STAT421: Final Exam

May 1, 1999

+ THIS IS A PRACTICE VERSION OF THE REAL FINAL. It's coverage in terms of content is very similar to the real final.

+ You are allowed **3 hours** for the real exam.

+ The real exam is closed book and closed notes.

+ There are 6 questions. The value of each question is shown. The parts (a), (b), etc. count equally.

Some abbreviations, notations:

> ACF = autocorrelation function.

> PACF = partial autocorrelation function.

> Wherever possible we will use the lower case variables in a density to indicate which random variables it is the density for. So for instance $f(x|y)$ is the conditional density of the random variable X given $Y = y$, and $f(\tilde{z}_t, \tilde{z}_{t+1}, \tilde{z}_{t+2}|a_t, a_{t-1})$ is the joint conditional density of $(\tilde{Z}_t, \tilde{Z}_{t+1}, \tilde{Z}_{t+2}$ given a_t, a_{t-1}.

1. 50 points Consider an MA(2) process Z_t defined as follows:

$$\tilde{Z}_t = a_t + (7/6)a_{t-1} + (1/3)a_{t-2},$$

where

$$\tilde{Z}_t = Z_t - 10,$$

and a_t is an i.i.d. $N(0, 25)$ sequence.
(a) Verify that Z_t is invertible.
(b) Compute $\mu = E[Z_t]$, $\sigma_Z^2 = \text{Var}[Z_t]$, and the autocorrelation function ρ_h, $h = 0, 1, \ldots$.
(c) Suppose you are given that $a_{-1} = 6$ and $a_0 = -9$. Compute forecast values for Z_1 and Z_2.
(d) Compute the forecast variance $\text{Var}[Z_1|a_0, a_{-1}]$ and $\text{Var}[Z_2|a_0, a_{-1}]$ for your results in part (c).
(e) What are the forecasts and forecast variances for Z_t, $t > 2$ given a_0 and a_{-1}?

2. 30 points A statistician buys a new car and records his gas mileage on the first 118 fillups (REAL DATA!). Figure 1 on the next page shows a time series plot, sample ACF, and sample PACF for the original series. Figure 2 shows the analogous plots for the series when the first 40 observations are deleted (only the last 78 observations are kept). The objective is to predict mileage for fillups 119, 120 , ...
(a) For each series (full series, last 78 observations), describe a tentative ARMA model identification based on the information provided. Justify your answers.
(b) Would you recommend using the full series or just the last 78 observations for model fitting and then forecasting? Justify your answer.

(3) 36 points (Forecasts for MA processes.) Consider an MA(1) process

$$Z_t = a_t - \theta a_{t-1}$$

where the a_t's are i.i.d. $N(0, \sigma_a^2)$.
(a) Show that $E[Z_1|a_0]$ is $-\theta a_0$.
(b) Show that $E[Z_2|Z_1, a_0]$ is $-\theta[Z_1 - \theta a_0]$.
(c) What is $E[Z_3|Z_2, Z_1, a_0)$?
(d) Show how one can recursively compute $E[Z_t|Z_{t-1}, Z_{t-2}, \ldots, Z_1, a_0)$.
(e) Of course, in practice we won't observe a_0. Show that $E[Z_t|Z_1, \ldots, Z_{t-1}]$ is given by

$$E[Z_t|Z_1, \ldots, Z_{t-1}] = \int E\left[Z_1, \ldots Z_{t-1}, a_0\right] \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left[-a_0^2/(2\sigma_a^2)\right] da_0.$$

(f) Describe an algorithm for computing the forecasts for Z_{n+1} ,Z_{n+2}, ..., $Z_{n+\ell}$ given data $(z_1, \ldots, z_n)$ modeled by the MA(1) process above.

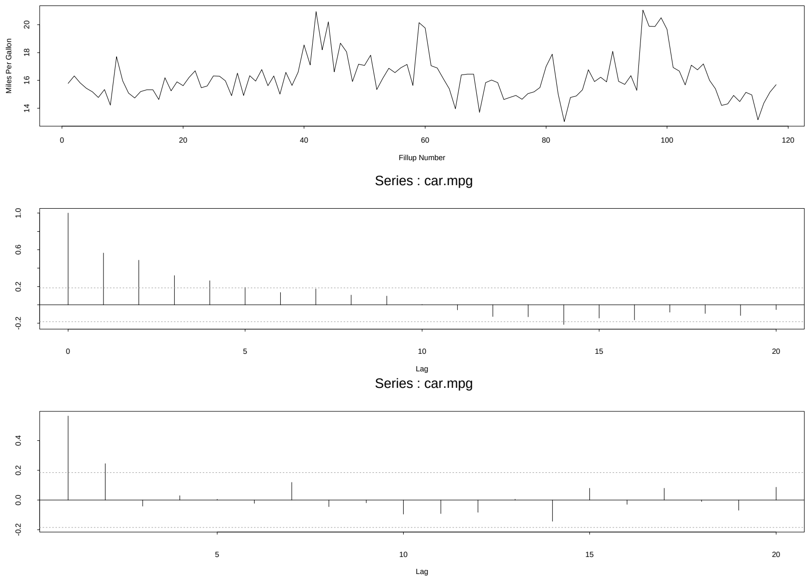

Figure 1: Time Series Plot, sample ACF, and sample PACF for the full series in Problem 2.

