

## Review materials for Exam II

Things you should be familiar with:

- Wold's Theorem relating stationarity to linearity; invertibility; how we manipulate polynomials in the backshift operator in such a way as to represent a series as linear and/or invertible.
- Calculations of ACFs (pretty simple for an MA; derive and solve the Yule-Walker equations for an AR). What is the important property of the ACF of an MA?
- Calculation of PACFs - define the PACF, then do the required minimization. What is the important property of the PACF of an AR?

- Forecasting - properties of conditional expectation; applying the Double Expectation Theorem;  $E[Y|X]$  as the minimum MSE forecast of  $Y$  using information in  $X$ , etc.
- Maximum likelihood estimation: The end result of our theoretical discussion was that the MLEs of the AR and MA parameters  $\phi$  and  $\theta$  are the minimizers of the sum of squares of the residuals:

$$S(\phi, \theta) = \sum w_t^2(\phi, \theta).$$

For a pure AR this is accomplished by a linear regression of the series on its own past. If there are MA terms then an iterative procedure (Gauss-Newton) is necessary. As an example, you should be able to show that, under the usual assumption (what is it?), for an MA(1) model you get

$$w_t(\theta) = \sum_{s=0}^{t-1} (-\theta)^s X_{t-s},$$

so that  $S(\theta)$  is a polynomial in  $\theta$ , of degree  $2(n-1)$ , to be minimized.

How would the iterations be carried out? Your application of Gauss-Newton should yield the iterative procedure

$$\theta_{k+1} = \theta_k - \frac{\sum_t w_t(\theta_k) \dot{w}_t(\theta_k)}{\sum_t \dot{w}_t^2(\theta_k)}.$$

- The usual estimate of the noise variance, obtained by modifying the MLE, is (what?).
- What is the “Information Matrix”, and what role does it play in making inferences about the parameters in an ARMA model?

---

You might wish to work through the following analysis on your own; it uses many of the techniques with which you are expected to be familiar.

Recall SOI and Recruits series. We earlier attempted to predict Recruits from SOI by regressing Recruits on time and lagged SOI ( $m = 3, \dots, 12$ ) - 12 parameters, including the intercept (Figure 1).

