10. Model diagnosis and prediction

Outline:

- Introduction
- Autocorrelation tests
- Zero mean, homoscedasticity and normality tests
- Model stability test
- Predictions

Recommended readings:

▷ Chapter 11 of D. Peña (2008).

▷ Chapter 5 of P.J. Brockwell and R.A. Davis (1996).

▷ Chapter 4 of J.D. Hamilton (1994).

Introduction

▷ The diagnosis of the model requires confirming that basic hypotheses made with respect to the residuals are true:

1. Marginal mean equal to zero.
2. Constant marginal variance.
3. No correlation for any lag.
4. Normal distribution.

▷ Moreover, these properties must be verified not only with respect to the marginal distributions but also to the distributions conditional on any set of information of past values in the series.

▷ For example, for the mean the condition is:

$$E(a_t|z_{t-1}, ..., z_1) = E(a_t|a_{t-1}, ..., a_1) = 0.$$

which is much stronger than the condition of zero marginal mean.

Introduction

▷ We also assume that:

$$Var(a_t|z_{t-1}, ..., z_1) = Var(a_t|a_{t-1}, ..., a_1) = \sigma^2$$

which generalizes the condition of constant marginal variance for any conditional distribution.

▷ Of the four conditions established for marginal distributions

- Condition (1) is not very restrictive. It is possible for a model to be very incorrect, and yet (1) is verified.
- The condition (2) of marginal variance is stronger.
- Condition (3), lack of correlation for any lag, is central to ensuring that the model is suitable.
- Finally, the condition of normality is useful, because it guarantees us that the non-correlation implies independence, and that we are not leaving information to be modelled.

Introduction

▷ The diagnosis is related, but not identical, to the model selection studied in the previous section:

- It is possible that the best model selected within a class leads to residuals that do not verify the above hypotheses, and thus we will have to search for a new one in a wider class.
- It is also possible to have various models whose residuals do verify the above hypotheses, and we can then select the best one using a selection criterion.
- We see, therefore, that the diagnosis of the model is a complementary step to selecting the best model from within a class using a selection criterion.

▷ Finally, in this section we will study the calculation of predictions using an ARIMA model when the parameters have been estimated in a sample and there is some uncertainty as well regarding the correct model followed by the data.

Autocorrelation tests

▷ The first test to run is to see whether the estimated residuals are uncorrelated. To do this, we calculate their *ACF* by means of:

$$\widehat{r}_k = \frac{\sum_{t=1}^{T-k} (\widehat{a}_t - \overline{a})(\widehat{a}_{t+k} - \overline{a})}{\sum_{t=1}^{T} (\widehat{a}_t - \overline{a})^2} \tag{206}$$

where $\overline{a}$ is the mean of the T residuals.

▷ If the residuals are independent, the coefficients, $\widehat{r}_k$, for a k which is not very small, are approximately random variables with zero mean, asymptotic variance $1/T$ and normal distribution.

▷ The asymptotic variance is valid for large k, but not for the first lags.

Autocorrelation tests

▷ For example, it can be proved that if the series follows an AR(1) process the asymptotic standard deviation of the first order autocorrelation for the residuals, $\widehat{r}_1$, is $\sqrt{(1-\phi^2)/T}$, which may be much less than $1/\sqrt{T}$.

▷ As a result, the value $1/\sqrt{T}$ must be considered as a maximum limit of the standard deviation of the residual autocorrelations.

▷ The usual procedure is to plot two parallel lines at a distance of $2/\sqrt{T}$ from the origin in its autocorrelation functions or partial autocorrelation functions, and check to see whether all the coefficients $\widehat{r}_k$ are within the confidence limits.

▷ Since these intervals are, approximately, 95%, on average one out of every twenty estimated autocorrelation coefficients will lie outside, thus the appearance of a significant value in a high lag is to be expected.

▷ Nevertheless, since according to the above these limits overestimate the variance in small lags, a value close to the confidence limits $\pm 2/\sqrt{T}$ in the initial lags should be considered a clear indication that the model is unsuitable.

Autocorrelation tests

The Ljung-Box test

▷ A global test that the first h coefficients are zero (h must be large) is the Ljung-Box test.

▷ If the residuals are really white noise, then the estimated correlation coefficients are asymptotically normal, with zero mean and variance $(T-k)/T(T+2)$.

▷ Therefore, the statistic

$$Q(h) = T(T+2)\sum_{j=1}^{h}\frac{\widehat{r}_j^2}{T-j} \tag{207}$$

is distributed, asymptotically, as a χ^2 with degrees of freedom equal to the number of coefficients in the sum (h) minus the number of estimated parameters, n.

The Ljung-Box autocorrelation tests

▷ For non-seasonal models $n = p + q + 1$, or $n = p + q$, according to whether the model has a constant or not.

▷ For seasonal models, which usually do not have a constant, $n = P+p+Q+q$.

▷ We will conclude that the model is unsuitable if the value of $Q(h)$ obtained using (207) is greater than the 0.95 percentile of the χ^2 distribution with $h-n$ degrees of freedom, which we will denote by $\chi^2_{.95}(h-n)$.

▷ In general, we reject the hypothesis of non-correlation of the residuals when the probability:

$$\Pr\left(\chi^2(h-n) > Q(h)\right)$$

is small (less than 0.05 or 0.01).