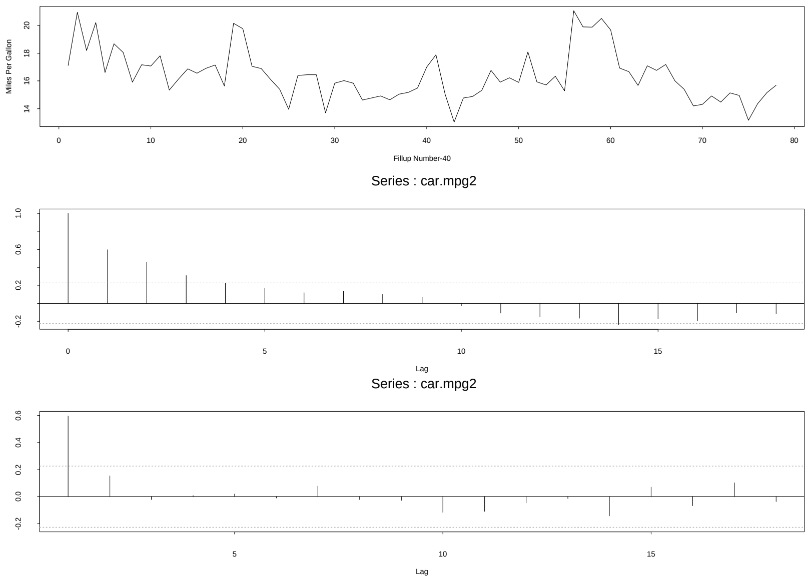

Figure 2: Time Series Plot, sample ACF, and sample PACF for the last 78 observations of the series in Problem 2.

(4) 20 points Consider an MA(4) process of the form

$$Z_t \;=\; a_t \,+\, \theta_4 a_{t-4}.$$

Write this as an AR(∞) process of the form

$$Z_t \;=\; \sum_{k=1}^{\infty} \pi_k Z_{t-k} \;+\; a_t.$$

What is the range of θ_4 values for which the process is invertible?

(5) 20 points To the same data set, one statistician fits an MA(2) model of the form

$$Z_t \;=\; a_t - 1.0 a_{t-1} + .09 a_{t-2},$$

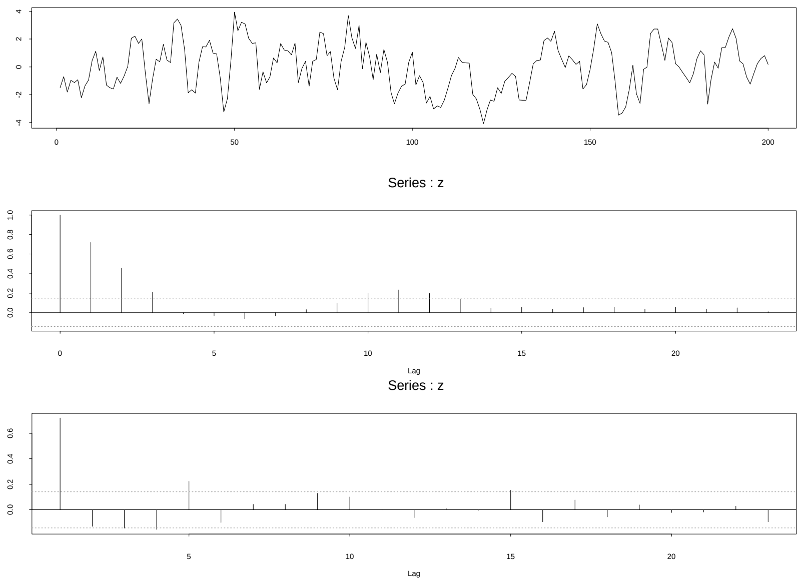

Figure 3: Time Series plot, sample ACF, and sample PACF in Problem 6.

while another statistician fits an ARMA(1,1) model of the form

$$Z_t \;=\; -.1Z_{t-1} + a_t - .9a_{t-1}.$$

The diagnostics for the fit models do not show any signficant lack of fit for either, and the two statisticians spend hours arguing that their own models are the "right ones."

