

Comparison of Various Tests For Randomness

DH Lee, Youngsoo Baek

2017.5.16.

Contents

1	Introduction	1
2	Goodness of Fit Tests for Time Series	1
3	Theoretical Background: Box-Pierce derivation	2
3.1	Box-Pierce Derivation	2
3.2	Modification: Ljung-Box Test	6
4	Simulations	7
4.1	Simulation: R Code	7
4.2	IID Normal	9
4.3	AR(1) with $\phi = 0.5$	10
4.4	AR(1) with $\phi = 0.1$	11
4.5	AR(1) with $\phi = 0.9$	12
4.6	ARMA(1,1) with $\phi = 0.5, \theta = 0.5$	13
4.7	AR(1) with a linear trend, $\phi = 0.5$	14
4.8	AR(1) with seasonality, $\phi = 0.9$	15
5	Conclusion	16
6	References	16
7	Appendix	17
7.1	Additional Plots	17
7.2	Tables of Acceptance Rates	23
7.3	R codes	30

1 Introduction

The objective of our project was comparing performance of various goodness of fit tests that test the randomness of time series. Our project was motivated by how different tests like Ljung-Box and difference sign tests can give wildly different results depending on certain characteristics of the series being tested. This project mainly consists of two parts. In the first portion, we examined the theory behind Box-Pierce derivation and the two parametric test statistics that employ this procedure, Box-Pierce and Ljung-Box. In the second portion, we simulated different time series models and compared performances of several parametric and non-parametric tests, from which we will make inference on their powers and noticeable characteristics.

2 Goodness of Fit Tests for Time Series

When we have fit a model to a series, a useful way of examining the adequacy of this fit model is applying a diagnostic method to the residuals. There are several tests that can be applied to the residuals. Each of these tests checks the null hypothesis that the residuals are a series of independent (Normal) variates.

- Box-Pierce test
- Ljung-Box test
- McLeod-Li test
- Difference sign test
- Turning point test
- Rank test

Box-Pierce test, Ljung-Box test, and McLeod-Li test are parametric tests, and their test statistic distribution is similarly derived from the Box-Pierce derivation procedure (which is discussed in the next section). difference sign test is a non-parametric test that checks whether the sign of difference between given two points is positive or negative. The difference sign test checks the randomness of the residuals by testing whether there is a constant pattern of increase or decrease in the series. Turning point test is a non-parametric test that checks whether a middle point in given three points has the highest or lowest value. The turning point test checks the randomness of the residuals by testing whether there are sufficient fluctuations in the series to consider each point independent from another. Rank test is a more elaborate form of test that checks the rank of a series instead of sign of difference between points. For simplification purposes, this test is not considered in our project.

The three non-parametric tests have existed as a more general form of testing independence of a sequence of random variables as early as the 18th century. Until Box and Pierce published their paper in 1970, they were the sole methods of testing the randomness of time series. The `test()` function of `itsmr` package still contains these tests as they make few assumptions about the distribution under test and have general applicability. Our project aims at comparing the performance of three parametric tests, difference sign test, and turning point test across various time series simulated with a pre-set ARMA model. We focus on the general comparison of test power across the five tests, and whether certain test performs better than the other for a series generated from a certain model.

3 Theoretical Background: Box-Pierce derivation

The motivation for looking at the Box-Pierce derivation is to get a better understanding of what parametric tests are actually doing when they tests a series. As mentioned above, the three main parametric tests follow from or slightly modify the Box-Pierce derivation procedure. All three test statistics use the distribution of the sum squared of residual autocorrelations to check whether the residuals are independent. Box-Pierce derivation shows how we can approximate this distribution by establishing a linear relationship between model iid autocorrelations and sample residual autocorrealtions. Ljung-Box test statistic, as seen later in this section, is a modification to Box-Pierce statistic that theoretically gives more accurate test results at smaller sample size. It should be noted that though McLeod-Li test is not included in this section, it also is a slight modification to the Box-Pierce test.

3.1 Box-Pierce Derivation

3.1.1 Distribution of iid noise autocorrelations

The general causal ARIMA(p, d, q) model can be stated as the following simple form:

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t$$

where $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$, and $\{Z_t\} \sim WN(0, \sigma^2)$, and where the roots of polynomial $\phi(z)$ and $\theta(z)$ lie outside the unit circle. Since the tests cannot check more exceptional cases of uncorrelated but not independent residuals, here we assume that Z_t is an iid noise.

In this model, the White Noise sequence Z_t has the following autocorrelation at lag k :

$$\rho(k) = \frac{\sum Z_t Z_{t-k}}{\sum Z_t^2}$$

The empirically known distribution of $\rho(k)$ is a Gaussian distribution centered at mean 0 with the following variance.

$$V(\rho(k)) = \frac{n-k}{n(n-2)} \approx \frac{1}{n} \quad (\text{for large } n)$$

Suppose we have a correlation vector $\rho = (\rho(1) \dots \rho(m))'$ for sufficiently large m . Then, ρ has a multivariate Normal distribution as below:

$$\rho \sim N(\mathbf{0}, (1/n)I)$$

Note we have approximated $V(\rho(k))$ for large n to simplify the covariance matrix of ρ . Therefore, we have the distribution of the sum squared of iid noise autocorrelations:

$$n \sum_{k=1}^m \rho^2(k) \sim \chi_m^2$$

The insight of Box-Pierce derivation is that if the residual series \hat{Z}_t is iid noise, then it should display similar behaviors as the model iid noise Z_t . Therefore, the below procedure attempts to approximate the distribution of the sum squared of **sample** residual autocorrelations $\hat{\rho}(k)$.

3.1.2 Distribution of AR process residual autocorrelations

For our project, we confine our interest to causal AR processes without loss of generality. Box and Pierce (1970) have shown that MA and ARMA processes behave similar behavior as AR processes and therefore the same test can be applied to the residuals in those cases.

The general causal AR(p) model is stated as:

$$\phi(B)X_t = Z_t$$

where the roots of polynomial $\phi(z)$ lie outside the unit circle. Also, we can express Z_t in terms of X_t as

$$Z_t = \psi(B)X_t$$

where $\psi(B) = \phi^{-1}(B)$.

To approximate the distribution of the sum squared sample residual autocorrelations $\sum \hat{\rho}^2(k)$, we can try establishing a linear relationship between $\rho(k)$ and $\hat{\rho}(k)$. We can approximate $\hat{\rho}(k)$ by a first order Taylor expansion about point $\hat{\phi} = \phi$. This is possible because we know the mean squared error of $\hat{\phi}$, $E(\phi - \hat{\phi})^2$, is of order $1/n$. Below, we use notation $\dot{\phi}$ for expressing variable of function $\hat{\rho}(k)$ since ϕ is a fixed model parameter. Notation \dot{Z}_t is used in similar context.

$$\begin{aligned} \hat{\rho}(k) &= \rho(k) + \sum_{j=1}^p (\phi_j - \hat{\phi}_j) \hat{\delta}_{jk} + O_p(1/n) \\ \text{where } \hat{\delta}_{jk} &= - \left. \frac{\partial \rho(k)}{\partial \phi_j} \right|_{\phi=\hat{\phi}} \end{aligned}$$

Now

$$\hat{\delta}_{jk} = -\frac{\partial}{\partial \dot{\phi}_j} \frac{\Sigma \dot{Z}_t \dot{Z}_{t-k}}{\Sigma \dot{Z}_t^2} \Big|_{\dot{\phi}=\hat{\phi}}$$

But at $\dot{\phi} = \hat{\phi}$, the variance of \dot{Z}_t is $\Sigma \hat{Z}_t^2$. This value is independent of variable $\dot{\phi}$. Therefore,

$$\hat{\delta}_{jk} = -\frac{1}{\Sigma \hat{Z}_t^2} \frac{\partial}{\partial \dot{\phi}_j} \Sigma \dot{Z}_t \dot{Z}_{t-k} \Big|_{\dot{\phi}=\hat{\phi}}$$

where

$$\begin{aligned} \Sigma \dot{Z}_t \dot{Z}_{t-k} &= \Sigma (\dot{\phi}(B)X_t)(\dot{\phi}(B)X_{t-k}) \\ &= \Sigma_t \Sigma_{i=0}^p \Sigma_{j=0}^p \dot{\phi}_i \dot{\phi}_j X_{t-i} X_{t-i-j} \end{aligned}$$

From the above expression, it follows that

$$\begin{aligned} \hat{\delta}_{jk} &= -\frac{1}{\Sigma \hat{Z}_t^2} \frac{\partial}{\partial \dot{\phi}_j} \Sigma_t \Sigma_{i=0}^p \Sigma_{j=0}^p \dot{\phi}_i \dot{\phi}_j X_{t-i} X_{t-i-j} \\ &= -\frac{\Sigma X_t^2}{\Sigma \hat{Z}_t^2} \Sigma_{i=0}^p \hat{\phi}_i [\hat{\rho}_X(k-i+j) + \hat{\rho}_X(k+i-j)] \\ &= -\frac{\Sigma_{i=0}^p \hat{\phi}_i [\hat{\rho}_X(k-i+j) + \hat{\rho}_X(k+i-j)]}{\Sigma_{i=0}^p \Sigma_{j=0}^p \hat{\phi}_i \hat{\phi}_j \rho_X(i-j)} \end{aligned}$$

where $\hat{\rho}_X(k) = \frac{\Sigma X_t X_{t-k}}{\Sigma X_t^2}$ is the sample autocorrelation of X_t at lag k .

Note that this partial derivative $\hat{\delta}_{jk}$ is in sample terms. If we replace $\hat{\phi}$ and $\hat{\rho}_X$ with theoretical parameter and autocorrelation, ϕ and ρ_X , we get the expression for δ_{jk} . But according to Bartlett's formula, $\hat{\rho}_X(k) = \rho_X(k) + O_p(1/\sqrt{n})$, and we know that $\hat{\phi} = \phi + O_p(1/\sqrt{n})$. This means that, in the Taylor expansion of $\hat{\rho}(k)$ in terms of $\rho(k)$, the partial derivative term $\hat{\delta}_{jk}$ is interchangeable with δ_{jk} .

Thus

$$\delta_{jk} = -\frac{\Sigma_{i=0}^p \phi_i [\rho_X(k-i+j) + \rho_X(k+i-j)]}{\Sigma_{i=0}^p \Sigma_{j=0}^p \phi_i \phi_j \rho_X(i-j)}$$

We can further simplify this expression by using the fact that, for an AR(p) process,

$$\rho_X(\nu) - \phi_1 \rho_X(\nu-1) - \dots - \phi_p \rho_X(\nu-p) = \phi(B) \rho_X(\nu) = 0$$

For example, in AR(1) case this holds true because $\phi(B) \rho_X(\nu) = (1 - \phi B) \rho_X(\nu) = \phi^\nu - (\phi)(\phi^{\nu-1}) = 0$.