Autocorrelation tests

The determinant test

▷ The Ljung-Box test has the drawback of giving, approximately, the same weight to all the autocorrelation coefficients and being invariant to permutations of these coefficients.

▷ Nevertheless, intuitively we should give more weight to the low order coefficients than to high. Peña and Rodríguez (2003) have proposed a more powerful test than that of Ljung-Box which has this property.

▷ The test is based on the autocorrelation matrix of the residuals:

$$\mathbf{R}_m = \begin{bmatrix} 1 & \widehat{r}_1 & ... & \widehat{r}_{m-1} \\ \widehat{r}_1 & 1 & ... & \widehat{r}_{m-2} \\ ... & ... & 1 & ... \\ \widehat{r}_{m-1} & \widehat{r}_{m-2} & ... & 1 \end{bmatrix}$$

The determinant autocorrelation test

▷ The statistic of the test is:

$$D_m = -\frac{T}{m+1}\log|\hat{\mathbf{R}}_m|,$$

which follows, asymptotically, a gamma distribution with parameters

$$\alpha = \frac{3(m+1)\left\{m-2\left(p+q\right)\right\}^2}{2\left\{2m\left(2m+1\right)-12(m+1)\left(p+q\right)\right\}}$$

$$\beta = \frac{3\left(m+1\right)\left\{m-2\left(p+q\right)\right\}}{2m\left(2m+1\right)-12\left(m+1\right)\left(p+q\right)}.$$

▷ The distribution has mean $\alpha/\beta = (m+1)/2 - (p+q)$ and variance:

$$\alpha/\beta^2 = (m+1)(2m+1)/3m - 2(p+q).$$

▷ The percentiles of D_m are easily obtained by calculating the parameters of the gamma with the above formulas and using the tables of this distribution (many computer programs include this function).

The determinant autocorrelation test

▷ Alternatively, the variable

$$ND_m^* = (\alpha/\beta)^{-1/4}\left(4/\sqrt{\alpha}\right)\left((D_m)^{1/4} - (\alpha/\beta)^{1/4}\left(1 - \frac{1}{6\alpha}\right)\right), \qquad (208)$$

distributed, approximately, as a normal standard variable.

▷ It can be proved that this statistic can be written as:

$$D_m = T\sum_{i=1}^{m}\frac{(m+1-i)}{(m+1)}\hat{\pi}_i^2, \qquad (209)$$

where $\hat{\pi}_i^2$ is the square of the partial correlation coefficient of order i.

▷ This test can be seen as a modified Ljung-Box test, where instead of utilizing the auto correlation coefficients we use the partial AC, but with weightings $\frac{(m+1-i)}{m}$. These weightings decrease linearly with the lag, such that $\hat{\pi}_1^2$ has weight one and $\hat{\pi}_m^2$ weight $1/m$.

Overfitting tests based on BIC criterion

▷ A complementary technique to the above tests is the overfitting technique, which consists in estimating a model of an order higher than the one being analyzed and checking whether significant estimated coefficients are obtained.

▷ With this it is possible to pick up small remaining structures that can improve the predictions, but that might not have been clearly detected in the analysis of residuals.

▷ In general, if we have fitted an ARIMA(p, d, q) that seems suitable, the overfitting is applied estimating the ARIMA $(p+r, d, q)$ and ARIMA $(p, d, q+r)$ models for a low value of r, normally 1 or 2, and checking whether the additional parameters are significant.

▷ It is not advisable to expand the AR and MA parts at the same time, since this might produce a compensation of effects.

Overfitting tests based on BIC criterion

▷ If the model fitted initially is:

$$\phi(B) z_t = \theta(B) a_t \tag{210}$$

we will obtain an equally good fit with:

$$\phi^*(B) z_t = \theta^*(B) a_t \tag{211}$$

with $\phi^*(B) = \phi(B)(1 - \phi B)$ and $\theta^*(B) = \theta(B)(1 - \theta B)$ and ϕ being approximately equal to θ.

▷ Therefore, if the correct model is (210) and we estimate (211) we obtain all the significant parameters and we will only notice the over-parameterizing when factorizing the AR and MA operators.

▷ As a result, it is always advisable to obtain the roots of the AR and MA operators in mixed models and check that there are no factors that cancel each other out on both sides.

Overfitting tests based on BIC criterion

▷ An automatic way of carrying out the overfitting is to adjust AR models up to a $p_{\max}$ order preset to the residuals of the model and select the best AR model by means of the BIC criterion.

- If the best model selected is an AR(0), that is, a white noise, we accept that the residuals are uncorrelated.

- In the opposite case, we reject the model as inadequate and the degree of the best model selected tells us how we should modify the current one.

▷ Since the BIC criterion is consistent, that is for large sample sizes it tends to select the correct model with a probability that tends to zero, this procedure works well in practice if we have large samples.

▷ With smaller samples sizes, it is always recommendable to run a determinant test on the estimated residuals.

Zero mean, homoscedasticity and normality tests

Test for zero mean

▷ The estimated residuals of an ARIMA model are not in principle subject to the restriction $\sum_{t=1}^{T} \widehat{a}_t = 0$.

▷ This condition is only imposed when we have an AR model and its parameters (including a constant) have been estimated by least squares, but in the exact maximum likelihood estimation this restriction does not exist.

