sample autocorrelation function of the first difference of the logarithm of the time series corresponding to the master card applications. **Right:** The sample partial autocorrelation function of the same.

Exercise 5.24 (master card applications)

For this problem we estimate an ARIMA model for the master card application data set. In the MATLAB script `prob_5_24.m` we load this data set and perform many of the steps above to determine the model. When this script is run it produces plots of the original data, the log transformed data, the first difference of this data, the sample autocorrelation function, and the sample partial autocorrelation function. These later two plots are shown in Figure 29. Based on this we will specify a ARIMA(2,1,1)

$$(1 - B)(1 - \phi_1 B - \phi_2 B^2)(z_t - \mu) = (1 - \theta B)a_t \,.$$

for this data set. Using the R function `prob_5_24.R` developed for this exercise we estimate these coefficients and find

$$\hat{\phi}_1 = 0.2642\,(0.1720) \quad \hat{\phi}_2 = -0.1518\,(0.1502) \quad \hat{\theta} = +0.743\,(0.1187)\,,$$

and $\hat{\sigma}^2 = 0.042$. Thus the only statistics significant coefficient is θ. We therefore drop the two AR coefficients and estimate only a MA(1) model. When we do this we find

$$\hat{\theta} = 0.6666\,(0.1316)\,,$$

and $\hat{\sigma}^2 = 0.0453$. Using the R function `predict` with the ARIMA developed here we predict the next four months to have values of

```
9.280366, 9.280366, 9.280366, 9.280366
```

The same value for all future predictions as would be expected from a ARIMA(0,1,1) model.

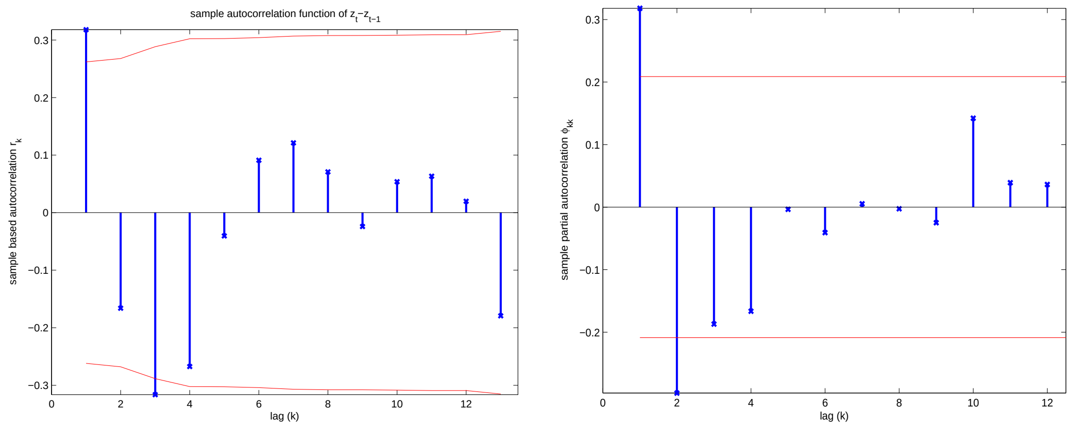


Figure 30: **Left:** The sample autocorrelation function of the first difference of the annual farm parity data set. **Right:** The sample partial autocorrelation function of this same data set.

Exercise 5.25 (annual farm parity ratios)

For this problem we estimate an ARIMA model for the annual farm parity ratio data set given with this exercise. In the MATLAB script `prob_5_25.m` we load this data set and perform many of the steps above to determine the model. When this script is run it produces plots of the original data, the first difference of this data, the sample autocorrelation function, and the sample partial autocorrelation function. These later two plots are shown in Figure 30. Based on this we might specify a ARIMA(2,1,0) model or

$$(1 - B)(1 - \phi_1 B - \phi_2 B^2)(z_t - \mu) = a_t\,,$$

for this data set. Using the R function `prob_5_25.R` developed for this exercise we estimate these coefficients and find

$$\hat{\phi}_1 = 0.5001\,(0.1248) \quad \hat{\phi}_2 = -0.3407\,(0.1231)\,,$$

and $\hat{\sigma}^2 = 39.01$. Using the R function `predict` we can predict forward five years to obtain

```
 74.84082 74.92062 74.67407 74.52358 74.53231
```

with standard errors given by

```
6.245749, 11.260293, 14.293271, 16.121295, 17.565683
```

Exercise 5.26 (modeling computer sales)

For this problem we estimate an ARIMA model for the computer sales data set given in Chapter 3 of this book. In the MATLAB script `prob_5_26.m` we load this data set and

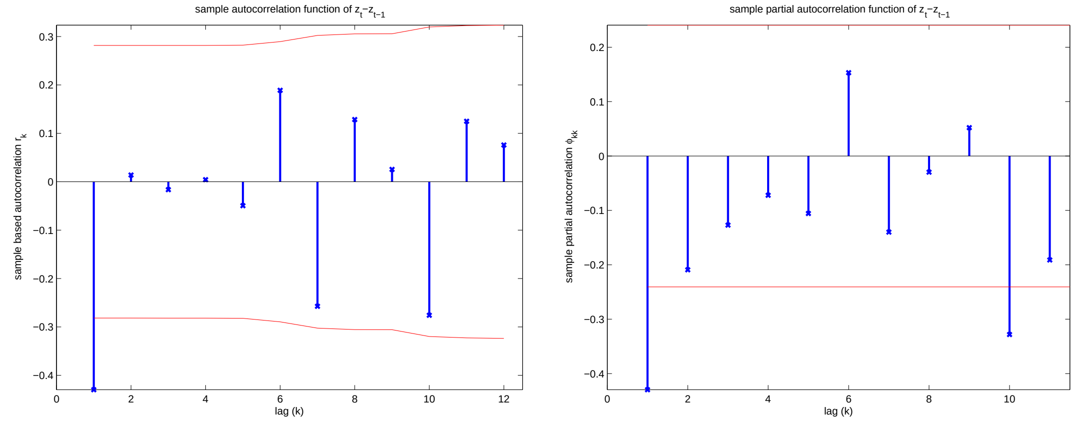


Figure 31: **Left:** The sample autocorrelation function of the difference between time series elements for the software sales data set. **Right:** The sample partial autocorrelation function of the same data.

perform many of the steps above to determine the model. When this script is run it produces plots of the original data, the first difference of this data, the sample autocorrelation function, and the sample partial autocorrelation function. These later two plots are shown in Figure 31. Based on this we will specify a ARIMA(0,1,1)

$$(1 - B)(z_t - \mu) = (1 - \theta_1 B)a_t \, .$$

for this data set. Using the R function `prob_5_26.R` we estimate these coefficients and find

$$\hat{\theta}_1 = 0.4647\,(0.1162) \quad \hat{\sigma}^2 = 2300 \, ,$$

Predicting ahead three months gives

```
412.3376 412.3376 412.3376
```

Exercise 5.27 (the random walk hypothesis)

For this problem we estimate an ARIMA model for the weekly closing prices of the SPY exchange traded fund (ETF). This represents a broad selection of stocks in the US stock market and is often thought to represent the "market". From yahoo finance we download these weekly closing prices and save them in the file `weekly_spy_prices.csv`. These prices are then loaded into MATLAB with the command `load_spy_data.m`. In the MATLAB script `prob_5_27.m` we load this data set and perform many of the steps above to determine the model. When this script is run it produces plots of the original data, the first difference of this data, the sample autocorrelation function, and the sample partial autocorrelation function. These later two plots are shown in Figure 32. These plots indicate that the weekly prices seem to follow a random walk model

$$z_t = z_{t-1} + a_t \, ,$$

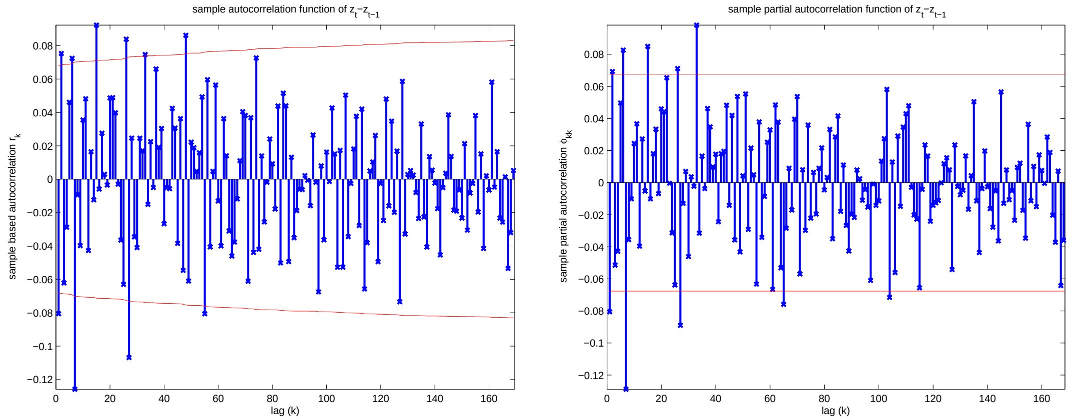


Figure 32: **Left:** The sample autocorrelation function for the first difference of SPY weekly closing prices. **Right:** The sample partial autocorrelation function of the same data set.

with a_t a white-noise process. There are some autocorrelations that appear to be significant but these might just be due to sample variation. If we choose to look at a more recent time frame the results are much the same in that there seems to be no easy linear model of this type that predicts this time series.

Chapter 6: Seasonal Autoregressive Integrated Moving Average Models

Notes On The Text

Examples of Seasonal ARIMA Models

Here and in the following subsections to verify understanding we will duplicate many of the seasonal ARIMA example presented in the book. Rather than do all of the work ourselves we will perform the initial data exploration by hand and then use the R function `arima` to find the parameter estimates needed to implement the hypothetical model. The R function `arima` is nice in that it allows one to construct *seasonal* arima models directly which are needed for this chapter.

Note, however, that when calling the `arima` function one can obtain *different* results for the SARIMA model parameter estimates if we use the option for producing automatic difference when calling this function or if we explicitly compute the differences ourselves before calling the function. For example in the `gas usage` data set (assuming the variable `Y` contains the time series data of interest) a call like

```
ord <- c(0,0,1)
ssn <- list(order=c(0,1,1),period=12)
arima( Y, order=ord, seasonal=ssn, method="ML" )
```

will produce *different* parameter estimates than the function call

```
Yd  <- diff(Y,lag=12,differences=1)
ord <- c(0,0,1)
ssn <- list(order=c(0,0,1),period=12)
arima( Yd, order=ord, seasonal=ssn, include.mean=TRUE, method="ML" )
```

On this data set the second example produces parameter estimates that more closely match the ones presented in the book, but since the first example is easier to use directly in the R function `predict` we will often perform model verification (plotting and verifying the values of the sample autocorrelation function of the model fit residuals are insignificant) using it.

Examples of Seasonal ARIMA Models: Gas Usage

Here duplicate the results presented in the book on modeling the time series of gas usage presented in the book. In the MATLAB script `model_gas_section_6_7_2.m`, using the MATLAB function `load_gas_usage.m` developed in this chapter we load the data and plot many

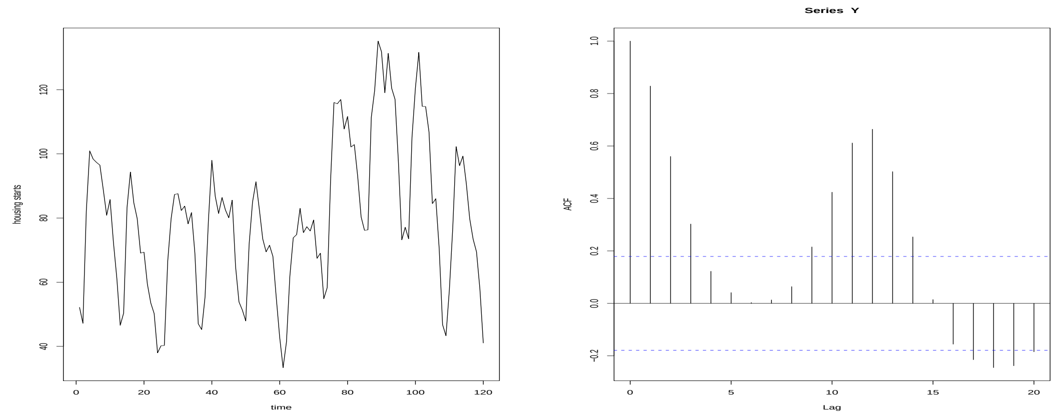


Figure 33: **Left:** The raw data for the monthly U.S. housing starts of single-family structures for the period January 1965 to December 1974. **Right:** The direct sample autocorrelation function for the U.S. housing starts data set.

of the figures provided in the book. We then use the R function `model_gas_section_6_7_2.R` to estimate the specific coefficients of the given model and observe that after doing so the residuals of the given model are all insignificant (the sign of a good model). Running the above codes demonstrates these conclusions.

Examples of Seasonal ARIMA Models: Housing Starts

Here we duplicate the results presented in the book on modeling housing starts. Rather than use a combined MATLAB/R approach we will instead perform all modeling in R. In the R script `model_housing_starts_section_6_7_3.R`, we begin by loading and plotting the data in Figure 33 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 33 (right). This autocorrelation has the characteristic shape of a periodic series that needs to be differenced before it will become stationary. Note that from the sample autocorrelation function plot the period appears to be 12, corresponding to monthly variations. We take the difference $1 - B^{12}$ and replot the autocorrelation function of this difference in Figure 34 (left). There we see the characteristic slow decay of a non-stationary time series. Because of this we next consider the first difference of the series $(1-B^{12})z_t$ plots the corresponding autocorrelation in Figure 34 (middle). This later difference is stationary. From the pattern of the autocorrelation function for $(1 - B)(1 - B^{12})z_t$ we will assume a single moving average (MA(1)) term and a single seasonal moving average (SMA(1)) term. Thus, if we assume a model of the form

$$(1 - B)(1 - B^{12})z_t = \theta_0 + (1 - \theta B)(1 - \Theta B^{12})a_t \,,$$

we find coefficients of this model (and their standard errors given by)

$$\theta = 0.27(0.08) \quad \Theta = 0.99(0.32) \quad \theta_0 = -0.07(0.14) \quad \sigma^2 = 38.26 \,,$$

dropping the insignificant mean component θ_0 and refitting gives parameter estimates that are almost identical to the ones given above. These parameters also agree relatively well with

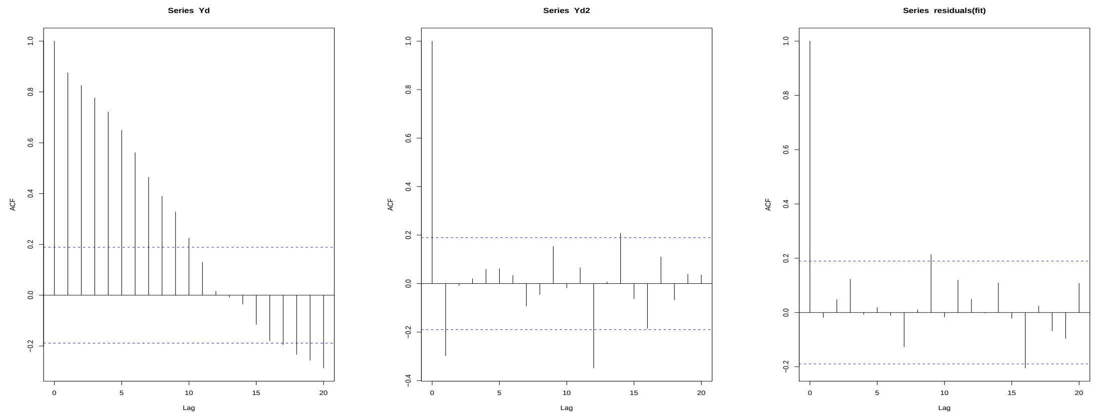


Figure 34: **Left:** The sample autocorrelation function of $(1 - B^{12})z_t$ for the housing starts data set. Note the strong long range correlation indicating another difference needs to be taken to make this series stationary. **Center:** The sample autocorrelation function of $(1-B)(1-B^{12})z_t$ for the housing starts data set. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the model found for the housing starts data set.

the ones presented in the book. Finally, for model validation we plot the autocorrelations of residuals of the fitted model in Figure 34 (right).

Examples of Seasonal ARIMA Models: Car Sales

Here we duplicate the results presented in the book on the car sales data set. In the R script `model_car_sales_section_6_7_3.R`. We load and plot the data in Figure 35 (left). When we plot the sample autocorrelation of this time series directly we get the plot shown in Figure 35 (right). This has the characteristic shape of a periodic series that needed to be seasonally differenced before it will become stationary. Note that the period again appears to be 12 corresponding to monthly variations. We next compute the differences $(1 - B^{12})z_t$ and replot the autocorrelation function in Figure 36 (left). The significant autocorrelations are present at the lags $k = 1$, 2, 12, and 17. To model these we might try to first eliminate the autocorrelation at the lags $k = 1$, 2, and 12 first. To do this we could attempt to fit either an MA(2) or a AR(2) model to $(1-B^{12})z_t$ along with a SMA(1). By stationarity and invertablilty these two models are similar if the coefficients AR coefficients (ϕ_l) and moving average coefficients (θ_l) are small. Trying both models in R we find that both models are quite similar as far as their log likelihoods are concerned. As the AR(2) model is the one considered in the book we present our results here. For the model

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})z_t = \theta_0 + (1 - \Theta B^{12})a_t \,,$$

we find coefficients of this model (and their standard errors given by)

$$\theta_0 = 0.99(0.17) \quad \phi_1 = 0.26(0.10) \quad \phi_2 = 0.20(0.11) \quad \Theta = 0.49(0.15) \quad \sigma^2 = 2.218 \,,$$

which are somewhat different values than the book quotes.

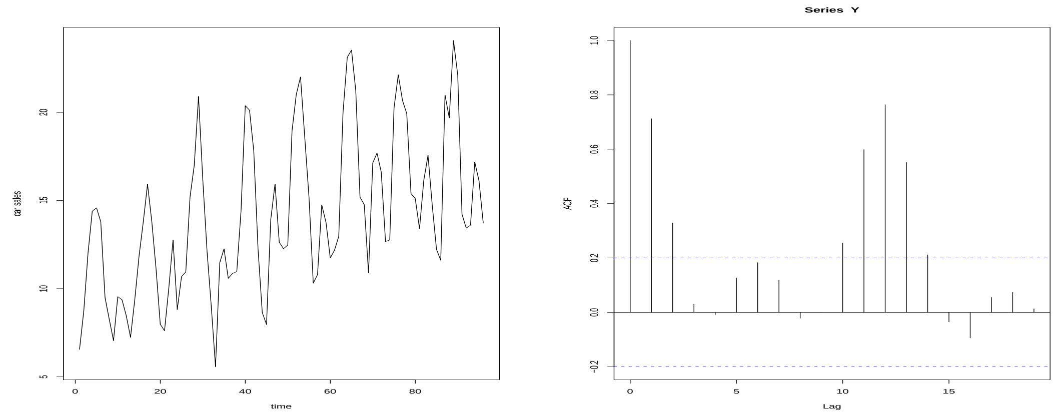


Figure 35: **Left:** The raw data for new car sales in Quebec from January 1960 to December 1968. **Right:** The direct sample autocorrelation function for the new car sales in Quebec data set.

