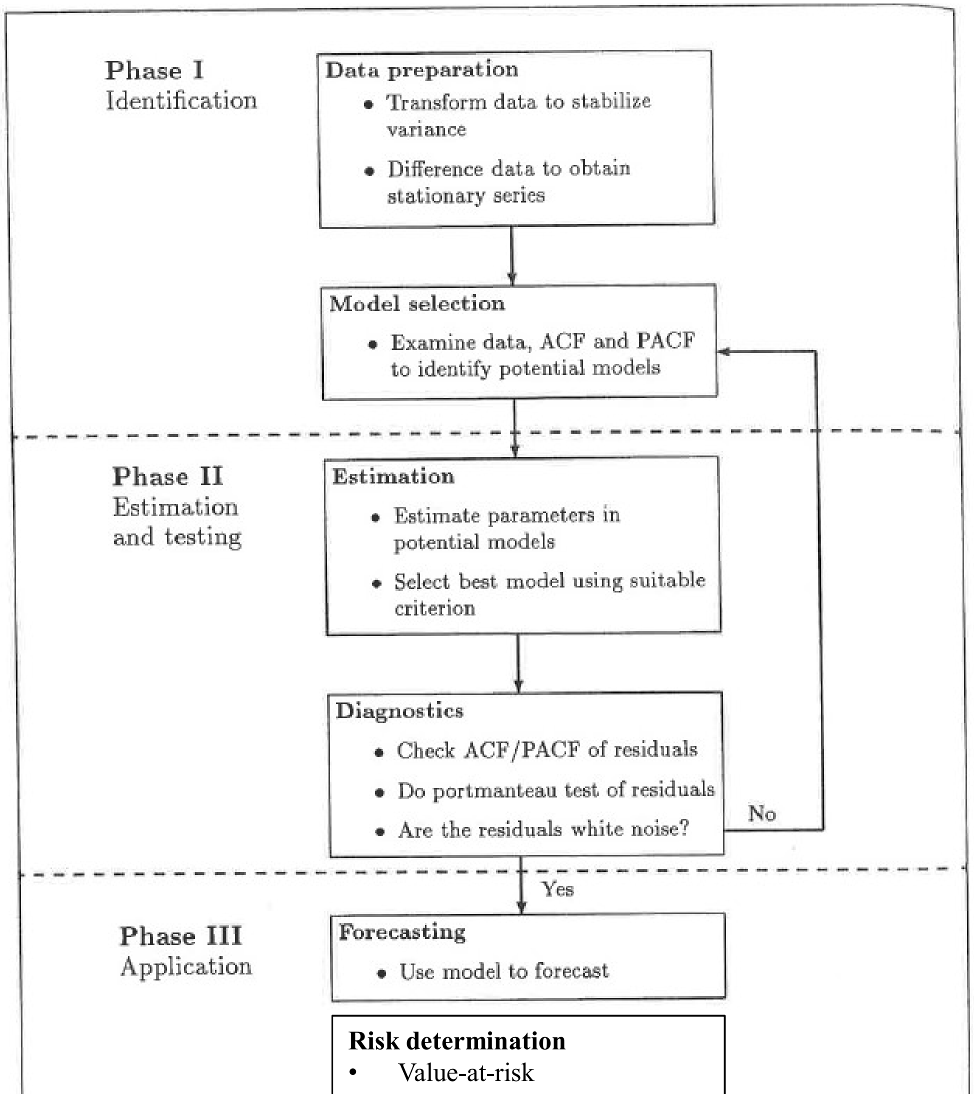
FIN307 MATLAB

FINANICAL TOOLBOX CHAPTER 1 – Financial Data Analysis (PART04)

Box-Jenkins Methodology



Schematic representation of the Box-Jenkins methodology for time series modeling.

Hypothesis test

Method 1

Approach to Hypothesis Testing with Fixed Probability of Type I Error	**1.** State the null and alternative hypotheses. **2.** Choose a fixed significance level α. **3.** Choose an appropriate test statistic and establish the critical region based on α. **4.** Reject H_0 if the computed test statistic is in the critical region. Otherwise, do not reject. **5.** Draw scientific or engineering conclusions.

