

STAT 720 sp 2019 Lec 08

Karl Gregory

3/7/2019

This lecture will make use of the `tscourse` package, which is installed with the following R code:

```
library(devtools)
devtools::install_github("gregorkb/tscourse")
library(tscourse)
```

Two tests for white noise

Given a realization X_1, \dots, X_n from a time series $\{X_t, t \in \mathbb{Z}\}$, we are interested in testing the hypotheses

$$H_0: \{X_t, t \in \mathbb{Z}\} \text{ are white noise} \quad \text{versus} \quad H_1: \{X_t, t \in \mathbb{Z}\} \text{ are not white noise.}$$

We are often interested in these hypotheses when the time series $\{X_t, t \in \mathbb{Z}\}$ represents residuals after removing a trend and a seasonal component from a time series or when $\{X_t, t \in \mathbb{Z}\}$ are prediction errors based on a time series model, such as an $\text{ARMA}(p, q)$ model. If our residuals behave like white noise, then our model has accounted for temporal dependence. We consider two tests, one based on the periodogram and one based on the sample autocorrelation function.

The Ljung-Box test

The Ljung-Box test is based on the asymptotic distribution of the sample autocorrelations, and it specifically tests the hypotheses

$$H_0: \{X_t, t \in \mathbb{Z}\} \text{ are iid} \quad \text{versus} \quad H_1: \{X_t, t \in \mathbb{Z}\} \text{ are not iid.}$$

Result leading to the Ljung-Box test

For any positive integer k , let $\hat{\boldsymbol{\rho}}_k = (\hat{\rho}_n(1), \dots, \hat{\rho}_n(k))^T$ and $\boldsymbol{\rho}_k = (\rho(1), \dots, \rho(k))^T$, where $\hat{\rho}_n(\cdot)$ and $\rho(\cdot)$ are the sample and true autocorrelation functions of the times series $\{X_t, t \in \mathbb{Z}\}$, respectively. Then Theorem 7.2.2 of B&D Theory gives

$$\sqrt{n}(\hat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k) \rightarrow \text{Normal}(0, \mathbf{W}_k) \quad \text{in distribution,}$$

as $n \rightarrow \infty$, where \mathbf{W}_k is the $k \times k$ matrix with entries given by

$$w_{ij} = \sum_{l=1}^{\infty} [\rho(l+i) + \rho(l-i) - 2\rho(i)\rho(l)][\rho(l+j) + \rho(l-j) - 2\rho(j)\rho(l)], \quad 1 \leq i, j \leq k$$

Under H_0 , we have

$$w_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j, \end{cases}$$

so that \mathbf{W}_k is the $k \times k$ identity matrix. This implies that under H_0 , we have

$$n \sum_{l=1}^k \hat{\rho}_n(l)^2 \rightarrow \chi_k^2 \quad \text{in distribution,}$$

as $n \rightarrow \infty$. This suggests use of the test statistic

$$Q = n \sum_{l=1}^k \hat{\rho}_n(l)^2$$

and the rejection rule which rejects H_0 at level α when Q exceeds the upper α quantile of the χ_k^2 distribution. Ljung and Box (1978) present a modification of this test statistic under which the test has closer-to-nominal size.

Test statistic and decision rule for the Ljung-Box test

The Ljung-Box test statistic is

$$Q_{LB} = n(n+2) \sum_{l=1}^k \hat{\rho}_n(l)^2 / (n-l),$$

and the rejection rule of the Ljung-Box test is

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if and only if } Q_{LB} > \chi_{k,\alpha}^2,$$

where $\chi_{k,\alpha}^2$ is the upper α quantile of the χ_k^2 distribution.