Show in fact that the two models are almost equal.

(6) 44 points A statician analyzes a series z_t with 200 observations in Splus. Below is the Splus code:

```
> par(mfrow=c(3,1))
> tsplot(z)
> acfz_acf(z)
> pacfz_acf(z,type="partial")
> par(mfrow=c(1,1))
> fit.ar1_arima.mle(z,model=list(ar=0))
> diag.ar1_arima.diag(fit.ar1)
> fit.ar4_arima.mle(z,model=list(ar=c(0,0,0,0)))
> diag.ar4_arima.diag(fit.ar4)
> fit.ma4_arima.mle(z,model=list(ma=c(0,0,0,0)))
> diag.ma4_arima.diag(fit.ma4)
```

Figure 3 shows the time series plot, sample ACF, and sample PACF. Figure 4 shows the output from the diagnostics of the AR(1) fit. Figure 5 shows the output from diagnostics fo the AR(4) fit. Figure 6 shows the output from the diagnostics of the MA(4) fit.

Based on this information, which of the three fits do you believe is the best? Do you believe any of the three fits is adequate? Justify your answers.

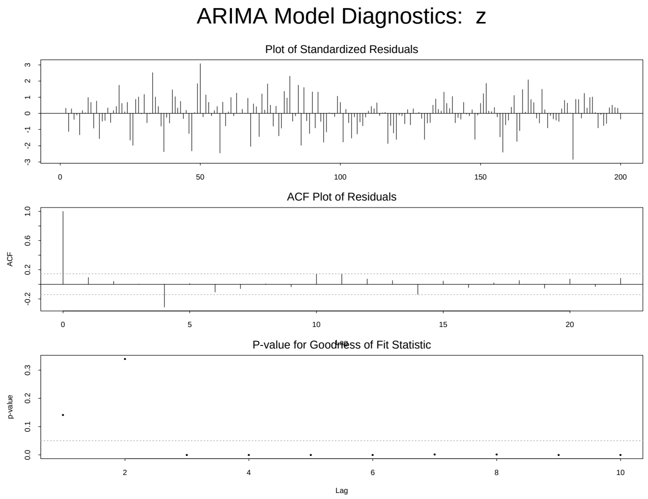

Figure 4: Diagnostic plots for AR(1) model fit to series in Problem 6.

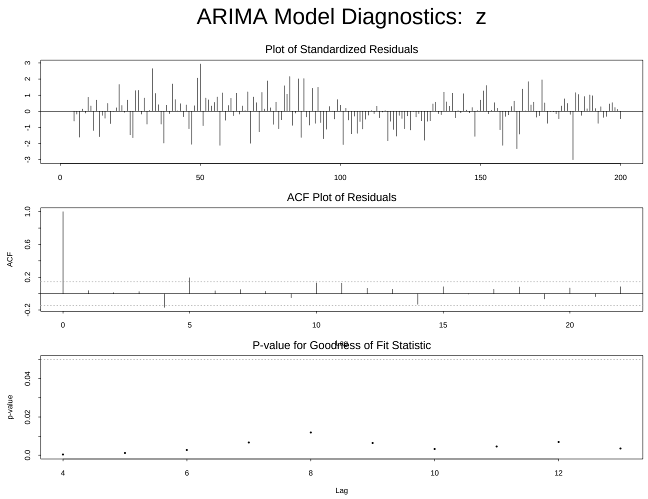

Figure 5: Diagnostic plots for AR(4) model fit to series in Problem 6.

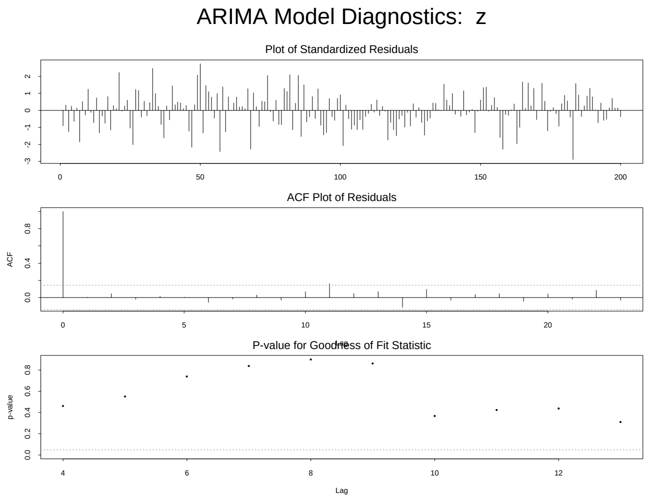

Figure 6: Diagnostic plots for MA(4) model fit to series in Problem 6.