Method 2

Significance Testing (P-Value Approach)	**1.** State null and alternative hypotheses. **2.** Choose an appropriate test statistic. **3.** Compute the P-value based on the computed value of the test statistic. **4.** Use judgment based on the P-value and knowledge of the scientific system.

MATLAB: Hypothesis Test

MathWorks®

Example:

- multiple regression

```
Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4
```

Build in functions: **fitlm**

Significance Testing (P-Value Approach)	**1.** State null and alternative hypotheses. **2.** Choose an appropriate test statistic. **3.** Compute the P-value based on the computed value of the test statistic. **4.** Use judgment based on the P-value and knowledge of the scientific system.

$$t = \frac{b_j - \beta_{j0}}{s\sqrt{c_{jj}}}$$

*p-value =2*P(t = 2.0827)*

```
Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	62.405	70.071	0.8906	0.39913
x1	1.5511	0.74477	2.0827	0.070822
x2	0.51017	0.72379	0.70486	0.5009
x3	0.10191	0.75471	0.13503	0.89592
x4	-0.14406	0.70905	-0.20317	0.84407

```
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 2.45
R-squared: 0.982,  Adjusted R-Squared 0.974
F-statistic vs. constant model: 111, p-value = 4.76e-07
```

Hypothesis test

Step (1):

H0: coefficient x1 =0

H1: coefficient x1≠ 0

Step (2):

t = 2.0827

Step (3):

P-value = 0.070822

Step (4):

p-value = 0.070822 > 0.05. There is no sufficient evidence to show that the x1≠0.

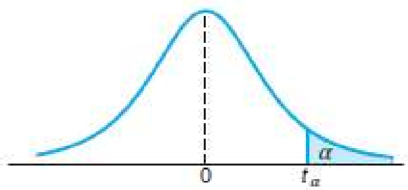
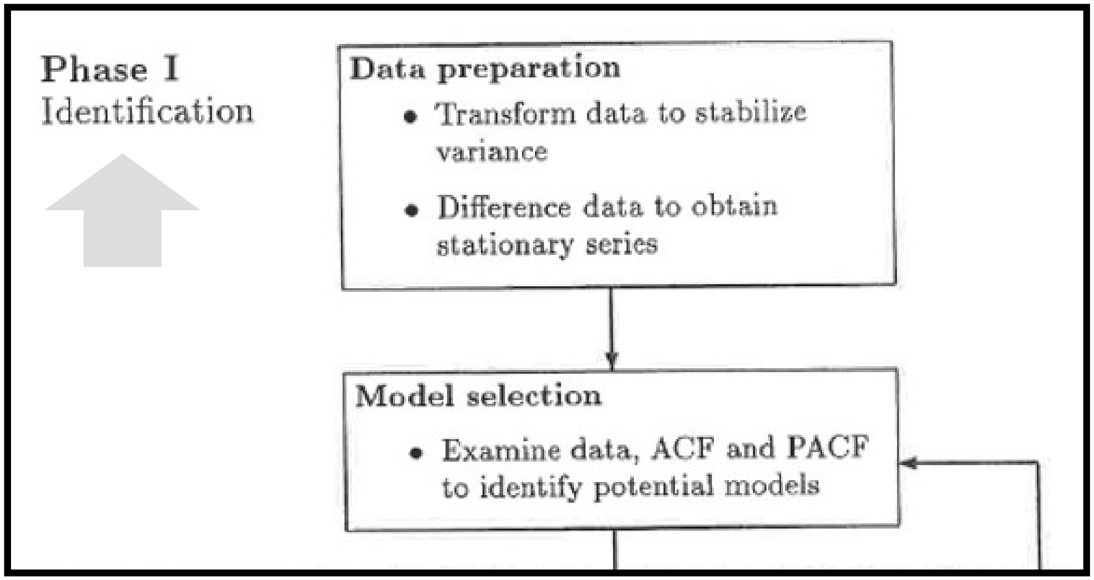


Table A.4 Critical Values of the t-Distribution

	α						
v	**0.40**	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262

*p-value=2*P(t = 2.0827)*
= 0.070822

Box-Jenkins Methodology



Does AR,MA, ARMA suitable?
1. ACF and PACF
2. Serial correlation test

Identify AR,MA, ARMA
1. ACF and PACF

Does ARCH effect exist?
1. ARCH test

Serial Correlation based on ACF and PACF

MathWorks®

Plot the sample autocorrelation function (ACF) and partial autocorrelation function (PACF).