The following R code defines a function for computing the Ljung-Box test and illustrates it on a dataset:

```
WNtest.LB <- function(X,k=1,nparms=0)
{
  n <- length(X)

  if(k > n-1 )
    stop("k must be less than n")
  if(k <= nparms)
    stop("k must be greater than nparms")

  rho.k <- sample.acf(X)$rho.hat[2:(k+1)]
  Q.LB <- n*(n+2) * sum( rho.k^2 / (n - 1:k))
  pval <- 1 - pchisq(Q.LB,k-nparms)

  return(pval)
}

n <- 100
X <- rnorm(n)

WNtest.LB(X)
```

```
## [1] 0.3900936
```

We have not addressed how to choose the value of k . The following simulation computes the size and power of the Ljung-Box test under the AR(1) model under different values of the AR(1) parameter ϕ and under the choices $k = 1, \dots, 10$:

```
phi <- c(0,.1,.2,.3,.4,.5)
S <- 500
B <- 100
n <- 50
```

```

K <- 10
alpha <- 0.05
prej <- matrix(NA,K,length(phi))

for(j in 1:length(phi))
{
  if(j==1)
  {
    XX <- matrix(rnorm(n*S),S,n)

  } else {
    XX00 <- get.ARMA.data(phi=phi[j],theta=NULL,sigma=1,n=S*(B+n))
    XX0 <- matrix(XX00,S,B+n,byrow=TRUE)
    XX <- XX0[,1:n]

  }

  for(k in 1:K)
  {
    prej[k,j] <- mean(apply(XX,1,WNtest.LB,k=k) < alpha)

  }

}

rownames(prej) <- paste( "k =",1:K)
colnames(prej) <- paste( "phi =",phi)

library(knitr)
kable(prej)

```

	phi = 0	phi = 0.1	phi = 0.2	phi = 0.3	phi = 0.4	phi = 0.5
k = 1	0.048	0.084	0.228	0.468	0.720	0.922
k = 2	0.056	0.070	0.180	0.372	0.656	0.856
k = 3	0.050	0.072	0.144	0.334	0.552	0.826
k = 4	0.046	0.068	0.144	0.298	0.534	0.794
k = 5	0.050	0.068	0.140	0.276	0.518	0.760
k = 6	0.052	0.072	0.134	0.280	0.496	0.726
k = 7	0.068	0.066	0.124	0.276	0.476	0.722
k = 8	0.064	0.070	0.134	0.274	0.456	0.696
k = 9	0.064	0.068	0.124	0.274	0.436	0.684
k = 10	0.052	0.086	0.124	0.276	0.416	0.684

The table shows the proportion of times the Ljung-Box test rejected the null hypothesis over the 500 simulation runs. Based on the simulation output, $k = 1$ seems to be the best choice under these settings, as it controls the Type I error rate and results in the greatest power.

Bartlett's test

Bartlett's test is based on the periodogram, and it is a test of the hypotheses

$$H_0: \{X_t, t \in \mathbb{Z}\} \text{ are iid Normal} \quad \text{versus} \quad H_1: \{X_t, t \in \mathbb{Z}\} \text{ are not iid Normal.}$$

It is based on considering how the periodogram ordinates would behave if $\{X_t, t \in \mathbb{Z}\}$ were iid $\text{Normal}(0, \sigma^2)$ random variables. Then it asks, “how different are the periodogram ordinates of the data from those we would expect under iid Normality?” The test is based on the following two results:

The first result is from Proposition 10.2.1 of B&D Theory: Suppose $\{X_t, t \in \mathbb{Z}\}$ are iid $\text{Normal}(0, \sigma^2)$ random variables and let $I_n(\lambda_1), \dots, I_n(\lambda_n)$ be the periodogram ordinates based on X_1, \dots, X_n , where $\lambda_1, \dots, \lambda_n$ are the Fourier frequencies $(2\pi/n)k$, $k = -\lfloor (n-1)/2 \rfloor, \dots, \lfloor n/2 \rfloor$. Then the joint distribution of the random variables

$$Y_j = \frac{\sum_{k=1}^j I_n(\lambda_k)}{\sum_{k=1}^q I_n(\lambda_k)}, \quad j = 1, \dots, q-1,$$

where $q = \lfloor (n-1)/2 \rfloor$, is the same as the joint distribution of the order statistics of a random sample of size $q-1$ from the $\text{Uniform}(0, 1)$ distribution.