The simplified form of δ_{jk} is

$$\delta_{jk} = \frac{\Sigma_{i=0}^p \phi_i \rho_X(k-j+i)}{\Sigma_{i=0}^p \phi_i \rho_X(i)} = \delta_{k-j}$$

Immediately it is clear that δ_{jk} only depends on the “lag” of $k-j$.

There are three properties of δ_{jk} we can see from this expression.

- $\delta_0 = 1$
- $\delta_\nu = 0$ for $\nu < 0$
- $\phi(B) \delta_\nu = \Sigma_{i=0}^p \phi_i [\phi(B) \rho_X(\nu+i)] / \Sigma_{i=0}^p \phi_i \rho_X(i) = 0$

What is this δ_{jk} ? We know that for causal AR processes $Z_t = \psi(B)X_t$ where $\psi(B) = \phi^{-1}(B)$. We can set $\psi_\nu = 0$ for $\nu < 0$. At $\nu = 0$, we have no information on previous X_t , so ψ_0 can be set 1.

Also, since, $\psi(B)\phi(B) = 1$, we can expand each polynomial and get the following relation:

$$\psi_\nu = \begin{cases} \phi_1\psi_{\nu-1} + \dots + \phi_{\nu-1}\psi_1 + \phi_\nu & \text{if } \nu < p \\ \phi_1\psi_{\nu-1} + \dots + \phi_p\psi_{\nu-p} & \text{if } \nu \geq p \end{cases}$$

Multiplying polynomial $\phi(B)$ on either side, we find that $\phi(B)\psi_\nu = 0$ for $\nu \neq 0$.

This indicates that ψ_ν has the exact same properties as δ_ν . Therefore, we can now fully rewrite the Taylor expansion of $\hat{\rho}(k)$ about $\hat{\phi} = \phi$ and establish a linear relationship between $\rho(k)$ and $\hat{\rho}(k)$.

$$\hat{\rho}(k) = \rho(k) + \sum_{j=1}^p (\phi_j - \hat{\phi}_j)\psi_{k-j} + O_p(1/n)$$

Let us express this linear relationship in terms of matrix multiplication, with vectors ρ and $\hat{\rho}$ containing autocorrelations at different lags. To the order of $1/n$, we can write

$$\hat{\rho} = \rho + \Psi(\phi - \hat{\phi})$$

where

$$\Psi = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ \psi_1 & 1 & \dots & \dots & \vdots \\ \vdots & \psi_1 & \ddots & & \vdots \\ \vdots & & & \ddots & 1 \\ \vdots & \vdots & & & \vdots \\ \psi_{m-1} & \psi_{m-2} & \dots & \dots & \psi_{m-p} \end{pmatrix}$$

Note that here, $\hat{\rho}'\Psi = \mathbf{0}$. This follows from the fact that residuals \hat{Z}_t and X_{t-j} are uncorrelated for $1 \leq j \leq p$, that is,

$$\sum_{t=p+1}^n \hat{Z}_t X_{t-j} = 0$$

and thus

$$\begin{aligned} 0 &= \sum_t \sum_k \hat{\psi}_k \hat{Z}_t \hat{Z}_{t-k-j} \\ &= \sum_k \hat{\psi}_k \hat{\rho}(k+j) \\ &= \sum_k \psi_k \hat{\rho}(k+j) + O_p(1/n) \end{aligned}$$

Thus, $\hat{\rho}'\Psi = \mathbf{0}$. Equivalently, $\Psi'\hat{\rho} = \mathbf{0}$.

$Q = \Psi(\Psi'\Psi)^{-1}\Psi'$ (notice it has an idempotent form). Multiplying the former equation of linear relationship between ρ and $\hat{\rho}$, we have

$$\begin{aligned} Q\hat{\rho} &= Q\rho + \Psi(\phi - \hat{\phi}) \\ \mathbf{0} &= Q\rho + \Psi(\phi - \hat{\phi}) \\ &= Q\rho + \hat{\rho} - \rho \\ \implies \hat{\rho} &= (I - Q)\rho \end{aligned}$$

Now, we can finally compare the distributions of model iid autocorrelations and sample residual autocorrelations thus:

$$\rho \sim N(0, (1/n)I)$$

$$\hat{\rho} \sim N(0, (1/n)(I - Q))$$

Q matrix is of rank p . Since the vector $\hat{\rho}$ contains autocorrelations at lags from 1 through arbitrary m , the covariance matrix of $\hat{\rho}$ is idempotent with rank $m - p$.

3.1.3 Box-Pierce test statistic

To look at the structure of this covariance matrix in more detail, let us compute the first few terms of covariance matrix of $\hat{\rho}$ in AR(1) case.

$$V(\hat{\rho}) = (1/n)(I - Q) = (1/n) \begin{pmatrix} \phi^2 & -\phi + \phi^3 & -\phi^2 + \phi^4 & \dots \\ -\phi + \phi^3 & 1 - \phi^2 + \phi^4 & -\phi^3 + \phi^5 & \dots \\ -\phi^2 + \phi^4 & -\phi^3 + \phi^5 & 1 - \phi^4 + \phi^6 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Notice here that since $|\phi| < 1$, the diagonal entries approach 1 and nondiagonal entries die out to 0. In fact, this shows that matrix $I - Q$ is similar to an identity matrix. In general, the covariance matrix of $\hat{\rho}$ is very similar to that of ρ but has a reduced rank of $m - p$ instead of m .

Now we have the distribution of the sum squared sample residual autocorrelations:

$$n \sum_1^m \hat{\rho}^2(k) \sim \chi_{m-p}^2$$

This is a remarkably similar distribution to that of $n \sum_1^m \rho^2(k)$ but with reduced degree of freedom. The **Box-Pierce test** statistic $Q(\rho(k))$ is defined as $n \sum_1^m \rho^2(k)$ with an approximate chi-squared distribution with degree of freedom $m - p$.

3.2 Modification: Ljung-Box Test

Box and Pierce (1970) thus provided us with a useful test statistic distribution that can be used to check how much the sample residuals behave similarly to an iid noise. Ljung and Box (1978), however, found that at small sample sizes Box-Pierce test can often give inaccurate results. This was caused by an approximation made to the variance of model iid autocorrelation $\rho(k)$:

$$V(\rho(k)) = \frac{n - k}{n(n - 2)} \approx \frac{1}{n}$$

It is not hard to notice that at small sample size n , there can be a sizable difference between the two terms. Thus, a new test statistic was proposed that uses a more precise expression of $V(\rho(k))$. The **Ljung-Box test** statistic is defined thus:

$$\tilde{Q}(\hat{\rho}(k)) = n(n - 2) \sum_{k=1}^m (n - k)^{-1} \hat{\rho}^2(k)$$

The Ljung-Box test is different from Box-Pierce test only because it makes a slight modification to the distribution of **model** iid noise autocorrelations. Therefore, the same linear relationship between $\rho(k)$ and $\hat{\rho}(k)$ still holds true.

The figure is a comparison of Monte Carlo distributions for $Q(\rho)$ (Box-Pierce) and $\tilde{Q}(\rho)$ (Ljung-Box) from Ljung and Box (1978). The smooth curve is a chi-squared distribution with degrees of freedom 30. It is immediately evident from the figure that the distribution of $Q(\rho)$ has a significant mean bias.

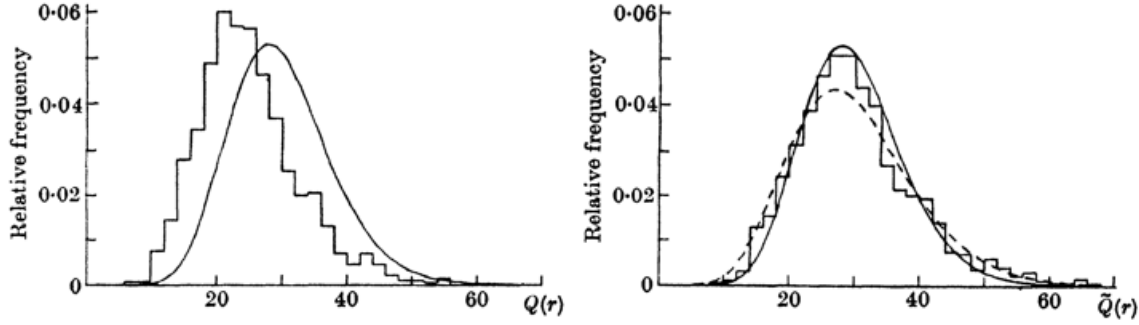


Fig. 1. Monte Carlo distributions of $Q(r)$ and $\tilde{Q}(r)$ and approximations; 1000 replications, $n = 100$ and $m = 30$; solid line, χ^2_{30} ; dashed line, $a\chi^2_b$ ($a = 1.52$, $b = 19.68$).

Figure 1: Ljung and Box (1978)

The Box-Pierce test is not without its advantage. Ljung and Box (1978) make clear that, while unclear in this figure, the distribution of $Q(\rho)$ has a smaller variance than $\tilde{Q}(\rho)$ that is much closer to the variance of chi-squared distribution $2m$. However, in most practical cases it is clear that the mean bias effect is so great that whatever gains from the smaller variance will be offset.

We have thus seen the theoretical background behind Box-Pierce and Ljung-Box tests. Theoretically, Ljung-Box test should show a better performance than Box-Pierce test when the sample size is relatively small. As the sample gets larger, the distribution of the two test statistics will become equivalent and the choice between two tests will be an indifferent one.

4 Simulations

In this portion of the project, we generated various time series from different AR (ARMA) models and ran randomness tests on those series. The objective was to compare whether a certain test performs better than the other in checking randomness of the series. In the following sections, we will compare our results for different models using a plot of acceptance rate of each test against increasing sample size. The different models we used are AR(1), AR(2), ARMA(1,1), ARMA(2,2), and generic AR (ARMA) models with linear trend and seasonality.

Different functions from different packages are used for our simulation. Box-Pierce and Ljung-Box test functions are included in `stat` package. McLeod-Li test function is included in `TSA` package. Difference sign and turning point tests are included in `randtest` package.