```
figure
subplot(2,1,1)
autocorr(Y)
subplot(2,1,2)
parcorr(Y)
```

Check serial correlation for

✓Stock market

✓FOREX

Serial Correlation based on ACF and PACF

Conditional Mean Model	ACF	PACF
AR(p)	Tails off gradually	Cuts off after p lags
MA(q)	Cuts off after q lags	Tails off gradually
ARMA(p,q)	Tails off gradually	Tails off gradually

Identify the possible ARMA(p,q)?

ARMA(1,0) = AR(1) ARMA(1,1)

ARMA(0,1) = MA(1) ARMA(2,2)

MathWorks®

Return equation:

$$r_t = \mu_t + e_t$$

$$\mu_t = ARMA \qquad error, noise, residual$$

MATLAB: the NYSE index

TEST 1: Ljung-Box Q-Test

Conduct the Ljung-Box Q-test for autocorrelation at lags 5, 10, and 15.

```
[h,p,Qstat,crit] = lbqtest(Y,'Lags',[5,10,15])
```

hypothesis   p-value   Q-statistic   Critical value

Approach to Hypothesis Testing with Fixed Probability of Type I Error

1. State the null and alternative hypotheses.
2. Choose a fixed significance level α.
3. Choose an appropriate test statistic and establish the critical region based on α.
4. Reject H_0 if the computed test statistic is in the critical region. Otherwise, do not reject.
5. Draw scientific or engineering conclusions.

Ljung-Box Q-test based on AutoCorrelation Function (ACF): $r_t = \mu_t + e_t$

STEP 1:

$$\text{null hypothesis } H_o : \rho_1 = \cdots = \rho_m = 0$$

$$\text{alternative hypothesis } H_a : \rho_i \neq 0 \text{ for some } i \in \{1, \ldots, m\}$$

$$\rho(i) = autocorrelation\ at\ lag - i$$

Ljung-Box Q-Test

Simulation studies suggest that the choice of $m \approx \ln(T)$

Eg. Data size, T=500

m = ln(500) = 6.21 = 6 or 7

STEP 2: Assume for degree of freedom, *v*=6, 5% level of significance

Table A.5 Chi-Squared Distribution Probability Table 739

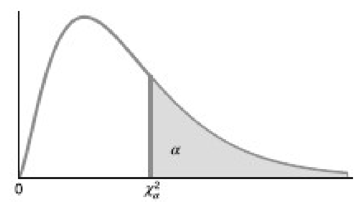
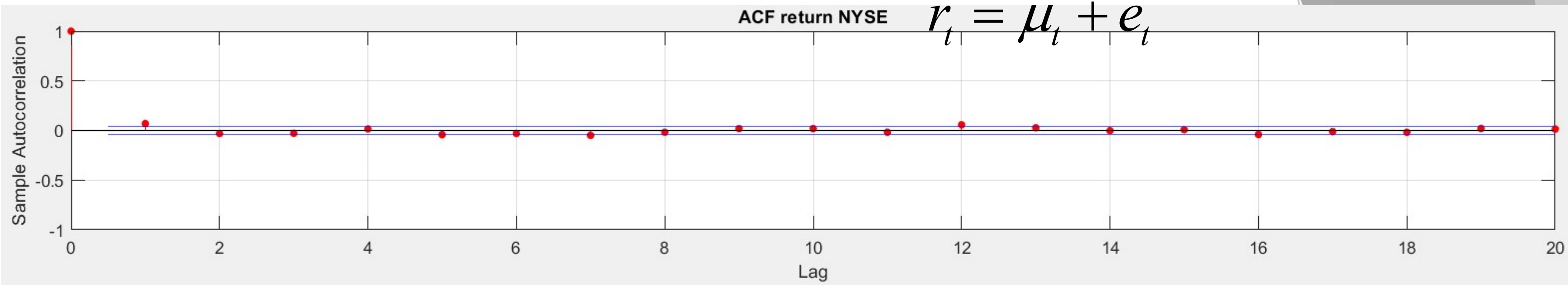


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.06[illegible]	16.013	16.622	18.475	20.278	24.321

$$critial\ \chi^2_{df=6} = 12.592$$



$r_t = \mu_t + e_t$

autocorrNYSE =

lag0	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8
1.0000	0.0696	-0.0300	-0.0283	0.0159	-0.0424	-0.0299	-0.0487	-0.0175

STEP 3:

$$Q(m) = N(N+2)\sum_{h=1}^{m} \frac{\rho_h^2}{N-h}$$

$$Q(6) = 3027(3027+2)$$

$$= 28.7908$$

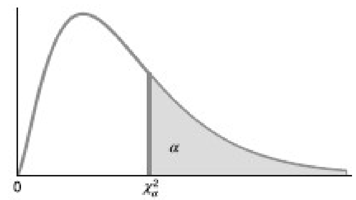


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.30**	**0.25**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.06[illegible]	16.013	16.622	18.475	20.278	24.321