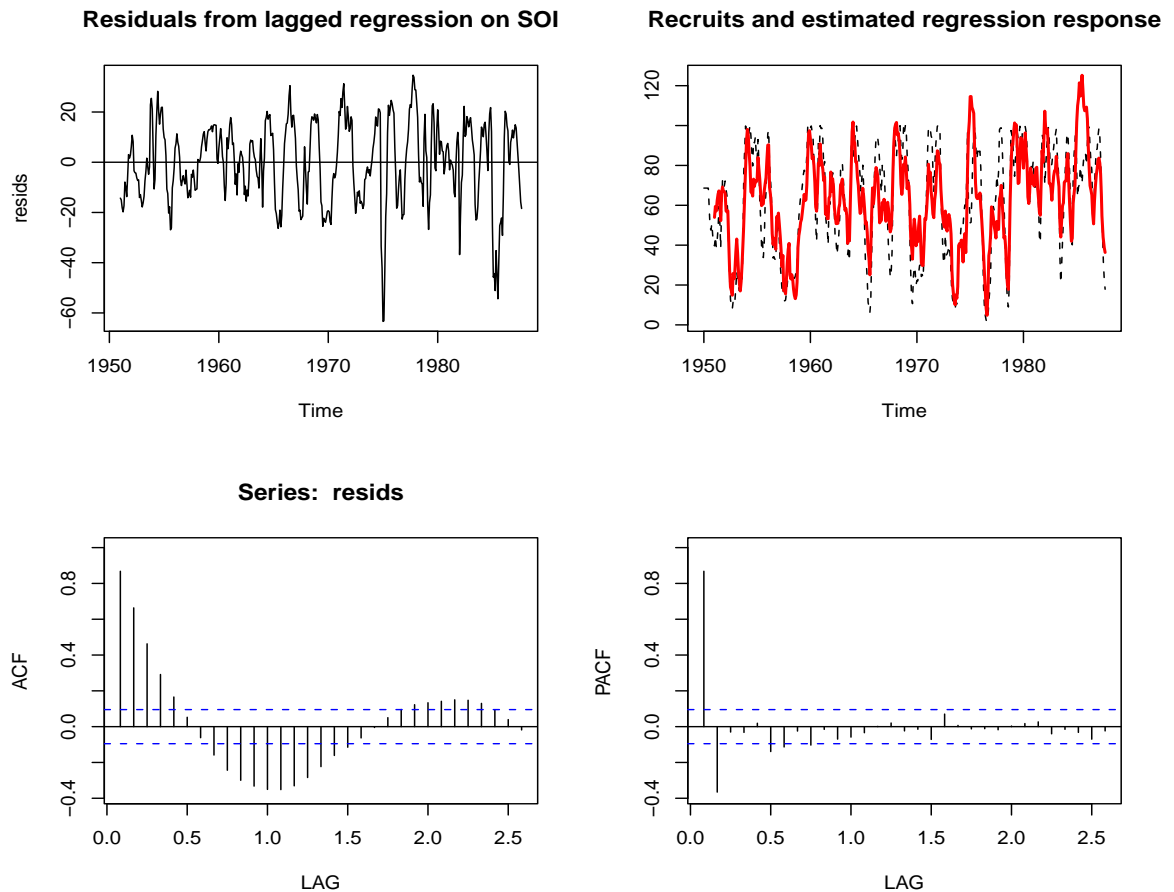


Figure 1. Recruits predicted by SOI;  $R^2 = .67$ .

Let's look for a better, and perhaps more parsimonious, model. Let  $X_t$  and  $Y_t$  denote the series  $\nabla_{12}SOI_t$  and  $\nabla_{12}Recruits_t$ . Then predicting  $Y_t$  from  $X_t$  amounts to predicting annual changes in one series from annual changes in the other.