4.1 Simulation: R Code

```
NormTest<-function(N){
  accept=as.data.frame(matrix(NA,ncol=5,nrow=1000))

  Result=as.data.frame(matrix(NA,ncol=5))
  for (j in N){
    for (i in 1:1000){
      Nseries<-rnorm(j,0,1)
      BP<-ifelse(Box.test(Nseries,lag=1,type=c("Box-Pierce"),fitdf=0)$p.value>0.05,1,0)
      LB<-ifelse(Box.test(Nseries,lag=1,type=c("Ljung-Box"),fitdf=0)$p.value>0.05,1,0)
      ML<-ifelse(McLeod.Li.test(y=Nseries,gof.lag=1,plot=FALSE)$p.value>0.05,1,0)
```

```

DS<-ifelse(difference.sign.test(Nseries)$p.value>0.05,1,0)
TP<-ifelse(turning.point.test(Nseries)$p.value>0.05,1,0)

accept[i,]<-cbind(BP, LB, ML, DS, TP)

}
result<-apply(accept, 2, sum)/1000
Result<-rbind(Result, result)
}
colnames(Result)<-c("BoxP", "LjungB", "McLeod", "DiffS", "TurnP")
output<-cbind(N, Result[-1,])
print(output)

with(output, plot(N, BoxP, "b", ylim=c(0,1), ylab="PropAccept",
                  main="Acceptance of IID H0", xlab="Sample Size"))
points(output$N, output$LjungB, "b", col="red")
points(output$N, output$McLeod, "b", col="blue")
points(output$N, output$DiffS, "b", col="magenta")
points(output$N, output$TurnP, "b", col="green")
legend('bottomleft', legend=c("BoxP", "LjungB", "McLeod", "DiffS", "TurnP"),
      col=c("black", "red", "blue", "magenta", "green"),
      lty=1, cex=0.8)

}

```

This function is one example of the functions we used to simulate time series realizations of a certain model. The function takes in a vector specifying sample size, an AR (ARMA) model, and other elements needed such as the order of linear polynomial trend (the above example simulates IID Normal noise instead of an ARMA model). The function creates realizations of the model and runs our five choice of goodness of fit tests. If the test p-value is greater than 0.05, the test is recorded as have failed to reject the null hypothesis. The whole process is repeated 1000 times. The proportions of “acceptances” (when the test fails to reject null hypothesis) out of 1000 runs are returned in the forms of matrices and plots against sample size. Though technically a test does not “accept” the null hypothesis even when it fails to reject it, here we will frequently say that the test “accepts” the null hypothesis for simplicity purposes.

Some of the plots that best serve our interests of comparing different test results across distinctive models are included in the following sections. For a full list of plots and tables for various models we simulated, see appendix.

4.2 IID Normal

We ran the tests on a Gaussian, iid noise as a general point of reference. As expected, all tests accept the series as iid noise around 95% of the total runs.

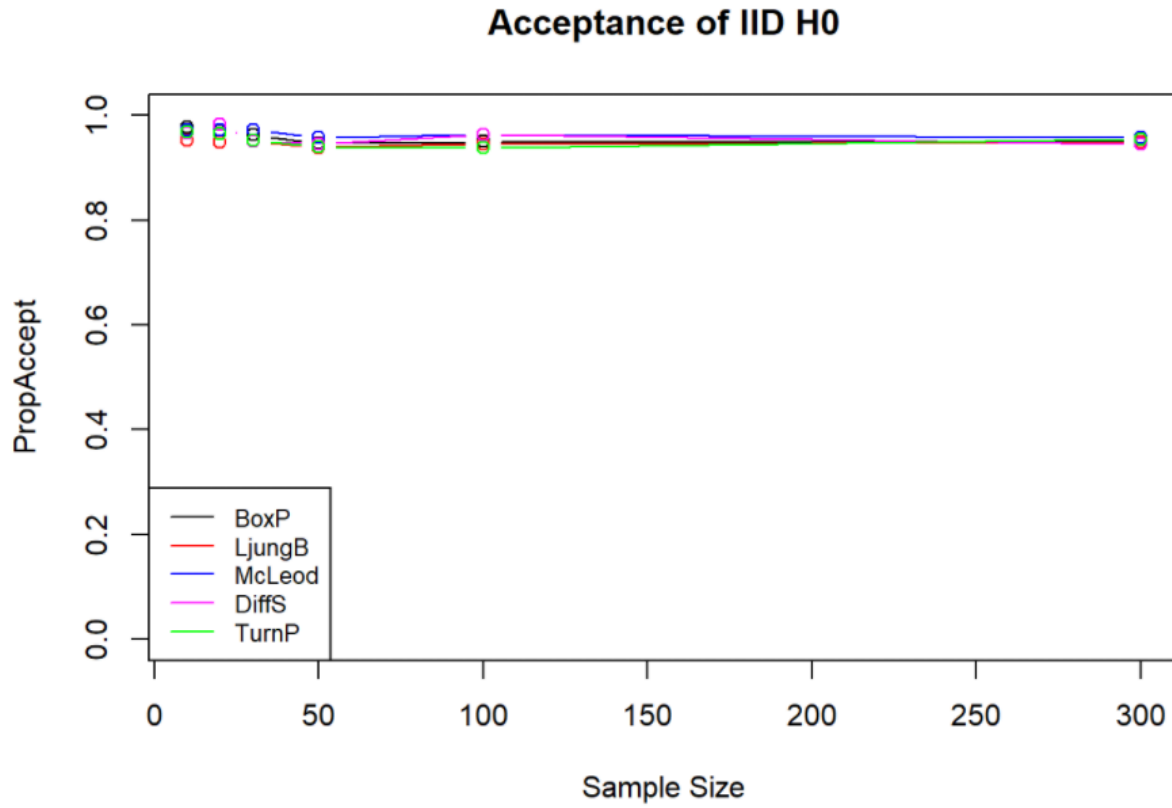


Figure 2: iid Gaussian noise

4.3 AR(1) with $\phi = 0.5$

For a fairly generic AR(1) model, with $\phi = 0.5$, most tests perform quite well. Box-Pierce and Ljung-Box acceptance rates drop quickly, rejecting almost 100% of the time at $n = 100$. Ljung-Box acceptance rates drops slightly faster than that of Box-Pierce in initially small n , although the two curves converge onto one another after $n = 100$. McLeod-Li and Turning point tests show relatively weak performance even when $n = 100$, but at $n = 300$ their acceptance rates drop down to about 5%. Difference sign test fails to distinguish this AR series from random noise.

What is noticeable here and the subsequent simulations is the extremely weak power of difference sign test even at sufficiently large sample sizes. This is because the nature of difference sign test makes it hard to distinguish series that nearly symmetrically oscillate around the center from random noise (Kendall and Stuart, 1958). Difference sign test has strong power when the series has a trend of constant increase or decrease, as will be shown below for series with linear trends.

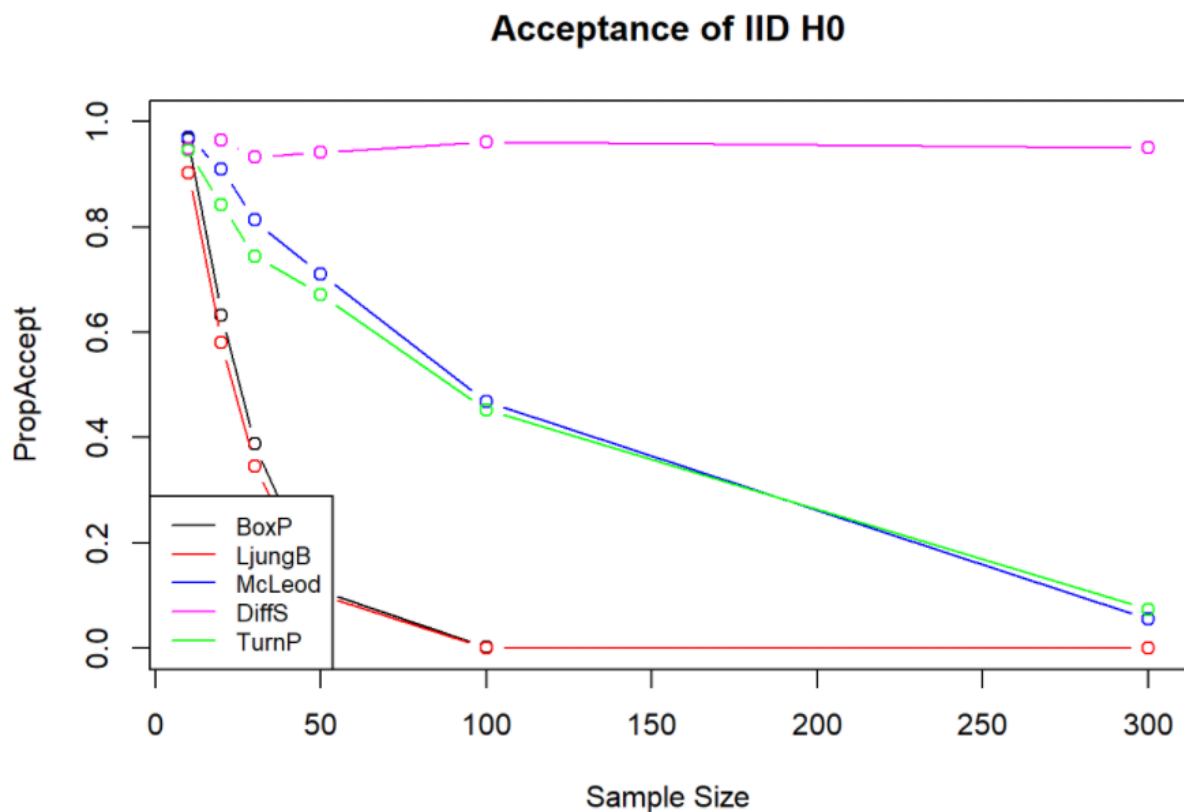


Figure 3: AR(1) model, $\phi=0.5$

4.4 AR(1) with $\phi = 0.1$

Next, an AR(1) model with small ϕ is set. Realizations of this model have very slight patterns and are hard to distinguish from random iid noise. Unsurprisingly, all tests perform poorly until $n = 100$, when Box-Pierce and Ljung-Box tests begin to reject in more cases with larger sample size. Turning point test performs slightly better than McLeod-Li test and difference sign test, but they all fail to reject too many times.