$$critial\ \chi^2_{df=6} = 12.592$$

STEP 4:

$$Q(6) = 28.7908 > 12.592$$

alternative hypothesis $H_a : \rho_i \neq 0$ for some $i \in \{1, \dots, m\}$

STEP 5: There is sufficient evidence that the NYSE return series has serial correlation at 5% level of significance.

Ljung-Box Q-Test

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are useful qualitative tools to assess the presence of autocorrelation at individual lags. The Ljung-Box Q-test is a more quantitative way to test for autocorrelation at multiple lags *jointly* [1]. The null hypothesis for this test is that the first *m* autocorrelations are jointly zero,

Step 1

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0.$$

The choice of *m* affects test performance. If *N* is the length of your observed time series, choosing $m \approx \ln(N)$ is recommended for power [2]. You can test at multiple values of *m*. If seasonal autocorrelation is possible, you might consider testing at larger values of *m*, such as 10 or 15.

The Ljung-Box test statistic is given by

Step 2

$$Q(m) = N(N+2)\sum_{h=1}^{m} \frac{\rho_h^2}{N-h}.$$

This is a modification of the Box-Pierce Portmanteau "Q" statistic [3]. Under the null hypothesis, Q(*m*) follows a χ_m^2 distribution.

Step 3

You can use the Ljung-Box Q-test to assess autocorrelation in any series with a constant mean. This includes residual series, which can be tested for autocorrelation during model diagnostic checks. If the residuals result from fitting a model with g parameters, you should compare the test statistic to a χ^2 distribution with $m - g$ degrees of freedom. Optional input arguments to `lbqtest` let you modify the degrees of freedom of the null distribution.

You can also test for conditional heteroscedasticity by conducting a Ljung-Box Q-test on a squared residual series. An alternative test for conditional heteroscedasticity is Engle's ARCH test (`archtest`).

References

[1] Ljung, G. and G. E. P. Box. "On a Measure of Lack of Fit in Time Series Models." *Biometrika*. Vol. 66, 1978, pp. 67–72.

[2] Tsay, R. S. *Analysis of Financial Time Series*. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc., 2010.

[3] Box, G. E. P. and D. Pierce. "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association*. Vol. 65, 1970, pp. 1509–1526.

TEST 2: Engle's ARCH Test

