

Exam 2 is Wed. Oct. 16. **You are allowed 6 sheets of notes and a calculator.** The exam covers HW2E, HW3-6, back of Q2, and Q3-6. Numbers refer to types of problems on exam.

**From Exam 1 review**, know pages 2-4. Will not have 1)-5) on Exam 2.

The output for an ARIMA(p,d,q) model is exactly like that on p. 4 of exam 1 review, except the default is to not give the “intercept” line if  $d \geq 1$ . If  $d = 0$ , let  $X_t = Y_t$ . If  $d = 1$ , then the first difference  $X_t = W_t = Y_t - Y_{t-1}$  for  $t = 2, \dots, n$ . If  $d = 2$ , then the second difference  $Z_t = Y_t - 2Y_{t-1} + Y_{t-2}$  for  $t = 3, \dots, n$ . For all 3 values of  $d$ , want  $X_t$  to follow a stationary and invertible ARMA(p,q) = ARIMA(p,0,q) model. Rarely use  $d = 2$  and hardly ever use  $d \geq 3$ .

11) **Know for final:** The 95% CI for  $\phi_k$  or  $\theta_k$  is (LCI, UCI) as in 9).

12) **Know for final:** The 4 step test of hypotheses for  $H_0: \phi_k = 0$  or  $H_0: \theta_k = 0$  is exactly the same as 10), except for the conclusion when testing AR parameters  $H_0: \phi_k = 0$ : if  $d = 1$ , replace  $Y_{t-k}$  by  $W_{t-k}$ , while if  $d = 2$ , replace  $Y_{t-k}$  by  $Z_{t-k}$ . See HW4 Bd).

13) **Know for final:** Use the ACF and PACF to identify  $p$  and  $q$  exactly as in 8). If  $d = 0$ , use the ACF and PACF of  $Y_t$  to get an ARMA(p,q) = ARIMA(p,0,q) model. If  $d = 1$ , use the ACF and PACF of  $W_t$  to get an ARIMA(p,1,q) model. If  $d = 2$ , use the ACF and PACF of  $Z_t$  to get an ARIMA(p,2,q) model. See exam 1 1), Q2 3), HW4 C) and HW2 E).

14) **Know for final:** Let  $X_t = Y_t$  if  $d = 0$ ,  $X_t = W_t$  if  $d = 1$ , and  $X_t = Z_t$  if  $d = 2$ . The *response plot* of  $\hat{X}_t$  vs  $X_t$  should scatter about the identity line with unit slope and zero intercept with no other pattern if the model is adequate. The vertical deviation of  $X_t$  from the identity line is the residual  $X_t - \hat{X}_t = \hat{\epsilon}_t$ . A *residual plot* of  $\hat{X}_t$  vs  $\hat{\epsilon}_t$  or of  $t$  vs  $\hat{\epsilon}_t$  should scatter about the  $\hat{\epsilon}_t = 0$  line, with no other pattern if the model is adequate. The evenly spaced  $t$  tends to make the residual plot of  $t$  vs  $\hat{\epsilon}_t$  to look “spiky.” Be able to recognize and sketch response and residual plots. See exam 1 4).

15) **Know for final:** Let  $X_t$  be as in 14). Hence  $\hat{X}_t = \hat{Y}_t$  for  $d = 0$ ,  $\hat{X}_t = \hat{W}_t$  for  $d = 1$ , and  $\hat{X}_t = \hat{Z}_t$  for  $d = 2$ . Given  $R$  output,  $X_{t-1}, \dots, X_{t-p}$ ,  $\hat{\epsilon}_{t-1}, \dots, \hat{\epsilon}_{t-q}$ , be able to find

$$\hat{X}_t = \hat{\tau} + \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p} + \hat{\theta}_1 \hat{\epsilon}_{t-1} + \dots + \hat{\theta}_q \hat{\epsilon}_{t-q}$$

for  $j \leq t \leq n+1$  where often  $j = \max(p, q) + 1$ . For  $d = 0$ , the MA(q) = ARIMA(0,0,q) model has  $\hat{\tau} = \hat{\mu}_Y$  given in the “intercept” line. For  $d = 0$  with  $p \geq 1$ , the “intercept” line gives  $\hat{\mu}_Y \neq \hat{\tau}$ , and  $\hat{\tau}$  needs to be given. For the ARIMA(p,d,q) model with  $d \geq 1$ , take  $\hat{\tau} = 0$  unless told otherwise. See HW4 Bg).

Different software produces different  $\hat{\epsilon}_t$  and  $\hat{X}_t$  for  $1 \leq t \leq \max(p, q)$ . When  $q \geq 1$ , and there is often a “burn in” period before the  $\hat{\epsilon}_t$  and  $\hat{X}_t$  are similar from the different programs.