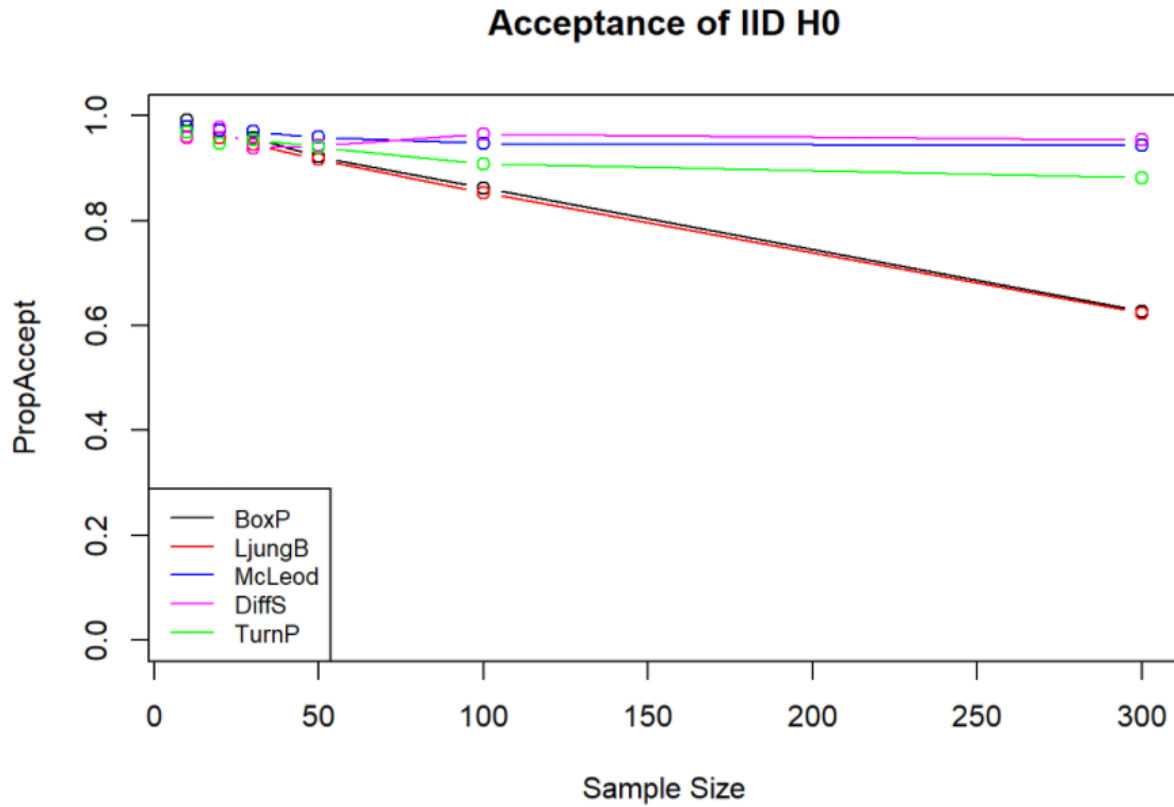


Figure 4: AR(1) model, $\phi=0.1$

4.5 AR(1) with $\phi = 0.9$

For an AR(1) model with $\phi = 0.9$, all tests except difference sign test shows a rapid decay of acceptance rates. All three parametric tests reject effectively 100% of the runs at $n = 100$, and Turning point acceptance rate converges to 0 at $n = 300$. This plot shows a generic pattern as the roots of polynomial $\phi(z)$ approach the unit circle.

When ϕ is smaller than .5, all tests suffer to distinguish the series from random noise. When ϕ is larger than .5, all tests except difference sign test shows a rapid decay of acceptance rates. Ljung-Box and Box-Pierce tests show the fastest decay of acceptance rates across all cases. For a similar simulations of AR(2) models, see appendix.

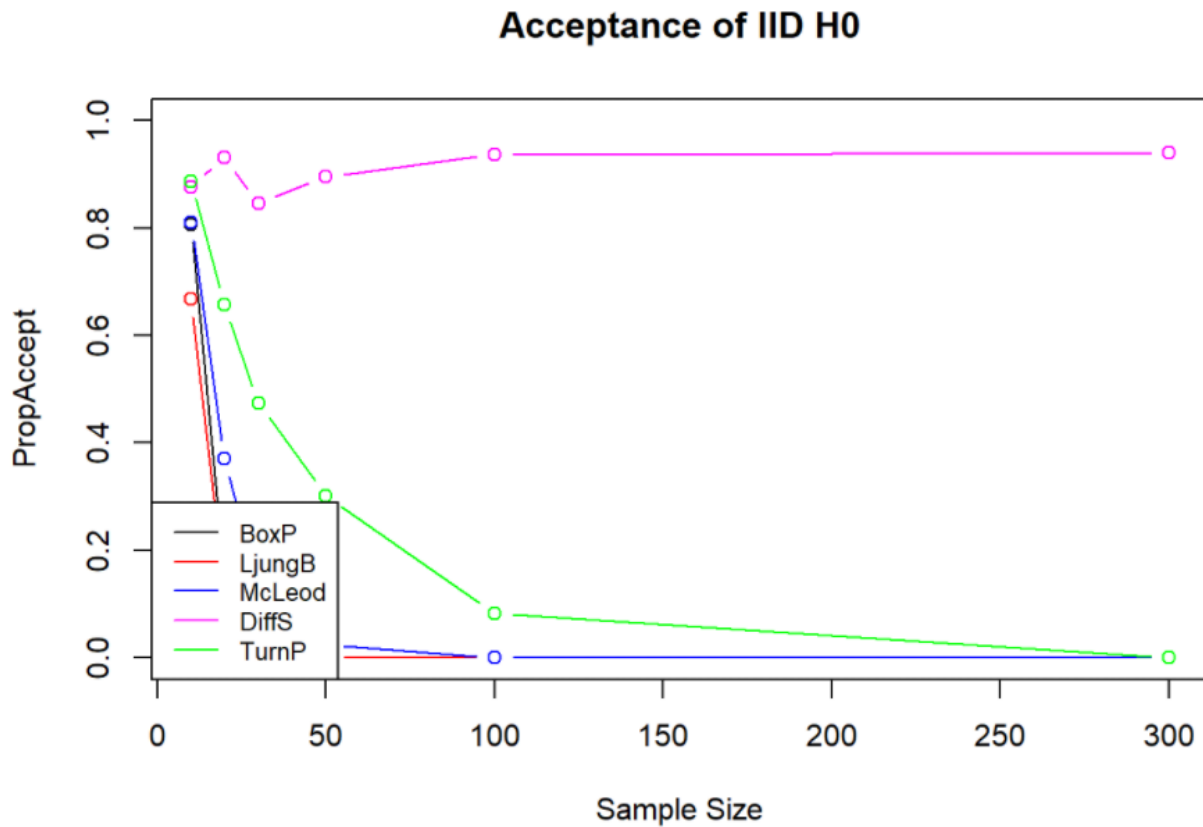


Figure 5: AR(1) model, $\phi=0.9$

4.6 ARMA(1,1) with $\phi = 0.5$, $\theta = 0.5$

For a generic ARMA(1,1) model where $\phi = \theta = .5$, all tests except difference sign test shows a rapid decay of acceptance rates, converging down to zero at $n = 100$. The curve for Ljung-Box again drops faster than that for Box-Pierce, and the two tests quickly converge onto one another at $n = 50$.

In the case of ARMA models, similarly as in AR models, the tests commit less Type II errors as the roots of polynomials $\phi(z)$ and $\theta(z)$ lie closer to the unit circle. As ϕ and θ get smaller, therefore, the roots will lie farther outside the unit circle and the tests will suffer distinguishing the series from random noise. A different example for ARMA(2,2) model is given in the appendix.

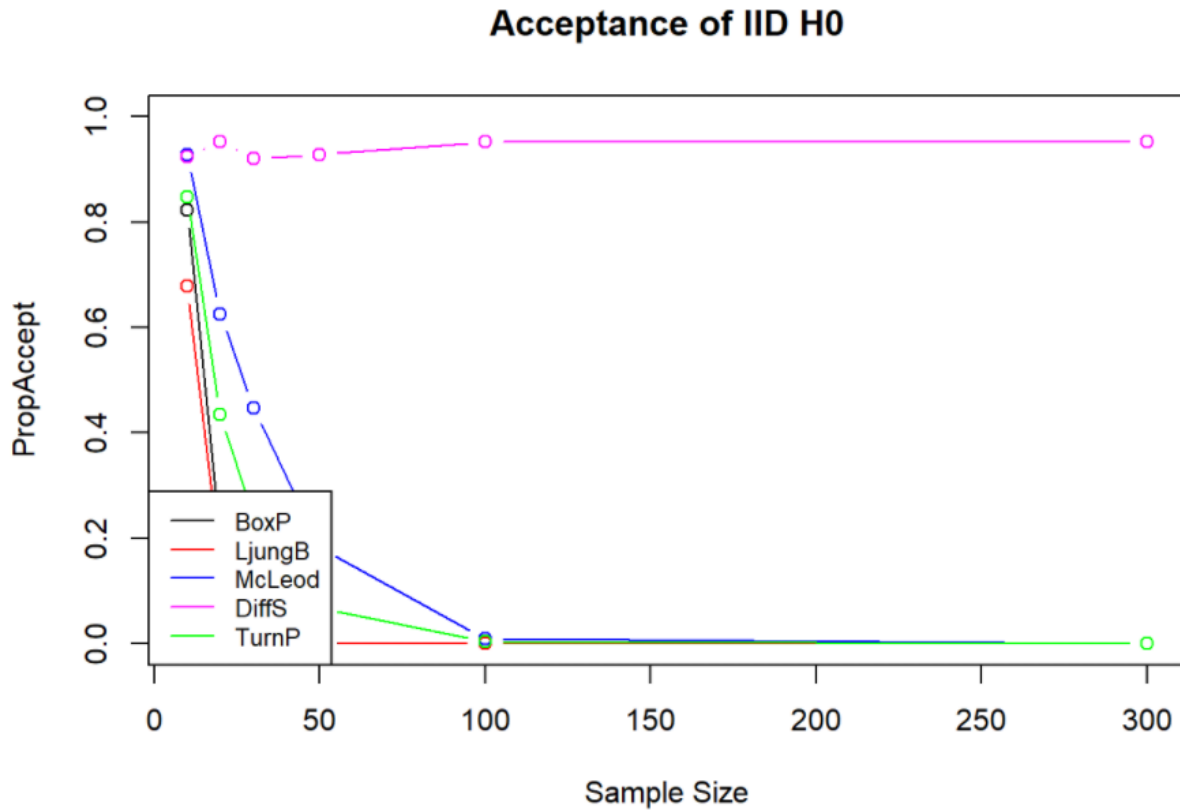


Figure 6: ARMA(1,1) model, $\phi=.5$, $\theta=.5$

4.7 AR(1) with a linear trend, $\phi = 0.5$

Next we applied tests to series with a linear trend. We set a generic AR(1) model with $\phi = 0.5$ and added a trend component, a first-degree polynomial $2 + 0.5t$.

Box-Pierce, Ljung-Box, and McLeod-Li tests are very sensitive to the trend, so they reject 100% of the cases at $n = 20$. Difference sign test performs better than turning point test, as its acceptance rate drops to 4% at $n = 50$, whereas the turning point test only reaches a similar point of 6% at $n = 100$. This finding aligns with our former remark that difference sign test has strong power against series with constant increase or decrease. It should be noted, however, that both nonparametric tests are performing very poorly given that the series has so obvious a trend. A similar plot for ARMA(1,1) cases is given in the appendix (curiously, though, difference sign test here performs worse than turning point test).