The second result comes from Feller (1948): Let U_1, \dots, U_m be a random sample from the $\text{Uniform}(0, 1)$ distribution and let $U_{(1)} < \dots < U_{(m)}$ be the order statistics. Then

$$\lim_{m \rightarrow \infty} P \left(\max_{1 \leq k \leq m} \sqrt{m} \left| U_{(k)} - \frac{k}{m} \right| \leq a \right) = \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2a^2 j^2).$$

Note that k/m is the height of the $\text{Uniform}(0, 1)$ cdf at k/m and $U_{(k)}$ is the height of the empirical distribution of U_1, \dots, U_m at k/m , so the quantity

$$\max_{1 \leq k \leq m} \sqrt{m} \left| U_{(k)} - \frac{k}{m} \right|$$

is a comparison between the empirical distribution of the sample U_1, \dots, U_m and the $\text{Uniform}(0, 1)$ cdf.

Test statistic and decision rule for Bartlett's test

The test statistic for Bartlett's test is constructed by comparing Y_1, \dots, Y_{q-1} to the heights of the $\text{Uniform}(0, 1)$ cdf at $1/(q-1), \dots, 1$. The test statistic is

$$B = \max_{1 \leq k \leq q-1} \sqrt{q-1} \left| Y_k - \frac{k}{q-1} \right|.$$

The rejection rule at level α is

$$\text{Reject } H_0 \text{ if and only if } \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2B^2 j^2) < \alpha.$$

Note that the above is just a comparison of the p -value to the significance level α . The following R code defines functions for computing Bartlett's test statistic and implements the test on simulated data. The plot printed is the empirical distribution of Y_1, \dots, Y_{q-1} , and the test rejects at the 0.05 and 0.01 levels if the points fall outside of the solid or dotted lines, respectively.

```
# This function computes the periodogram at the Fourier frequencies.
pgram <- function(X, plot=FALSE)
{
  n <- length(X)
```

```

lambda <- (-floor((n-1)/2):floor(n/2))/n*2*pi

E <- matrix(NA,n,n)
for(i in 1:n)
{
  E[i,] <- 1/sqrt(n) * exp(1i*i*lambda)
}

# compute discrete Fourier transform of X
D <- as.vector(t(Conj(E)) %*% X)
I <- Mod(D)^2

if(plot == TRUE)
{
  plot(I[lambda>0]~lambda[lambda>0],type="o")
}

# compute cumulative periodogram
Y0 <- cumsum(I[lambda < 0]) / sum(I[lambda < 0])
Y <- Y0[-length(Y0)]

output <- list( I = I,
               lambda = lambda,
               Y = Y)

return(output)
}

# This function performs Bartlett's test for whether the time series is iid Gaussian.
WNtest.Bartlett <- function(X,plot=FALSE)
{
  Y <- pgram(X)$Y
  q <- length(Y) + 1

  # compute test statistic
  dev <- sqrt(q-1) * abs(Y - c(1:(q-1))/(q-1))
  B <- max(dev)

  # get p-value
  j <- c(-100:100)
  pval <- 1 - sum( (-1)^j * exp(-2*B^2*j^2) )

  if(plot == TRUE)
  {
    plot(Y,
         main=paste("Bartlett test: p val = ",round(pval,3),sep=""),
         xlab="frequencies",
         ylab="cumulative periodogram",
         col = ifelse( dev > 1.36,"red","black" ),
         pch = ifelse( dev > 1.36,19,1 ))
    abline(-1.36/sqrt(q-1),1/(q-1))
  }
}