16) For ARIMA(p,d,q) models, the ARMA(p,q) model for  $X_t$  is stationary if the roots of the polynomial  $\phi(B) = 0$  each have modulus  $> 1$  and so are all outside the unit circle, and the ARMA(p,q) model for  $X_t$  is invertible if the roots of the polynomial  $\theta(B) = 0$  each have modulus  $> 1$  and so are all outside the unit circle. The polynomials are  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  and  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  where  $R$  replaces  $-\theta_j$

by  $\theta_j$ . For an AR(2) model (change 0 to d for an ARI(p,d) = ARIMA(p,d,0) model), output looks like

```
out <- arima(y,c(2,0,0)); p<-2
Mod(polyroot(c(1, -out$coef[-(p+1)])))
[1] 2.733461 2.733461
```

For an MA(2) = ARIMA(0,0,2) model (change 0 to d for an IMA(d,q) = ARIMA(0,d,q) model), output looks like (no negative sign because in  $R$ , the  $\hat{\theta}_i$  are  $-\hat{\theta}_i$  from those of the book)

```
out <- arima(y,c(0,0,2)); q<-2
Mod(polyroot(c(1, out$coef[-(q+1)])))
[1] 1.500266 4.255771
```

For an ARMA(2,2) = ARIMA(2,0,2) model (change 0 to d for an ARIMA(p,d,q) model), output looks like

```
out <- arima(y,c(2,0,2)); p<-2; q<-2;
#AR phi(B) = 0
Mod(polyroot(c(1, -out$coef[1:p])))
[1] 1.134829 2.632953
#MA theta(B) = 0
Mod(polyroot(c(1, out$coef[(p+1):(p+q)])))
[1] 1.000004 6.469955
```

The output gives the moduli of the roots of the polynomials  $\hat{\phi}(B) = 0$  and  $\hat{\theta}(B) = 0$ . The models are “good” in that all of the moduli are greater than 1. Since these are estimated roots, moduli near 1, (like 1.13 or 1.000004) may suggest that the ARMA(p,q) model for the time series  $X_t$  is not stationary or not invertible. The model would be bad if any of the moduli are less than 1. See HW4 A).

17) Let  $X_t = \nabla^d Y_t$  be the differenced time series. The first difference is  $X_t = W_t = \nabla Y_t = Y_t - Y_{t-1}$ . The second difference is  $X_t = Z_t = \nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla W_t = Y_t - 2Y_{t-1} + Y_{t-2}$ . If  $Y_t$  follows an ARIMA(p,d,q) model, want  $X_t$  to follow a stationary and invertible ARMA(p,q) = ARIMA(p,0,q) model.

18) ARIMA(p,d,q) models have  $\{Y_t\}$  nonstationary if  $d \geq 1$ .

- a) Often need  $d = 1$  if the ACF and/or PACF of  $Y_t$  do not decay to 0 fast: there is linear decay or the sinusoidal decay does not damp out quickly.
- b) Often need  $d = 1$  if the plot of the time series  $Y_t$  does not appear to have constant mean, especially if there is a stochastic linear trend.
- c) If the time series  $Y_t$  is nonstationary, but different large segments of the series behave much like the rest of the series after allowing for changes in level and/or slope, often need  $d = 1$ .

19) If first differencing  $d = 1$  works, then want  $W_t$  to be a stationary and invertible ARMA(p,q) model. So the plot of  $W_t$  should oscillate about a constant mean and the variability of long segments of the plot should be similar since stationary time series have constant mean and constant variance. The ACF and PACF of  $W_t$  should both have spikes

$r_k$  and  $\hat{\phi}_{kk}$  that go to 0 rapidly in that roughly 95% of the spikes are within the  $\pm 2$  SE bars after the first few lags. The residual ACF and PACF for the ARIMA(p,1,q) model should resemble those of white noise, so roughly 95% of spikes for lags  $k \geq 1$  should be within the  $\pm 2$  SE bars (horizontal lines).

20) Similar remarks to 19) apply with  $d = 2$  and  $Z_t$  instead of  $W_t$ .

21) Suppose  $X_t > 0$  for all  $t$  (if  $\tilde{X}_t = 0$  is possible, use  $X_t = \tilde{X}_t + 1$ ). Use  $Y_t = \log(X_t) = \ln(X_t)$  and first differences  $d = 1$  a) or b) hold.

a) Use  $Y_t = \log(X_t)$  if  $E(X_t) = \mu_t$  and  $\sqrt{V(X_t)} = \sigma\mu_t$ , ie if the standard deviation of  $X_t$  increases at the mean of  $X_t$  increases. Often the plot of  $t$  vs  $X_t$  will be linear with positive slope and the variability of  $X_t$  increases with  $t$ .

b) Use  $Y_t = \log(X_t)$  if  $E(X_t) \approx e^{a+bt}$  so  $Y_t \approx$  linear.