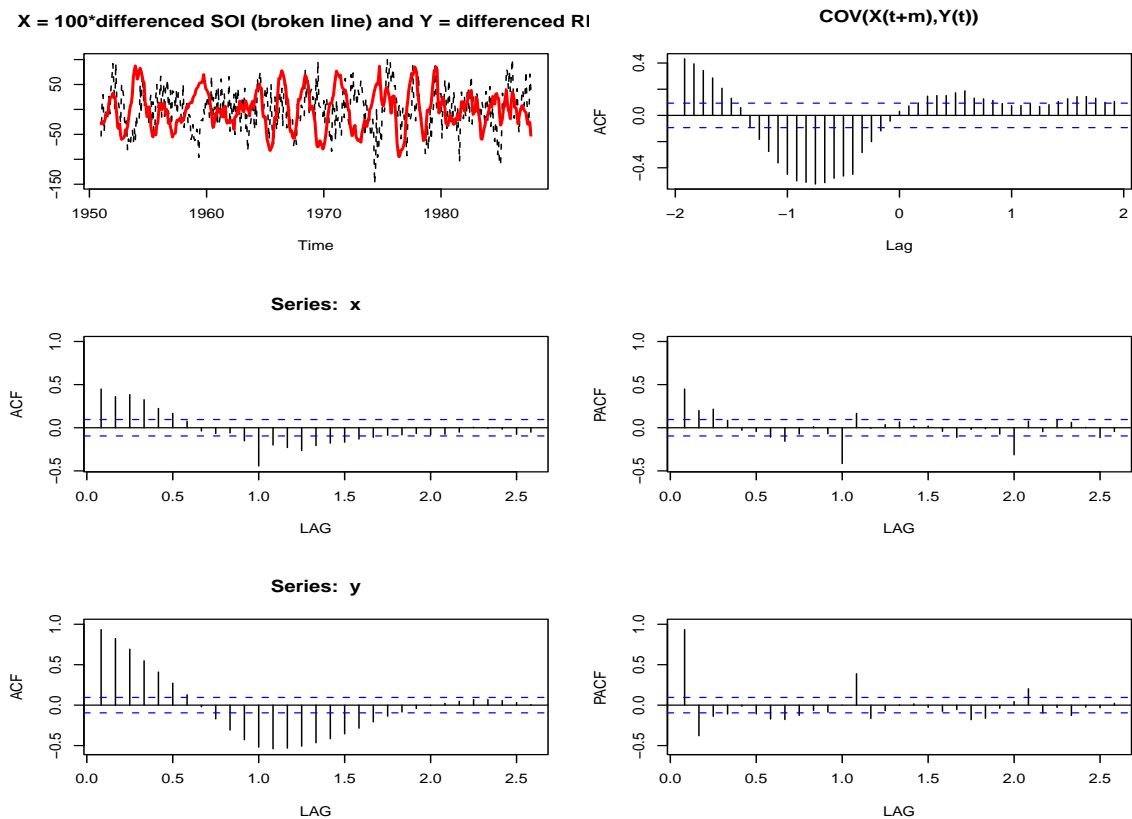


Figure 2.  $X_t = \nabla_{12}SOI_t$  and  $Y_t = \nabla_{12}Recruits_t$ .  
ACF, PACF and CCF plots.

Both series look stationary. (Why? What do we look for in the ACF/PACF plots in Figure 2?). Consider the CCF. The most extreme values of

$$COV[X_{t+m}, Y_t] = COV[X_t, Y_{t-m}]$$

with  $m \leq 0$  are at  $m = -8, -9, -10$ , indicating a linear relationship between  $X_t$  and  $Y_{t+8}, Y_{t+9}, Y_{t+10}$ . Thus we might regress  $Y_{t+10}$  on  $Y_{t+9}, Y_{t+8}$  and  $X_t$ ; equivalently  $Y_t$  on  $Y_{t-1}, Y_{t-2}, X_{t-10}$ :

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 X_{t-10} + Z_t, \quad (t = 11, \dots, 453)$$

for a series  $\{Z_t\}$  following a time series model (perhaps even white noise, if we're lucky).

```
y1 = lag(y, -1)
y2 = lag(y, -2)
x10 = lag(x, -10)
fit2 = dynlm(y ~ y1 + y2 + x10 - 1) #No intercept
z = fit2$resid
fit2$coef
```