```

```

    abline(+1.36/sqrt(q-1),1/(q-1))
    abline(-1.63/sqrt(q-1),1/(q-1),lty=3)
    abline(+1.63/sqrt(q-1),1/(q-1),lty=3)

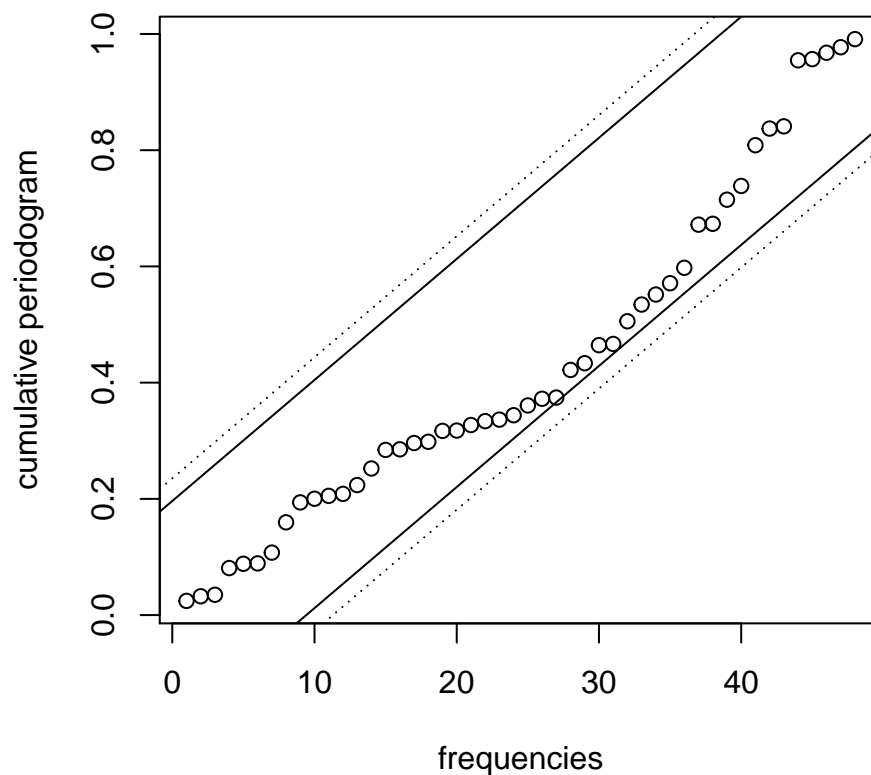
  }

  return(pval)
}

X <- rnorm(100)
WNtest.Bartlett(X,plot=TRUE)

```

Bartlett test: p val = 0.067



```
## [1] 0.06678687
```

Goodness of fit testing for ARMA models

When we fit a model to time series data, we would like to know whether we have correctly modeled the dependence structure. To see how well our model fits the data, we study the “residuals” of the fitted model. What are the residuals? Before defining the residuals of a fitted time series model, we first consider the one-step-ahead prediction errors, scaled to have unit variance.

Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary time series and let $\hat{X}_1, \dots, \hat{X}_n$ be the one-step-ahead predictions for X_1, \dots, X_n and denote by v_0, \dots, v_{n-1} the associated MSEs. Then we define the standardized one-step-ahead prediction

errors as

$$\varepsilon_j = v_{j-1}^{-1/2}(X_j - \hat{X}_j), \quad \text{for } j = 1, \dots, n.$$

If the predictors $\hat{X}_1, \dots, \hat{X}_n$ are chosen such that they completely account for the dependence among X_1, \dots, X_n , then $\varepsilon_1, \dots, \varepsilon_n$ will be uncorrelated random variables with mean zero and unit variance; that is, they will behave like a realization from a $\text{WN}(0, 1)$ time series. We will treat $\varepsilon_1, \dots, \varepsilon_n$ as our residuals.

Note that these “residuals” are different from those which we considered in the context of removing a trend or seasonal effect from a time series. Here we are considering a time series which is stationary, so that trends and seasonal effects have already been removed, and our residuals are one-step-ahead prediction errors rather than deviations from a mean.

Now, if $\{X_t, t \in \mathbb{Z}\}$ is an $\text{ARMA}(p, q)$ time series with the parameters $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^T$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^T$ and white noise variance σ^2 , then the one-step-ahead predictors and the MSEPs are functions of $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and σ^2 such that we may write $\hat{X}_j = \hat{X}_j(\boldsymbol{\theta}, \boldsymbol{\phi})$ and $v_{j-1} = v_{j-1}(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$, for $j = 1, \dots, n$, and thus also

$$\varepsilon_j = \varepsilon_j(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = v_{j-1}^{-1/2}(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)(X_j - \hat{X}_j(\boldsymbol{\phi}, \boldsymbol{\theta})), \quad \text{for } j = 1, \dots, n.$$