22) Assume that **all** of the values of the  $X_t > 0$ . A *power transformation* has the form  $Y_t = X_t^\lambda$  for  $\lambda \neq 0$  and  $Y = t_0(Z) = \log(Z)$  for  $\lambda = 0$  where

$$\lambda \in \Lambda = \{0, \pm 1/4, \pm 1/3, \pm 1/2, \pm 1\}.$$

The *modified power transformation* is

$$Y_t = m_\lambda(X_t) = \frac{X_t^\lambda - 1}{\lambda}$$

for  $\lambda \neq 0$  and  $Y_t = m_\lambda(X_t) = \log(X_t)$  for  $\lambda = 0$  where  $\lambda \in \Lambda$ .

23) In decreasing order of importance are  $\lambda = 1$  (no transformation),  $\lambda = 0$  (log transformation),  $\lambda = 1/2$  (square root transformation useful for Poisson count data),  $\lambda = -1$  and  $\lambda = -1/2$ .

24) Could plot  $m_\lambda(X_t)$  for  $\lambda \in \Lambda$  and take  $Y_t = m_{\hat{\lambda}}(X_t)$  where  $\hat{\lambda} \in \lambda$  gives the best plot.

25) Assume  $X_t > 0$  for all  $t$ . Divide  $\{1, \dots, n\}$  into groups  $G_i$  of approximately the same length  $k_i \approx k$ , ie  $k = 12$  or  $k = 20$ . Compute the sample mean  $\bar{X}_i = \frac{1}{k_i} \sum_{t \in G_i} X_t$  and sample range  $R_i = \max(X_t) - \min(X_t), t \in G_i$ . ( $R_i \approx aSD(X_t)$  where  $a \in [2, 3]$ .) Then plot  $\bar{X}_i$  vs  $R_i$ . If the plotted points scatter about a horizontal line, no transformation is needed. If the plotted points scatter about a line with positive slope, use  $Y_t = \log(X_t)$ . If the plotted points scatter about a curve with roughly exponential growth (that is at first slow then fast), use  $Y_t = 1/X_t$ . If the plotted points scatter about logarithmic type curve (fast growth the slow growth), use  $Y_t = \sqrt{\bar{X}_i}$ .

26) Let  $I$  be a time series model. The  $AIC(I)$  statistic is used as an to pick a model from several ARIMA(p,d,q) models. The model  $I_{min}$  with the smallest AIC is always of interest but often overfits: has too many unnecessary parameters. Imagine fitting ARIMA(p,d,q) where  $d = 0, 1$  or  $2$  is fixed and  $p$  and  $q$  run from  $0$  to  $j$  for small  $j$ . The number of parameters in the model for fixed  $d$  is  $p + q + 2$  where  $\sigma = \sqrt{\rho_0} = \sqrt{V(X_t)}$ ,  $\tau$ ,  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  are the parameters.  $AIC(I)$  tends to be large when the model does not have enough terms, to drop as needed terms are added, and then to rise as unnecessary terms are added. If  $\Delta(I) = AIC(I) - AIC(I_{min})$ , then models with  $\Delta(I) \leq 2$  are good, models with  $4 \leq \Delta(I) \leq 7$  are borderline. The initial model to look at is the model  $I_I$  with the smallest number of predictors such that  $\Delta(I_I) \leq 2$ , and also examine submodels  $I$  with fewer predictors than  $I_I$  with  $\Delta(I) \leq 7$ .

27) The `aicmatrix` computes  $AIC(I) - AIC(I_{min})$  for ARIMA(p,d,q) models where  $d$  is fixed, and  $p$  and  $q$  run from 0 to  $j$  for small  $j = pmax$  such as  $pmax = 5$ . Here  $I_{min}$  is the ARIMA( $p_m, d, q_m$ ) model with the smallest AIC(I). This model will have a 0.00 in the `aicmatrix`. Look for model  $I_I$  with  $p_I + q_I \leq p_m + q_m$  as small as possible such that the `aicmatrix` entry  $\leq 2$ . It is possible that  $I_I = I_{min}$ . Also look at models  $I$  with  $p+q \leq p_I+q_I$  with `aicmatrix` entries  $\leq 7$ , especially models with entries  $\leq 4$ .  $AIC(I) - AIC(I_{min})$  is also shown for ARI(p,d) and IMA(d,q) models where  $1 \leq p, q \leq 15$ . Negative entries means the ARI(p,d) or IMA(d,q) model had  $AIC(I) < AIC(I_{min})$  from the `aicmatrix`. Check that the selected model  $I$  does not fail to reject  $H_0$  for  $H_0: \phi_p = 0$  or  $H_0: \theta_q = 0$ . Make the usual checks of plotting the time series, ACF, PACF, response and residual plots, ACF and PACF of the residuals, and `tsdiag(outI)`. From the output below,  $I_{min}$  is the ARIMA(5,1,4) model and  $I_I$  is the ARIMA(3,1,0) model. Interesting models have  $p + q \leq 3$  with entries  $\leq 7$ . So look at the ARIMA(2,1,1), ARIMA(1,1,2), and ARIMA(1,1,1) models after examining the  $I_I = \text{ARIMA}(3,1,0)$  model.