```
[h,pValue,stat,cValue] = archtest(res)
[h,pValue,stat,cValue] = archtest(res,param1,val1,param2,val2,...)
```

Return equation: $r_t = \mu_t + e_t$

$\mu_t = ARMA$ $\quad$ $error, noise, residual$

$$residual-squared: \quad e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_m e_{t-m}^2 + u_t$$

e_t^2 depends on $e_{t-1}^2, e_{t-2}^2, e_{t-3}^2, \ldots, e_{t-m}^2$

AutoRegressive Conditional Heteroscedastic (ARCH) effect

Step 1

The alternative hypothesis for Engle's ARCH test is autocorrelation in the squared residuals, given by the regression

$$H_a : e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_m e_{t-m}^2 + u_t,$$

where u_t is a white noise error process. The null hypothesis is

$$H_0 : \alpha_0 = \alpha_1 = \ldots = \alpha_m = 0.$$

To conduct Engle's ARCH test using `archtest`, you need to specify the lag m in the alternative hypothesis. One way to choose m is to compare loglikelihood values for different choices of m. You can use the likelihood ratio test (`lratiotest`) or information criteria (`aicbic`) to compare loglikelihood values.

To generalize to a GARCH alternative, note that a GARCH(P,Q) model is locally equivalent to an ARCH($P + Q$) model. This suggests also considering values $m = P + Q$ for reasonable choices of P and Q.

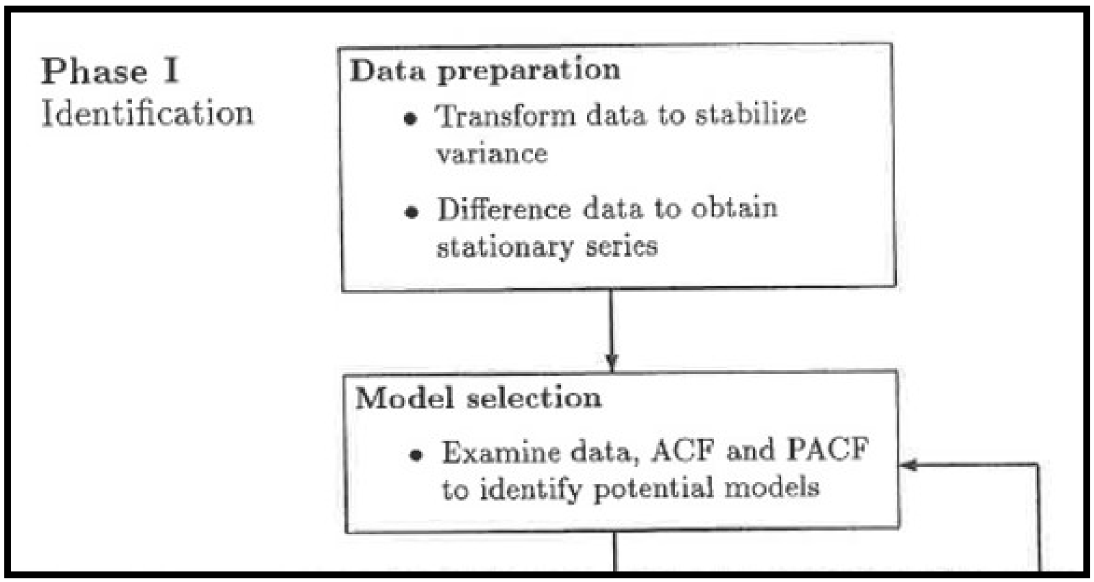
The test statistic for Engle's ARCH test is the usual F statistic for the regression on the squared residuals. Under the null hypothesis, the F statistic follows a χ^2 distribution with m degrees of freedom. A large critical value indicates rejection of the null hypothesis in favor of the alternative.

References

[1] Engle, Robert F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica*. Vol. 50, 1982, pp. 987–1007.

MathWorks®

Box-Jenkins Methodology



Procedures:

Load data

Convert price to return series

Check for Autocorrelation

Test the Significance of Autocorrelations

Test for Significant ARCH Effects


THE
END