▷ To test the hypothesis of zero mean in the general case, assuming T residuals and $p+q$ parameters, we calculate the mean $\overline{a} = T^{-1} \sum_{t=1}^{T} \widehat{a}_t$ and and variance $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (\widehat{a}_t - \overline{a})^2$, and we conclude that $E[\widehat{a}_t] \neq 0$, if

$$\frac{\overline{a}}{\widehat{\sigma}/\sqrt{T}}$$

is significantly large.

▷ This test must be applied after checking that the residuals are uncorrelated, to ensure that $\widehat{\sigma}^2$ is a reasonable estimator of the variance.

Zero mean, homoscedasticity and normality tests

Test for homoscedasticity

▷ The stability of the marginal variance of the residuals is confirmed by studying the graph of the residuals over time.

▷ If in view of the estimated residuals there seems to be a change of variance from a point $t = n_1$ on, we can divide the sample interval into two parts and apply a test of variances.

▷ In the hypothesis under which both sections have the same variance, the statistic

$$F = \frac{\sum_{t=1}^{n_1} \widehat{a}_t^2/n_1}{\sum_{t=n_1+1}^{T} \widehat{a}_t^2/(T-n_1)} = \frac{s_1^2}{s_2^2}$$

will be distributed approximately as an F with n_1 and $T - n_1$ degrees of freedom.

Zero mean, homoscedasticity and normality tests

Test for homoscedasticity

▷ In the same way, if we suspect h changes in variance in the periods $n_1, ..., n_h$ the test of variance equality is

$$\lambda = T \log \widehat{\sigma}^2 - \sum_{i=1}^{h} n_i \log s_i^2$$

where $\widehat{\sigma}^2$ is the variance of the residuals in the entire sample and s_i^2 is the variance in section i of length n_i observations.

▷ Under the hypothesis that the variance is the same in all the sections it is proved that this statistic is, asymptotically, a chi-square with $h - 1$ degrees of freedom.

▷ To apply this test it is advisable to have at least 10 observations in each section.

Zero mean, homoscedasticity and normality tests

Test for normality

▷ The hypothesis that the residuals have a normal distribution is checked using any of the usual tests.

▷ A simple test is to calculate the coefficient of asymmetry

$$\alpha_1 = \frac{\sum(\widehat{a}_t - \overline{a})^3}{\widehat{\sigma}^3}$$

and kurtosis

$$\alpha_2 = \frac{\sum(\widehat{a}_t - \overline{a})^4}{\widehat{\sigma}^4}$$

of the residuals and use the condition that, under the hypothesis of normality, the variable:

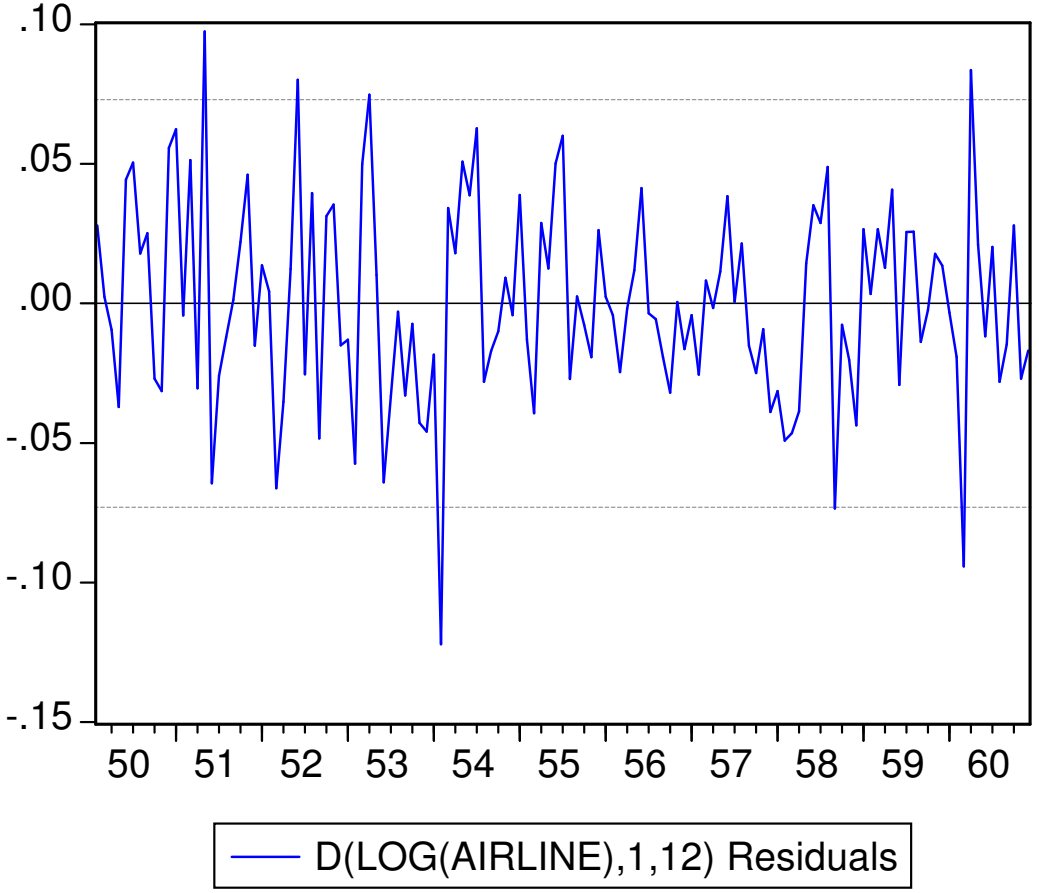
$$X = \frac{T\alpha_1^2}{6} + \frac{T(\alpha_2 - 3)^2}{24}$$

is a χ^2 with two degrees of freedom.