These are rules of thumb: they do not always work but often lead to a good model. If  $I_I$  is the ARIMA(1,0,1) model, might take an AR(3) or MA(3) model even though these have 1 more parameter. If  $I_{min}$  is located at the bottom right corner, the ARIMA(pmax,d,pmax) model, then  $pmax$  may not be large enough. If the `aicmat` function has an error, decrease  $j = pmax$  by 1 until the function works. May also need to decrease  $k$  from 15 for the function to work. Note that the `aicmatrix` entries are  $\Delta(I)$ .

```
aicmat(WWWusage,dd=1,pmax=5,k=15)
```

```
$aics
```

	q						
p	0	1	2	3	4	5	
0	119.86	38.67	8.74	9.13	8.24	7.72	
1	18.10	3.16	5.11	3.44	3.96	5.14	
2	11.04	5.15	6.22	4.63	2.10	6.95	
3	0.85	2.80	4.48	3.27	3.62	5.29	
4	2.79	1.74	5.04	7.94	4.26	6.99	
5	4.72	6.50	2.40	10.50	0.00	1.63	

Find  $I_I$  by looking at models at models on and above the diagonal through (5,4) and (4,5) which have  $p+q \leq 9$ . Interesting models are on or above the diagonal through (3,0), (2,1), (1,2) and (0,3) since they have  $p+q \leq 3$ .

```
$ariaics
```

```
[1] 18.10 11.04 0.85 2.79 4.72 6.31 6.41 6.99 8.06 10.00 11.95 10.46
[13] 8.77 9.97 11.67
```

```
$imaaics
```

```
[1] 38.67 8.74 9.13 8.24 7.72 7.12 2.33 3.61 4.59 6.58 7.96 7.14
[13] 8.77 8.90 9.90
```

28) **Know for final:** Be able to find  $I_{min}$ , the initial model  $I_I$  to look at, and other models to look at using 26) and 27) and output.

29) **Know for Final:** Fix  $d$ . An RR plot is a plot of residuals  $\hat{e}_{t1}$  from model 1 vs residuals  $\hat{e}_{t2}$  from model 2. An FF plot is a plot of fitted values  $\hat{X}_{t1}$  from model 1 vs fitted values  $\hat{X}_{t2}$  from model 2. If the plotted points in both plots look like the identity line, the simpler model may be better.

30) Assuming  $\tau = 0$ , the ARIMA(p,d=1,q) model is  $W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t$  or  $Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \dots + \phi_p(Y_{t-p} - Y_{t-p-1}) - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t$ . So  $Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + \dots + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t$ .

31) The ARIMA(0,0,0) model is  $Y_t = e_t$  where  $\{e_t\}$  is a white noise.

32) The ARIMA(0,d,q) model is the IMA(d,q) model, and the ARIMA(p,d,0) model is the ARI(p,d) model.

33) The ARIMA(0,1,1) model = IMA(1,1) model is  $W_t = -\theta_1 e_{t-1} + e_t$  or  $Y_t = Y_{t-1} - \theta_1 e_{t-1} + e_t$ . Want  $|\theta_1| < 1$  so that  $W_t$  follows a stationary and invertible ARMA(0,1) = MA(1) model without an intercept.

34) The ARIMA(0,2,2) = IMA(2,2) model is  $Z_t = -\theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$  or  $Y_t = 2Y_{t-1} - Y_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$ . Want  $|\theta_2| < 1$  and  $\theta_2 \pm \theta_1 < 1$  so that  $Z_t$  follows a stationary and invertible ARMA(0,2) = MA(2) model without an intercept.

35) The ARIMA(1,1,0) = ARI(1,1) model is  $W_t = \phi_1 W_{t-1} + e_t$  or  $Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t$ . Want  $|\phi_1| < 1$  so that  $W_t$  follows a stationary and invertible ARMA(1,0) = AR(1) model without an intercept.