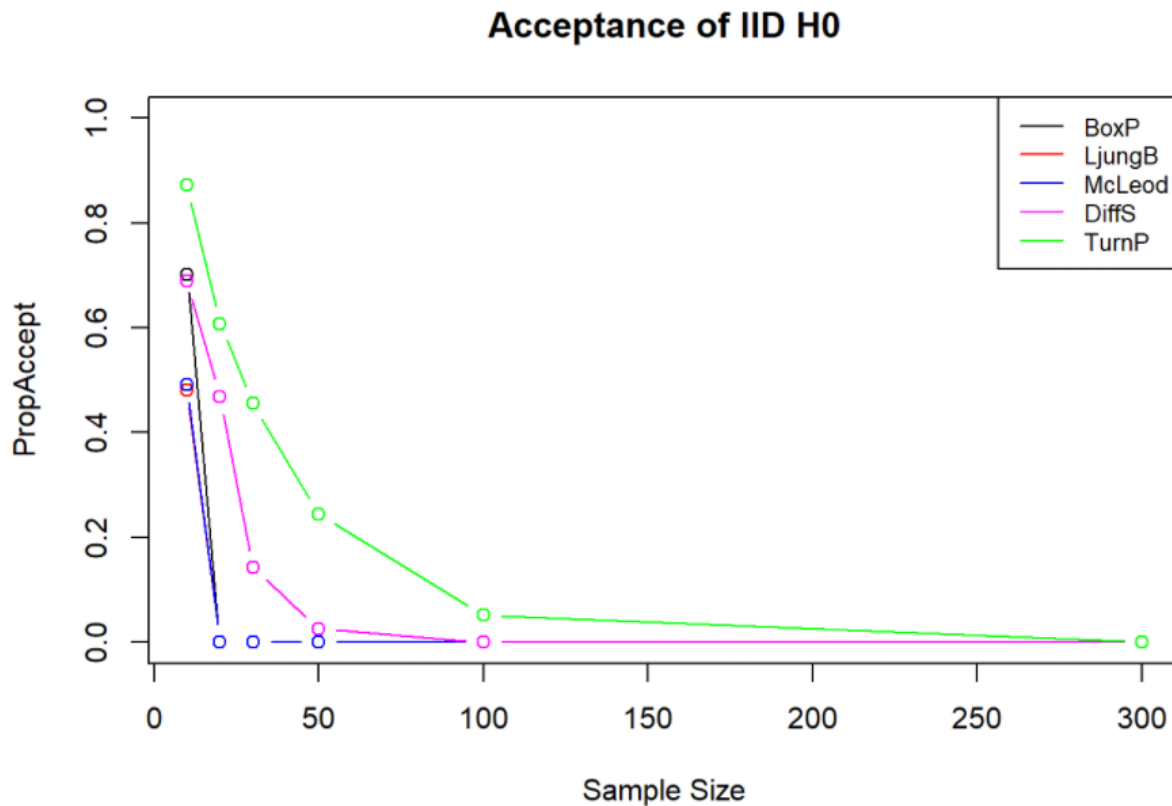


Figure 7: AR(1) model with linear trend

4.8 AR(1) with seasonality, $\phi = 0.9$

We also simulated the same AR(1) model but with seasonality, a pattern determined by sinusoidal curve $k \sin \frac{2\pi t}{d}$. Here, the pattern is annual, with period $d = 12$.

The three parametric tests and turning point tests all show a quick decay of acceptance rates, while the difference sign test is showing poor performance, as expected from a seasonal, fluctuating series.

Theoretically, turning point test has relatively strong power against series with seasonality rather than trend (Kendall and Stuart, 1958). In this plot, however, it is hard to discern this fact because all other three parametric tests are very sensitive to obvious seasonal deviations of the series.

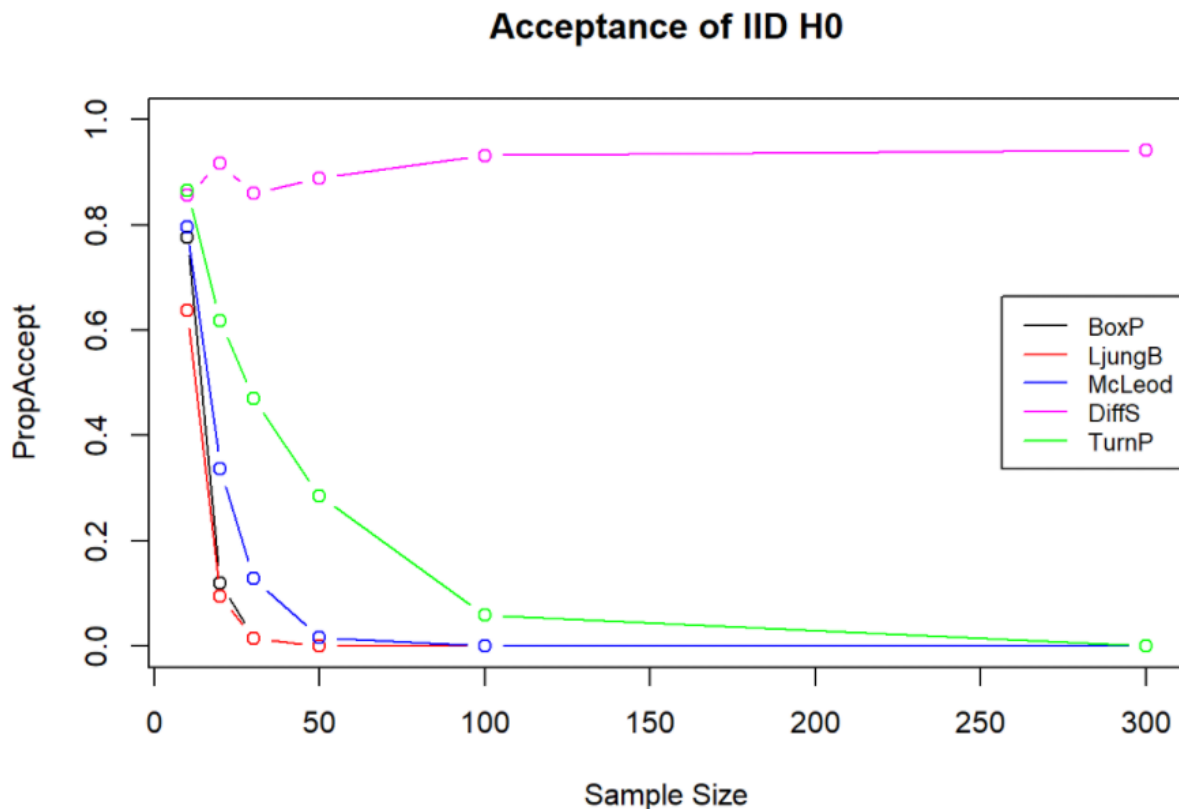


Figure 8: AR(1) model with seasonality

5 Conclusion

Several useful observations can be drawn from our simulations.

- All tests cannot accurately distinguish AR or ARMA series when the roots of polynomial $\phi(z)$ and $\theta(z)$ lie close to the unit circle. Specifically, in AR(1) case, the tests show commit less Type II errors as the magnitude of ϕ increases.
- Ljung-Box and Box-Pierce tests perform better than other tests across all simulated models.
- Ljung-Box performs slightly better than Box-Pierce at a sample size less than 100, but both tests show nearly identical results as the sample size increases.
- McLeod-Li test performs worse than Ljung-Box and Box-Pierce tests, but better than the two non-parametric tests in most cases.
- Turning point and difference sign tests perform relatively worse than the parametric tests across almost all cases.
- Difference sign test cannot distinguish most ARMA models that fluctuate around the center from random iid noise.

Based on our results, we will advise the readers to give more weights to Ljung-Box and Box-Pierce tests over others in most practical cases, as they have the strongest powers. The difference between these two tests is practically null in large sample sizes. One should be cautious of relying on non-parametric tests, particularly difference sign test, which should be primarily reserved for testing trend of the data. When the difference sign test rejects a null hypothesis, this in fact is a strong evidence against the null hypothesis. Also, a note of caution is also needed not to purely rely on the parametric tests, since our simulations in this project all compare the performance of tests based on a model we previously set. This indicates that, while Ljung-Box and Box-Pierce tests have strong powers, they are slightly more prone to Type I errors than other tests. Therefore, as long as a series does not display any visible pattern deviating from iid noise, non-parametric tests can still serve as a point of reference for subjective judgment.

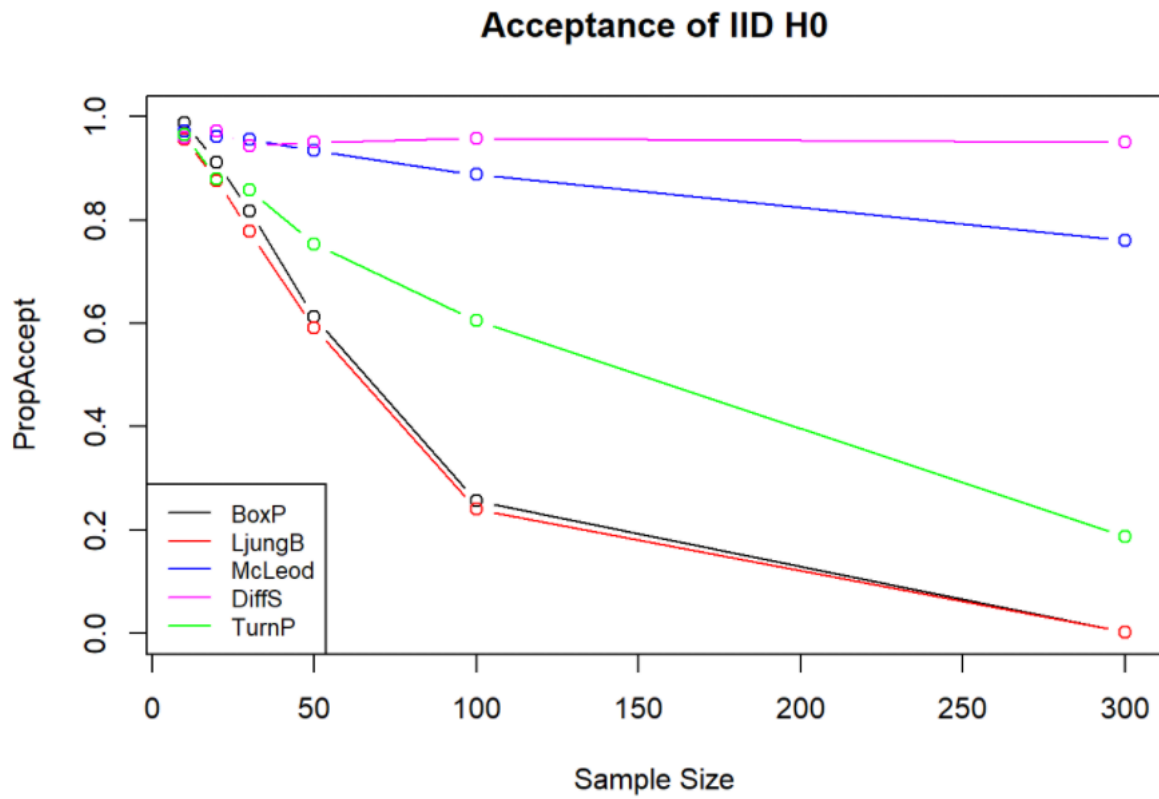
6 References

- Box, G.E.P. and Pierce, D., "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65 (December 1970) 1509-1526.
- Ljung, G.M. and Box, G.E.P., "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65 (1978) 297-304.
- Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics, Volume 3: Design and Analysis, and Time Series*, London:Griffin, 1958.