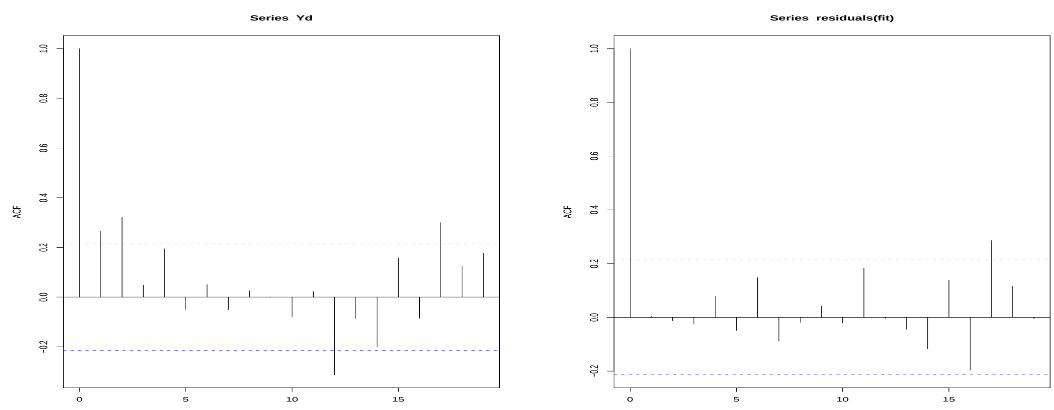


Figure 36: **Left:** The sample autocorrelation function of $(1 - B^{12})z_t$ for the car sales data set. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the model found for the car sales data set. Note there still seems to be a significant autocorrelation at lag $k = 17$.

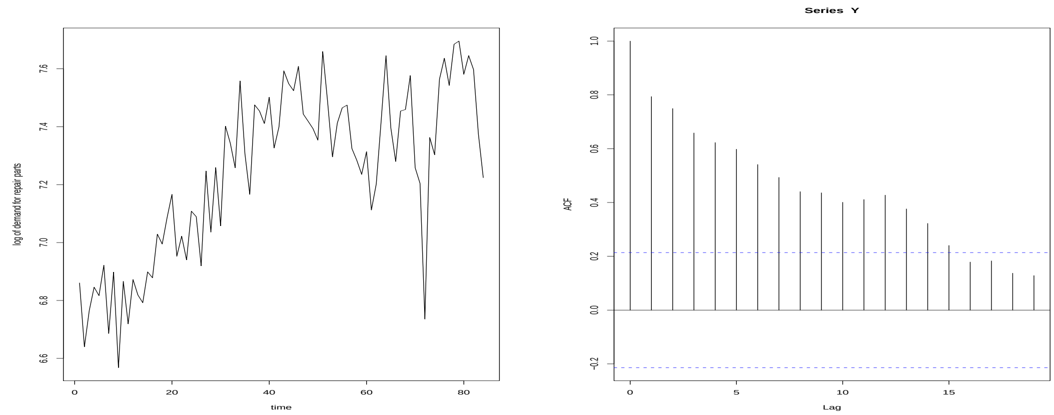


Figure 37: **Left:** The raw data for the log of the demand for repair parts data set. **Right:** The direct sample autocorrelation function for the log of the parts demand data set.

dropping the insignificant mean component and refitting gives parameter estimates that are almost identical to the ones given above. These parameters also agree very well with the ones presented in the book. Finally, for model validation we plot the autocorrelations of residuals of the fitted model in Figure 36 (right).

Examples of Seasonal ARIMA Models: Demand for Repair Parts

Here we duplicate the results presented in the book on the car sales data set. In the R script `model_demand_section_6_7_5.R`, we load and plot the log of the raw data for this example in Figure 37 (left). We next plot the sample autocorrelation of this time series directly and display it in Figure 37 (right). This has the characteristic shape of a periodic series that needed to be differenced before it will become stationary. We next take the difference $1-B$ and replot the autocorrelation function in Figure 38 (left). This autocorrelation structure looks like it could be fit well with a MA(1) model. If we assume a model of the form

$$(1-B)\log(z_t) = \theta_0 + (1-\theta B)b_t \,,$$

we find coefficients of this model (and their standard errors given)

$$\theta = 0.59(0.12) \quad \theta_0 = 0.00(0.00) \quad \sigma^2 = 0.025 \,,$$

dropping the insignificant mean component and refitting gives parameter estimates that are almost identical to the ones given above. We next plot the sample autocorrelation of the residuals b_t for the above fit in Figure 38 (center). Note the significant value of the autocorrelation at lag $k = 12$. To model this we will assume a seasonal MA(1) model for the residuals b_t. That is we assume that b_t satisfies

$$b_t = (1 - \Theta B^{12})a_t \,,$$

for some white noise process a_t. This in tern means that the process $\log(z_t)$ satisfies

$$(1-B)\log(z_t) = (1-\theta B)(1-\Theta B^{12})a_t \,,$$

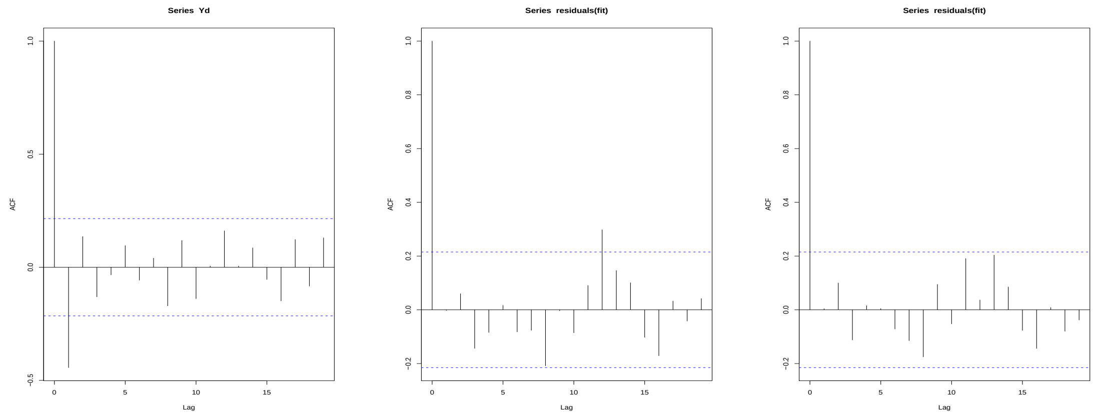


Figure 38: **Left:** The sample autocorrelation function of $(1 - B)\log(z_t)$ for the demand for repair parts. **Center:** The same autocorrelation function $r_{\hat{a}}$ for the non-seasonal model found for the demand for repair parts sales data set. Note there still seems to be a significant autocorrelation at lag $k = 12$. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the seasonal model $(1 - B)\log(z_t) = (1 - \theta B)(1 - \Theta B^{12})a_t$ found for the demand for repair parts sales data set. Note that there no longer are any significant autocorrelations.

or a SARIMA model of $(0, 1, 1)(0, 0, 1)_{12}$. Fitting this model in the R script `model_demand_6_7_5.R` and plotting the autocorrelation function for the resulting residuals gives the plot in Figure 38 (right). In this case, all autocorrelations are insignificant.

Appendix 6 (computing the autocorrelation of an $(0, d, 1)(1, D, 1)_{12}$ model)

In this section of these notes we will perform some of the steps found in the appendix to this chapter aimed at computing the autocorrelation function for the $(0, d, 1)(1, D, 1)_{12}$ model. To do this we will derive two equations relating γ_0 and γ_{12} and solve these to obtain explicit expressions for γ_0 and γ_{12}. We will then do the same procedure for γ_1 and γ_{11}, for γ_2 and γ_{10}, etc down to γ_6. After these are obtained we will then derive the requested expressions for γ_k for $k \geq 13$. To begin this process we recognize that when written out in full the given model for the stationary difference $w_t \equiv (1 - B)^d (1 - B^s)^D z_t$ is given by

$$w_t = \Phi w_{t-12} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13} \,. \tag{113}$$

Multiplying this expression by w_t and taking expectations we would have

$$E(w_t^2) = \Phi E(w_t w_{t-12}) + E(w_t a_t) - \theta E(w_t a_{t-1}) - \Theta E(w_t a_{t-12}) + \theta\Theta E(w_t a_{t-13}) \,,$$

or written in terms of the autocorrelation functions $\gamma_k \equiv E(w_t w_{t-k})$ we have

$$\gamma_0 = \Phi\gamma_{12} + E(w_t a_t) - \theta E(w_t a_{t-1}) - \Theta E(w_t a_{t-12}) + \theta\Theta E(w_t a_{t-13}) \,. \tag{114}$$

Now to further evaluate this expression note that if we had the Wold decomposition of the stationary difference w_t, that is an expansion of w_t in terms of a linear filter representation as

$$w_t = a_t + \sum_{l=1}^{\infty} \psi_l a_{t-l} = \sum_{l=0}^{\infty} \psi_l B^l a_t \,,$$

then it is easy to compute the remaining expectations, $E(w_t a_{t-k})$, above since they are given by

$$E(w_t a_{t-k}) = \psi_k E(a_t^2) = \psi_k \sigma^2 \, .$$

Since in Equation 114 the above expansion we only need to determine some of the expectations $E(w_t a_{t-k})$ (that is we only need these values for $E(w_t a_t)$, $E(w_t a_{t-1})$ and $E(w_t a_{t-12})$ we don't need to compute ψ_k for all k. An easy way to compute ψ_k for a few values of k is to recognized that in the Wold decomposition ψ_k is exactly equal to the kth coefficient in the Taylor series expansion of the back-shift operator, B, that when applied to a_t produces w_t or $\frac{\pi(B)}{\phi(B)}$. Thus for the specific model given here we see that

$$w_t = \left(\frac{\pi(B)}{\phi(B)}\right) a_t = \left(\frac{1 - \theta B - \Theta B^{12} + \theta\Theta B^{13}}{1 - \Phi B^{12}}\right) a_t \, .$$

Therefore to evaluate $E(w_t a_{t-k})$ for $k = 0, 1, 12$ we need to find is the kth coefficient with respect to the variable B in the Taylor expansion of

$$\frac{1 - \theta B - \Theta B^{12} + \theta\Theta B^{13}}{1 - \Phi B^{12}} \, .$$

The values can be explicitly computed from derivatives of this expression. From Taylor's theorem we have

$$E(w_t a_{t-k}) = \frac{1}{k!} \frac{d^k}{dB^k} \left(\frac{1 - \theta B - \Theta B^{12} + \theta\Theta B^{13}}{1 - \Phi B^{12}}\right)\Bigg|_{B=0} \sigma^2 \, . \tag{115}$$

We can use an algebraic manipulation package like Mathematica to compute the specific values of $E(w_t a_{t-k})$ we need. In the MATHEMATICA file `appendix_6.nb` we find

$$\begin{aligned}
E(w_t a_t) &= \sigma^2 \\
E(w_t a_{t-1}) &= -\theta\sigma^2 \\
E(w_t a_{t-12}) &= (\Phi - \Theta)\sigma^2 \\
E(w_t a_{t-13}) &= \theta(-\Phi + \Theta)\sigma^2 \, .
\end{aligned}$$

Thus when we put these into Equation 114 we obtain

$$\begin{aligned}
\gamma_0 &= \Phi\gamma_{12} + \sigma^2 + \theta^2\sigma^2 - \Theta(\Phi - \Theta)\sigma^2 + \theta^2\Theta(-\Phi + \Theta)\sigma^2 \\
&= \Phi\gamma_{12} + \sigma^2(1 + \theta^2)(1 + \Theta(\Theta - \Phi)) \, ,
\end{aligned}$$

which agrees with the result in the book.

The above equation has two unknowns γ_0 and γ_{12}. To obtain another equation that relates them we multiply Equation 113 by w_{t-12} to get

$$w_t w_{t-12} = \Phi w_{t-12}^2 + a_t w_{t-12} - \theta a_{t-1} w_{t-12} - \Theta a_{t-12} w_{t-12} + \theta\Theta a_{t-13} w_{t-12} \, .$$

Taking expectations of this we get

$$\gamma_{12} = \Phi\gamma_0 + E(a_t w_{t-12}) - \theta E(a_{t-1} w_{t-12}) - \Theta E(a_{t-12} w_{t-12}) + \theta\Theta E(a_{t-13} w_{t-12}) \, . \tag{116}$$

Again from the Wold decomposition for w_t we have

$$w_t = \psi(B)a_t = a_t + \sum_{l=1}^{\infty} \psi_l a_{t-l} \,,$$

so that shifting time by twelve units gives

$$w_{t-12} = a_{t-12} + \sum_{l=1}^{\infty} \psi_l a_{t-12-l} \,.$$

From this we can compute that

$$\begin{aligned}
E(w_{t-12}\, a_t) &= 0 \\
E(w_{t-12}\, a_{t-1}) &= 0 \\
E(w_{t-12}\, a_{t-12}) &= \sigma^2 \\
E(w_{t-12}\, a_{t-13}) &= \psi_1 \sigma^2 \,.
\end{aligned}$$

Now from before $\psi_1 = -\theta$ so Equation 116 becomes

$$\gamma_{12} = \Phi\gamma_0 - \Theta\sigma^2(1+\theta^2) \,,$$

which is the second equation for γ_0 and γ_{12}. Solving these two equations gives values for γ_0 and γ_{12} that are given in the book.

Next to derive equations for γ_1 and γ_{11} compute the expectation of the product $w_t w_{t-1}$ as

$$\begin{aligned}
\gamma_1 &= E(w_t w_{t-1}) \\
&= E(w_{t-1}(\Phi w_{t-12} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13})) \\
&= \Phi\gamma_{11} + E(w_{t-1}a_t) - \theta E(w_{t-1}a_{t-1}) - \Theta E(w_{t-1}a_{t-12}) + \theta\Theta E(w_{t-1}a_{t-13}) \,.
\end{aligned}$$

To continue this evaluation we need to compute $E(w_{t-1}a_t)$, $E(w_{t-1}a_{t-1})$, $E(w_{t-1}a_{t-12})$, and $E(w_{t-1}a_{t-13})$. We can evaluate many of these expectations by shifting the time index. We find

$$\begin{aligned}
E(w_{t-1}a_t) &= E(w_t a_{t+1}) = 0 \\
E(w_{t-1}a_{t-1}) &= E(w_t a_t) = \sigma^2 \\
E(w_{t-1}a_{t-13}) &= E(w_t a_{t-12}) = (\Phi - \Theta)\sigma^2 \,.
\end{aligned}$$

The one new expression we need to evaluate is $E(w_{t-1}a_{t-12}) = E(w_t a_{t-11})$. In the MATHEMATICA file we find this to be zero. Thus with these expectations evaluated we obtain

$$\gamma_1 = \Phi\gamma_{11} - \theta\sigma^2 + \Theta\theta(\Phi - \Theta)\sigma^2 \,,$$

which is equation 5 in the book. Now γ_{11} is given by

$$\begin{aligned}
\gamma_{11} &= E(w_t w_{t-11}) \\
&= E((\Phi w_{t-12} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13})w_{t-11}) \\
&= \Phi\gamma_1 - \Theta E(a_{t-12}w_{t-11}) + \theta\Theta E(a_{t-13}w_{t-11}) \\
&= \Phi\gamma_1 - \Theta E(a_{t-1}w_t) + \theta\Theta E(a_{t-2}w_t) \\
&= \Phi\gamma_1 + \Theta\theta\sigma^2 \,,
\end{aligned}$$

since $E(a_{t-2}w_t) = 0$. This is equation 6 in the book. These two equations can be solved for γ_1 and γ_{11} giving the results found in the book.

Computing the values of γ_2 and γ_{10} in the same way as the others thus far we find

$$\begin{aligned}
\gamma_2 &= E(w_t w_{t-2}) \\
&= E(w_{t-2}(\Phi w_{t-12} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13})) \\
&= \Phi\gamma_{10} + E(w_{t-2}a_t) - \theta E(w_{t-2}a_{t-1}) - \Theta E(w_{t-2}a_{t-12}) + \theta\Theta E(w_{t-2}a_{t-13}) \\
&= \Phi\gamma_{10} \,,
\end{aligned}$$

and

$$\begin{aligned}
\gamma_{10} &= E(w_t w_{t-10}) \\
&= E(w_{t-10}(\Phi w_{t-12} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13})) \\
&= \Phi\gamma_2 - \Theta E(w_{t-10}a_{t-12}) + \theta\Theta E(w_{t-10}a_{t-13}) \\
&= \Phi\gamma_2 \,,
\end{aligned}$$

so $\gamma_2 = \gamma_{10} = 0$. From these final arguments the remaining claims given in the book that $\gamma_3 = \gamma_4 = \cdots = \gamma_{10} = 0$ and $\gamma_k = \Phi\gamma_{k-12}$ for $k > 13$ is seem reasonable.

Exercise Solutions

Exercise 6.1 (computing the ACF)

This problem is aimed as comparing non-multiplicative models like that in Part (a) with multiplicative models like that in Part (b).

Computing the autocorrelation function in this case is easy since z_t is already written in a Wold decomposition, thus

$$\begin{aligned}
z_{t-k} &= (1 - \theta_1 B - \theta_{12}B^{12} - \theta_{13}B^{13})a_{t-k} \\
&= (B^k - \theta_1 B^{k+1} - \theta_{12}B^{k+12} - \theta_{13}B^{k+13})a_t \,,
\end{aligned}$$

so we can compute $E(z_t z_{t-k})$ by multiplying the polynomial

$$1 - \theta_1 B - \theta_{12}B^{12} - \theta^{13}B^{13}$$

(representing z_t) with

$$B^k - \theta_1 B^{k+1} - \theta_{12}B^{k+12} - \theta_{13}B^{k+13}$$

representing z_{t-k} and then read off the coefficients that would be produced when the expectation is taken. In the MATHEMATICA file `prob_6_1.nb` we perform this multiplication and find (everything should be multiplied by σ^2)

$$\begin{aligned}
\gamma_0 &= 1 \\
\gamma_1 &= -2\theta_1 \\
\gamma_2 &= \theta_1^2 \\
\gamma_{12} &= -2\theta_{12} \\
\gamma_{13} &= 2\theta_1\theta_{12} - 2\theta_{13} \\
\gamma_{14} &= 2\theta_1\theta_{13} \\
\gamma_{24} &= \theta_{12}^2 \\
\gamma_{25} &= 2\theta_{12}\theta_{13} \\
\gamma_{26} &= \theta_{13}^2 \,,
\end{aligned}$$

and zero otherwise. Since γ_0 (normalized by σ^2) is one these are also the expressions for ρ_k.