36) Suppose that  $\tau \neq 0$  in the ARIMA(p,d,q) model with  $d \geq 0$ . Want  $X_t = \nabla^d Y_t$  to follow a stationary and invertible ARMA(p,q) model with intercept:  $X_t = \tau + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t$  or  $(X_t - \mu_X) = \phi_1(X_{t-1} - \mu_X) + \dots + \phi_p(X_{t-p} - \mu_X) - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t$  where  $\tau = \mu_X(1 - \phi_1 - \dots - \phi_p)$  and  $\mu_X = \frac{\tau}{1 - \phi_1 - \dots - \phi_p}$ .

If  $X_t$  is as above, then  $E(X_t) = \mu_X \neq 0$ , and  $Y_t = \mu_t + \tilde{Y}_t$  where  $\mu_t$  is a deterministic polynomial of  $t$  of degree  $d$  and  $\tilde{Y}_t$  follows an ARIMA(p,d,q) model with  $E[\tilde{Y}_t] = 0$ .

37) Let  $\rho_k = \text{corr}(Y_t, Y_{t-k})$ , and let  $r_k = \hat{\rho}_k$  be what is plotted in the ACF. For a stationary and invertible ARMA(p,q) model where  $\{e_t\}$  is a white noise, then for any fixed  $m$ ,

$$\sqrt{n} \begin{pmatrix} r_1 - \rho_1 \\ \vdots \\ r_m - \rho_m \end{pmatrix} \xrightarrow{D} N_m(\mathbf{0}, \mathbf{\Sigma})$$

where  $\mathbf{\Sigma} = (\sigma_{ij})$ . Hence  $r_k$  is a  $\sqrt{n}$  consistent asymptotically normal estimator of  $\rho_k$ , and if  $\rho_k \rightarrow 0$  rapidly as  $k$  increases, so should the  $r_k$  if  $n > \max(100, 4k, 10(p+q))$ .

38) For a white noise,  $V(r_k) \approx 1/n$  and  $\text{corr}(r_k, r_j) \approx 0$  for  $k \neq j$ .

39) For a stationary and invertible ARMA(p,q) model,  $V(r_k) \approx \sigma_{kk}^2/n$  and  $\text{corr}(r_k, r_j) \approx \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}} \approx \text{constant} \neq 0$  for large  $n$ . These correlations can make the sample ACF have ripples and trends that are not present in the theoretical ACF.

40) Similar results hold for the sample PACF, but  $V(\hat{\phi}_{kk}) \approx 1/n$  for  $k > p$  if  $\{Y_t\}$  follows a stationary AR(p) model. Also,  $V(\hat{\phi}_{kk}) \approx 1/n$  for a white noise.

41) Suppose  $X_t = \nabla^d Y_t$  is a stationary and invertible time series,  $X_t \sim \text{ARMA}(p,q)$  if  $Y_t \sim \text{ARIMA}(p,d,q)$ . Then estimate  $\phi_i$  and  $\theta_j$  using  $X_t$  and techniques for ARMA(p,q) models. Usually use an intercept if  $d = 0$ , but do not use an intercept if  $d \geq 1$ .

42) Know that the Yule-Walker equations give the method of moments estimators for AR(p) models (and ARIMA(p,d,0) models).

43) Let  $\mathbf{x} = (X_1, \dots, X_p)^T$ . Let  $\boldsymbol{\rho}_{\mathbf{x}} = (\text{corr}(X_i, X_j))$  be the pop. correlation matrix of  $\mathbf{x}$ . Let  $\mathbf{\Sigma}_{\mathbf{x}} = (\text{Cov}(X_i, X_j)) = E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T]$  be the population covariance

matrix of  $\mathbf{x}$ . Let  $\mathbf{w} = (\mathbf{x}^T, Y)^T$ . Let  $\Sigma_{\mathbf{x}, Y} = Cov(\mathbf{x}, Y) = E[(\mathbf{x} - E(\mathbf{x}))(Y - E(Y))] = (Cov(X_1, Y), \dots, Cov(X_p, Y))^T$ . Let  $\Sigma_Y = V(Y)$  and  $\rho_Y = 1$ . Let  $\rho_{\mathbf{x}, Y} = (corr(X_1, Y), \dots, corr(X_p, Y))^T$ . Let the pop. covariance matrix and correlation matrix of  $\mathbf{w}$  be

$$\Sigma_{\mathbf{w}} = \begin{pmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{x}, Y} \\ \Sigma_{Y, \mathbf{x}} & \Sigma_Y \end{pmatrix} \quad \text{and} \quad \rho_{\mathbf{w}} = \begin{pmatrix} \rho_{\mathbf{x}} & \rho_{\mathbf{x}, Y} \\ \rho_{Y, \mathbf{x}} & \rho_Y \end{pmatrix}.$$