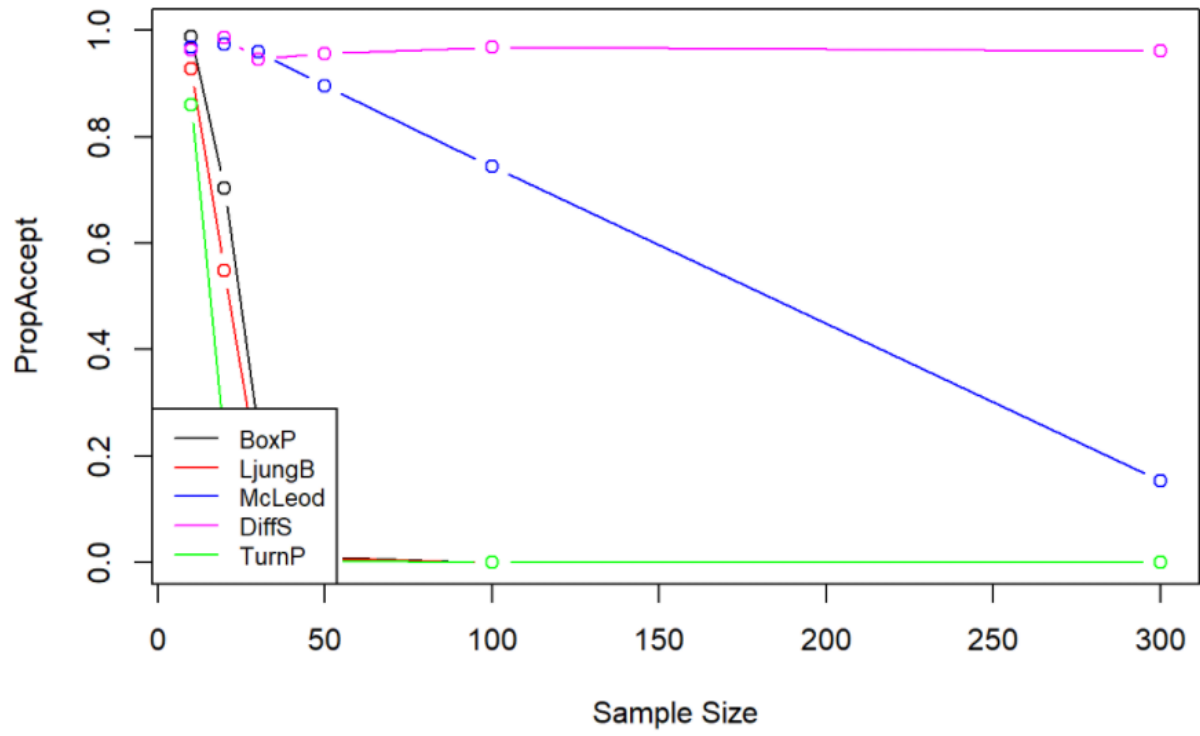
7 Appendix

7.1 Additional Plots

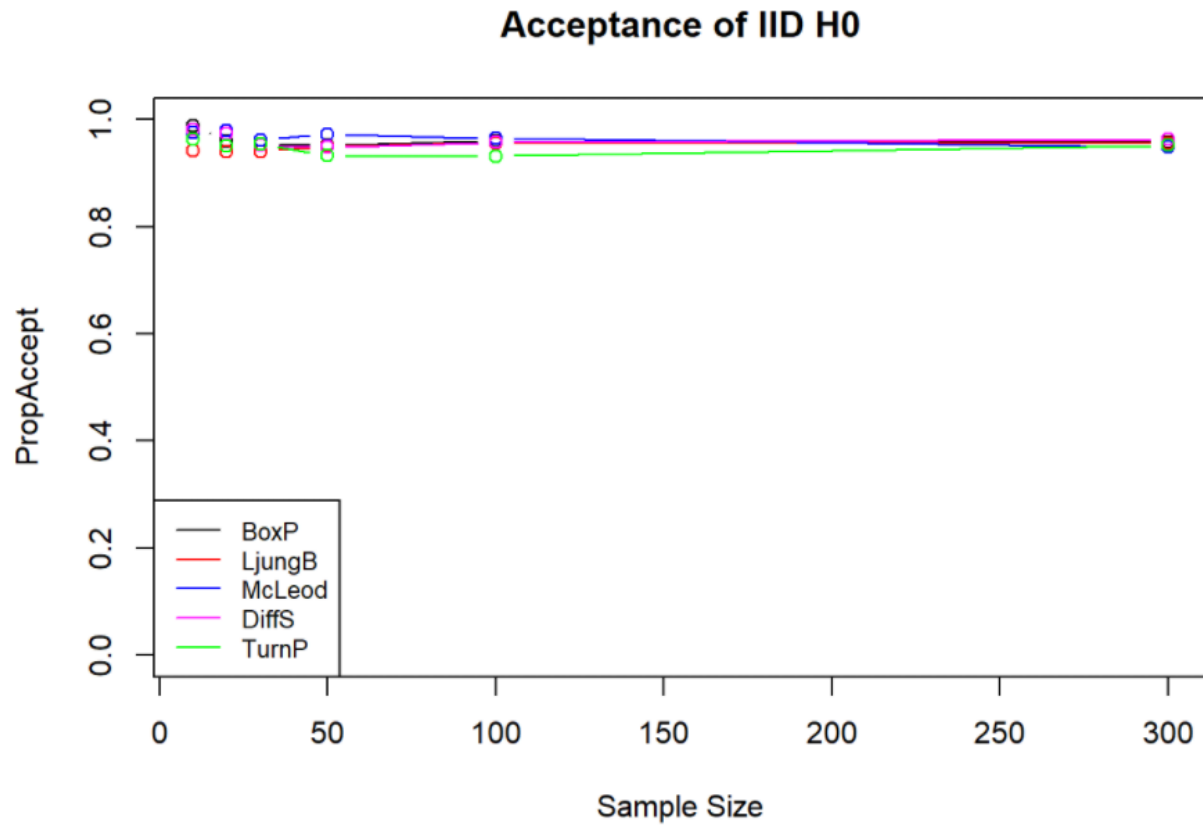
7.1.1 AR(2) models



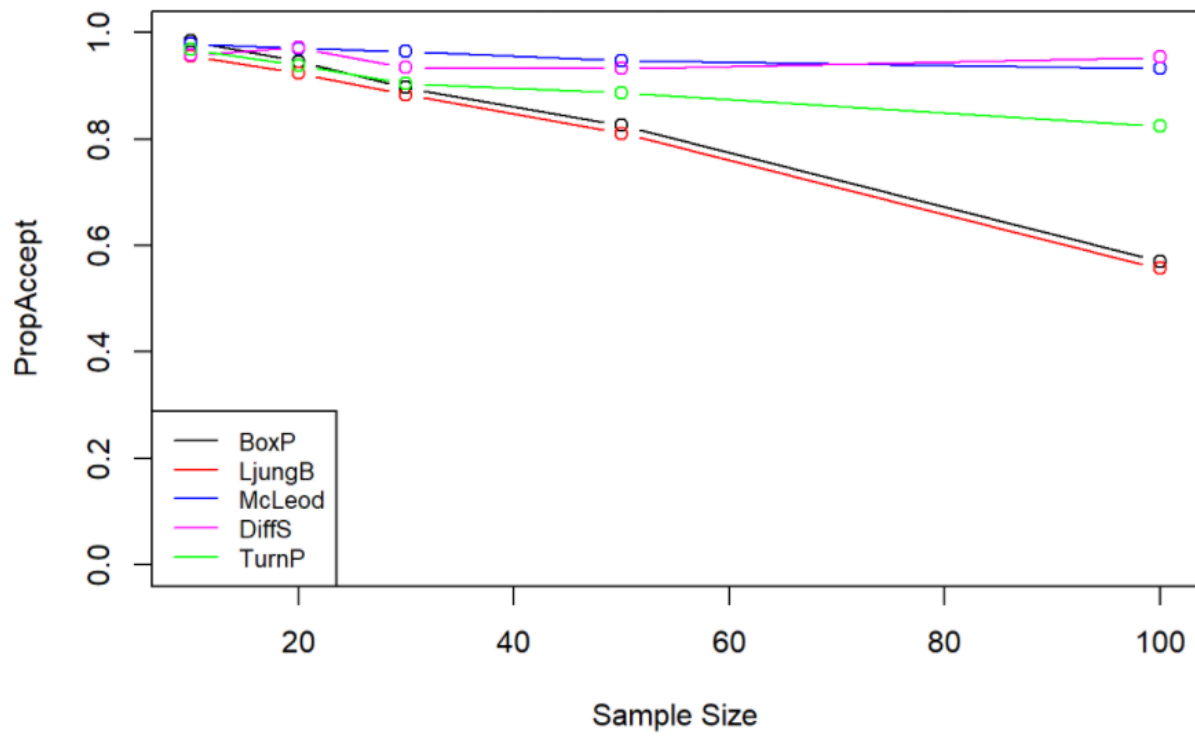
Acceptance of IID H0



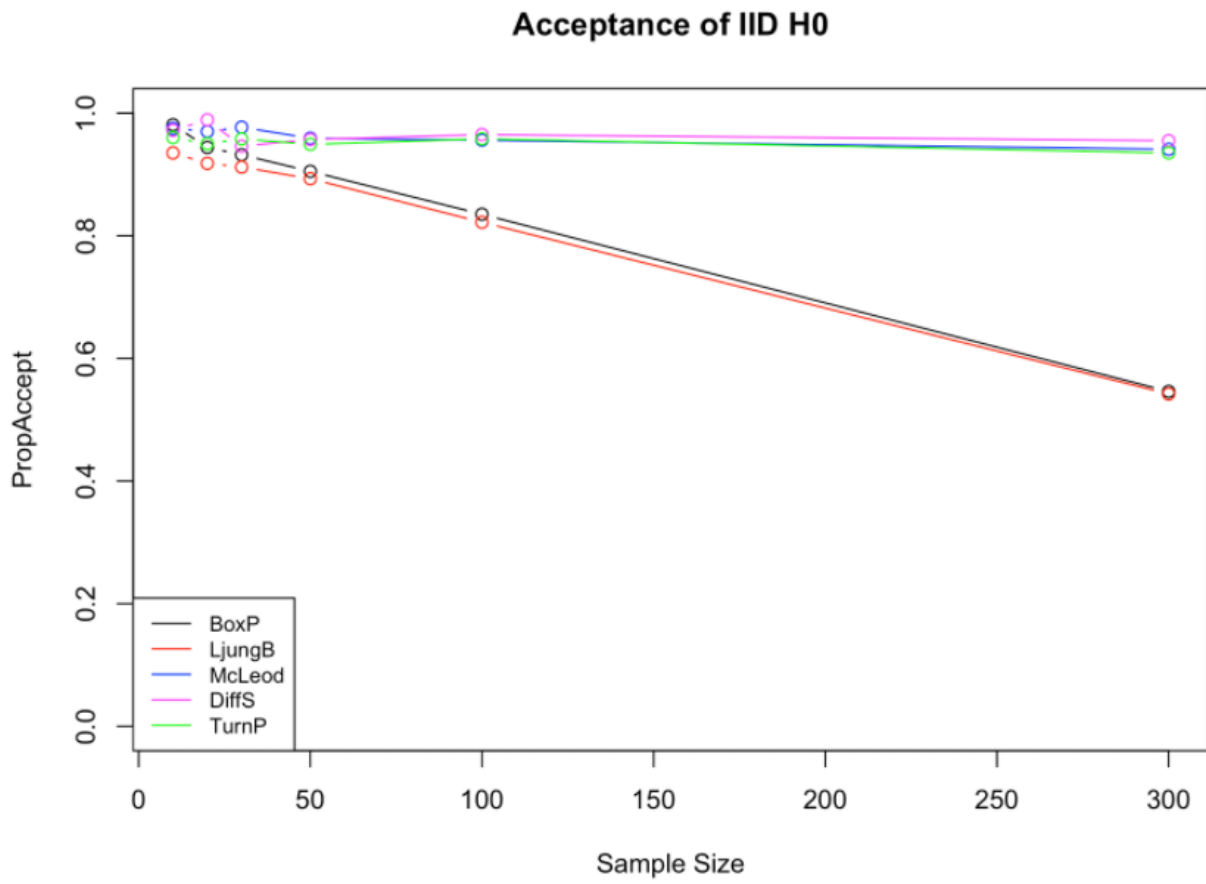
7.1.2 ARMA(1,1) models with different coefficients



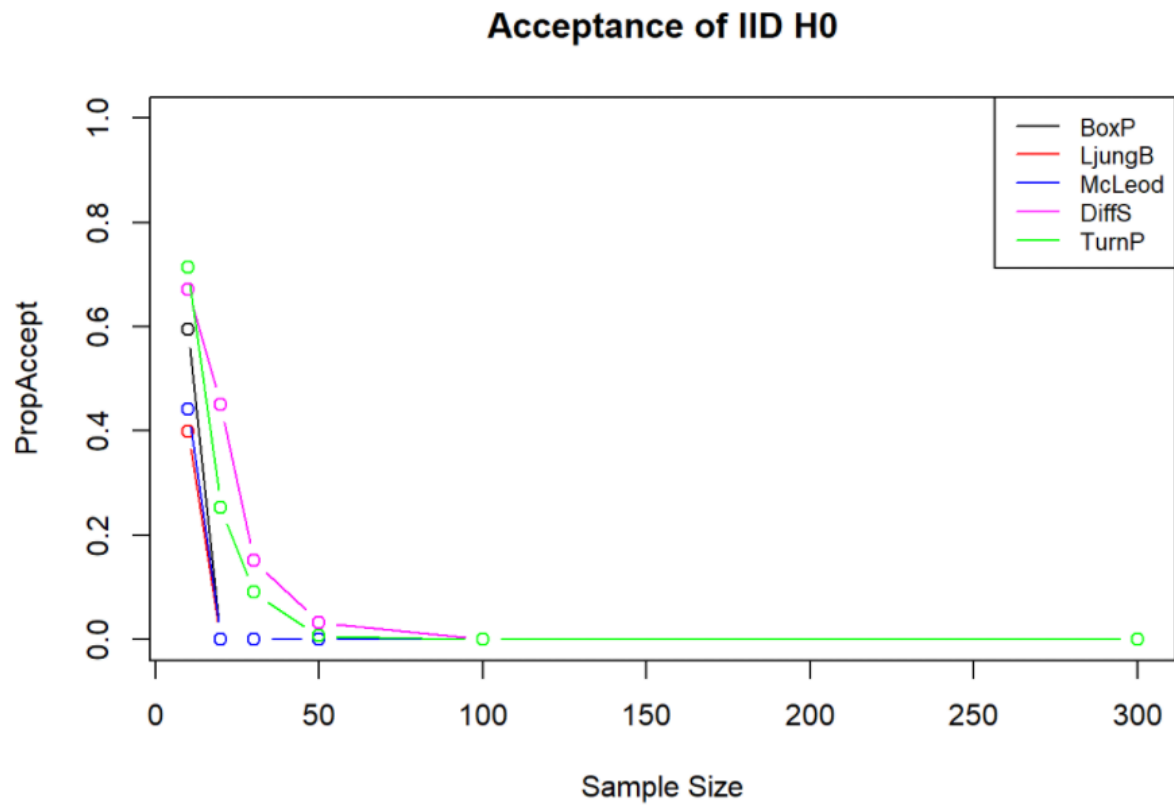
Acceptance of IID H0



7.1.3 ARMA(2,2) model



7.1.4 ARMA(1,1) model and linear trend



7.2 Tables of Acceptance Rates

7.2.1 IID Normal

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.978	0.952	0.972	0.967	0.968
## 3	20	0.970	0.949	0.972	0.983	0.965
## 4	30	0.963	0.951	0.972	0.951	0.953
## 5	50	0.947	0.939	0.958	0.945	0.941
## 6	100	0.950	0.946	0.964	0.964	0.938
## 7	300	0.949	0.949	0.958	0.945	0.954

Figure 9: Acceptance rates for iid Normal Series

7.2.2 AR(1) with $\phi = 0.1$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.992	0.959	0.980	0.958	0.969
## 3	20	0.973	0.958	0.973	0.978	0.947
## 4	30	0.958	0.946	0.970	0.939	0.955
## 5	50	0.922	0.917	0.959	0.944	0.941
## 6	100	0.862	0.853	0.947	0.966	0.909
## 7	300	0.627	0.624	0.944	0.955	0.882

Figure 10: Acceptance rates for AR1 with $\phi = 0.1$

7.2.3 AR(1) with $\phi = 0.5$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.967	0.902	0.971	0.949	0.945
## 3	20	0.633	0.580	0.910	0.965	0.842
## 4	30	0.388	0.346	0.813	0.933	0.745
## 5	50	0.110	0.101	0.711	0.942	0.671
## 6	100	0.002	0.001	0.469	0.962	0.453
## 7	300	0.000	0.000	0.055	0.950	0.074

Figure 11: Acceptance rates for AR1 with $\phi = 0.5$

7.2.4 AR(1) with $\phi = 0.9$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.807	0.668	0.810	0.876	0.887
## 3	20	0.153	0.123	0.370	0.932	0.657
## 4	30	0.013	0.009	0.128	0.845	0.474
## 5	50	0.001	0.001	0.024	0.895	0.302
## 6	100	0.000	0.000	0.000	0.937	0.082
## 7	300	0.000	0.000	0.000	0.941	0.000

Figure 12: Acceptance rates for AR1 with $\phi = 0.9$

7.2.5 AR(2) with $\phi_1 = 0.3, \phi_2 = -0.1$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.989	0.956	0.973	0.959	0.966