▷ Finally, it is *always* recommendable to study the graph of the estimated residuals $\widehat{a}_t$ over time.

Model diagnosis - Examples

Example 91. *Lets begin with a classical example: The airline time series. The figure gives the graph of the residuals of the series estimated using an ARIMA*$(0,1,1)\times(0,1,1)_{12}$ *model.*



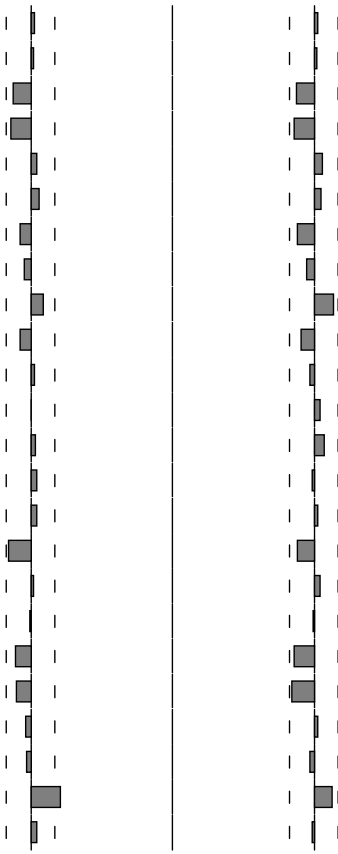
Example 91. *The figure gives* ACF *of the residuals.*

▷ *No coefficient is significant and also Q statistic is not significant for all lags.*

▷ *Thus we conclude that, with this test, we find no evidence of serial dependence in the residuals.*

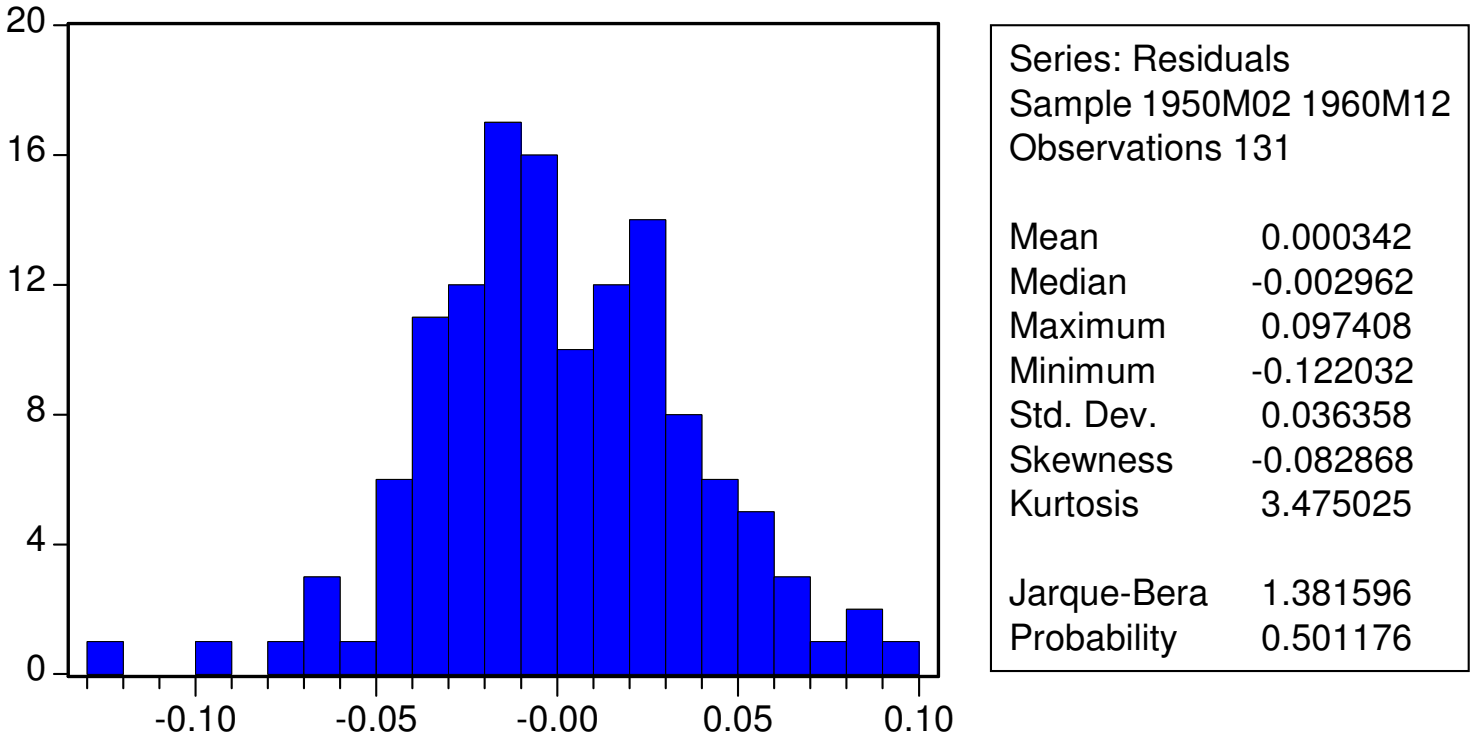
Correlogram of Residuals

Date: 02/20/08 Time: 10:57
Sample: 1950M02 1960M12
Included observations: 131
Q-statistic probabilities adjusted for 2 ARMA term(s)