Part (b): Note that this model is the special case of the model $(0,0,1)(0,0,1)_{12}$ so we can consider model number 3 from this chapter. There it was found that model has an autocorrelation function given by

$$\rho_k = \begin{cases} 1 & k = 0 \\ -\frac{\theta}{1+\theta^2} & k = 1 \\ \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} & k = 11 \\ -\frac{\Theta}{1+\Theta^2} & k = 12 \\ \rho_{11} & k = 13 \\ 0 & \text{otherwise} \end{cases}$$

Exercise 6.2 (more autocorrelation functions)

As in Exercise 6.1 since these models are already given in the Wold decomposition form, computing the autocorrelation functions for them is relatively easy.

Part (a): Using the MATHEMATICA file `prob_6_2.nb` *some* of the coefficients are

$$\begin{aligned}
\gamma_0 &= 1 \\
\gamma_1 &= -2\theta_1 \\
\gamma_2 &= \theta_1^2 - 2\theta_2 \\
\gamma_3 &= 2\theta_1\theta_2 \\
\gamma_4 &= \theta_2^2 \\
\gamma_{12} &= -2\Theta \\
\gamma_{13} &= 4\Theta\theta_1 \,,
\end{aligned}$$

the remaining coefficients are found in the above MATHEMATICA file.

Part (b): For this part of the problem the value of the autocorrelation function for this model can found in the above MATHEMATICA file.

Exercise 6.3 (observations of the sample ACF)

Part (a): If the model $y_t \equiv (1 - B^{12})z_t = \alpha_t$, where the residual series α_t has an autocorrelation function that dies down and a sample partial autocorrelation function with one spike at lag 1, then possibly an AR(1) model could be used to model the residual α_t. Doing this the model for z_t would become

$$(1 - \phi B)(1 - B^{12})z_t = a_t \,,$$

or $(1, 0, 0)(0, 1, 0)_{12}$ model.

Part (b): If after this modification our model now has an autocorrelation function with a spike at lag 12 then maybe we need to add a seasonal MA(1) component. That is the residual above, a_t, may satisfy $a_t = (1 - \Theta B^{12})b_t$, so the resulting model for z_t would be

$$(1 - \phi B)(1 - B^{12})z_t = (1 - \Theta B^{12})b_t \,,$$

or a $(1, 0, 0)(0, 1, 1)_{12}$ model.

Exercise 6.4 (the $\beta_{1,2}^{(t+1)}$ updating equations)

To solve this expression we will evaluate the given expression for z_{t+l}

$$z_{t+l} = \beta_1^{(t)} \sin(\frac{2\pi l}{12}) + \beta_2^{(t)} \cos(\frac{2\pi l}{12}) + e_t(l) \,,$$

at z_{t+l+1} in two ways. The first is based on considering z_{t+l+1} as an increment in t and not in l. Thus we evaluate z_{t+l+1} centered in time on the point $t+1$ to get

$$z_{t+l+1} = \beta_1^{(t+1)} \sin(\frac{2\pi l}{12}) + \beta_2^{(t+1)} \cos(\frac{2\pi l}{12}) + e_{t+1}(l) \,, \tag{117}$$

this is basically just evaluating $\beta_{1,2}^{(t)}$ at the time $t+1$. The second way is to expand z_{t+l+1} about the point t (with l incremented by one) to get

$$z_{t+l+1} = \beta_1^{(t)} \sin(\frac{2\pi(l+1)}{12}) + \beta_2^{(t)} \cos(\frac{2\pi(l+1)}{12}) + e_t(l+1) \,. \tag{118}$$

Equating Equation 117 with Equation 118 for $l = 1, 2$ gives

$$\begin{aligned}
\beta_1^{(t+1)} \sin(\frac{\pi}{6}) + \beta_2^{(t+1)} \cos(\frac{\pi}{6}) + e_{t+1}(1) &= \beta_1^{(t)} \sin(\frac{\pi}{3}) + \beta_2^{(t+1)} \cos(\frac{\pi}{3}) + e_t(2) \\
\beta_1^{(t+1)} \sin(\frac{\pi}{3}) + \beta_2^{(t+1)} \cos(\frac{\pi}{3}) + e_{t+1}(2) &= \beta_1^{(t)} \sin(\frac{\pi}{2}) + \beta_2^{(t+1)} \cos(\frac{\pi}{2}) + e_t(3) \,,
\end{aligned}$$

or in matrix form

$$\begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} \beta_1^{(t+1)} \\ \beta_2^{(t+1)} \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^{(t)} \\ \beta_2^{(t)} \end{bmatrix} + \begin{bmatrix} e_t(2) - e_{t+1}(1) \\ e_t(3) - e_{t+1}(2) \end{bmatrix} .$$

Inverting the matrix on the left hand side gives the update equation of

$$\begin{bmatrix} \beta_1^{(t+1)} \\ \beta_2^{(t+1)} \end{bmatrix} = \begin{bmatrix} -1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3}/2 & 1/2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1^{(t)} \\ \beta_2^{(t)} \end{bmatrix} + \begin{bmatrix} -1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \begin{bmatrix} e_t(2) - e_{t+1}(1) \\ e_t(3) - e_{t+1}(2) \end{bmatrix} .$$

Now recall that the error in the l-step-ahead prediction, $e_t(l)$, is given by

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \cdots + \psi_{l-2} a_{t+2} + \psi_{l-1} a_{t+1} .$$

Evaluating this expression for $e_{t+1}(l)$ and $e_t(l+1)$ we find

$$\begin{aligned} e_t(l+1) &= a_{t+l+1} + \psi_1 a_{t+l} + \psi_2 a_{t+l-1} + \cdots + \psi_{l-2} a_{t+3} + \psi_{l-1} a_{t+2} + \psi_l a_{t+1} \\ e_{t+1}(l) &= a_{t+1+l} + \psi_1 a_{t+l} + \psi_2 a_{t+l-1} + \cdots + \psi_{l-2} a_{t+3} + \psi_{l-1} a_{t+2} . \end{aligned}$$

Thus the *difference* $e_t(l+1) - e_{t+1}(l)$ is therefore given simply by

$$e_t(l+1) - e_{t+1}(l) = \psi_l a_{t+1} .$$

Since for this model the first two linear representation coefficients ψ_l are given by $\psi_1 = \sqrt{3} - \theta_1$ and $\psi_2 = 2 - \sqrt{3}\theta_1 - \theta_2$ the above becomes

$$\begin{aligned} \begin{bmatrix} \beta_1^{(t+1)} \\ \beta_2^{(t+1)} \end{bmatrix} &= \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} \beta_1^{(t)} \\ \beta_2^{(t)} \end{bmatrix} + \begin{bmatrix} -1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} - \theta_1 \\ 2 - \sqrt{3}\theta_1 - \theta_2 \end{bmatrix} a_{t+1} \\ &= \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} \beta_1^{(t)} \\ \beta_2^{(t)} \end{bmatrix} + \begin{bmatrix} \sqrt{3} - 2\theta_2 - \sqrt{3}\theta_2 \\ 1 + \theta_2 \end{bmatrix} a_{t+1} , \end{aligned}$$

which is the result quoted in the book.

Exercise 6.5 (more $\beta_i^{(t+1)}$ updating equations)

For the model is given by equation 6.15 in the book

$$z_{t+l} = \beta_0^{(t)} + \beta_{26}^{(t)}(-1)^l + \sum_{j=1}^{5} \left(\beta_{1j}^{(t)} \sin(\frac{2\pi j l}{12}) + \beta_{2j}^{(t)} \cos(\frac{2\pi j l}{12}) \right) + e_t(l) , \quad (119)$$

by incrementing t to $t+1$ we obtain the representation in terms of the coefficients $\beta^{(t+1)}$ as

$$z_{t+l+1} = \beta_0^{(t+1)} + \beta_{26}^{(t+1)}(-1)^l + \sum_{j=1}^{5} \left(\beta_{1j}^{(t+1)} \sin(\frac{2\pi j l}{12}) + \beta_{2j}^{(t+1)} \cos(\frac{2\pi j l}{12}) \right) + e_{t+1}(l) . \quad (120)$$

We next evaluate Equation 119 at $l+1$ to get

$$z_{t+l+1} = \beta_0^{(t)} - \beta_{26}^{(t)}(-1)^l + \sum_{j=1}^{5} \left(\beta_{1j}^{(t)} \sin(\frac{2\pi j(l+1)}{12}) + \beta_{2j}^{(t)} \cos(\frac{2\pi j(l+1)}{12}) \right) + e_t(l+1) . \quad (121)$$

Equating Equations 120 and 121 for $l = 1, 2, \cdots, 12$ provides a relationship that relates $\beta^{(t+1)}$ to $\beta^{(t)}$. If we introduce a vector $\mathbf{v}^{(t)}$ the values of which are $\beta^{(t)}$ in in then each equation above can be viewed as the dot product of a certain row vector $\mathbf{r}(l)$ and $\mathbf{v}^{(t)}$. The components of $\mathbf{v}$ and $\mathbf{r}(l)$ in this case look like

$$\mathbf{v}^{(t)} = \begin{bmatrix} \beta_0^{(t)} \\ \beta_{26}^{(t)} \\ \beta_{11}^{(t)} \\ \beta_{21}^{(t)} \\ \beta_{12}^{(t)} \\ \beta_{22}^{(t)} \\ \beta_{13}^{(t)} \\ \beta_{23}^{(t)} \\ \beta_{14}^{(t)} \\ \beta_{24}^{(t)} \\ \beta_{15}^{(t)} \\ \beta_{25}^{(t)} \end{bmatrix} \quad \text{and} \quad \mathbf{r}(l) = \begin{bmatrix} 1 \\ (-1)^l \\ \sin(\frac{2\pi}{12}l) \\ \cos(\frac{2\pi}{12}l) \\ \sin(\frac{4\pi}{12}l) \\ \cos(\frac{4\pi}{12}l) \\ \sin(\frac{6\pi}{12}l) \\ \cos(\frac{6\pi}{12}l) \\ \sin(\frac{8\pi}{12}l) \\ \cos(\frac{8\pi}{12}l) \\ \sin(\frac{10\pi}{12}l) \\ \cos(\frac{10\pi}{12}l) \end{bmatrix} .$$

Then as a matrix system when we equate Equations 120 and 121 for $l = 1, 2, 3, \cdots, 12$ we obtain the system

$$\begin{bmatrix} \mathbf{r}(1)' \\ \mathbf{r}(2)' \\ \mathbf{r}(3)' \\ \vdots \\ \mathbf{r}(12) \end{bmatrix} \begin{bmatrix} \beta_0^{(t+1)} \\ \beta_{26}^{(t+1)} \\ \beta_{11}^{(t+1)} \\ \beta_{21}^{(t+1)} \\ \vdots \\ \beta_{14}^{(t+1)} \\ \beta_{24}^{(t+1)} \\ \beta_{15}^{(t+1)} \\ \beta_{25}^{(t+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}(1+1)' \\ \mathbf{r}(2+1)' \\ \mathbf{r}(3+1)' \\ \vdots \\ \mathbf{r}(12+1) \end{bmatrix} \begin{bmatrix} \beta_0^{(t)} \\ \beta_{26}^{(t)} \\ \beta_{11}^{(t)} \\ \beta_{21}^{(t)} \\ \vdots \\ \beta_{14}^{(t)} \\ \beta_{24}^{(t)} \\ \beta_{15}^{(t)} \\ \beta_{25}^{(t)} \end{bmatrix} + \begin{bmatrix} e_t(2) - e_{t+1}(1) \\ e_t(3) - e_{t+1}(2) \\ e_t(4) - e_{t+1}(3) \\ e_t(5) - e_{t+1}(4) \\ \vdots \\ e_t(10) - e_{t+1}(9) \\ e_t(11) - e_{t+1}(10) \\ e_t(12) - e_{t+1}(11) \\ e_t(13) - e_{t+1}(12) \end{bmatrix} .$$

As in the previous exercise since

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-2} a_{t+2} + \psi_{l-1} a_{t+1} ,$$

we have

$$\begin{aligned} e_t(l+1) - e_{t+1}(l) &= a_{t+l+1} + \psi_1 a_{t+l} + \psi_2 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+2} + \psi_l a_{t+1} \\ &- (a_{t+1+l} + \psi_1 a_{t+l} + \psi_2 a_{t+l-1} + \cdots + \psi_{l-2} a_{t+3} + \psi_{l-1} a_{t+2}) \\ &= \psi_l a_{t+1} . \end{aligned}$$

Recalling that for this model $\psi_l = 1 - \Theta$ for $j = 12, 24, \cdots$ and is zero otherwise the above system becomes

$$\begin{bmatrix} \mathbf{r}(1)' \\ \mathbf{r}(2)' \\ \mathbf{r}(3)' \\ \vdots \\ \mathbf{r}(12)' \end{bmatrix} \begin{bmatrix} \beta_0^{(t+1)} \\ \beta_{26}^{(t+1)} \\ \beta_{11}^{(t+1)} \\ \beta_{21}^{(t+1)} \\ \vdots \\ \beta_{14}^{(t+1)} \\ \beta_{24}^{(t+1)} \\ \beta_{15}^{(t+1)} \\ \beta_{25}^{(t+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{r}(1+1)' \\ \mathbf{r}(2+1)' \\ \mathbf{r}(3+1)' \\ \vdots \\ \mathbf{r}(12+1)' \end{bmatrix} \begin{bmatrix} \beta_0^{(t)} \\ \beta_{26}^{(t)} \\ \beta_{11}^{(t)} \\ \beta_{21}^{(t)} \\ \vdots \\ \beta_{14}^{(t)} \\ \beta_{24}^{(t)} \\ \beta_{15}^{(t)} \\ \beta_{25}^{(t)} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 1-\Theta \end{bmatrix} a_{t+1} \, .$$

Solving this expression for $\mathbf{v}^{(t+1)}$ gives gives the following update equations

$$\begin{aligned}
\beta_0^{(t+1)} &= \beta_0^{(t)} + \frac{1}{12}(1-\Theta)a_{t+1} \\
\beta_{26}^{(t+1)} &= -\beta_{26}^{(t)} + \frac{1}{12}(1-\Theta)a_{t+1} \\
\beta_{11}^{(t+1)} &= \frac{\sqrt{3}}{2}\beta_{11}^{(t)} - \frac{1}{2}\beta_{21}^{(t)} \\
\beta_{21}^{(t+1)} &= \frac{1}{2}\beta_{11}^{(t)} + \frac{\sqrt{3}}{2}\beta_{21}^{(t)} + \frac{1}{6}(1-\Theta)a_{t+1} \\
\beta_{21}^{(t+1)} &= \frac{1}{2}\beta_{21}^{(t)} - \frac{\sqrt{3}}{2}\beta_{22}^{(t)} \\
\beta_{22}^{(t+1)} &= \frac{\sqrt{3}}{2}\beta_{21}^{(t)} + \frac{1}{2}\beta_{22}^{(t)} + \frac{1}{6}(1-\Theta)a_{t+1} \\
\beta_{31}^{(t+1)} &= -\beta_{32}^{(t)} \\
\beta_{32}^{(t+1)} &= \beta_{31}^{(t)} + \frac{1}{6}(1-\Theta)a_{t+1} \\
\beta_{41}^{(t+1)} &= -\frac{1}{2}\beta_{41}^{(t)} - \frac{\sqrt{3}}{2}\beta_{42}^{(t)} \\
\beta_{42}^{(t+1)} &= \frac{\sqrt{3}}{2}\beta_{41}^{(t)} - \frac{1}{2}\beta_{42}^{(t)} + \frac{1}{6}(1-\Theta)a_{t+1} \\
\beta_{51}^{(t+1)} &= -\frac{\sqrt{3}}{2}\beta_{51}^{(t)} - \frac{1}{2}\beta_{52}^{(t)} \\
\beta_{52}^{(t+1)} &= \frac{1}{2}\beta_{51}^{(t)} - \frac{\sqrt{3}}{2}\beta_{52}^{(t)} + \frac{1}{6}(1-\Theta)a_{t+1} \, ,
\end{aligned}$$

which is the same expression given in the book. Much of the algebra involved in the above matrix calculations are performed in the MATHEMATICA file `prob_6_5.nb`.