## 3	20	0.911	0.877	0.962	0.973	0.880
## 4	30	0.818	0.779	0.957	0.943	0.859
## 5	50	0.612	0.591	0.934	0.951	0.754
## 6	100	0.256	0.240	0.888	0.958	0.606
## 7	300	0.002	0.002	0.761	0.950	0.187

Figure 13: Acceptance rates for AR2 with $\phi = (0.3, -0.1)$

7.2.6 AR(2) with $\phi_1 = 0.758, \phi_2 = -0.758$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.989	0.928	0.969	0.963	0.860
## 3	20	0.704	0.549	0.974	0.987	0.228
## 4	30	0.251	0.175	0.960	0.945	0.058
## 5	50	0.011	0.009	0.896	0.957	0.004
## 6	100	0.000	0.000	0.744	0.969	0.000
## 7	300	0.000	0.000	0.153	0.961	0.000

Figure 14: Acceptance rates for AR2 with $\phi = (0.758, -0.758)$

7.2.7 ARMA(1,1) with $\phi = 0.01, \theta = 0.01$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.988	0.942	0.976	0.981	0.964
## 3	20	0.959	0.940	0.980	0.974	0.950
## 4	30	0.952	0.941	0.962	0.952	0.955
## 5	50	0.952	0.950	0.973	0.949	0.933
## 6	100	0.960	0.957	0.965	0.958	0.932
## 7	300	0.958	0.957	0.949	0.964	0.952

Figure 15: Acceptance rates for ARMA(1,1) with $\phi = 0.01, \theta = 0.01$

7.2.8 ARMA(1,1) with $\phi = 0.1, \theta = 0.1$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.984	0.957	0.979	0.958	0.968
## 3	20	0.946	0.924	0.971	0.972	0.938
## 4	30	0.897	0.883	0.965	0.934	0.905
## 5	50	0.826	0.811	0.947	0.933	0.887
## 6	100	0.570	0.557	0.933	0.954	0.824

Figure 16: Acceptance rates for ARMA(1,1) with $\phi = 0.1, \theta = 0.1$

7.2.9 ARMA(1,1) with $\phi = 0.5, \theta = 0.5$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.823	0.678	0.927	0.924	0.847
## 3	20	0.205	0.150	0.625	0.953	0.435
## 4	30	0.023	0.021	0.447	0.921	0.238
## 5	50	0.002	0.001	0.183	0.928	0.067
## 6	100	0.000	0.000	0.009	0.952	0.004
## 7	300	0.000	0.000	0.000	0.952	0.000

Figure 17: Acceptance rates for ARMA(1,1) with $\phi = 0.5, \theta = 0.5$

7.2.10 ARMA(2,2) with $\phi_1 = \theta_1 = -0.057, \phi_2 = \theta_2 = -0.029$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.981	0.954	0.971	0.980	0.977
## 3	20	0.932	0.905	0.972	0.987	0.964
## 4	30	0.939	0.927	0.979	0.957	0.972
## 5	50	0.884	0.872	0.959	0.953	0.943
## 6	100	0.832	0.821	0.955	0.970	0.948
## 7	300	0.526	0.522	0.945	0.950	0.923