44) The Yule Walker equations can be written in two equivalent forms using  $\rho_k = \gamma_k/\gamma_0$ . Let  $\phi = (\phi_1, \dots, \phi_p)^T$ . Let  $\mathbf{x} = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})^T$  and  $Y = Y_t$ . The Yule Walker equations are

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \dots + \phi_p\rho_{p-1} \quad \text{or} \quad \gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1 + \phi_3\gamma_2 + \dots + \phi_p\gamma_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \dots + \phi_p\rho_{p-2} \quad \text{or} \quad \gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0 + \phi_3\gamma_1 + \dots + \phi_p\gamma_{p-2} \\ &\vdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \dots + \phi_p \quad \text{or} \quad \gamma_p = \phi_1\gamma_{p-1} + \phi_2\gamma_{p-2} + \phi_3\gamma_{p-3} + \dots + \phi_p\gamma_0. \end{aligned}$$

In matrix form

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{or} \quad \rho_{\mathbf{x}, Y} = \rho_{\mathbf{x}}\phi$$

or

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \dots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{or} \quad \Sigma_{\mathbf{x}, Y} = \Sigma_{\mathbf{x}}\phi.$$

Hence  $\phi_{YW} = \Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x}, Y} = \rho_{\mathbf{x}}^{-1}\rho_{\mathbf{x}, Y}$ . Then  $\hat{\phi}_{YW} = \hat{\Sigma}_{\mathbf{x}}^{-1}\hat{\Sigma}_{\mathbf{x}, Y} = \tilde{\rho}_{\mathbf{x}}^{-1}\tilde{\rho}_{\mathbf{x}, Y}$ . Here the estimators are found by replacing  $\rho_k$  by  $r_k = \hat{\rho}_k$  and by replacing  $\gamma_k$  by  $\hat{\gamma}_k$ . Note that  $\rho_{\mathbf{x}}$ ,  $\rho_{\mathbf{x}, Y}$ ,  $\Sigma_{\mathbf{x}}$  and  $\Sigma_{\mathbf{x}, Y}$  have the desired form since  $\mathbf{x} = (Y_{t-1}, \dots, Y_{t-p})^T$  and  $Y = Y_t$ . Hence  $cov(X_i, Y) = cov(Y_{t-i}, Y_t) = \gamma_i$ ,  $cov(X_i, X_j) = cov(Y_{t-i}, Y_{t-j}) = \gamma_{|i-j|}$ ,  $corr(X_i, X_j) = corr(Y_{t-i}, Y_{t-j}) = \rho_{|i-j|}$ , and  $corr(X_i, Y) = corr(Y_{t-i}, Y_t) = \rho_i$ .

45) Suppose equations  $Y_t = (1, \mathbf{x}_t^T)\beta + e_t$  can be put in matrix form  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$  where  $\mathbf{X}$  is of full rank with more rows than columns  $p+1$  and  $\beta = (\phi_0, \phi^T)^T = (\phi_0, \phi_1, \dots, \phi_p)^T$ . Then the least squares estimator  $\hat{\beta}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ ,  $\hat{\phi}_{0, LS} = \bar{Y} - \hat{\beta}_{LS}^T\bar{\mathbf{x}}$ , and  $\hat{\phi}_{LS} = \hat{\Sigma}_{\mathbf{x}}^{-1}\hat{\Sigma}_{\mathbf{x}, Y}$ . The population parameters are  $\phi_0 = E(Y) - \beta^TE(\mathbf{x})$  and  $\phi_{LS} = \Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x}, Y}$ . The stationary AR(p) model can be put in this form, with  $\mathbf{x}_t = (Y_{t-1}, \dots, Y_{t-p})^T$ ,  $Y = Y_t$ , and  $\phi_{LS} = \phi_{YW}$ . Here  $\hat{\Sigma}_{\mathbf{x}}$  and  $\hat{\Sigma}_{\mathbf{x}, Y}$  are the usual estimated covariance matrices used when  $\mathbf{w}_i = (\mathbf{x}_i, Y_i)^T$  are iid from some population. Write the AR(p) equations  $Y_t = \phi_0 + \phi_1Y_{t-1} + \dots + \phi_pY_{t-p} + e_t$  in matrix form  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$  or

$$\begin{bmatrix} Y_{p+1} \\ Y_{p+2} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & Y_p & Y_{p-1} & \dots & Y_1 \\ 1 & Y_{p+1} & Y_p & \dots & Y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & Y_{n-1} & Y_{n-2} & \dots & Y_{n-p} \end{bmatrix} \begin{bmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} e_{p+1} \\ e_{p+2} \\ \vdots \\ e_n \end{bmatrix}.$$

46) For the AR(p) model, assume a normal white noise. Then the maximum likelihood estimator is approximately the least squares estimator.