Exercise 6.6 (the $(0,1,1)(0,1,1)_{12}$ model)

The $(0,1,1)(0,1,1)_{12}$ model is specifically given by

$$(1-B)(1-B^{12}) = (1-\theta B)(1-\Theta B^{12})a_t\,,$$

or in autoregressive form by

$$\frac{(1-B)(1-B^{12})}{(1-\theta B)(1-\Theta B^{12})}z_t = a_t\,.$$

Consider the first fraction above

$$\begin{aligned}\frac{1-B}{1-\theta B} &= (1-B)\sum_{k=0}^{\infty}\theta^k B^k = \sum_{k=0}^{\infty}\theta^k B^k - \sum_{k=0}^{\infty}\theta^k B^{k+1} = \sum_{k=0}^{\infty}\theta^k B^k - \sum_{k=1}^{\infty}\theta^{k-1}B^k \\ &= 1+(\theta-1)\sum_{k=0}^{\infty}\theta^k B^{k+1}\,.\end{aligned}$$

In the same way we find the second fraction given by

$$\frac{1-B^{12}}{1-\Theta B^{12}} = 1+(\Theta-1)\sum_{k=0}^{\infty}\Theta^k B^{12(k+1)}\,.$$

Thus their product is given by

$$\begin{aligned}\left(\frac{1-B}{1-\theta B}\right)\left(\frac{1-B^{12}}{1-\Theta B^{12}}\right) &= \left(1+(\theta-1)\sum_{k=0}^{\infty}\theta^k B^{k+1}\right)\left(1+(\Theta-1)\sum_{k=0}^{\infty}\Theta^k B^{12(k+1)}\right) \\ &= 1+(\theta-1)\sum_{k=0}^{\infty}\theta^k B^{k+1}+(\Theta-1)\sum_{k=0}^{\infty}\Theta^k B^{12(k+1)} \\ &+ (\theta-1)(\Theta-1)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\theta^{k_1}\Theta^{k_2}B^{12k_2+k_1+13}\,.\end{aligned}$$

These expressions give explicit representation of the coefficients ϕ_k in the decomposition of z_t into its pure autoregressive formulation $z_t = a_t + \sum_{k=1}^{\infty}\phi_k z_{t-k}$ in that using the above but moving every term with a shift (i.e. a B coefficient) to the right hand side of the equals sign we find

$$\begin{aligned}z_t &= a_t - (\theta-1)\sum_{k=0}^{\infty}\theta^k z_{t-(k+1)} - (\Theta-1)\sum_{k=0}^{\infty}\Theta^k z_{t-12(k+1)} \\ &- (\theta-1)(\Theta-1)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\theta^{k_1}\Theta^{k_2}z_{t-(12k_2+k_1+13)}\,.\end{aligned}$$

Then we can predict the value of z_{t+1} using the conditional expectation formula

$$\begin{aligned}z_t(1) &= E(z_{t+1}|z_t, z_{t-1},\cdots) \\ &= E(a_{t+1}+\sum_{k=1}^{\infty}\phi_k z_{t+1-k}|z_t,z_{t-1},\cdots) = \sum_{k=1}^{\infty}\phi_k z_{t+1-k}\,.\end{aligned}$$

Using the just found autoregressive formulation for the $(0,1,1)(0,1,1)_{12}$ model the expression for $z_t(1)$ becomes

$$
\begin{aligned}
z_t(1) &= (1-\theta)\sum_{k=0}^{\infty}\theta^k z_{t-k} + (1-\Theta)\sum_{k=0}^{\infty}\Theta^k z_{t-12k-11} \\
&- (1-\theta)(1-\Theta)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\theta^{k_1}\Theta^{k_2} z_{t-(12k_2+k_1+12)} \\
&= \text{EWMA}(z_t) + \text{SEWMA}(z_{t-11}) \\
&- (1-\theta)(1-\Theta)\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\theta^{k_1}\Theta^{k_2} z_{t-(12k_2+k_1+12)} \,.
\end{aligned}
$$

Using the definitions of the two weighted averages SEWMA and EWMA and recognizing that the seasonal exponential weighted moving average operator $\text{SEWMA}(z_t)$ is linear, we can write the expression presented in this exercise for the value of $z_t(1)$ as

$$
\begin{aligned}
z_t(1) &= \text{EWMA}(z_t) + \text{SEWMA}(z_{t-11}) - \text{SEWMA}(\text{EWMA}(z_{t-12})) \\
&= (1-\theta)\sum_{j\geq 0}\theta^j z_t + (1-\Theta)\sum_{j\geq 0}\Theta^j z_{t-12j-11} - \text{SEWMA}\left((1-\theta)\sum_{j\geq 0}\theta^j z_{t-12-j}\right) \\
&= (1-\theta)\sum_{j\geq 0}\theta^j z_t + (1-\Theta)\sum_{j\geq 0}\Theta^j z_{t-12j-11} \\
&- (1-\theta)(1-\Theta)\sum_{j\geq 0}\sum_{k\geq 0}\Theta^k\theta^j z_{t-12k-12-j} \,,
\end{aligned}
$$

since this expression is equivalent to the one we derived above we have shown the desired relationship.

Exercise 6.7 (quarterly modeling)

Part (a): The stationary difference is $(1-B)(1-B^4)z_t$. Since this equals $(1-\Theta)a_t$ it would have an autocorrelation structure with a single peak at lag $k=4$.

Part (b): The given model is

$$
(1 - B - B^4 + B^5)z_t = (1-\Theta B^4)a_t \,,
$$

or

$$
z_t = z_{t-1} + z_{t-4} - z_{t-5} + a_t - \Theta a_{t-4} \,.
$$

so incrementing t by l we get

$$
z_{t+l} = z_{t+l-1} + z_{t+l-4} - z_{t+l-5} + a_{t+l} - \Theta a_{t+l-4} \,.
$$

and our predictions are given by

$$
\begin{aligned}
z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \ldots) \\
&= E(z_n + z_{n-3} - z_{n-4} + a_{n+1} - \Theta a_{n-3}|z_n, z_{n-1}, \ldots) \\
&= z_n + z_{n-3} - z_{n-4} - \Theta a_{n-3} \,. \\
z_n(2) &= z_n(1) + z_{n-2} - z_{n-3} - \Theta a_{n-2} \\
z_n(3) &= z_n(2) + z_{n-1} - z_{n-2} - \Theta a_{n-1} \\
z_n(4) &= z_n(3) + z_n - z_{n-1} - \Theta a_n \\
z_n(5) &= z_n(4) + z_n(1) - z_n \\
z_n(6) &= z_n(5) + z_n(2) - z_n(1) \\
z_n(7) &= z_n(6) + z_n(3) - z_n(2) \\
&\vdots \\
z_n(l) &= z_n(l-1) + z_n(l-4) - z_n(l-5) \quad \text{for} \quad l \geq 6 \,,
\end{aligned}
$$

or

$$
(1 - B)(1 - B^4) z_n(l) = 0 \,,
$$

for the eventual forecast function

Part (c): To compute $z_n(1)$ we assume we have been measuring z_n and making predictions for some time. Then recalling that an estimate for a_n can be given by $\hat{a}_n \approx z_n - z_n(1)$, so

$$
z_n(1) = z_n + z_{n-3} - z_{n-4} - \hat{\Theta}\hat{a}_{n-3} \,,
$$

is how to predict one-step-ahead when we have observed the time series up to and including z_n.

Exercise 6.8 (a comparison of three seasonal ARIMA models)

For the model $(0,1,1)(1,0,0)_{12}$ given by

$$
(1 - \Phi B^{12})(1 - B) z_t = (1 - \theta B) a_t \,,
$$

we have z_t given by

$$
z_t = z_{t-1} + \Phi z_{t-12} - \Phi z_{t-13} + a_t - \theta a_{t-1} \,.
$$

Incrementing t by l gives

$$
z_{t+l} = z_{t+l-1} + \Phi z_{t+l-12} - \Phi z_{t+l-13} + a_{t+l} - \theta a_{t+l-1} \,.
$$

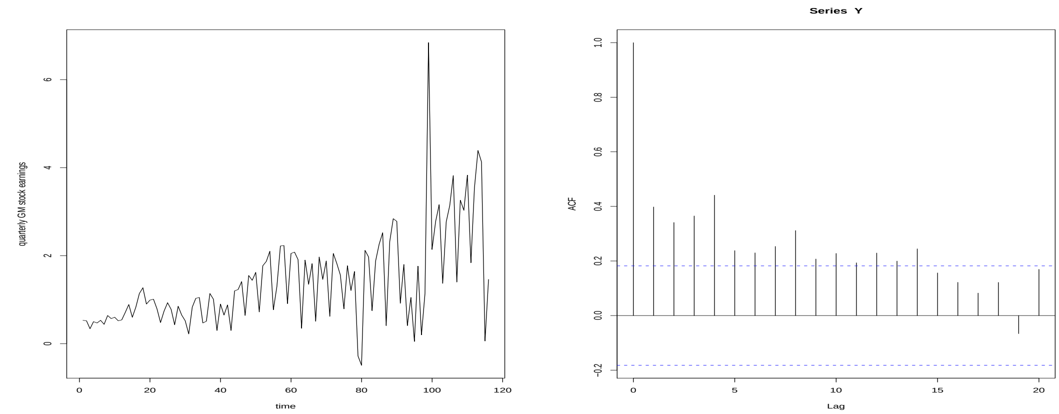


Figure 39: **Left:** A time series plot of the raw data for the quarterly earnings per share for General Moters (GM) stock. **Right:** The direct sample autocorrelation function for the GM quarterly earnings data set.

From this our forecasts are given by

$$
\begin{aligned}
z_n(1) &= E(z_{n+1}|z_n, z_{n-1}, \dots) \\
&= z_n + \Phi z_{n-11} + \Phi z_{n-12} - \theta a_n \,. \\
z_n(2) &= z_n(1) + \Phi z_{n-10} + \Phi z_{n-11} \\
z_n(3) &= z_n(2) + \Phi z_{n-9} + \Phi z_{n-10} \\
&\vdots \\
z_n(10) &= z_n(9) + \Phi z_{n-2} + \Phi z_{n-3} \\
z_n(11) &= z_n(10) + \Phi z_{n-1} + \Phi z_{n-2} \\
z_n(12) &= z_n(11) + \Phi z_n + \Phi z_{n-1} \\
z_n(13) &= z_n(12) + \Phi z_n(1) + \Phi z_n \\
z_n(14) &= z_n(13) + \Phi z_n(2) + \Phi z_n(1) \\
&\vdots \\
z_n(l) &= z_n(l-1) + \Phi z_n(l-12) + \Phi z_n(l-13) \quad \text{for} \quad l \geq 14 \,,
\end{aligned}
$$

So the eventual forecast function is given by

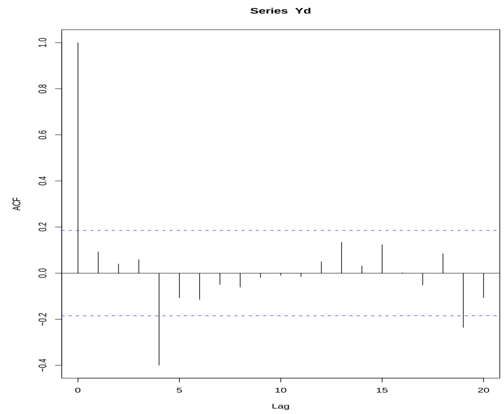
$$
z_n(l) = z_n(l-1) + \Phi z_n(l-12) + \Phi z_n(l-13) \,,
$$

or

$$
(1 - \Phi B^{12})(1 - B) z_n(l) = 0 \,.
$$

Exercise 6.9 (quarterly earnings per share)

In the R script `prob_6_9.R`, we begin by loading and plotting the data in Figure 39 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 39 (right).



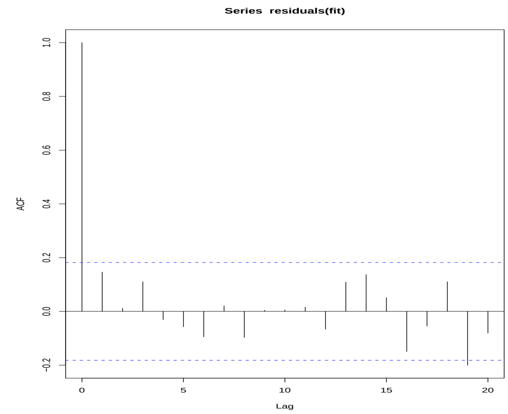


Figure 40: **Left:** The sample autocorrelation function of $(1 - B^4)z_t$ for the GM quarterly earnings data set. **Right:** The autocorrelation function of the residuals, $r_{\hat{a}}$, for the model found in the text below for the GM stock earnings data set.

This autocorrelation has the characteristic shape of a periodic series that needs to be differenced before it will become stationary. From the sample autocorrelation function plot the period appears to be 4, corresponding to quarterly variations. We thus take the difference $1 - B^4$ and replot the autocorrelation function of this difference in Figure 40 (left). This later difference is stationary. From the pattern of the autocorrelation function for $(1-B^4)z_t$ we will assume a single seasonal moving average (SMA(1)) term to model this. Thus, if we assume a model of the form

$$(1 - B^4)z_t = \theta_0 + (1 - \Theta B^4)a_t \, ,$$

we can find coefficients for this model (and their standard errors given by) in one of two ways as discussed on Page 120. Taking explicit differences first and then using the R command `arima` gives considerably different results than letting the `arima` command do the seasonal differences itself. For example, if `Y` contains the initial time series and I execute the following R commands

```
ord <- c(0,0,0)
ssn <- list(order=c(0,1,1),period=4)
arima( Y, order=ord, seasonal=ssn, method="ML" )
```

gives for estimates of the coefficients (some output trimmed)

```
Coefficients:
         sma1
      -0.6578
s.e.   0.0883

sigma^2 estimated as 0.8548:  log likelihood = -151.27,  aic = 306.54
```

while the call where we compute the explicit seasonal differences of `Y` ourselves

```
Yd  <- diff(Y,lag=4,differences=1)
ord <- c(0,0,0)
ssn <- list(order=c(0,0,1),period=4)
arima( Yd, order=ord, seasonal=ssn, include.mean=TRUE, method="ML" )
```

gives

```
Coefficients:
         sma1  intercept
      -1.0000     0.0751
s.e.   0.2952     0.0092

sigma^2 estimated as 0.6805:  log likelihood = -144.1,  aic = 294.2
```

The problem with this last results is that if $\Theta = 1.0$ (the sign convention of the moving average component in R is the opposite one of the book) then the suggested model becomes (assuming θ_0 is insignificant)

$$(1 - B^4)z_t = (1 - \Theta B^4)a_t = (1 - B^4)a_t \Rightarrow z_t = a_t \, .$$

This makes me think that when using the function `arima` it is better to let the function compute the seasonal differences itself rather than explicitly doing it. In either case, plots of the sample autocorrelation function help guide the model building process. To complete this exercise, using the first specification of the `arima` command we obtain the sample autocorrelations of the residuals of the proposed $(0,0,0)(0,1,1)_4$ model in Figure 40 (right). The next four predictions (and their standard errors) under this model would be

```
3.130336 3.452201 1.265497 2.399056
0.924579 0.924579 0.924579 0.924579
```

Exercise 6.10 (monthly arrivals of U.S. citizens)

See the R script `prob_6_10.R` for the code to do this problem. In that R script, we load and plot the data in Figure 41 (left). When we plot the sample autocorrelation directly we get the plot in Figure 41 (right). This has the characteristic shape of a periodic series that needed to be differenced before it will become stationary. Note that the period appears to be 12 corresponding to monthly variations. We take the difference $1 - B^{12}$ and replot the autocorrelation function in Figure 42 (left). There we see the characteristic slow decay of a non-stationary time series. We then consider the first difference of this series and plots the corresponding autocorrelation in Figure 42 (middle). This later difference is stationary.

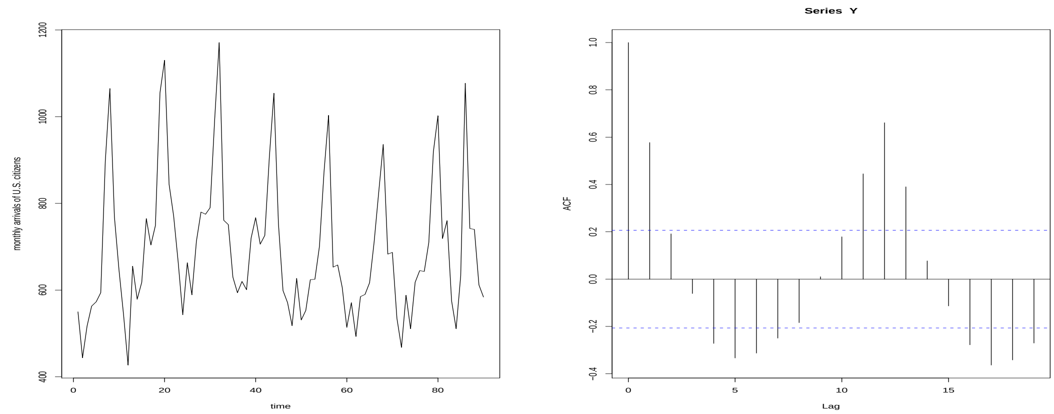


Figure 41: **Left:** A time series plot of the raw data for the monthly arrival of U.S. citizens. **Right:** The direct sample autocorrelation function for the monthly arrival data set.

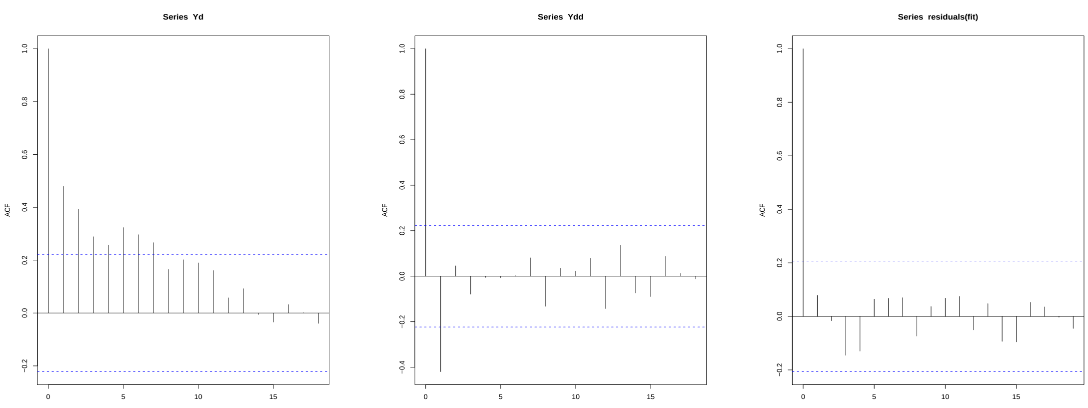


Figure 42: **Left:** The sample autocorrelation function of $(1-B^{12})z_t$ for the monthly arrivals of U.S. citizen data set. **Center:** The sample autocorrelation function of $(1-B)(1-B^{12})z_t$ for the monthly arrivals of U.S. citizen data set. **Right:** The autocorrelation function of the residuals, $r_{\hat{a}}$, for the model found in the text for the monthly arrivals of U.S. citizen data set.

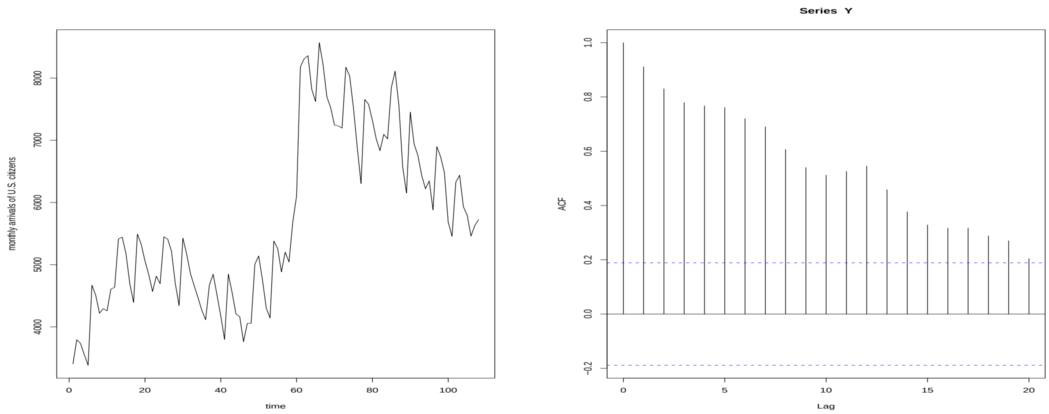


Figure 43: **Left:** The raw data for the unemployed U.S. civilian labor force (in thousands) for the period 1970-1978. **Right:** The direct sample autocorrelation function for this U.S. civilian labor force data set.