Autocorrelation	Partial Correlation		Q-Stat	Prob
		1	0.1204	
		2	0.1804	
		3	2.2844	0.131
		4	5.2684	0.072
		5	5.5496	0.136
		6	6.0438	0.196
		7	6.8403	0.233
		8	7.1001	0.312
		9	8.4610	0.294
		10	9.3103	0.317
		11	9.4385	0.398
		12	9.4432	0.491
		13	9.6473	0.562
		14	9.9598	0.619
		15	10.258	0.673
		16	14.120	0.441
		17	14.175	0.512
		18	14.179	0.585
		19	16.012	0.523
		20	17.708	0.475
		21	17.896	0.529
		22	17.993	0.588
		23	25.775	0.215
		24	26.077	0.248

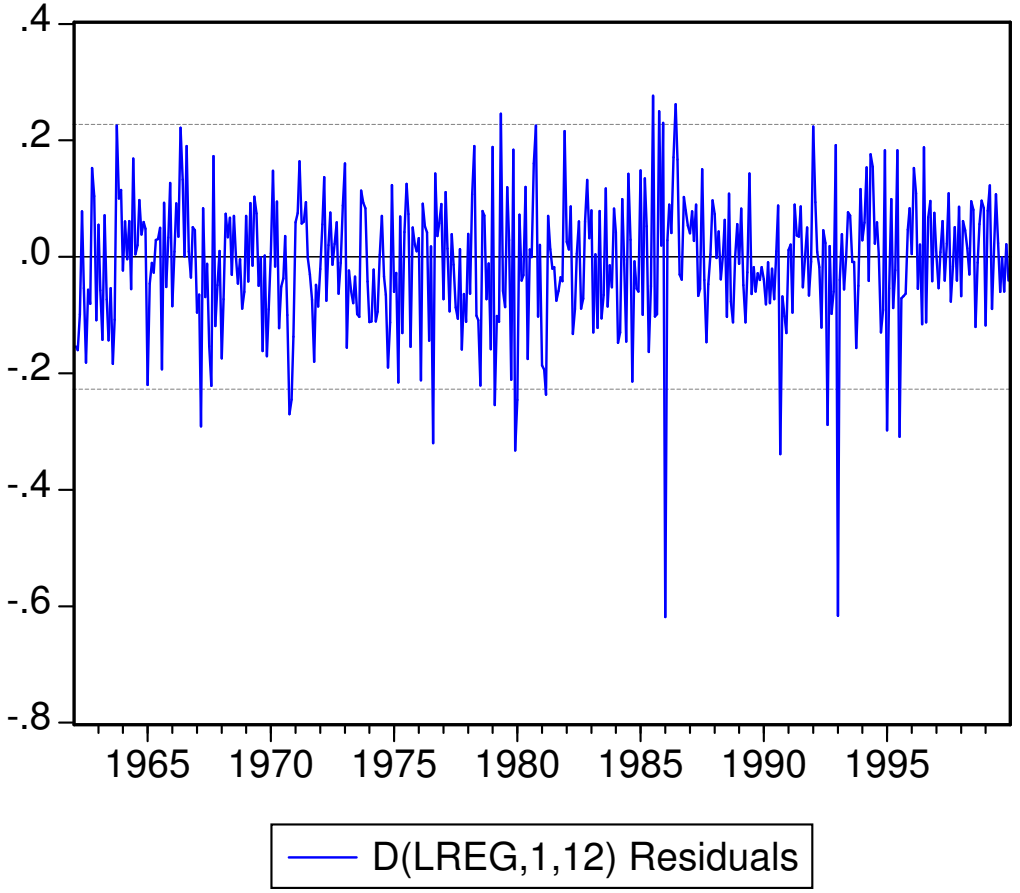
▷ The mean of the residuals is not significantly different from zero and the variability of the residuals, except for one possible outlier, seems constant over time.



