Remark: Problem 1 is the most important problem on this assignment (it will prepare you for your project). Problem 2 was taken largely from last year's final exam. Problem 3 consists of a bunch of rambling on my part; I got tired of rambling, so I just turned it into a problem.

1. In Homework 4, you were asked to identify a small set of candidate ARIMA(p, d, q) models (perhaps some of you even made a "final choice") for each of the following data sets:

- `ibm`: daily closing IBM stock prices (dates not given)
- `internet`: number of users logged on to an Internet server each minute (dates/times not given)
- `gasprices`: average price (US dollars per gallon) for regular gasoline in the United States; there are $n = 145$ weekly observations collected from 1/5/2009 to 10/10/2011 (**Source:** Rajon Coles, Fall 2011).

Remembering your candidate models for each data set, fit and diagnose your model selections. That is, use the methods from Chapter 7 to fit your chosen models (for uniformity, you could just use maximum likelihood for each model fit). Then, diagnose your fitted model(s) by doing a thorough analysis of the residuals and implementing the overfitting technique (Chapter 8).

For each data set and for each model you entertained in Homework 4, what do you think now? Would you like to suggest another model for further investigation? Or, are you satisfied with your Homework 4 model selections?

- If your original model choices in Homework 4 are "reasonable," convince me that they are.
- If your original model choices are deemed "not reasonable," use the information from your diagnoses to specify another model. Then, evaluate the merit of this new model using the methods from Chapter 7 and Chapter 8.

Your goal is to come up with **one final model** for each data set−the "best" one. Convince me that your final model does a good job at explaining the variability in the data, but also adhere to the Principle of Parsimony. There are no "right" answers here, but there are certainly bad answers (stay away from these).

Important: Remember your "final" model for each data set!! On Homework 6 (which I will count as extra credit), you will be forecasting future values!

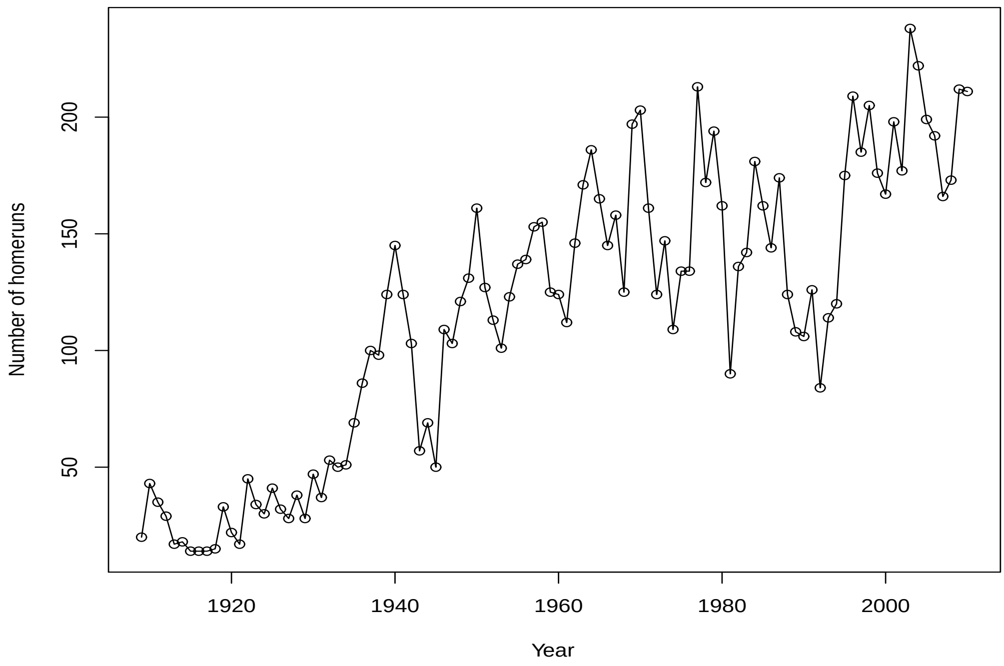
2. In class, we looked at the number of homeruns hit by the Boston Red Sox each year during 1909-2010. Denote this process by $\{Y_t\}$. I have displayed the data in Figure 1. Note that these data are available on the course web site (`homeruns`).