Now, suppose that an $\text{ARMA}(\hat{p}, \hat{q})$ model is fit to some data X_1, \dots, X_n , where \hat{p} and \hat{q} may or may not be the correct AR and MA orders, resulting in the estimated parameter values $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \dots, \hat{\phi}_{\hat{p}})^T$, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_{\hat{q}})^T$, and $\hat{\sigma}^2$. Then we define the residuals of the fitted model as

$$\hat{\varepsilon}_j = \varepsilon_j(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2) = v_{j-1}^{-1/2}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2)(X_j - \hat{X}_j(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})), \quad \text{for } j = 1, \dots, n.$$

If the data X_1, \dots, X_n are a realization from an $\text{ARMA}(p, q)$ model such that $p = \hat{p}$ and $q = \hat{q}$, so that a model with the correct AR and MA orders has been fit, then the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ should behave somewhat like a realization from a $\text{WN}(0, 1)$ time series (not exactly, since they are a function of some estimators, but approximately for a large-enough sample size n). In order to check whether the $\text{ARMA}(\hat{p}, \hat{q})$ model is appropriate for the data, we therefore check whether $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ behave like white noise.

Ljung-Box test for ARMA model goodness of fit

We can use the Ljung-Box test on the sample autocorrelation function of the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ of a fitted $\text{ARMA}(\hat{p}, \hat{q})$ model to see if the model fits the data. Specifically, we test the hypotheses

$$H_0: \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n \text{ are iid} \quad \text{versus} \quad H_1: \hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n \text{ are not iid.}$$

Let $\hat{\rho}_{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n}(\cdot)$ be the sample autocorrelation function of the the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$. Then the test statistic for the Ljung-Box goodness of fit test is

$$Q_{\text{LB}} = n(n+2) \sum_{j=1}^k \hat{\rho}_{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n}(j)/(n-j),$$

which is the same as it was before, and the rejection rule is

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if and only if } Q_{\text{LB}} > \chi_{k-\hat{p}-\hat{q}, \alpha}^2.$$

Note that the degrees of freedom of the chi-squared distribution is now $k - \hat{p} - \hat{q}$. This accounts for the estimation of the $p + q$ parameters $\phi_1, \dots, \phi_{\hat{p}}$ and $\theta_1, \dots, \theta_{\hat{q}}$. We see that we must choose $k > \hat{p} + \hat{q}$.

If we fail to reject H_0 , then we may argue that our model fits the data well.