▷ Is normally distributed? **Yes**

Model diagnosis - Examples

Example 92. *The figure gives the graph of the residuals of the vehicle registration series estimated using an ARIMA*$(0,1,1)\times(1,1,1)_{12}$ *model. Some noticeable outliers are observed.*



Example 92. *The figure gives* ACF *of the residuals.*

Correlogram of Residuals

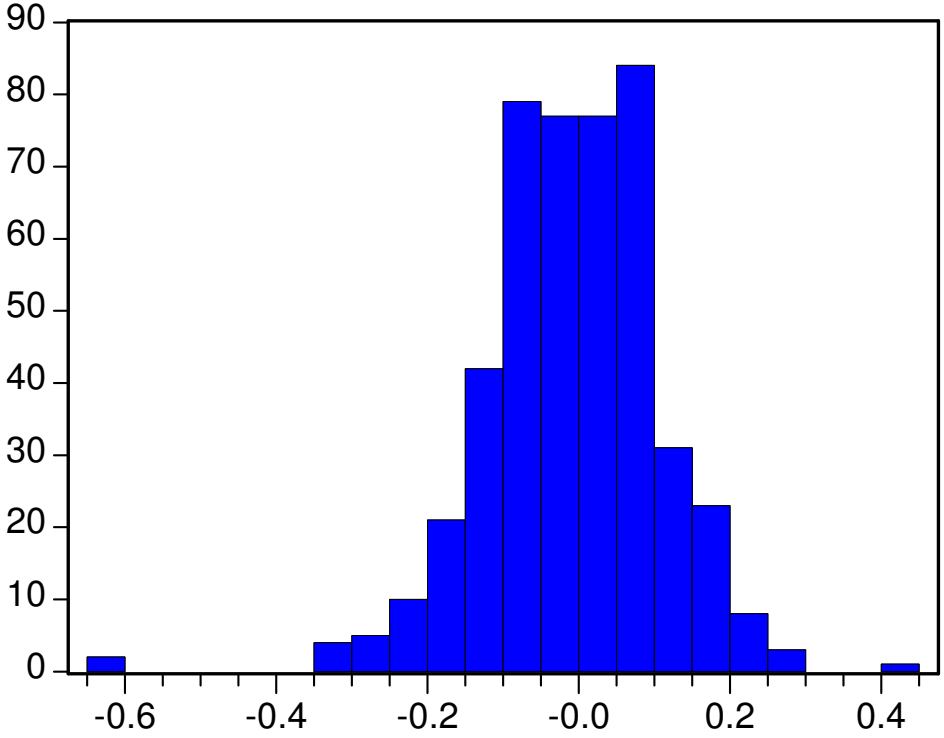
Date: 02/18/08 Time: 14:49
Sample: 1962M02 1999M12
Included observations: 455
Q-statistic probabilities adjusted for 3 ARMA te

Autocorrelation		AC	Q-Stat	Prob
	1	-0.049	1.1092	
	2	-0.004	1.1163	
	3	0.075	3.7297	
	4	-0.098	8.1915	0.004
	5	0.005	8.2039	0.017
	6	0.029	8.5845	0.035
	7	-0.083	11.783	0.019
	8	0.042	12.607	0.027
	9	0.031	13.066	0.042
	10	-0.081	16.169	0.024
	11	0.060	17.839	0.022
	12	0.025	18.135	0.034
	13	-0.023	18.383	0.049
	14	0.007	18.409	0.073
	15	-0.005	18.423	0.103
	16	-0.068	20.632	0.081
	17	0.029	21.019	0.101
	18	-0.053	22.371	0.098
	19	0.011	22.430	0.130
	20	-0.029	22.826	0.155
	21	-0.051	24.092	0.152
	22	-0.000	24.092	0.193
	23	0.071	26.540	0.149
	24	0.024	26.822	0.177

▷ *No coefficient is clearly significant but the Q statistic is significant for lags 4 to 14.*

▷ *Thus, we conclude that, with this test, we reject the serial independence of the residuals.*

▷ The mean of the residuals is not significantly different from zero and the variability of the residuals, except for some possible outliers, seems constant over time.

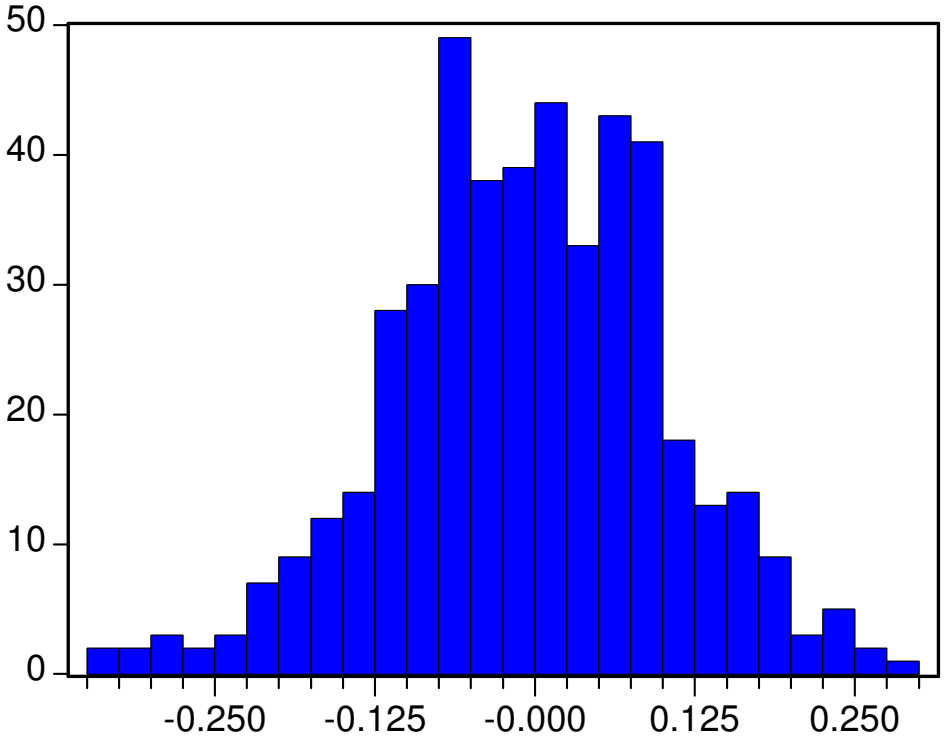


Series: RESID	
Sample 1960M01 1999M12	
Observations 467	
Mean	-0.007667
Median	-0.002161
Maximum	0.432698
Minimum	-0.618389
Std. Dev.	0.116063
Skewness	-0.626473
Kurtosis	6.050075
Jarque-Bera	211.5672
Probability	0.000000