From the pattern of the autocorrelation function for $(1-B)(1-B^{12})z_t$ we will assume a model for this series consisting of a single moving average (MA(1)) term. Thus, if we assume a model of the form

$$(1-B)(1-B^{12})z_t = \theta_0 + (1-\theta B)a_t\,,$$

we find the coefficient θ_0 insignificant, and

$$\theta = 0.74(0.08) \quad \sigma^2 = 7251\,,$$

Finally, for model validation we plot the autocorrelations of residuals of the fitted model in Figure 42 (right).

Exercise 6.11 (modeling unemployed U.S. civilians)

In the R script `prob_6_11.R`, we begin by loading and plotting the data in Figure 43 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 43 (right). This autocorrelation has the characteristic shape of a periodic series that needs to be differenced before it will become stationary. Because of the great number of adjacent significant values of r_k we take the first difference $1-B$ and replot the autocorrelation function of this difference in Figure 44 (left). There we see a strong periodic component at the lag $k=12$. Because of this we next consider the first seasonal difference of the series or $(1-B^{12})(1-B)z_t$. The autocorrelation for this time series is plotted in Figure 44 (right). Based on the oscillations observed we hypothesis that our model could be improved with the inclusion of a AR(2) component. Thus the model at this point would be $(2,1,0)(0,1,0)_{12}$. Plots of the sample autocorrelation for such a model are given in Figure 45 (left) Based on this looks like we are now missing a SMA(1) term which when we include we get the following model

$$(1-\phi_1 B-\phi_2 B^2)(1-B^{12})(1-B)z_t = (1-\Theta B)a_t\,.$$

Finally, for model validation we plot the autocorrelations of residuals of the fitted model in Figure 45 (right).

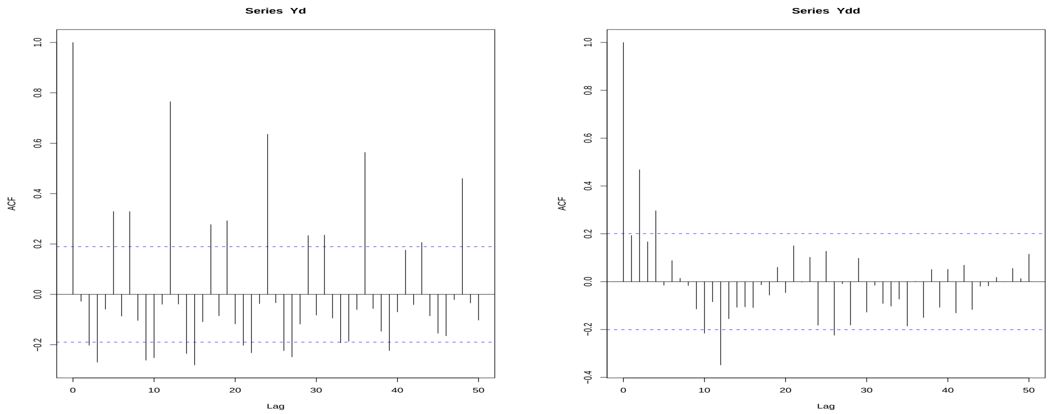


Figure 44: **Left:** The sample autocorrelation function of $(1-B)z_t$ for the U.S. labor force data set. **Right:** The sample autocorrelation function of $(1-B^{12})(1-B)z_t$ for the U.S. labor force data set.

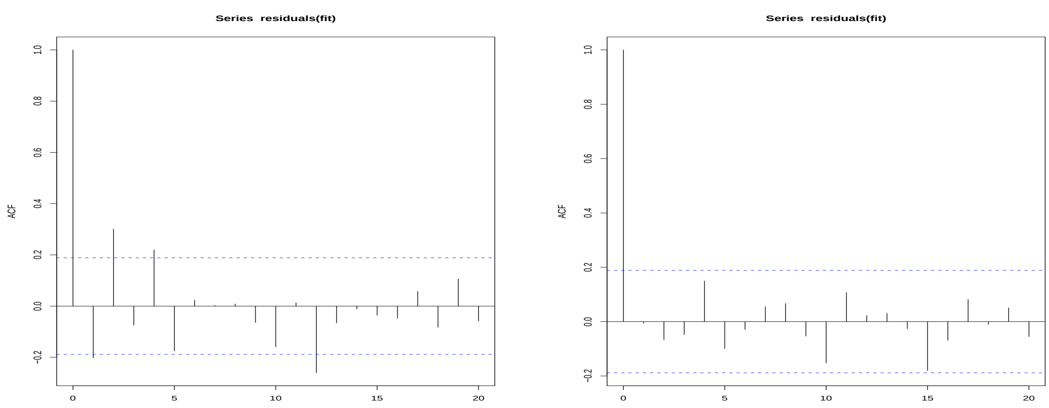


Figure 45: **Left:** The sample autocorrelation function of the residuals for the model $(1-\phi_1 B-\phi_2 B^2)(1-B^{12})(1-B)z_t$ for the U.S. labor force data set. **Right:** The sample autocorrelation function $r_{\hat{a}}$ for the model found for the U.S. labor force data set.

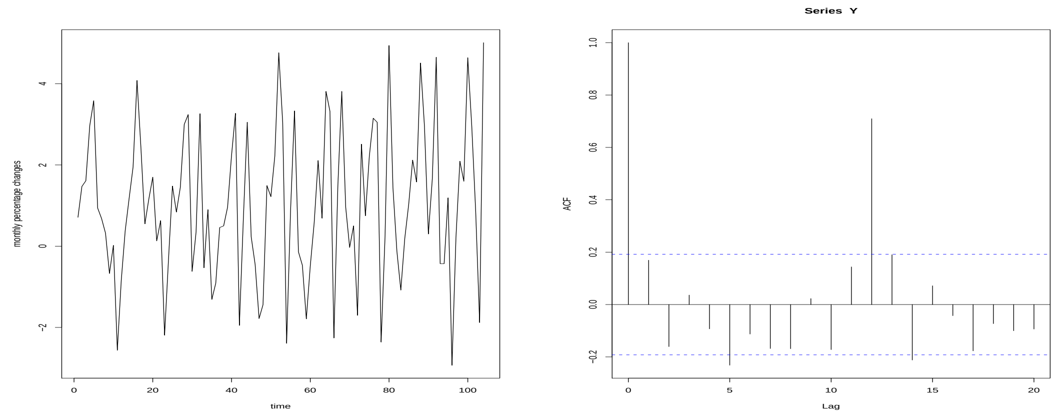


Figure 46: **Left:** The raw data for the monthly Canadian wage change data set. **Right:** The direct sample autocorrelation function for Canadian wage change data.

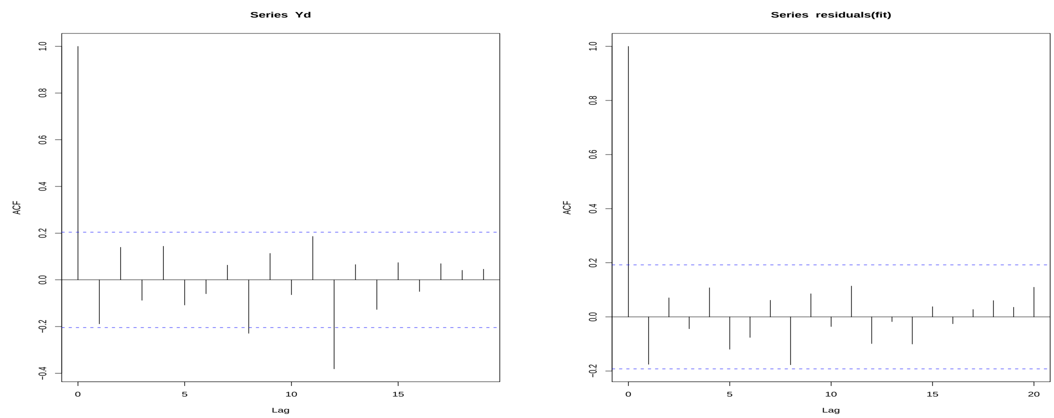


Figure 47: **Left:** The sample autocorrelation function of $(1-B^{12})z_t$ for the monthly Canadian wage change data set. This indicates a SMA(1) model maybe applicable. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the model found for monthly Canadian wage change data set.

Exercise 6.12 (modeling monthly Canadian wage change)

In the R script `prob_6_12.R`, we begin by loading and plotting the data in Figure 46 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 46 (right). Based on the large spike at lag $k = 12$ we will take the first seasonal difference of this data $(1 - B^{12})z_t$, and replot the autocorrelation function of this difference in Figure 47 (left). From this plot we choose to model this data as having a single seasonal moving average SMA(1) as

$$(1 - B^{12})z_t = \theta_0 + (1 - \Theta B^{12})a_t \,.$$

We find coefficients of this model (and their standard errors given by)

$$\Theta = 0.43(0.10) \quad \theta_0 = 0.07(0.06) \quad \sigma^2 = 1.00 \,.$$

Using the model above with these parameters produces residuals with autocorrelations presented in Figure 47 (right).

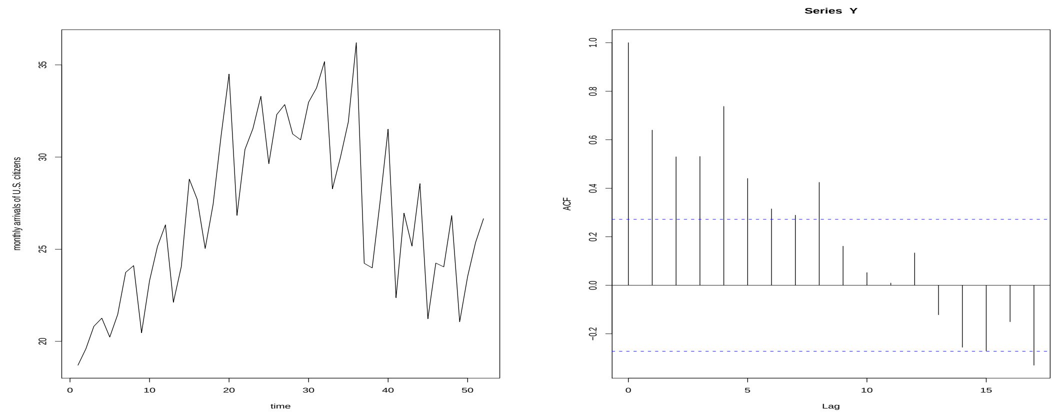


Figure 48: **Left:** The raw data for the beer shipments data set. **Right:** The direct sample autocorrelation function for the beer shipments data set.

Exercise 6.13 (modeling beer shipments)

In the R script `prob_6_13.R`, we begin by loading and plotting the data in Figure 48 (left). When next plot the sample autocorrelation of this data directly to get the plot in Figure 48 (right). This autocorrelation has the characteristic shape of a non-stationary series that needs to be differenced before it will become stationary. We take the difference $1 - B$ and replot the autocorrelation function of this difference in Figure 49 (left). There we see the characteristic profile of a periodic problem. Because the autocorrelation appears to exponentially decay with a period of four we first tried a single seasonal autoregressive (SAR(1)) term to model this data. When we fit that model and plot the residual autocorrelations we obtain the plot in Figure 49 (middle). Note that this autocorrelation still has a significant $k = 1$ lag value. Thus in assuming a model of the form

$$(1 - \Phi B^4)(1 - B)z_t = b_t \,,$$

we find the time series b_t satisfies $b_t = (1 - \theta B)a_t$ and so we should revise our original model to be

$$(1 - \Phi B^4)(1 - B)z_t = (1 - \theta B)a_t \,.$$

When this model is fit to the beer shipment data set the residuals of the fitted model are plotted in Figure 49 (right).

Exercise 6.14 (modeling traffic fatalities)

In the R script `prob_6_14.R`, we begin by loading and plotting the data in Figure 50 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 50 (right). This autocorrelation has the characteristic shape of a periodic series that needs to be differenced before it will become stationary. Note that from the sample autocorrelation function plot the period appears to be 12, corresponding to monthly variations. We take the difference $1 - B^{12}$ and replot the autocorrelation function of this difference in Figure 51 (left).

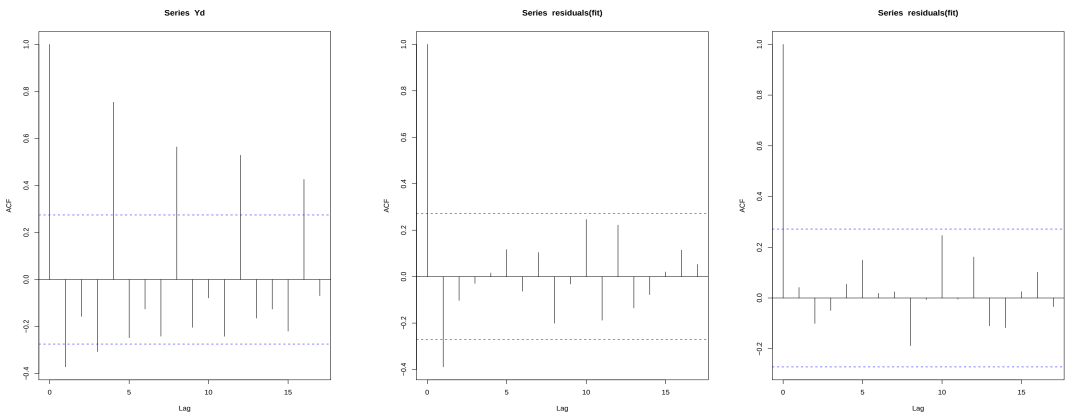


Figure 49: **Left:** The sample autocorrelation function of $(1 - B)z_t$ for the beer sales data set. Note the strong periodic component that remains and decays very slowly. **Middle:** The sample autocorrelation function of the residuals for the model $(1 - \Phi B^4)(1 - B)z_t$ for the beer sales data set. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the final model found for beer sales data set.

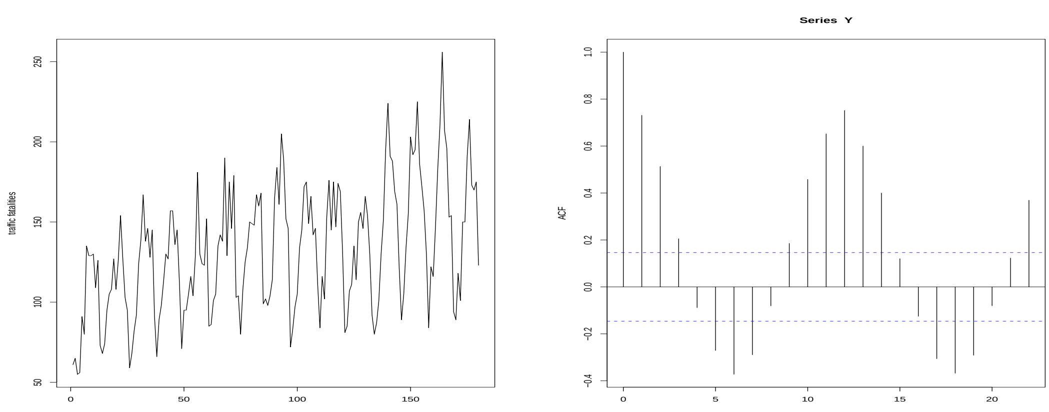


Figure 50: **Left:** The raw data for the monthly traffic fatalities for the period January 1960 to December 1968. **Right:** The direct sample autocorrelation function for the traffic fatalities data set.

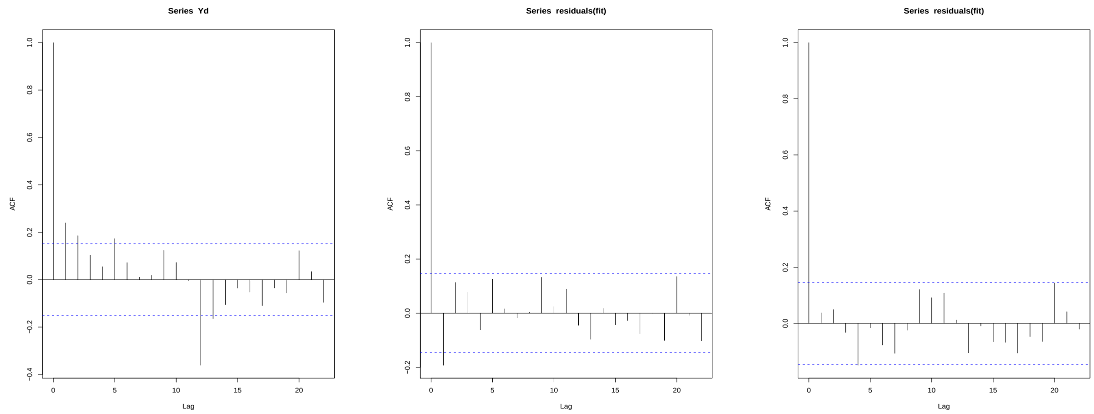


Figure 51: **Left:** The sample autocorrelation function of $(1-B^{12})z_t$ for the traffic fatalities data set. **Center:** The sample autocorrelation function of the residuals $r_{\hat{a}}$ for the model $(1-\theta B)(1-B^{12})z_t = (1-\Theta B^{12})a_t$ fit to the traffic fatalities data set. **Right:** The same autocorrelation function $r_{\hat{a}}$ for the model found for the housing starts data set.

There we see a exponential decay in the autocorrelation structure and a spike at the $k = 12$ lag. These observations lead us to try a first order autoregressive model AR(1) coupled with a first order seasonal moving average model. The autocorrelation of the residuals for such a model is plotted in in Figure 51 (middle). From this pattern of the autocorrelation function we still have a significant $k = 1$ term so we will modify the original model by adding a single moving average (MA(1)) term. Thus, our model now take the form

$$(1-\phi B)(1-B^{12})z_t = (1-\theta B)(1-\Theta B^{12})a_t \,.$$

Finally, for model validation we plot the autocorrelations of residuals of this fitted model in Figure 51 (right).