47) For the AR(p) model, the method of moments (Yule-Walker or Durbin Levinson), least squares, and maximum likelihood estimator of 46) are asymptotically equivalent. Under mild conditions on the white noise with zero mean and variance  $\sigma_e^2$  (normality is not needed),  $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N_p(\mathbf{0}, \sigma_e^2 \Sigma_{\mathbf{x}}^{-1})$ , and  $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N_{p+1}(\mathbf{0}, \sigma_e^2 \mathbf{C})$  where  $\lim_{n \rightarrow \infty} \frac{\mathbf{X}^T \mathbf{X}}{n} \rightarrow \mathbf{C}^{-1}$ . Hence  $\hat{\beta} \approx N_{p+1}(\beta, \hat{\sigma}_e^2 (\mathbf{X}^T \mathbf{X})^{-1})$ . So tests from ordinary multiple linear regression can be applied to AR(p) time series, and  $SE(\hat{\beta}_i) = \sqrt{\hat{\sigma}_e^2 (\mathbf{X}^T \mathbf{X})_{ii}^{-1}}$ .

48) The results for AR(p) models also apply to ARIMA(p,d,0) models for known d where the AR(p) estimator is applied to  $X_t = \nabla^d Y_t$ .

49) For ARIMA(p,d,q) models, assume  $X_t = \nabla^d Y_t$  is a stationary and invertible ARMA(p,q) model. If  $q \geq 1$ , then the method of moments estimator applied to  $X_t$  is inefficient. The least squares and maximum likelihood estimators are asymptotically equivalent and asymptotically normal under mild conditions on the white noise. For d = 0, these estimators try to minimize  $\sum (Y_t - \tau - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q})^2$  which is a nonlinear least squares equation since the  $e_{t-j}$  are unknown. Hence the large sample confidence intervals and tests for  $\phi_i$  and  $\theta_j$  have coverage  $\rightarrow 1 - \alpha = 0.95$  and type I error  $\rightarrow \alpha = 0.05$  if  $\alpha = 0.05$ .

50) Model  $I_1$  is a submodel of model  $I_2$  if  $p_1 \leq p_2$ ,  $q_1 \leq q_2$  and  $d_1 = d_2$ .

51) In the RR and FF plots of 29), suppose that  $I_1$  is a submodel of  $I_2$  or that  $I_2 = I_I$  of 27) where  $I_1$  has no more parameters than  $I_I$ . Assume  $n \geq 10(p_2 + q_2)$ . If both the RR and FF plots resemble the identity line (with correlation  $\geq 0.97$ , say), the two models are about the same, with respect to residuals and fitted values. Then the plots suggest that  $I_1$  is better than  $I_2$  if  $I_1$  is a submodel of  $I_2$ . If  $I_2 = I_I$ , consider using  $I_1$  if  $I_1$  has fewer parameters than  $I_I$  or if  $I_1$  has nearly the same number of parameters but is simpler or makes more sense to subject matter experts (eg  $I_1$  is an AR(3) model and  $I_I$  is an ARMA(1,1) model, and the PACF suggests an AR(3) model). Of course you need to check that the chosen model is good.

52) If in 51) there is not high correlation in both plots, may need to use model  $I_2$ , assuming  $n \geq 10(p_2 + q_2)$ . However, high correlation in the FF plot ( $\geq 0.97$ , say), suggests that predictions from the 2 models are about the same.

53) **Getting a good ARIMA(p,d,q) model:** Assume that there is a good ARIMA(p,d,q) model and that  $n$  is large enough to find the model. So  $n \geq 10(p_I + q_I)$ .

i) Plot the time series. If it is clearly nonstationary, eg linear with increasing variability, take log transformations if necessary, first differences  $W_t$ , and plot  $W_t$ .

ii) If the time series plot is not clearly nonstationary, make the ACF and PACF. If the ACF and PACF suggest nonstationarity, take first differences  $W_t$  and repeat i) and ii). Sometimes need second differences  $Z_t$ . Sometimes need to compare models for  $Y_t$  and  $W_t$  or models for  $W_t$  and  $Z_t$ .

iii) Get the AICs, find  $I_I$  and other models, if any, to look at.

iv) For model  $I_I$ , check that you reject  $H_0 : \phi_p = 0$  and  $H_0 : \theta_q = 0$ . Want the model to be stationary and invertible. Watch out for  $|\phi_p|$  and  $|\theta_q|$  near 1. Want good response and residual plots. Want the ACF and PACF of the residuals to resemble those of a

white noise. Want the pvalues of the Ljung-Box statistic to all be above the horizontal line at 0.05. Want the correlation matrix of the predictors to have absolute off diagonal entries  $< 0.9$ .