y1	y2	x10
1.3086448	-0.4198637	-4.1031903

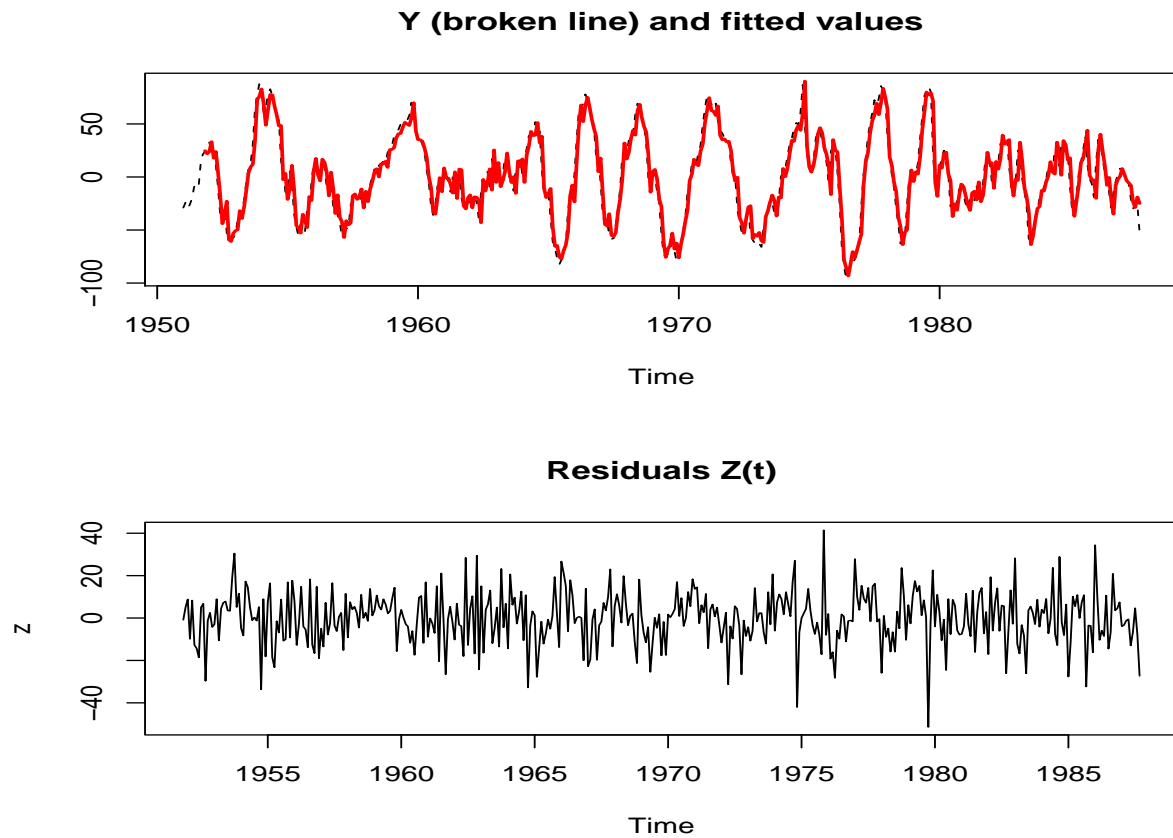


Figure 3. Top: Differenced recruits  $Y_t$  and fitted values  $\hat{Y}_t = 1.31Y_{t-1} - .42Y_{t-2} - 4.10X_{t-10}$ . Bottom: Residuals  $Z_t$  from this fit.  $R^2 = .90$ .

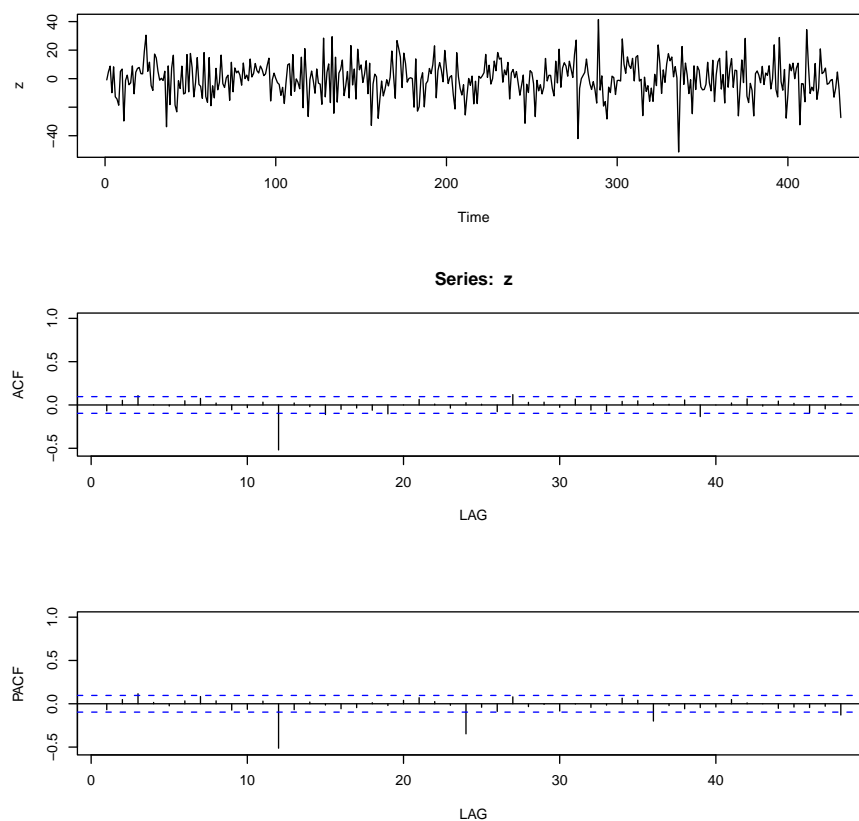


Figure 4.  $\{Z_t\}$ , ACF and PACF.

The residual series  $\{Z_t\}$  still exhibits some trends - see Figure 4. Seasonal  $\text{ARMA}(2, 1)_{12}$ ? Fitting a selection of models resulted in the 'best' model, as chosen by AIC (what does this mean? how is it chosen?)  $\text{ARMA}(0, 1) \times (1, 1)_{12}$ . But the values for the simpler model  $\text{ARMA}(0, 1) \times (0, 1)_{12}$  are almost identical, so I prefer this. This model gives satisfactory



results - see Figure 5 - but the normality is suspect ( $p \approx 0$  in Shapiro-Wilk test - what does this mean?).