Exercise 6.15 (modeling gasoline demand)

In the R script `prob_6_15.R`, we begin by loading and plotting the data in Figure 52 (left). When next plot the sample autocorrelation of this data to get the plot in Figure 52 (right). This autocorrelation has the characteristic shape of a series that needs to be differenced before it will become stationary. We therefore take the first difference $1 - B$ and replot the autocorrelation function of this difference in Figure 53 (left). From this plot we see a strong periodic component remains in the autocorrelation. These observations lead us to further take a seasonal difference. When we plot the sample autocorrelation function of the time series $(1-B^{12})(1-B)z_t$ we obtain Figure 53 (right). From this pattern of the autocorrelation function we still have several significant autocorrelations with considerable long range structure. Because of this we add an AR(2) model to the existing model giving a model

$$(1-\phi_1 B - \phi_2 B^2)(1-B)(1-B^{12})z_t = a_t \,.$$

The autocorrelation of the residuals from such a model are shown in Figure 54 (left). From

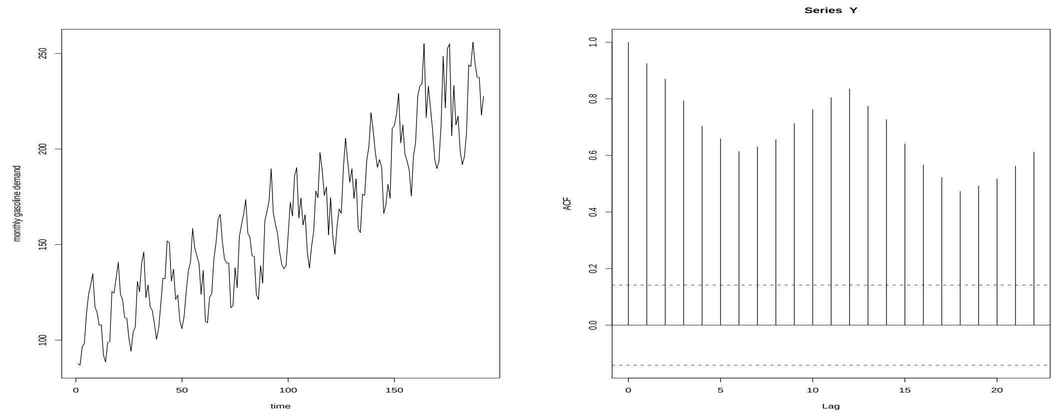


Figure 52: **Left:** The raw data for the monthly gasoline demand in Ontario (in millions of gallons) from January 1960 to December 1975. **Right:** The direct sample autocorrelation function for the gasoline demand data set.

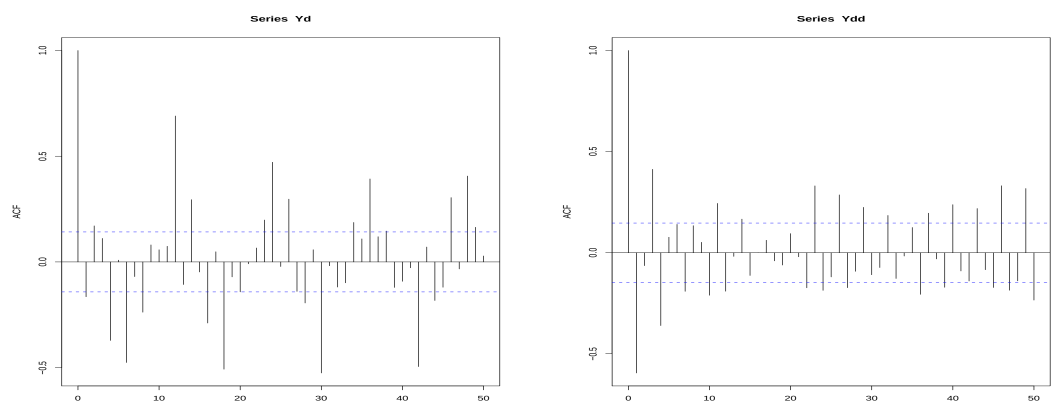


Figure 53: **Left:** The sample autocorrelation function of $(1-B)z_t$ for the gasoline demand data set. **Right:** The sample autocorrelation function of $(1-B^{12})(1-B)z_t$ for the gasoline demand data set.

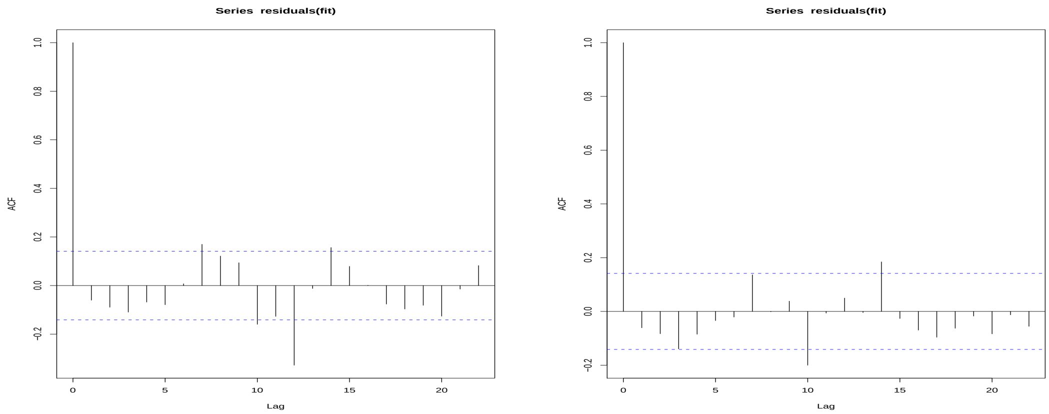


Figure 54: **Left:** The sample autocorrelation function of $(1-B)z_t$ for the gasoline demand data set. **Right:** The sample autocorrelation function of $(1-B^{12})(1-B)z_t$ for the gasoline demand data set.

this last plot we see that we still have a significant $k = 12$ term so we will modify the original model by adding a single seasonal moving average (SMA(1)) term. Thus, our final model now take the form

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})z_t = (1 - \Theta B)a_t \, .$$

The residuals of this final model are presented in Figure 54 (right).

Chapter 7: Relationships Between Forecasts From General Exponential Smoothing and Forecasts from ARIMA Time Series Models

Notes On The Text

Updating the coefficients in the eventual forecast function

If we specify f^* to be the fitting functions determined such that they satisfy

$$\varphi(B)z_n(l) = \phi(B)(1-B)^d z_n(l) = 0\,,$$

that is f^* are the $p+d$ independent solutions to the eventual forecast equation. Then coefficient updates $\beta^{(n)}$ must satisfy

$$f^*(l)'\beta^{(n+1)} = f^*(l+1)'\beta^{(n)} + \psi_l(z_{n+1} - z_n(1)) \quad \text{for} \quad l > q-p-d\,.$$

If we evaluate this expression for lookaheads "l" taken to be $l, l+1, l+2, \cdots, l+p+d-1$ we see that the *total* number of equations we will have specified is

$$(l+p+d-1) - l + 1 = p+d\,,$$

the exact number needed given the number of solutions $p+d$ to the eventual forecast difference equation,

$$\varphi(B)z_n(l) = (1-B)^d\phi(B)z_n(l) = 0\,,$$

These equations specifically look like

$$\begin{aligned}
f^*(l)'\beta^{(n+1)} &= f^*(l+1)'\beta^{(n)} + \psi_l(z_{n+1} - z_n(1)) \\
f^*(l+1)'\beta^{(n+1)} &= f^*(l+2)'\beta^{(n)} + \psi_{l+1}(z_{n+1} - z_n(1)) \\
f^*(l+2)'\beta^{(n+1)} &= f^*(l+3)'\beta^{(n)} + \psi_{l+2}(z_{n+1} - z_n(1)) \\
&\vdots \\
f^*(l+p+d-1)'\beta^{(n+1)} &= f^*(l+p+d)'\beta^{(n)} + \psi_{l+p+d-1}(z_{n+1} - z_n(1))\,.
\end{aligned}$$

Now define the matrix F_l^* to have its rows made up of the values of $f^*(l)', f^*(l+1)', \dots f^*(l+p+d-1)$, we see that the above system can be written as the matrix system

$$F_l^*\beta^{(n+1)} = F_{l+1}^*\beta^{(n)} + \begin{bmatrix} \psi_l \\ \psi_{l+1} \\ \psi_{l+2} \\ \vdots \\ \psi_{l+p+d-1} \end{bmatrix} (z_{n+1} - z_n(1))\,.$$

Multiplying both sides of this equation by F_l^{*-1} to get

$$\beta^{(n+1)} = F_l^{*-1}F_{l+1}^*\beta^{(n)} + F_l^{*-1} \begin{bmatrix} \psi_l \\ \psi_{l+1} \\ \psi_{l+2} \\ \vdots \\ \psi_{l+p+d-1} \end{bmatrix} (z_{n+1} - z_n(1))\,,$$

which is the coefficient update equation and is equation 7.18 in the book.

Illustrative Examples: The 12-point sinusoidal model

Note that trigonometric fitting functions of the form $\sin(xl)$ and $\cos(xl)$ the roots G of the polynomial $\phi_m(B)$ are given by the complex exponentials $e^{\pm ix}$, so that in the 12-point sinusoidal model example given here where $m = 3$, $f_1(l) = 1$, $f_2(l) = \sin(\frac{2\pi}{12}l)$, $f_3(l) = \sin(\frac{2\pi}{12}l)$ we have our roots given by

$$G_{\pm} = e^{\pm i\frac{2\pi}{12}} = \cos(\frac{\pi}{6}) \pm i\sin(\frac{\pi}{6}) = \frac{\sqrt{3}}{2} \pm \frac{i}{2}\,.$$

Thus the reducing polynomial $\phi(B)$ that results from the trigonometric terms look like

$$\left(1 - \left(\frac{\sqrt{3}}{2} + \frac{i}{2}\right) B\right)\left(1 - \left(\frac{\sqrt{3}}{2} - \frac{i}{2}\right) B\right) = 1 - \sqrt{3}B + B^2\,,$$

which is the polynomial given on the left-hand-side of equation 7.2.7 in the book.

Exercise Solutions

Exercise 7.1 (specifying appropriate ARIMA models)

For each part of this exercise, we must find an order m polynomial, $\phi_m(B)$, the roots, G_j, of which satisfy

$$\phi_m(B) f_i(l) = \prod_{j=1}^{m}(1 - G_j B) f_i(l) = 0\,,$$

for $1 \le i \le m$. That is the polynomial $\phi_m(B)$ (when considered as a difference operator) annihilates each $f_i(\cdot)$. Then with this polynomial $\phi_m(B)$ the ARIMA model that gives rise to the same forecasts as general exponential smoothing with a discount coefficient of ω is given by

$$\prod_{j=1}^{m}(1 - G_j B) z_t = \prod_{j=1}^{m}\left(1 - \frac{\omega}{G_j}B\right) a_t\,. \tag{122}$$

This equation will be used in the parts below.

Part (a): If $f(l) = (1, l)'$ then the polynomial $\phi_m(B)$ that annihilates the functions $f_1(l) = 1$ and $f_2(l) = l$ is $\phi_2(B) = (1 - B)^2$. From Equation 122 the equivalent ARIMA model that gives rise to an eventual forecast function that agrees with exponential smoothing with a discount coefficient ω is

$$(1 - B)^2 z_t = (1 - \omega B)^2 a_t\,.$$

Part (b): In this case we need $\phi_3(B) = (1-B)^3$ so the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is

$$(1-B)^3 z_t = (1-\omega B)^3 a_t\,.$$

Part (c): Exponential smoothing with $m = 2$ fitting functions has $\phi_2(B) = (1-B)(1-\phi B)$ so

$$(1-B)(1-\phi B)z_t = (1-\omega B)\left(1-\frac{\omega B}{\phi}\right)a_t\,,$$

is the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω.

Part (d): For these trigonometric fitting functions the roots G are given by $G = e^{\pm i\frac{\pi}{2}} = \pm i$. Thus

$$\phi_3(B) = (1-B)(1-iB)(1+iB) = (1-B)(1+B^2)\,.$$

As a check lets verify that $1+B^2$ annihilates the fitting functions $\sin(\frac{\pi}{2}l)$ and $\cos(\frac{\pi}{2}l)$ as it should

$$\begin{aligned}
(1+B^2)\sin(\frac{\pi}{2}l) &= \sin(\frac{\pi}{2}l) + \sin(\frac{\pi}{2}(l-2)) \\
&= \sin(\frac{\pi}{2}l) + \sin(\frac{\pi}{2}l)\cos(\pi) - \cos(\frac{\pi}{2}l)\sin(\pi) = 0\,.
\end{aligned}$$

and

$$\begin{aligned}
(1+B^2)\cos(\frac{\pi}{2}l) &= \cos(\frac{\pi}{2}l) + \cos(\frac{\pi}{2}(l-2)) \\
&= \cos(\frac{\pi}{2}l) + \cos(\frac{\pi}{2}l)\cos(\pi) + \sin(\frac{\pi}{2}l)\sin(\pi) = 0\,.
\end{aligned}$$

Thus the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is

$$(1-B)(1+B^2)z_t = (1-\omega B)(1+\omega^2 B^2)a_t\,.$$

Part (e): This is a similar result as in Part (d) but now we need a second factor of $(1-B)$. Thus we see that $\phi(B) = (1-B)^2(1+B^2)$ and the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is

$$(1-B)^2(1+B^2)z_t = (1-\omega B)^2(1+\omega^2 B^2)a_t\,.$$

Part (f): In case we have added the two additional fitting functions $l\sin(\frac{\pi}{2}l)$ and $l\cos(\frac{\pi}{2}l)$, which will be annihilated by a second powers of the expression $(1-G_{\pm}B)$, where $G_{\pm}$ is the complex roots used in the modeling of the trigonometric terms $\sin(\frac{\pi}{2}l)$ and $\cos(\frac{\pi}{2}l)$. Thus we now have

$$\begin{aligned}
\phi(B) &= (1-B)^2\left(1-\left(\frac{\sqrt{3}}{2}+\frac{i}{2}\right)B\right)^2\left(1-\left(\frac{\sqrt{3}}{2}-\frac{i}{2}\right)B\right)^2 \\
&= (1-B)^2(1-\sqrt{3}B+B^2)^2\,.
\end{aligned}$$

and the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is given by applying Equation 122.

Part (g): In this case we have trigonometric fitting functions with two different frequencies. For the fitting function $\sin(\frac{2\pi}{12}l)$ and $\cos(\frac{2\pi}{12}l)$ we already found roots given by

$$G_{\pm} = e^{i\left(\frac{2\pi}{12}\right)} = \frac{\sqrt{3}}{2} \pm \frac{i}{2}.$$

For the second set of fitting functions $\sin(\frac{4\pi}{12}l)$ and $\cos(\frac{4\pi}{12}l)$ we have roots given by

$$G_{\pm} = e^{i\left(\frac{4\pi}{12}\right)} = \cos(\frac{\pi}{3}) \pm i\sin(\frac{\pi}{3}) = \frac{1}{2} \pm \frac{\sqrt{3}}{2}.$$

Thus the polynomial $\phi(B)$ looks like

$$\begin{aligned}
\phi(B) &= (1-B)^2 \\
&\times \left(1 - \left(\frac{\sqrt{3}}{2} + \frac{i}{2}\right)B\right)\left(1 - \left(\frac{\sqrt{3}}{2} - \frac{i}{2}\right)B\right) \\
&\times \left(1 - \left(\frac{1}{2} + \frac{i\sqrt{3}}{2}\right)B\right)\left(1 - \left(\frac{1}{2} - \frac{i\sqrt{3}}{2}\right)B\right) \\
&= (1-B)^2(1-\sqrt{3}B+B^2)(1-B+B^2).
\end{aligned}$$

Thus the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is

$$(1-B)^2(1-\sqrt{3}B+B^2)(1-B+B^2)z_t = (1-\omega B)^2(1-\sqrt{3}\omega B+\omega^2B^2)(1-\omega B+\omega^2B^2)a_t.$$

Part (h): Since this is a subset of the Part (g) above, the polynomial $\phi(B)$ in this case is

$$\begin{aligned}
\phi(B) &= (1-B)^2\left(1 - \left(\frac{1}{2} + \frac{i\sqrt{3}}{2}\right)B\right)\left(1 - \left(\frac{1}{2} - \frac{i\sqrt{3}}{2}\right)B\right) \\
&= (1-B)^2(1-B+B^2).
\end{aligned}$$

Thus the ARIMA model that gives rise to the equivalent forecasts as general exponential smoothing with a discount coefficient ω is

$$(1-B)^2(1-B+B^2)z_t = (1-\omega B)^2(1-\omega B+\omega^2B^2)a_t.$$

Chapter 8: Special Topics

Notes On The Text

Transfer Function Analysis

To estimate the values of the totality of parameters in the type of transfer function-noise model considered here

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_t \,,$$

or equivalently by multiplying by $\delta(B)\phi(B)$ we obtain a form without any fractions

$$\delta(B)\phi(B)y_t = \phi(B)\omega(B)x_{t-b} + \delta(B)\theta(B)a_t \,.$$

Taking polynomial representations for $\delta(\cdot)$, $\phi(\cdot)$, $\omega(\cdot)$, and $\theta(\cdot)$ of the type discussed in the book where

$$\begin{aligned}
\delta(B) &= 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r \\
\omega(B) &= \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_s B^s \\
\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\
\theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \,.
\end{aligned}$$

Now we see that from the order of the above polynomials the order of the polynomial *products* $\delta(B)\phi(B)$ is $r+p$, that of $\phi(B)\omega(B)$ is $p+s$, and that of $\delta(B)\theta(B)$ is $r+q$. Thus the model above becomes (introducing by notation some new coefficients d_k, c_k, and b_k)

$$\begin{aligned}
y_t &+ d_1 y_{t-1} + d_2 y_{t-2} + \cdots + d_{r+p} y_{t-r-p} \\
&- c_0 x_{t-b} - c_1 x_{t-1-b} - c_2 x_{t-2-b} - \cdots - c_{p+s} x_{t-p-s-b} \\
&+ a_t - b_1 a_{t-1} - b_2 a_{t-2} - b_3 a_{t-3} - \cdots b_{r+q} a_{t-r-q} \,,
\end{aligned}$$

which is equivalent to the books equation 8.23.