▷ Is normally distributed?

▷ Is normally distributed? **No**.

▷ But, if we omit three atypical observations, we obtain the following residual statistics:

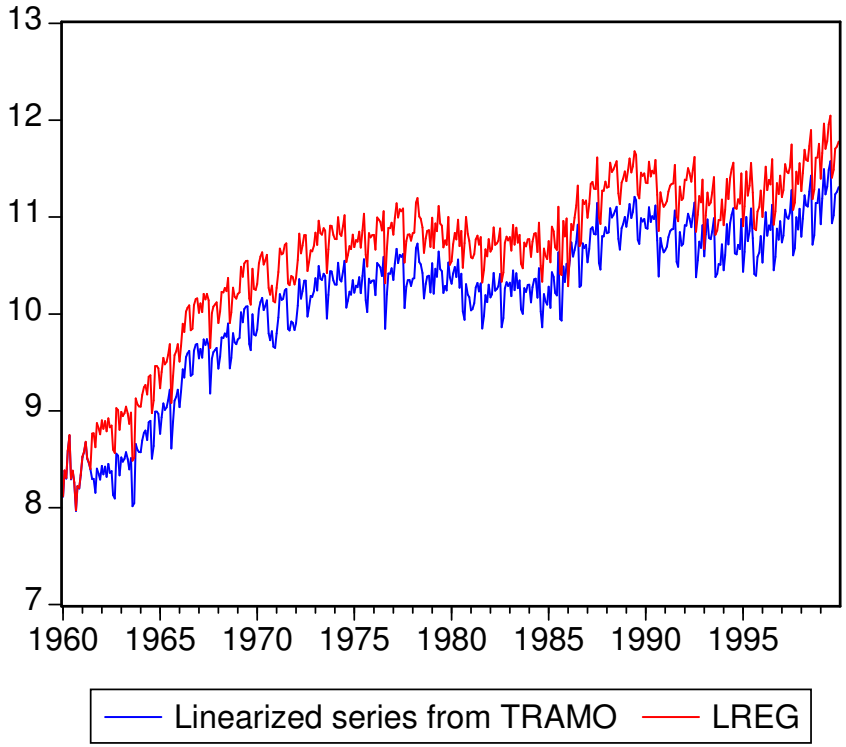


Series: RESID	
Sample 1960M01 1999M12	
Observations 464	
Mean	-0.005988
Median	-0.002014
Maximum	0.276686
Minimum	-0.338763
Std. Dev.	0.107378
Skewness	-0.200384
Kurtosis	3.208056
Jarque-Bera	3.942107
Probability	0.139310

▷ Thus we conclude that outliers can influence the autocorrelation and normality test's results.

Example 92. *The outliers detection will be studied in the Module 2, but here we will use the TRAMO/SEATS automatic detection:*

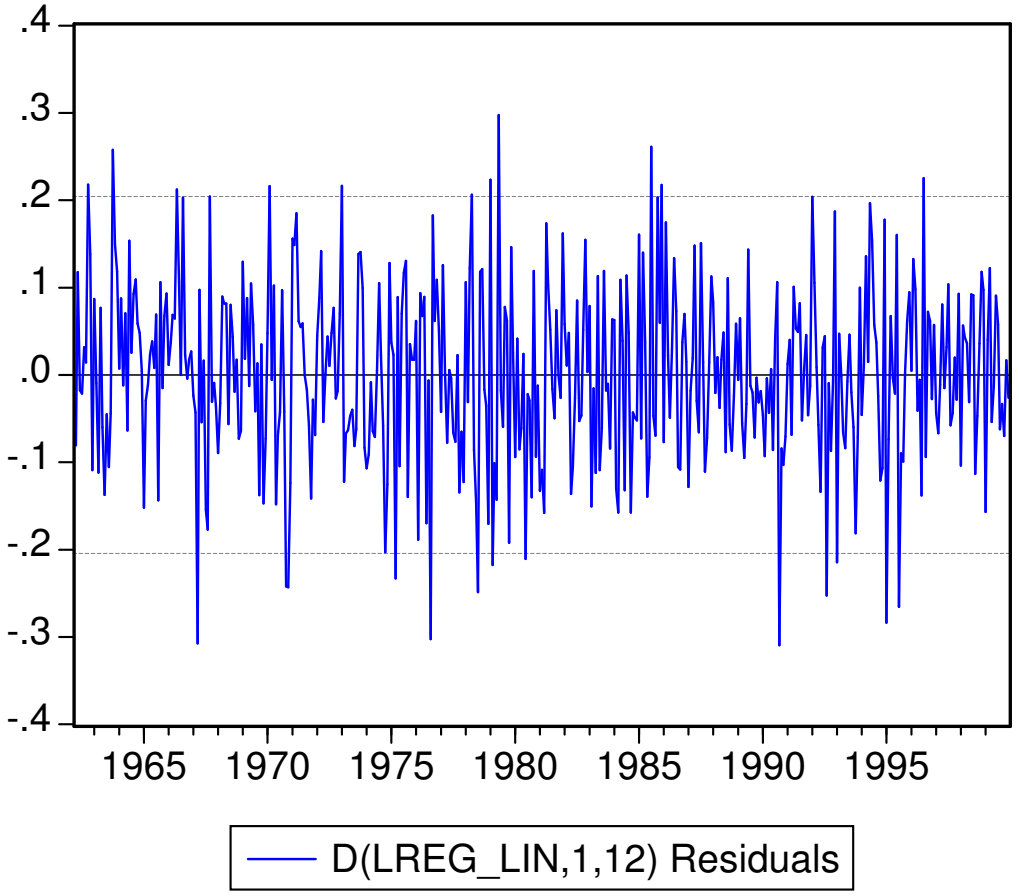
```
type: {'TC'  'TC'  'TC'  'LS'}
date: {'01-86'  '01-93'  '04-60'  '07-61'}
```