I used R to fit the model

$$(1 - B)Z_t = e_t - \theta e_{t-1},$$

where $Z_t = \sqrt{Y_t}$. Here is the output:

```
> arima(sqrt(homeruns),order=c(0,1,1),method='ML') # maximum likelihood
```

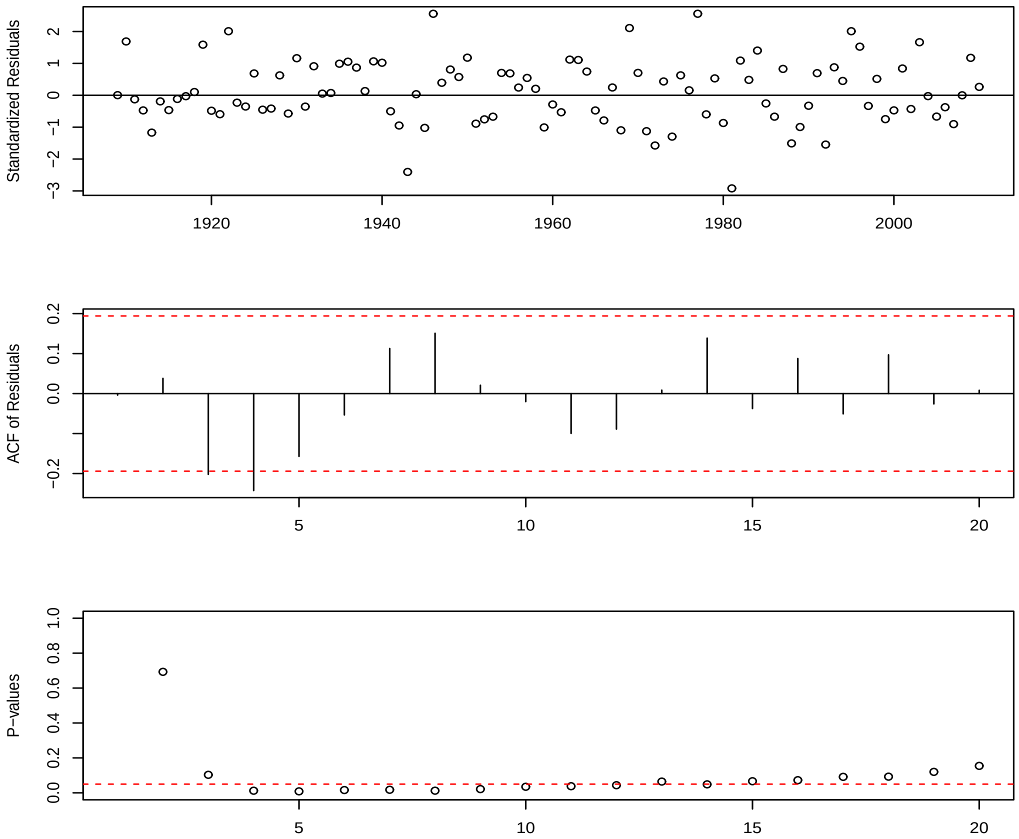



```
Coefficients:
          ma1
      -0.2488
s.e.   0.1072
sigma^2 estimated as 1.437:  log likelihood = -161.64, aic = 325.28
```

(a) Why do you think I used the square-root transformation? Why do you think I used a nonstationary model?
(b) Based on the model that I fit (judged to be a "reasonable" candidate model during the model specification phase), what do you think the sample ACF of $\{Z_t\}$ looked like? the sample ACF of $\{\nabla Z_t\}$? Try to answer these questions without looking at the ACFs in R.
(c) Write an approximate 95 percent confidence interval for θ based on the model fit. Interpret the interval.
(d) I have displayed below the `tsdiag` output from fitting the model above to the Boston Red Sox homerun data. I have also performed the Shapiro-Wilk and runs tests for the standardized residuals; see the R output below:

```
> shapiro.test(rstandard(homerun.fit)
W = 0.9884, p-value = 0.5256

> runs(rstandard(homerun.fit))
$pvalue
[1] 0.378
```



Based on the information available, what do you think of the adequacy of the model that I fit to these data? Can you suggest a better model?

3. I have put the following paper on the course web site:

- Fisher, T. and Gallagher, C. (2012). New weighted Portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association* **107**, 777-787.

As I mentioned in class (30 Oct 2013), Fisher and Gallagher proposed new goodness-of-fit (GOF) statistics for ARIMA models (and also for nonlinear models, too). Read Sections 1 and 2 of this paper. Don't worry if you cannot understand all of the technical details; this journal is one of the most prestigious journals in statistics and some articles can be quite mathematical.