Figure 18: Acceptance rates for ARMA(2,2) with $\phi = (-0.057, -0.029), \theta = (-0.057, -0.029)$

7.2.11 AR(1) with a linear trend, $\phi = 0.5$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.702	0.481	0.491	0.689	0.872
## 3	20	0.000	0.000	0.000	0.469	0.608
## 4	30	0.000	0.000	0.000	0.142	0.456
## 5	50	0.000	0.000	0.000	0.026	0.245
## 6	100	0.000	0.000	0.000	0.000	0.052
## 7	300	0.000	0.000	0.000	0.000	0.000

Figure 19: Acceptance rates for AR(1) with a linear trend

7.2.12 ARMA(1,1) with a linear trend, $\phi = 0.5$, $\theta = 0.5$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.594	0.399	0.442	0.672	0.715
## 3	20	0.000	0.000	0.000	0.451	0.253
## 4	30	0.000	0.000	0.000	0.152	0.091
## 5	50	0.000	0.000	0.000	0.033	0.007
## 6	100	0.000	0.000	0.000	0.000	0.000
## 7	300	0.000	0.000	0.000	0.000	0.000

Figure 20: Acceptance rates for ARMA(1,1) with a linear trend

7.2.13 AR(1) with seasonality, $\phi = 0.5$

##	N	BoxP	LjungB	McLeod	DiffS	TurnP
## 2	10	0.776	0.638	0.796	0.856	0.865
## 3	20	0.120	0.095	0.336	0.917	0.618
## 4	30	0.015	0.014	0.129	0.860	0.471
## 5	50	0.000	0.000	0.016	0.889	0.285
## 6	100	0.000	0.000	0.000	0.932	0.059
## 7	300	0.000	0.000	0.000	0.942	0.000

Figure 21: Acceptance rates for AR(1) with seasonality

7.3 R codes

7.3.1 Simulation: iid Normal

7.3.2 Simulation: AR(1) series, $\phi=0.1$

```
ARMATest<-function(N,a){
  accept=as.data.frame(matrix(NA,ncol=5,nrow=1000))

  Result=as.data.frame(matrix(NA,ncol=5))
  for (j in N){
    for (i in 1:1000){
      Nseries<-sim(a,j)
      BP<-ifelse(Box.test(Nseries,lag=1,type=c("Box-Pierce"),fitdf=0)$p.value>0.05,1,0)
      LB<-ifelse(Box.test(Nseries,lag=1,type=c("Ljung-Box"),fitdf=0)$p.value>0.05,1,0)
      ML<-ifelse(McLeod.Li.test(y=Nseries,gof.lag=1,plot=FALSE)$p.value>0.05,1,0)
      DS<-ifelse(difference.sign.test(Nseries)$p.value>0.05,1,0)
      TP<-ifelse(turning.point.test(Nseries)$p.value>0.05,1,0)

      accept[i,]<-cbind(BP, LB, ML, DS, TP)

    }
    result<-apply(accept,2,sum)/1000
    Result<-rbind(Result,result)
  }
  colnames(Result)<-c("BoxP", "LjungB", "McLeod", "DiffS", "TurnP")
  output<-cbind(N, Result[-1,])

  with(output,plot(N,BoxP,"b",ylim=c(0,1),ylab="PropAccept",main="Acceptance of IID H0",xlab="Sample Size",
    points(output$N,output$LjungB,"b",col="red")
    points(output$N,output$McLeod,"b",col="blue")
    points(output$N,output$DiffS,"b",col="magenta")
    points(output$N,output$TurnP,"b",col="green")
    legend('bottomleft', legend=c("BoxP", "LjungB", "McLeod", "DiffS", "TurnP"),col=c("black", "red", "blue", "magenta"))

  print(output)

}
AR1_1<-specify(ar=0.1,ma=0,sigma2=1)
TestAR1_1<-ARMATest(c(10,20,30,50,100,300),AR1_1)
```

7.3.3 Simulation: AR(1) series, $\phi=0.5$

```
AR1_2<-specify(ar=0.5,ma=0,sigma2=1)
TestAR1_2<-ARMATest(c(10,20,30,50,100,300),AR1_2)
s1<-sim(AR1_2,300)
plotc(s1)
```

7.3.4 Simulation: AR(1) series, $\phi=0.9$

```
AR1_3<-specify(ar=0.9,ma=0,sigma2=1)
TestAR1_3<-ARMAtest(c(10,20,30,50,100,300),AR1_3)
```

7.3.5 Simulation: AR(2) series, roots= $(-2,5)$, $\phi=(0.3,-0.1)$

```
AR2_2<-specify(ar=c(0.3,-0.1),ma=0,sigma2=1)
TestAR2_2<-ARMAtest(c(10,20,30,50,100,300),AR2_2)
```

\subsubsection{ Simulation: AR(2) series, roots $(-1.1,-1.2)$, $\phi=(0.758,-0.758)$ |

```
AR2_1<-specify(ar=c(0.758,-0.758),ma=0,sigma2=1)
TestAR2_1<-ARMAtest(c(10,20,30,50,100,300),AR2_1)
```

7.3.6 Simulation: ARMA(1,1) series, $\phi=0.01$, $\theta=0.01$

```
ARMA1_1<-specify(ar=0.01,ma=0.01,sigma2=1)
TestARMA1_1<-ARMAtest(c(10,20,30,50,100,300),ARMA1_1)
```

7.3.7 Simulation: ARMA(1,1) series, $\phi=0.1$, $\theta=0.1$

```
ARMA1_2<-specify(ar=0.1,ma=0.1,sigma2=1)
TestARMA1_2<-ARMAtest(c(10,20,30,50,100),ARMA1_2)
```

7.3.8 Simulation: ARMA(1,1) series, $\phi=0.5$, $\theta=0.5$

```
ARMA1_3<-specify(ar=0.5,ma=0.5,sigma2=1)
TestARMA1_3<-ARMAtest(c(10,20,30,50,100,300),ARMA1_3)
s2<-sim(ARMA1_3,300)
plotc(s2)
```

7.3.9 Simulation: ARMA(2,2) series, roots= $(5,-7)$, $\phi=\theta=(-0.057,-0.029)$

```
ARMA2<-specify(ar=c(-2/35,-1/35),ma=c(-2/35,-1/35),sigma2=1)
TestARMA2<-ARMAtest(c(10,20,30,50,100,300),ARMA2)
```

7.3.10 Simulation: AR(1) model with linear trend

```
TrendTest<-function(N,a,m,n){
  accept=as.data.frame(matrix(NA,ncol=5,nrow=1000))

  Result=as.data.frame(matrix(NA,ncol=5))
  for (j in N){
    for (i in 1:1000){
```

```

index<-1:j
Nseries<-sim(a,j)+m*index+n
BP<-ifelse(Box.test(Nseries,lag=1,type=c("Box-Pierce"),fitdf=0)$p.value>0.05,1,0)
LB<-ifelse(Box.test(Nseries,lag=1,type=c("Ljung-Box"),fitdf=0)$p.value>0.05,1,0)
ML<-ifelse(McLeod.Li.test(y=Nseries,gof.lag=1,plot=FALSE)$p.value>0.05,1,0)
DS<-ifelse(difference.sign.test(Nseries)$p.value>0.05,1,0)
TP<-ifelse(turning.point.test(Nseries)$p.value>0.05,1,0)

accept[i,]<-cbind(BP,LB,ML,DS,TP)

}
result<-apply(accept,2,sum)/1000
Result<-rbind(Result,result)
}
colnames(Result)<-c("BoxP","LjungB","McLeod","DiffS","TurnP")
output<-cbind(N,Result[-1,])

with(output,plot(N,BoxP,"b",ylim=c(0,1),ylab="PropAccept",main="Acceptance of IID H0",xlab="Sample Size",
points(output$N,output$LjungB,"b",col="red")
points(output$N,output$McLeod,"b",col="blue")
points(output$N,output$DiffS,"b",col="magenta")
points(output$N,output$TurnP,"b",col="green")
legend('topright', legend=c("BoxP","LjungB","McLeod","DiffS","TurnP"),col=c("black","red","blue","magenta"))

print(output)

}

TestAR_trend1<-TrendTest(c(10,20,30,50,100,300),AR1_2,.5,2)
index1<-1:300
s3<-sim(AR1_2,300)+0.5*index1+2
plotc(s3)

```

7.3.11 Simulation: ARMA(1,1) model with linear trend

```

TestAR_trend2<-TrendTest(c(10,20,30,50,100,300),ARMA1_3,.5,2)
index<-1:300
s4<-sim(ARMA1_3,300)+0.5*index1+2
plotc(s4)

```

7.3.12 Simulation: AR(1) model with annual seasonality

```

SeasonTest<-function(N,a,d,k){

  accept=as.data.frame(matrix(NA,ncol=5,nrow=1000))

  Result=as.data.frame(matrix(NA,ncol=5))
  for (j in N){
    for (i in 1:1000){
      index<-1:j
      Nseries<-sim(a,j)+k*sin(2*pi*index/d)
    }
  }
}

```



```

BP<-ifelse(Box.test(Nseries,lag=1,type=c("Box-Pierce"),fitdf=0)$p.value>0.05,1,0)
LB<-ifelse(Box.test(Nseries,lag=1,type=c("Ljung-Box"),fitdf=0)$p.value>0.05,1,0)
ML<-ifelse(McLeod.Li.test(y=Nseries,gof.lag=1,plot=FALSE)$p.value>0.05,1,0)
DS<-ifelse(difference.sign.test(Nseries)$p.value>0.05,1,0)
TP<-ifelse(turning.point.test(Nseries)$p.value>0.05,1,0)

accept[i,]<-cbind(BP, LB, ML, DS, TP)

}
result<-apply(accept,2,sum)/1000
Result<-rbind(Result,result)
}
colnames(Result)<-c("BoxP", "LjungB", "McLeod", 'DiffS', "TurnP")
output<-cbind(N,Result[-1,])

with(output,plot(N,BoxP,"b",ylim=c(0,1),ylab="PropAccept",main="Acceptance of IID H0",xlab="Sample Size",
points(output$N,output$LjungB,"b",col="red")
points(output$N,output$McLeod,"b",col="blue")
points(output$N,output$DiffS,"b",col="magenta")
points(output$N,output$TurnP,"b",col="green")
legend('right', legend=c("BoxP", "LjungB", "McLeod", "DiffS", "TurnP"),col=c("black", "red", "blue", "magenta"))

print(output)

}
TestAR_season<-SeasonTest(c(10,20,30,50,100,300),AR1_3,12,.5)

```