Notes on 8.1.2: Forecasting with Transfer Function Models

To derive the forecast equations we begin by recalling the transfer function Model 8.4 as

$$Y_n = \nu(B)X_n + N_n \,. \qquad (123)$$

Defining $x_n = \nabla^d X_n$ and $y_n = \nabla^d Y_n$ the stationary differences of the series X_n and Y_n obtained by taking $\nabla^d = (1-B)^d$ of Equation 123 we get

$$y_n = \nu(B)x_n + n_n \,,$$

with $n_n = \nabla^d N_n$. Assuming we have found a transfer function model of $\nu(B) = \frac{\omega(B)}{\delta(B)} B^b$ and an ARMA noise model of $n_n = \frac{\theta(B)}{\phi(B)} a_n$ this becomes the

$$y_n = \frac{\omega(B)}{\delta(B)} x_{t-b} + \frac{\theta(B)}{\phi(B)} a_n \,, \tag{124}$$

which is equation 8.21 in the book. Expressed with a unit coefficient in front of the white-noise series a_n and the original variables X_n and Y_n we have

$$\frac{\phi(B)}{\theta(B)} \nabla^d Y_n = \frac{\phi(B)}{\theta(B)} \frac{\omega(B)}{\delta(B)} B^b \nabla^d X_n + a_n \,.$$

From this expression if we then Taylor expand the functions in front of Y_n and X_n as

$$\frac{\phi(B)}{\theta(B)} \nabla^d = \frac{\phi(B)}{\theta(B)} (1-B)^d = 1 - \pi_1 B - \pi_2 B^2 - \cdots - \pi_k B^k - \cdots$$

and

$$\frac{\phi(B)}{\theta(B)} \frac{\omega(B)}{\delta(B)} B^d \nabla^d = \frac{\phi(B)}{\theta(B)} \nu(B)(1-B)^d = \nu_0^* + \nu_1^* B + \nu_2^* B^2 + \cdots \,,$$

the model above could be written

$$Y_n = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots + \nu_0^* X_n + \nu_1^* X_{t-1} + \nu_2^* X_{t-2} + \cdots + a_n \,. \tag{125}$$

From this expression we expect that our l-step-ahead predictions, $Y_n(l)$, can be written as a linear combination of α_n the whitened x_n signal and a_n the whitened error signal

$$Y_n(l) = (\eta_0^{(1)} Y_n + \eta_1^{(1)} Y_{n-1} + \eta_2^{(1)} Y_{n-2} + \cdots) + (\eta_0^{(2)} X_n + \eta_1^{(2)} X_{n-1} + \eta_2^{(3)} X_{n-2} + \cdots) \,.$$

As an equivalent expression for $Y_n(l)$ recall that $\nabla^d X_n = \frac{\theta_x(B)}{\phi_x(B)} \alpha_n$ so Y_n satisfy

$$\begin{aligned}
(1-B)^d Y_n &= \nu(B) \nabla^d X_n + \frac{\theta(B)}{\phi(B)} a_n \\
&= \nu(B) \frac{\theta_x(B)}{\phi_x(B)} \alpha_n + \frac{\theta(B)}{\phi(B)} a_n \,,
\end{aligned}$$

by using the input whitening $\frac{\phi_x(B)}{\theta_x(B)}$ of x_t. Solving for Y_n in this expression we obtain

$$Y_n = (1-B)^{-d} \nu(B) \phi_x^{-1}(B) \theta_x(B) \alpha_n + (1-B)^{-d} \theta(B) \phi^{-1}(B) a_n \,. \tag{126}$$

If we define functions $u(B)$ and $\psi(B)$ as

$$\begin{aligned}
u(B) &= (1-B)^{-d} \nu(B) \phi_x^{-1}(B) \theta_x(B) \\
\psi(B) &= (1-B)^{-d} \phi^{-1}(B) \theta(B) \,,
\end{aligned}$$

in Equation 126 we have the α, a representation for Y_{n+l}, by setting $n \to n + l$.

As an example of using these equations for the illustration model given in the text

$$(1-B)(1-\delta B) Y_n = (1-B) \omega_0 X_n + (1-\theta B)(1-\delta B) a_n \,,$$

expanding the products gives

$$(1 - \delta B - B + \delta^2 B^2) Y_n = \omega_0 X_n - \omega_0 X_{n-1} + a_n - (\delta + \theta) a_{n-1} + \delta\theta a_{n-2} .$$

So solving for Y_n we get

$$Y_n = (1+\delta) Y_{n-1} - \delta^2 Y_{n-2} + \omega_0 X_n - \omega_0 X_{n-1} + a_n - (\delta+\theta) a_{n-1} + \delta\theta a_{n-2} .$$

Evaluating this expression at $n \to n + l$ gives

$$Y_{n+l} = (1+\delta) Y_{n+l-1} - \delta^2 Y_{n+l-2} + \omega_0 X_{n+l} - \omega_0 X_{n+l-1} + a_{n+l} - (\delta+\theta) a_{n+l-1} + \delta\theta a_{n+l-2} ,$$

or the equation in the book. Now to derive the forecast error variances we need to know the coefficients u_j and ψ_j in the expression for Y_{n+l} in terms of α_{n+l} and a_{n+l}, where

$$\begin{aligned} Y_n(l) &= u(B)\alpha_{n+l} + \psi(B) a_{n+l} \\ &= u_0 \alpha_{n+l} + u_1 \alpha_{n+l-1} + \cdots + u_{l-1}\alpha_{n+1} + u_l \alpha_n + u_{l+1}\alpha_{n-1} + \cdots \\ &+ \ a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1} + \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots \end{aligned}$$

For the model given here since $x_n = (1 - \theta_x B)\alpha_n$ we have

$$(1 - B) Y_n = \frac{\omega_0}{1 - \delta B}(1 - \theta_x B)\alpha_n + (1 - \theta B) a_n ,$$

or

$$Y_n = \frac{\omega_0}{(1 - B)(1 - \delta B)}(1 - \theta_x B)\alpha_n + \frac{1 - \theta B}{1 - B} a_n .$$

Thus we see that the functions $u(B)$ and $\psi(B)$ are

$$\begin{aligned} u(B) &= \frac{\omega_0}{(1 - B)(1 - \delta B)}(1 - \theta_x B) \\ \psi(B) &= \frac{1 - \theta B}{1 - B} . \end{aligned}$$

From which via Taylor's theorem we can easily compute the coefficients $u_0, u_1, \cdots, \psi_1, \psi_2, \cdots$ needed in the expression for the error variance $V[e_n(l)]$.

Notes on ARIMA time series modeling

In this section we show using the method of "repeated substitution", that the given state-space representation of an ARIMA model is equivalent to the usual definition. If we write the given example system (equation 8.50) in terms of its individual equations we get the following

$$\begin{aligned} S_{t+1,1} &= \phi_1 S_{t,1} + S_{t,2} + a_{t+1} \\ S_{t+1,2} &= \phi_2 S_{t,1} + S_{t,3} - \theta_1 a_{t+1} \\ S_{t+1,3} &= \phi_3 S_{t,1} + S_{t,4} - \theta_2 a_{t+1} \\ &\vdots \\ S_{t+1,k-1} &= \phi_{k-1} S_{t,1} + S_{t,k} - \theta_{k-2} a_{t+1} \\ S_{t+1,k} &= \phi_k S_{t,1} - \theta_{k-1} a_{t+1} . \end{aligned}$$

We insert the second equation in the first equation to find

$$S_{t+1,1} = \phi_1 S_{t,1} + \phi_2 S_{t-1,1} + S_{t-1,3} - \theta_1 a_t + a_{t+1} .$$

Using the equation for $S_{t+1,3}$ in the above we find $S_{t+1,1}$ given by

$$S_{t+1,1} = \phi_1 S_{t,1} + \phi_2 S_{t-1,1} + \phi_3 S_{t-2,1} + S_{t-2,4} - \theta_2 a_{t-1} - \theta_1 a_t + a_{t+1} .$$

Using the equation for $S_{t+1,4}$ in the above we find $S_{t+1,1}$ given by

$$S_{t+1,1} = \phi_1 S_{t,1} + \phi_2 S_{t-1,1} + \phi_3 S_{t-2,1} + \phi_4 S_{t-3,1} + S_{t-3,5} - \theta_3 a_{t-2} - \theta_2 a_{t-1} - \theta_1 a_t + a_{t+1} .$$

Continuing in this manner we can see how the solution for $S_{t,1}$ of this linear system (or the state-space representation) in the same as the solution of the more standard ARIMA(p,q) state space equation

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} .$$

Notes on the Section Entitled: Bayesian Forecasting

From the given linear expression for $\beta_{t+l} = H_t \beta_t + \varepsilon_t$ and using $\hat{\beta}_{t+l|t} = A^l \hat{\beta}_{t|t}$ to be the expectation of $\beta_{t+l|t}$ we have

$$\begin{aligned}
P_{t+l|t} &= E\left((\beta_{t+l} - \hat{\beta}_{t+l|t})(\beta_{t+l} - \hat{\beta}_{t+l|t})' \right) \\
&= E\left(\left[A^l(\beta_t - \hat{\beta}_{t|t}) + a_{t+l} + A a_{t+l-1} + \cdots + A^{l-1} a_{t+1} \right] \right. \\
\times &\qquad \left. \left[A^l(\beta_t - \hat{\beta}_{t|t}) + a_{t+l} + A a_{t+l-1} + \cdots + A^{l-1} a_{t+1} \right]' \right) \\
&= A^l P_{t|t} (A')^l + E(a_{t+l} a'_{t+l}) + A E(a_{t+l-1} a'_{t+l-1}) A' + \cdots + A^{l-1} E(a_{t+1} a'_{t+1}) (A')^{l-1} \\
&= A^l P_{t|t} (A')^l + R_2 + A R_2 A' + \cdots + A^{l-1} R_2 (A')^{l-1} ,
\end{aligned}$$

using the properties of zero mean and uncorrelated of the random shocks a's.

Notes on the Section: Models with Time Varying Coefficients

When the system is given by the following set of equations

$$\begin{aligned}
y_t &= x'_t \beta_t + \varepsilon_t \\
\beta_{t+1} &= \beta_t + a_{t+1} ,
\end{aligned}$$

To cast this in the Kalman filtering framework we will take the vector β_t as the "state", the scalar y_t as the measurement, and the vector x'_t as the measurement matrix H_t. Since y_t is a scalar, to parameterize this model we can take $R_2 = V(a_t) = \sigma^2 \Omega$ where $\sigma^2 = V(\varepsilon_t) = R_1$ (thus we have "factored" out the scalar variance σ^2 from the definition of the variance of

a_t). To complete the setup for the Kalman filter we assume a parametric model of the state β of the following normal form

$$
\begin{aligned}
(\beta_{t+1}|Y_t) &\sim N(\hat{\beta}_{t+1|t}, \sigma^2 P_{t+1|t}) \\
(\beta_{t+1}|Y_{t+1}) &\sim N(\hat{\beta}_{t+1|t+1}, \sigma^2 P_{t+1|t+1}) \, .
\end{aligned}
$$

A direct application of the Kalman filtering equations specified to this specific problem is then given by

$$
\begin{aligned}
\hat{\beta}_{t+1|t} &= \hat{\beta}_{t|t} \\
\sigma^2 P_{t+1|t} &= \sigma^2 P_{t|t} + \sigma^2 \Omega \Rightarrow P_{t+1|t} = P_{t|t} + \Omega \\
\hat{\beta}_{t+1|t+1} &= \hat{\beta}_{t+1|t} + k_{t+1}(y_{t+1} - x'_{t+1}\hat{\beta}_{t+1|t}) \\
\sigma^2 P_{t+1|t+1} &= \sigma^2 P_{t+1|t} - k_{t+1} x'_{t+1} \sigma^2 P_{t+1|t} \Rightarrow P_{t+1|t+1} = P_{t+1|t} - k_{t+1} x'_{t+1} P_{t+1|t} \\
k_{t+1} &= \sigma^2 P_{t+1|t} x'_{t+1} (x'_{t+1} \sigma^2 P_{t+1|t} x_{t+1} + \sigma^2)^{-1} = (x'_{t+1} P_{t+1|t} x_{t+1} + 1)^{-1} P_{t+1|t} x'_{t+1} \, .
\end{aligned}
$$

These match the books equation 8.58. Since there are only simple modifications due to the time update (drift), if we simplify notation by letting $\hat{\beta}_t = \hat{\beta}_{t|t}$ and $P_t = P_{t|t}$ we can combine the above equations into just two equations giving

$$
\begin{aligned}
\hat{\beta}_{t+1} &= \hat{\beta}_t + [1 + x'_{t+1}(P_t + \Omega)x_{t+1}]^{-1}(P_t + \Omega)x_{t+1}(y_{t+1} - x'_{t+1}\hat{\beta}_t) && (127) \\
P_{t+1} &= P_t + \Omega - [1 + x'_{t+1}(P_t + \Omega)x_{t+1}]^{-1}(P_t + \Omega)x_{t+1}x'_{t+1}(P_t + \Omega) \, , && (128)
\end{aligned}
$$

which matches the books equations 8.59. As a further simplification if we assume our state β has *no* time dependence then $\beta_{t+1} = \beta_t = \beta$ (a constant) and $V(a_t) = 0 = \sigma^2 \Omega$ or $\Omega = 0$ and the above two equations simplify to

$$
\begin{aligned}
\hat{\beta}_{t+1} &= \hat{\beta}_t + [1 + x'_{t+1} P_t x_{t+1}]^{-1} P_t x_{t+1}(y_{t+1} - x'_{t+1}\hat{\beta}_t) && (129) \\
P_{t+1} &= P_t - [1 + x'_{t+1} P_t x_{t+1}]^{-1} P_t x_{t+1} x'_{t+1} P_t \, . && (130)
\end{aligned}
$$

which match the books equations 8.60.

Exercise Solutions

Exercise 8.1 (impulse response weights for various transfer function models)

With a model for Y_t given by

$$
Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots = v(B) X_t \, ,
$$

where the function $v(B) = v_0 + v_1 B + v_2 B^2 + \cdots$ is call a *transfer* function where the coefficients $v_0, v_1, \cdots$ are called the *impulse response weights*. For this exercise to find these values we will compute the Taylor series for each of the given functions and plot their values as a function of their index in a "stem plot". See the MATHEMATICA file `prob_8_1.nb` where this is done.

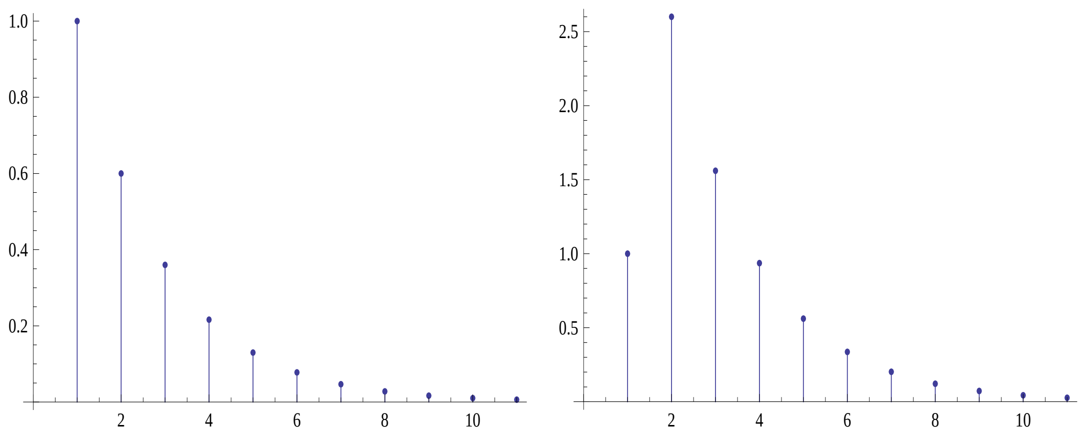


Figure 55: **Left:** The impulse response weights for the transfer function $\omega_0/(1-\delta_1 B)$. **Right:** The impulse response weights for the transfer function $(\omega_0 - \omega_1 B)/(1 - \delta_1 B)$.

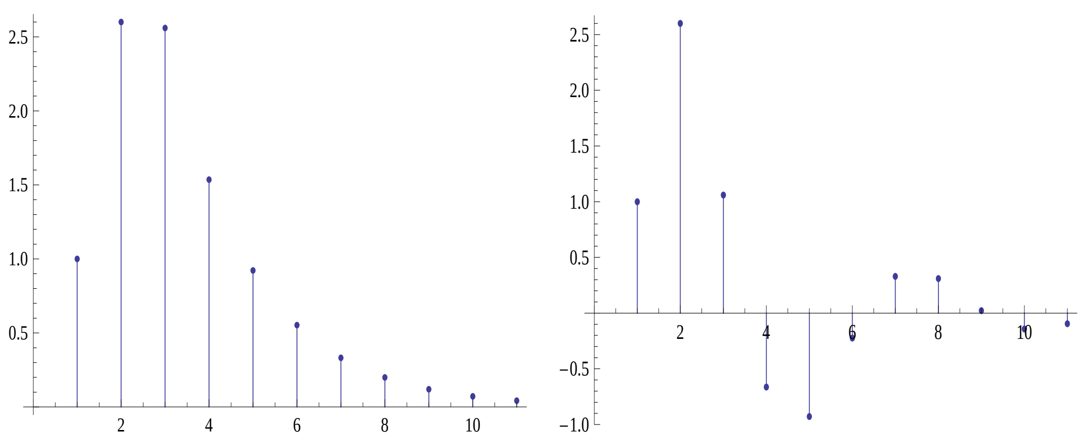


Figure 56: **Left:** The impulse response weights for the transfer function $(\omega_0 - \omega_1 B - \omega_2 B^2)/(1 - \delta_1 B)$. **Right:** The impulse response weights for the transfer function $(\omega_0 - \omega_1 B)/(1 - \delta_1 B - \delta_2 B^2)$.