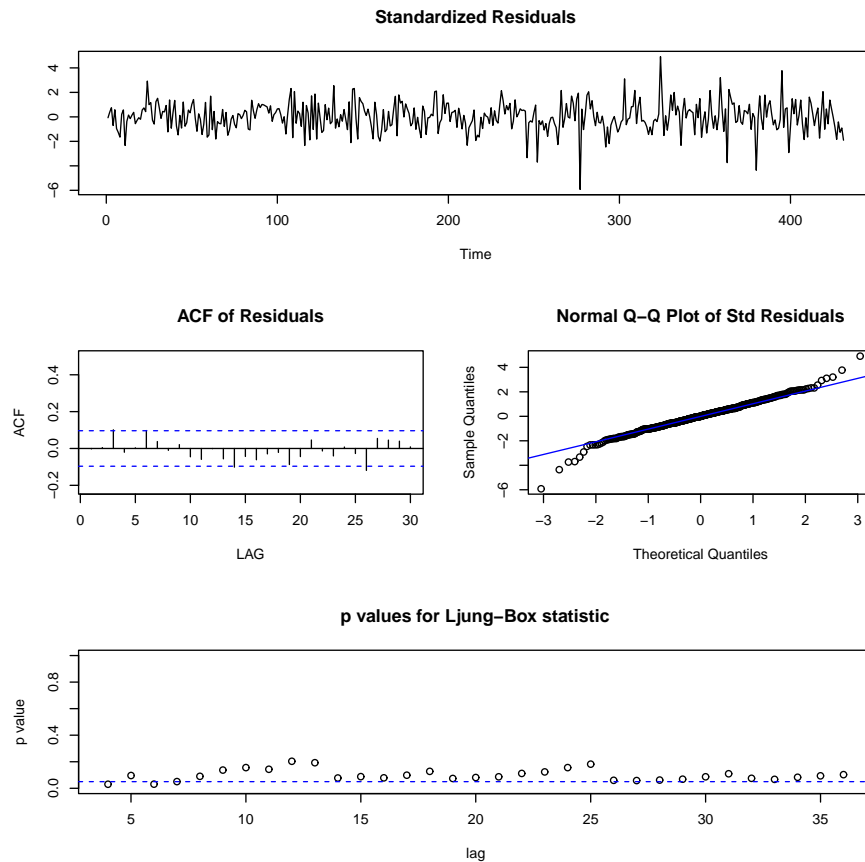


Figure 5. Residuals from  $\text{ARMA}(0, 1) \times (0, 1)_{12}$  fit to  $\{Z_t\}$ .

Coefficients:

	ma1	sma1	xmean
	-0.1125	-1.0000	0.0259
s.e.	0.0465	0.0397	0.0343

sigma^2 estimated as 74.8

The interpretation of this output is that

$$\dot{Z}_t = (1 + \theta B) (1 + \Theta B^{12}) w_t,$$

with  $\dot{Z}_t = Z_t - \beta_0$ , where  $\hat{\beta}_0 = .0259$ ,  $\hat{\theta} = -.1125$ ,  $\hat{\Theta} = -1$ . But  $\hat{\beta}_0$  is less than one standard error away from 0, so we will assume that  $\beta_0 = 0$ . Thus the estimated model is

$$Z_t = w_t + \hat{\theta}w_{t-1} - w_{t-12} - \hat{\theta}w_{t-13}.$$

**Predictions:** You should be sufficiently familiar with the prediction theory covered in class that you can follow the following development, although I wouldn't expect you to come up with something like this on your own in an exam. (The point of knowing the theory is so that you can still do something sensible when the methods you know don't quite apply!)

The best forecast of  $Y_{t+l}$  is  $\hat{Y}_{t+l}^t = E[Y_{t+l}|Y^t]$  (estimated), in the case of a single series  $\{Y_t\}$ . (Why? - you should be able to apply the Double Expectation Theorem, so as to derive this result. And what does “best” mean, in this context?)