In Section 8.2 of the notes, we presented the Ljung-Box statistic

$$Q_* = n(n+2)\sum_{k=1}^{K} \frac{\widehat{r}_k^2}{n-k}$$

which can be used to test

$$H_0 : \text{the ARMA}(p, q) \text{ model is appropriate}$$
versus
$$H_1 : \text{the ARMA}(p, q) \text{ model is not appropriate}$$

(perhaps after transformation, differencing, or both). When H_0 is true, that is, the null model is correct, Q_* follows (approximately) a χ^2 distribution with degrees of freedom equal to $K-p-q$. Note that on page 777 of the paper by Fisher and Gallagher, they denote the Ljung-Box statistic by $\widetilde{Q}$. Also note that we use K to denote the "maximum lag." Fisher and Gallagher use m.

The authors' new ARIMA GOF statistics are in Section 2 of their paper. In particular, their new statistics are $\widetilde{Q}_W$ and $\widetilde{M}_W$; see Equations (5) and (6), respectively. The $\widetilde{Q}_W$ statistic is very similar to the Ljung-Box statistic. The $\widetilde{M}_W$ statistic depends on the sample **partial autocorrelations** of the residuals (note that Fisher and Gallagher use $\widehat{\pi}_k$ notation to denote the sample partial autocorrelations). As the authors show in their Theorem 1 (on page 779), their new GOF statistics $\widetilde{Q}_W$ and $\widetilde{M}_W$ follow a different large-sample (asymptotic) distribution (under H_0) than the one we stated for the Ljung-Box test. The asymptotic distribution identified by the authors for their statistics is a χ^2 **mixture** (i.e., a linear combination of m χ^2 distributions). It is hard to write out this mixture distribution explicitly, so the authors approximate it using a gamma distribution with parameters α and β given in Equations (7) and (8), respectively (these approximations essentially come from method-of-moments-type arguments). The nice thing about using the gamma approximation is that the gamma probability distribution is "tabled" in R; type `help(pgamma)` in R and you will see.

Here is what I want you to do.

- Refit the square-root transformed IMA(1,1) model to the Boston Red Sox homerun data in Problem 2. Save the standardized residuals from this fit and calculate the first 10 sample autocorrelations (of the standardized residuals) and also the first 10 sample partial autocorrelations (of the standardized residuals).
- For each $m = 2, 3, 4, ..., 10$, calculate the values of $\widetilde{Q}_W$ and $\widetilde{M}_W$. You can do this by hand (may take a while) or you can get R to do this for you.
- For each m, calculate the probability value for the test of H_0 versus H_1 above using the gamma approximation. For example, suppose you have calculated $\widetilde{Q}_W$ with $m = 3$; call it `Q.W.3`. To find the probability value, type in

```
m = 3
alpha = (3/4)*(m*(m+1)^2)/(2*m^2+3*m+1-6*m) # Equation (7)
beta = (2/3)*(2*m^2+3*m+1-6*m)/(m*(m+1)) # Equation (8)

# p-value
1-pgamma(Q.W.3,shape = alpha, scale = beta)
```

  Note that in the formulas for α and β, we have $p + q = 0 + 1 = 1$, because we are fitting an IMA(1,1) to the (square-root transformed) process. The function `pgamma` gives a cumulative probability; therefore, `1-pgamma` gives the right-tail probability (like the Ljung-Box test, both competing GOF tests are one-sided, upper-tail tests).

- Therefore, you should calculate probability values for each value of m ($m = 2, 3, 4, ..., 10$) for both $\widetilde{Q}_W$ and $\widetilde{M}_W$. The gamma approximation described in the last bullet applies to both $\widetilde{Q}_W$ and $\widetilde{M}_W$.

- What do the competing GOF testing procedures say about the adequacy of the model that I fit to the Boston homerun data? Compare your conclusions to those reached from the Ljung-Box test (p-values for this test are given in the `tsdiag` output in Problem 2).

Discussion: Why is this paper so interesting? At least within an ARIMA modeling framework (which is the focus in our class), examine carefully the authors' Table 2. Do you see how the powers associated with $\widetilde{Q}_W$ and $\widetilde{M}_W$ are almost always larger than the power associated with the Ljung-Box test (statistic denoted by $\widetilde{Q}$)? This means that if the H_0 model is **not correct**, the testing procedures that use $\widetilde{Q}_W$ and $\widetilde{M}_W$ have a larger probability of rejecting H_0 when compared to the Ljung-Box test. Furthermore, the authors' Table 1 provides evidence that the new GOF tests confer higher power while still maintaining the nominal Type I Error rates.