v) Fit the ARIMA( $p_I, d, q_I$ ) model, the ARIMA( $p_I, d, q_I+1$ ) model and the ARIMA( $p_I+1, d, q_I$ ) model. Want  $\hat{\phi}_1, \dots, \hat{\phi}_{p_I}, \hat{\theta}_1, \dots, \hat{\theta}_{q_I}$  to be similar for all three models. Want to fail to reject  $H_0 : \phi_{p_I+1} = 0$  and  $H_0 : \theta_{q_I+1} = 0$ .

vi) To check model  $I$  vs  $I_I$ , want  $p + q \leq p_I + q_I$  and  $\Delta(I) \leq 7$  with preferably  $\Delta(I) \leq 4$ . Want iv) and v) to hold with  $p$  and  $q$  replacing  $p_I$  and  $q_I$ . Want good RR and FF plots as in 51).

vii) Make the usual checks on the selected model  $I$ : want  $\Delta(I) \leq 4$  with no more parameters than  $I_{min}$ , want good response and residual plots, want the model to be reasonable given the ACF and PACF, want the ACF and PACF of the residuals to resemble those of a white noise. Want all of the Ljung-Box pvalues to lie above the horizontal line at 0.05. The fitted model should be stationary and invertible. Want zero or very few coefficients where you fail to reject  $H_0 : \psi = 0$  where  $\psi$  is a  $\phi_i$  or  $\theta_j$ . The RR and FF plots using  $I$  and  $I_I$  should look good.

54) When looking at AICs, you can often eliminate model  $I_2$  if there is a model  $I_1$  with fewer parameters and a smaller  $\Delta(I)$ . In 27), the ARIMA(1,1,1) model  $I_1$  with  $\Delta(I_1) = 3.16$  is more interesting than the ARIMA(2,1,1) model  $I_2$  with  $\Delta(I_2) = 5.15$ , and more interesting than the ARIMA(1,1,2) model  $I_3$  with  $\Delta(I_3) = 5.11$ .

55) Want the ACF and PACF of the residuals to resemble those of a white noise, but residuals are not a white noise. Benchmarks for concern are  $|Z_k| > 1.25$  for  $k = 1, 2, 3$ , and  $|Z_k| > 1.6$  for  $k \geq 4$ . The  $\pm 2/\sqrt{n}$  lines correspond to  $|Z_k| = 2$ . These benchmarks result in more than 5% errors when  $\{\hat{e}_t\}$  is a good approximation to a white noise.

56) The QQ plot of the residuals is used to check normality of residuals. Normality is reasonable if a straight line fits the plotted points well. Small deviations at the end of the line are ok. Flattening at both ends suggests nonnormality due to too many cases in both tails. Cases above the “best fitting” line at the upper right of the plot suggest positive skew: there are too many cases in the upper right tail for normality.

57) The Ljung-Box test tests whether there is residual autocorrelation. Want all of the plotted points above the pvalue = 0.05 horizontal line. Want no autocorrelation since  $\hat{e}_t \approx e_t$  and  $\{e_t\}$  is a white noise for a good model.

58) Suppose an ARIMA( $p, d, q$ ) model appears to give a good fit. Then fit i) an ARIMA( $p+1, d, q$ ) model and ii) an ARIMA( $p, d, q+1$ ) model. If  $n > 10(p+q)$ , want  $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$  to be similar in all 3 models. For model i), want to fail to reject  $H_0 : \phi_{p+1} = 0$  and for model ii) want to fail to reject  $H_0 : \theta_{q+1} = 0$ .

59) If the true model is an ARIMA( $p, d, q$ ) model, you can fit an ARIMA( $p+i, d, q$ ) model or an ARIMA( $p, d, q+j$ ) model with little change in the true parameter estimates. For an ARIMA( $p+i, d, q+j$ ) model with  $i, j \geq 1$ , the parameters in the overfitted model are not unique, and the estimated parameters can change greatly. (The RR and FF plots should still look good if  $n > 10(p+q+i+j)$ .)

60) For forecasting, predict  $Y_{t+1}, \dots, Y_{t+L}$  given the past  $Y_1, \dots, Y_t$  (and  $\hat{e}_j, \dots, \hat{e}_t$  where often  $j = 1$  or  $j = \max(p, q) + 1$ ).

61) A large sample  $100(1-\alpha)\%$  prediction interval (PI) for  $Y_{t+j}$  is  $(L_n, U_n)$  where the coverage  $P(L_n \leq Y_{t+j} \leq U_n) = 1 - \delta_n \rightarrow 1 - \alpha$  as  $n \rightarrow \infty$ .