Exercise 8.2 (the gas furnace example)

For this exercise we are asked to evaluate $Y_{206}(l)$ and $V(e_{260}(l))$ for $l = 1, 2, \cdots, 6$. To solve this problem we first need to derive an expression for Y_{n+l} implied by the given model for Y_t. To begin we multiply by $(1 - 0.57B)(1 - 1.53B + 0.63B^2)$ to get an expression without fractions and find a left hand side (LHS) given by

$$\text{LHS} = (1 - 0.57B)(1 - 1.53B + 0.63B^2)(Y_t - 53.51)\,,$$

with a corresponding right hand side (RHS) given by

$$\text{RHS} = -(0.53 + 0.37B + 0.51B^2)(1 - 1.53B + 0.63B^2)(X_{t-3} + 0.057) + (1 - 0.57B)a_t\,.$$

Expanding the polynomial products above gives

$$\begin{aligned}
\text{LHS} &= (1 - 2.1B + 1.5021B^2 - 0.3591B^3)(Y_t - 53.51) \\
\text{RHS} &= (-0.53 + 0.4409B - 0.2778B^2 + 0.5472B^3 - 0.3213B^4)(X_{t-3} + 0.057) + (1 - 0.57B)a_t\,.
\end{aligned}$$

Solving for Y_t in the LHS expression and equating it to the RHS we arrive at an expression for Y_t in terms of past values of X_t and Y_t

$$\begin{aligned}
Y_t - 53.51 &= 2.1(Y_{t-1} - 53.51) - 1.5021(Y_{t-2} - 53.51) + 0.3591(Y_{t-3} - 53.51) \\
&- 0.53(X_{t-3} + 0.057) + 0.4409(X_{t-4} + 0.057) - 0.2778(X_{t-5} + 0.057) \\
&+ 0.5472(X_{t-6} + 0.057) - 0.3213(X_{t-7} + 0.057) \\
&+ a_t - 0.57a_{t-1}\,.
\end{aligned}$$

Using this we can write Y_{t+l} directly as

$$\begin{aligned}
Y_{t+l} &= -53.51 + 2.1(Y_{t+l-1} - 53.51) - 1.5021(Y_{t+l-2} - 53.51) + 0.3591(Y_{t+l-3} - 53.51) \\
&- 0.53(X_{t+l-3} + 0.057) + 0.4409(X_{t+l-4} + 0.057) - 0.2778(X_{t+l-5} + 0.057) \\
&+ 0.5472(X_{t+l-6} + 0.057) - 0.3213(X_{t+l-7} + 0.057) \qquad (131) \\
&+ a_{t+l} - 0.57a_{t+l-1}\,.
\end{aligned}$$

To compute the expected forecast values $Y_n(l)$ we recall its definition as the conditional expectation of Y_{n+l} given everything we know up until time $t = n$ or

$$Y_n(l) = E[Y_{n+l}|Y_n, Y_{n-1}, \ldots; X_n, X_{n-1}, \ldots]\,.$$

Using this definition and the notation $E[|\cdot]$ to simply the conditional expectation notation we have

$$\begin{aligned}
Y_n(l) &= -53.51 + 2.1(E[Y_{t+l-1}|\cdot] - 53.51) - 1.5021(E[Y_{t+l-2}|\cdot] - 53.51) + 0.3591(E[Y_{t+l-3}|\cdot] - 53.51) \\
&- 0.53(E[X_{t+l-3}|\cdot] + 0.057) + 0.4409(E[X_{t+l-4}|\cdot] + 0.057) - 0.2778(E[X_{t+l-5}|\cdot] + 0.057) \\
&+ 0.5472(E[X_{t+l-6}|\cdot] + 0.057) - 0.3213(E[X_{t+l-7}|\cdot] + 0.057) \\
&+ a_{t+l} - 0.57a_{t+l-1}\,.
\end{aligned}$$

Note that to evaluate the above we need the expressions $E[X_{t+l}|\cdot]$ which are the predictions of X_{t+l} using the model for X_{t+l} which is obtained from

$$X_{t+l} + 0.057 = 1.97(X_{t-1+l} + 0.057) - 1.37(X_{t-2+l} + 0.057) + 0.34(X_{t-3+l} + 0.057) + \alpha_{t+l}\,.$$

To use these expressions to make future predictions of Y_n and X_n one way we can imagine doing this by keeping four arrays that hold the predicted values. The four arrays would hold the values of any observed X_n and X's future predictions $X_n(l)$, observed values of Y_n and Y's future predictions $Y_n(l)$, a_n and a's expected future predictions which are zero, and finally α_n and α's expected future predictions which are again zero. As an example, to predict from the index 260 onward these arrays would be populated as (we introduce a fifth array n for convenience)

$$
\begin{aligned}
n &= [258, 259, 260, 261, 262, 263, \cdots] \\
\text{xa} &= [X_{258}, X_{259}, X_{260}, X_{260}(1), X_{260}(2), X_{260}(3), \cdots] \\
\text{ya} &= [Y_{258}, Y_{259}, Y_{260}, Y_{260}(1), Y_{260}(2), Y_{260}(3), \cdots] \\
\text{aa} &= [Y_{258} - Y_{257}(1), Y_{259} - Y_{258}(1), Y_{260} - Y_{259}(1), 0, 0, 0, \cdots] \\
\text{Aa} &= [X_{258} - X_{257}(1), X_{259} - X_{258}(1), X_{260} - X_{259}(1), 0, 0, 0, \cdots] \,.
\end{aligned}
$$

The update equations can be written in terms of these arrays. For example, to compute l steps ahead from Y_n can now be written as (assuming that the array index `n` corresponds to the index for 260 and that there is sufficient data in all arrays)

```
for( l=1, l <= L; ++l ){
  xa[n+l] = - 0.057 + 1.97 (xa[n+l-1] + 0.057) - 1.37 (xa[n+l-2] + 0.057) \
            + 0.34 (xa[n+l-3] + 0.057) + Aa[n+l]
  ya[n+l] = - 53.51 + 2.1 ( ya[n+l-1] - 53.51 ) - 1.5021 ( ya[n+l-2] - 53.51 ) \
            + 0.3591 ( ya[n+l-3] - 53.51 ) - 0.53 ( xa[n+l-3] + 0.057 ) \
            + 0.4409 ( xa[n+l-4] + 0.057 ) - 0.2778 ( xa[n+l-5] + 0.057 ) \
            + 0.5472 ( xa[n+l-6] + 0.057 ) - 0.3213 ( xa[n+l-7] + 0.057 ) \
            + aa[n+l] - 0.57 aa[n+l-1]
}
```

Predictions using this method give the following

The expression for the variance $V[e_n(l)]$ is given by

$$
V[e_n(l)] = \sum_{j=0}^{l-1} (u_j^2 \sigma_\alpha^2 + \psi_j^2 \sigma^2) = \sum_{j=0}^{l-1} (0.0353 u_j^2 + 0.0561 \psi_j^2) \,,
$$

where u_j and ψ_j are the coefficients in the linear filter model for Y_n i.e. when Y_n is written in the form

$$
Y_n = u(B)\alpha_n + \psi(B) a_n \,.
$$

This form can be obtained by solving for X_n in terms of α_n and putting the resulting expression into the expression for Y_t. For example in this problem we have X_t given by

$$
X_t = -0.057 + \frac{\alpha_t}{1 - 1.97B + 1.37B^2 - 0.34B^3} \,,
$$

so that using this the expression for Y_t becomes

$$
\begin{aligned}
Y_t - 53.51 &= -\frac{0.53 + 0.37B + 0.51B^2}{1 - 0.57B} \left(\frac{1}{1 - 1.97B + 1.37B^2 - 0.34B^3} \right) \alpha_{t-3} \\
&+ \frac{1}{1 - 1.53B + 0.36B^2} a_t \,.
\end{aligned}
$$

From the above our functions $u(B)$ and $\psi(B)$ are given by

$$
\begin{aligned}
u(B) &= -\frac{(0.53 + 0.37B + 0.51B^2)B^3}{(1 - 0.57B)(1 - 1.97B + 1.37B^2 - 0.34B^3)} \\
&= -0.53B^3 - 1.7162B^4 - 3.54791B^5 - 5.32746B^6 - 6.50813B^7 - 6.89408B^8 + O(B)^6 \\
\psi(B) &= \frac{1}{1 - 1.53B + 0.36B^2} \\
&= 1 + 1.97B + 2.5109B^2 + 2.58757B^3 + 2.32739B^4 + 1.89368B^5 + O(B)^6 \,.
\end{aligned}
$$

See the MATHEMATICA file `prob_8_1.nb` where some of these results are derived.

Exercise 8.7 (ARIMA(0,1,1) as a state-space model)

For the ARIMA(0,1,1) model $(1 - B)y_t = (1 - \theta B)a_t$ we have y_t given by

$$
y_t = y_{t-1} + a_t - \theta a_{t-1} \,,
$$

Thus we take our state vector S_t to be $S_t = \left[\begin{array}{c} y_t \\ -\theta a_t \end{array} \right]$ with $k = \max(p, q+1) = \max(0, 2) = 2$ our state update equations are given by

$$
\left[\begin{array}{c} S_{t+1,1} \\ S_{t+1,2} \end{array} \right] = \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right] \left[\begin{array}{c} S_{t,1} \\ S_{t,2} \end{array} \right] + \left[\begin{array}{c} 1 \\ -\theta \end{array} \right] a_{t+1} \,.
$$

The individual equations are then

$$
\begin{aligned}
S_{t+1,1} &= S_{t,1} + S_{t,2} + a_{t+1} \\
S_{t+1,2} &= -\theta a_{t+1} \,.
\end{aligned}
$$

We can check that that these are correct by putting the second equation into the first. When we do this we find

$$
S_{t+1,1} = S_{t,1} - \theta a_t + a_{t+1} \,,
$$

which is an ARIMA(0,1,1) model as claimed.

Exercise 8.8 (ARIMA(1,1,1) as a state-space model)

The ARIMA(1,1,1) model is $(1 - B)(1 - \phi B)z_t = (1 - \theta B)a_t$ or

$$
(1 - (1 + \phi)B + \phi B^2)z_t = (1 - \theta B)a_t \,,
$$

or

$$
z_t = (1 + \phi)z_{t-1} - \phi z_{t-2} + a_t - \theta a_{t-1} \,. \tag{132}
$$

Here we have $k = \max(p, q+1) = \max(2, 1+1) = 2$, $\phi_1 = 1+\phi$, and $\phi_2 = -\phi$ to get the system

$$\left[\begin{array}{c} S_{t+1,1} \\ S_{t+1,2} \end{array}\right] = \left[\begin{array}{cc} 1+\phi & 1 \\ -\phi & 0 \end{array}\right] \left[\begin{array}{c} S_{t,1} \\ S_{t,2} \end{array}\right] + \left[\begin{array}{c} 1 \\ -\theta \end{array}\right] a_{t+1} \,.$$

Lets check that this reduces to the requested ARIMA(1,1,1) model. In the equation form we have

$$\begin{aligned} S_{t+1,1} &= (1+\phi)S_{t,1} + S_{t,2} + a_{t+1} \\ S_{t+1,2} &= -\phi S_{t,1} - \theta a_{t+1} \,. \end{aligned}$$

We put the second equation into the first to get

$$S_{t+1,1} = (1+\phi)S_{t,1} - \phi S_{t-1,1} + a_{t+1} - \theta a_{t+1} \,,$$

which is the same as Equation 132 showing the desired equivalence.

Exercise 8.9 (the Kalman updating equations for simple linear regression)

For a regression model like this one where we assume a model of the form

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \,,$$

with β_0, β_1 constant, a state-space representation is obtained by taking $S_t = \left[\begin{array}{c} \beta_0 \\ \beta_1 \end{array}\right]$ with no process/plant noise and the identity for the system matrix A, so the Kalman system equation is simply $S_{t+1} = S_t$. Let $H_t = \mathbf{x}_t' = \left[\begin{array}{cc} 1 & x_t \end{array}\right]$, where x_t is the scalar measurement received at time t corresponding to the scalar response y_t. Then the Kalman updating equations become (given an initial guess at β_0 and β_1 which could be 0)

$$\begin{aligned} \hat{S}_{t+1|t} &= \hat{S}_{t|t} \\ P_{t+1|t} &= P_{t|t} \\ k_{t+1} &= P_{t+1|t} H_{t+1}' (H_{t+1} P_{t+1|t} H_{t+1}' + R_1)^{-1} \\ &= P_{t|t} \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] \left(\frac{1}{\left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_{t|t} \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] + \sigma_\varepsilon^2} \right) \\ \hat{S}_{t+1|t+1} &= \hat{S}_{t+1|t} + k_{t+1}(y_{t+1} - H_{t+1}\hat{S}_{t+1|t}) \\ P_{t+1|t+1} &= P_{t+1|t} - k_{t+1} H_{t+1} P_{t+1|t} \,. \end{aligned}$$

Now since there is no time component / dynamics associated with our state S_t we can drop the first two equations above which corresponding to state propagation and simply iterate the state and covariance update equations. These equations then become after dropping the conditional notation

$$\begin{aligned} k_{t+1} &= P_t H_{t+1}' (H_{t+1} P_t H_t' + R_1)^{-1} \\ \hat{S}_{t+1} &= \hat{S}_t + k_{t+1}(y_{t+1} - H_{t+1}\hat{S}_t) \\ P_{t+1} &= P_t - k_{t+1} H_{t+1} P_t \,. \end{aligned}$$

Again since $H_t = \mathbf{x}_t' = \left[\begin{array}{cc} 1 & x_t \end{array}\right]$ and $\hat{S}_t = \left[\begin{array}{c} \beta_0^{(t)} \\ \beta_1^{(t)} \end{array}\right]$, $R_1 = \sigma_\varepsilon^2$, and P_t is the covariance matrix of $\left[\begin{array}{c} \beta_0^{(t)} \\ \beta_1^{(t)} \end{array}\right]$. These vector equations become

$$\begin{aligned}
k_{t+1} &= P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] \frac{1}{\left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] + \sigma_\varepsilon^2} \\
\hat{S}_{t+1} &= \hat{S}_t + k_{t+1}(y_{t+1} - \left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] \hat{S}_t) \\
P_{t+1} &= P_t - P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] \left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_t \frac{1}{\left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] + \sigma_\varepsilon^2} \\
&= P_t - \left(\frac{1}{\left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] + \sigma_\varepsilon^2} \right) P_t \left[\begin{array}{cc} 1 & x_{t+1} \\ x_{t+1} & x_{t+1}^2 \end{array}\right] P_t \,.
\end{aligned}$$

Let P_t be given by

$$P_t = \left[\begin{array}{cc} P_{11}^{(t)} & P_{12}^{(t)} \\ P_{12}^{(t)} & P_{22}^{(t)} \end{array}\right] .$$

Then one of the required matrix products in the above expressions become

$$\begin{aligned}
\left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] &= \left[\begin{array}{cc} 1 & x_{t+1} \end{array}\right] \left[\begin{array}{c} P_{11}^{(t)} + x_{t+1} P_{12}^{(t)} \\ P_{12}^{(t)} + x_{t+1} P_{22}^{(t)} \end{array}\right] \\
&= P_{11}^{(t)} + x_{t+1} P_{12}^{(t)} + x_{t+1} P_{12}^{(t)} + x_{t+1}^2 P_{22}^{(t)} \\
&= P_{11}^{(t)} + 2 x_{t+1} P_{12}^{(t)} + x_{t+1}^2 P_{22}^{(t)} \,.
\end{aligned}$$

Using this our state and covariance update equations become

$$\begin{aligned}
\left[\begin{array}{c} \beta_0^{(t+1)} \\ \beta_1^{(t+1)} \end{array}\right] &= \left[\begin{array}{c} \beta_0^{(t)} \\ \beta_1^{(t)} \end{array}\right] + \frac{y_{t+1} - (\beta_0^{(t)} + \beta_0^{(t)} x_{t+1})}{P_{11}^{(t)} + 2 x_{t+1} P_{12}^{(t)} + x_{t+1}^2 P_{22}^{(t)} + \sigma_\varepsilon^2} \left(P_t \left[\begin{array}{c} 1 \\ x_{t+1} \end{array}\right] \right) \\
&= \left[\begin{array}{c} \beta_0^{(t)} \\ \beta_1^{(t)} \end{array}\right] + \left(\frac{y_{t+1} - (\beta_0^{(t)} + \beta_0^{(t)} x_{t+1})}{P_{11}^{(t)} + 2 x_{t+1} P_{12}^{(t)} + x_{t+1}^2 P_{22}^{(t)} + \sigma_\varepsilon^2} \right) \left[\begin{array}{c} P_{11}^{(t)} + x_{t+1} P_{12}^{(t)} \\ P_{12}^{(t)} + x_{t+1} P_{22}^{(t)} \end{array}\right] ,
\end{aligned}$$

and

$$\begin{aligned}
\left[\begin{array}{cc} P_{11}^{(t+1)} & P_{12}^{(t+1)} \\ P_{12}^{(t+1)} & P_{22}^{(t+1)} \end{array}\right] &= \left[\begin{array}{cc} P_{11}^{(t)} & P_{12}^{(t)} \\ P_{12}^{(t)} & P_{22}^{(t)} \end{array}\right] \\
&- \left(\frac{1}{P_{11}^{(t)} + 2 x_{t+1} P_{12}^{(t)} + x_{t+1}^2 P_{22}^{(t)} + \sigma_\varepsilon^2} \right) \\
&\times \left[\begin{array}{cc} P_{11}^{(t)} & P_{12}^{(t)} \\ P_{12}^{(t)} & P_{22}^{(t)} \end{array}\right] \left[\begin{array}{cc} 1 & x_{t+1} \\ x_{t+1} & x_{t+1}^2 \end{array}\right] \left[\begin{array}{cc} P_{11}^{(t)} & P_{12}^{(t)} \\ P_{12}^{(t)} & P_{22}^{(t)} \end{array}\right] .
\end{aligned}$$

One could simplify these expressions further to produce *scalar* update equations if desired.

Exercise 8.10 (the model $y_t = \mu_t + \varepsilon_t$ and $\mu_{t+1} = \mu_t + a_{t+1}$)

Part (a): Let the state be equal to μ_t. From this system the state propagation equation looks to be

$$\mu_{t+1} = 1\mu_t + a_{t+1}\,,$$

so $A = 1$ and our scalar state is $\beta_t = \mu_t$. The Kalman measurement equation in general from of

$$y_t = H_t\beta_t + \varepsilon_t\,,$$

will match our received value of y_t if we take $H_t \equiv 1$. Note this is an example of system like discussed on page 156.

Part (b): Now the Kalman update equations for a system like this were worked out on page 156. Here we have we have $\Omega = \frac{V(a_t)}{V(\varepsilon_t)} = \frac{\omega}{1} = \omega$, with $x_t' = H_t = 1$, so in steady-state where $P_{t+1} = P_t = P$, Equation 128 becomes

$$P = (P+\omega) - (1 + (P+\omega))^{-1}(P+\omega)^2\,,$$

or $\omega = \frac{(P+\omega)^2}{1+P+\omega}$ or

$$(P+\omega)^2 - \omega(P+\omega) - \omega = 0\,,$$

Solving this quadratic for $P+\omega$ gives for P the solution

$$P(\omega) = -\frac{\omega}{2} \pm \frac{1}{2}\sqrt{\omega^2 + 4\omega}\,, \tag{133}$$

Thus the update equation from Equation 127 is given by

$$\mu_{t+1} = \mu_t + \frac{P(\omega)+\omega}{1+P(\omega)+\omega}(y_{t+1} - \mu_t)\,,$$

which is the expression for simple exponential smoothing with a smoothing constant, α, given by

$$\alpha = \frac{P(\omega)+\omega}{1+P(\omega)+\omega}\,,$$

and $P(\omega)$ is given by Equation 133.

Exercise 8.11 (a linear growth model)

Part (a): Let the state for a Kalman model be equal to $S_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$, then the measurement equation has $H_t = \begin{bmatrix} 1 & 0 \end{bmatrix}$ so that $y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \varepsilon_t$ and the state propagation equation is given by

$$\begin{aligned} S_{t+1} &= \left[\begin{array}{c} \mu_{t+1} \\ \beta_{t+1} \end{array}\right] = \left[\begin{array}{c} \mu_t + \beta_{t+1} + a_{1,t+1} \\ \beta_t + a_{2,t+1} \end{array}\right] \\ &= \left[\begin{array}{c} \mu_t + \beta_t + a_{2,t+1} + a_{1,t+1} \\ \beta_t + a_{2,t+1} \end{array}\right] \\ &= \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} \mu_t \\ \beta_t \end{array}\right] + \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} a_{1,t+1} \\ a_{2,t+1} \end{array}\right], \end{aligned}$$

so $A = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right]$ and $a_{t+1} = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right] \left[\begin{array}{c} a_{1,t+1} \\ a_{2,t+1} \end{array}\right]$.

References

[1] W. G. Kelley and A. C. Peterson. *Difference Equations. An Introduction with Applications.* Academic Press, New York, 1991.