The following R code generates some data from an $\text{ARMA}(p, q)$ model, fits a model to the data using maximum likelihood, and then obtains the residuals; lastly, it performs the Ljung-Box test on the residuals to see if the fitted model fits the data. We can obtain the residuals quite easily using the `arima()` function, but we can also get them using the Durbin-Levinson algorithm based on the autocovariance function from the fitted model, which the following code also illustrates.

```

phi <- c(.8,.1)
theta <- c(.4)
p <- length(phi)
q <- length(theta)
sigma <- 1
n <- 100

# generate data from an ARMA(p,q) model and center it to have mean 0
X <- get.ARMA.data(phi,theta,sigma,n)
X.bar <- mean(X)
X.cent <- X - X.bar

# choose p and q; get mles and residuals using the arima() function
p.hat <- 2
q.hat <- 1
arima.out <- arima(X.cent,order=c(p.hat,0,q.hat),include.mean=FALSE)
arima.resid <- arima.out$resid

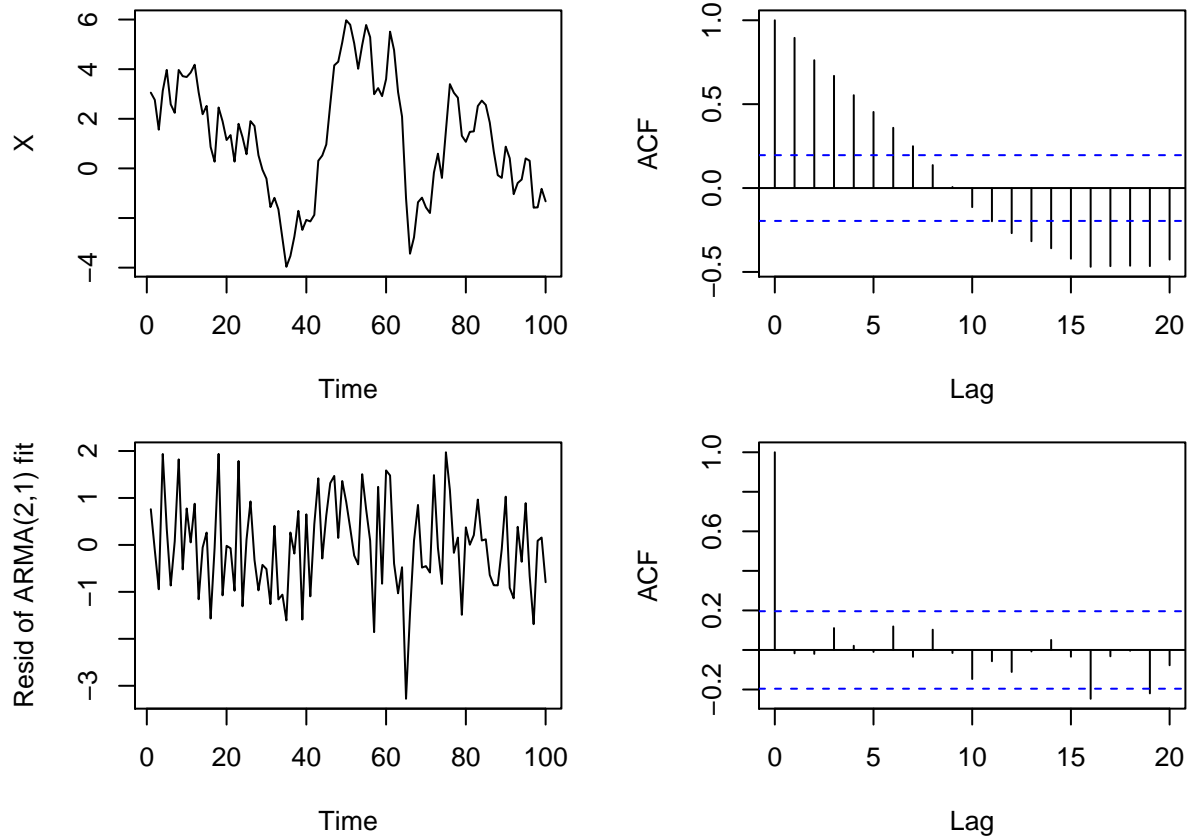
# construct residuals via the DL algorithm (they are the same)
phi.hat <- arima.out$coef[1:p.hat]
theta.hat <- arima.out$coef[-c(1:p.hat)]
sigma.hat <- sqrt(arima.out$sigma2)

gamma.hat <- ARMAacvf(phi.hat,theta.hat,sigma.hat,max.lag=n)
gamma.hat.0 <- gamma.hat[1]
gamma.hat.n <- gamma.hat[-1]

DL.1step.out <- DL.1step(X.cent,gamma.hat.0,gamma.hat.n)
resid <- (X.cent - DL.1step.out$X.pred[-(n+1)]) / sqrt(DL.1step.out$v[-(n+1)])

par(mfrow=c(2,2),mar=c(4.1, 4.1, 1.1, 2.1))
plot(X)
acf(X,main="")
plot(resid,ylab=paste("Resid of ARMA(",p.hat,",",q.hat,") fit",sep=""))
acf(resid,main="")

```

```
# perform Ljung-Box test on the residuals:
LB.pval <- WNtest.LB(resid,k=p.hat+q.hat+1,nparms=p.hat+q.hat)
LB.pval
```

```
## [1] 0.2361771
```

References

Feller, W. 1948. "On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions." *The Annals of Mathematical Statistics* 19 (2). Institute of Mathematical Statistics: 177–89.

Ljung, Greta M, and George EP Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (2). Oxford University Press: 297–303.