If we think of the data now as consisting of the union of the series  $\{X_t, Y_t\}$  then it follows that the predictions are

$$Y_{t+l}^t = E[Y_{t+l}|X^t, Y^t].$$

For  $l = 1, \dots, 10$ , we then get (using  $X_{t+l-10}^t = X_{t+l-10}$  - why?):

$$Y_{t+l}^t = \beta_1 Y_{t+l-1}^t + \beta_2 Y_{t+l-2}^t + \beta_3 X_{t+l-10} + Z_{t+l}^t.$$

These can be calculated recursively (how?) once the  $Z_{t+l}^t$  are obtained. These in turn can be gotten numerically - on R, for instance, using

`sarima.for(z, n.ahead = 12, 0,0,1, 0,0,1, 12).`

- You should be able to write down what would be computed to get  $Z_{t+l}^t$ , using the residual series  $\{Z_t\}$  as the data.

- To get prediction intervals we need the variances of the predictions; here more work is required. Let

$$\beta(B) = 1 - \beta_1 B - \beta_2 B^2$$

be the polynomial involved in the original regression, so that

$$\begin{aligned}\beta(B)Y_t &= \beta_3 X_{t-10} + Z_t, \\ Z_t &= (1 + \theta B) (1 + \Theta B^{12}) w_t \\ &= \alpha(B) w_t, \text{ say.}\end{aligned}$$

- The series  $\{Z_t\}$  is *linear*, hence *stationary*.
- $\{Z_t\}$  is not *invertible*. (Why not? What does invertibility mean? Representation?)
- The polynomial  $\beta(B)$  has zeroes  $B = 1.34, 1.77$ :

```

> a = c(1, -1.3086448, 0.4198637)
> zeroes = polyroot(a)
> zeroes
[1] 1.341769-0i 1.775063+0i
> abs(zeroes)
[1] 1.341769 1.775063

```

Thus  $1/\beta(B)$  has a series representation with square summable coefficients. Define  $\gamma(B) = 1/\beta(B)$  and  $\delta(B) = \gamma(B)\alpha(B)$ , both expanded as series. Then our two models

$$\begin{aligned}\beta(B)Y_t &= \beta_3 X_{t-10} + Z_t, \\ Z_t &= \alpha(B)w_t,\end{aligned}$$

can be combined as

$$\begin{aligned}Y_t &= \beta_3 \gamma(B) X_{t-10} + \gamma(B) Z_t \\ &= \beta_3 \gamma(B) X_{t-10} + \gamma(B) \alpha(B) w_t \\ &= \beta_3 \gamma(B) X_{t-10} + \delta(B) w_t.\end{aligned}$$

For  $l = 1, \dots, 10$  we have

$$Y_{t+l} = \beta_3 \gamma(B) X_{t+l-10} + \sum_{k=0}^{\infty} \delta_k w_{t+l-k},$$

$$Y_{t+l}^t = \beta_3 \gamma(B) X_{t+l-10} + \sum_{k \geq l} \delta_k w_{t+l-k},$$

and hence

$$Y_{t+l} - Y_{t+l}^t = \sum_{k=0}^{l-1} \delta_k w_{t+l-k},$$

with  $VAR[Y_{t+l} - Y_{t+l}^t] = \sigma_w^2 \sum_{k=0}^{l-1} \delta_k^2$ . This leads to prediction intervals

$$\hat{Y}_{t+l}^t \pm z_{\alpha/2} \hat{\sigma}_w \sqrt{\sum_{k=0}^{l-1} \hat{\delta}_k^2}.$$

To complete this we must express the  $\delta_k$  in terms of  $\beta_1, \beta_2, \phi, \Theta$  and then replace these with estimates. This starts with the representation

$$\delta(B) = \frac{\alpha(B)}{\beta(B)} = \frac{(1 + \theta B)(1 + \Theta B^{12})}{(1 - \beta_1 B - \beta_2 B^2)}.$$

(Then what?)

There are only 4 parameters now, rather than 12 in our original attempt at modelling Recruits in terms of SOI.

- R code for an analysis of the varve series is also available.