▷ So, we will repeat the analysis using the outlier-free series.

Model diagnosis - Examples

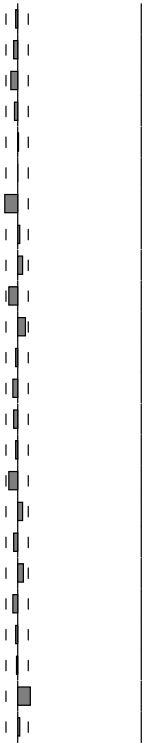
Example 93. *The figure gives the graph of the residuals of the corrected vehicle registration series estimated using an ARIMA*$(2,1,0)\times(1,1,1)_{12}$ *model.*



Example 93. *The figure gives* ACF *of the residuals.*

Correlogram of Residuals

Date: 02/18/08 Time: 15:12
Sample: 1962M04 1999M12
Included observations: 453
Q-statistic probabilities adjusted for 4 ARMA te

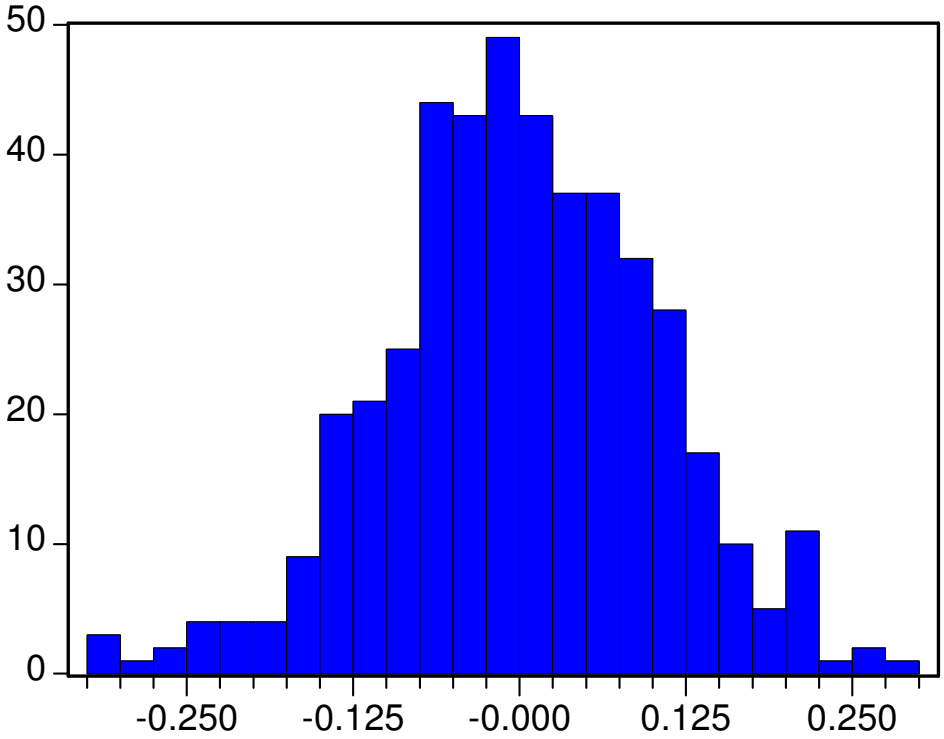


Autocorrelation		AC	Q-Stat	Prob
	1	-0.010	0.0479	
	2	-0.027	0.3897	
	3	-0.049	1.5024	
	4	-0.020	1.6842	
	5	0.009	1.7256	0.189
	6	0.001	1.7258	0.422
	7	-0.101	6.4085	0.093
	8	0.020	6.5934	0.159
	9	0.045	7.5193	0.185
	10	-0.068	9.6546	0.140
	11	0.074	12.188	0.095
	12	-0.009	12.223	0.142
	13	-0.035	12.787	0.172
	14	-0.033	13.288	0.208
	15	-0.009	13.331	0.272
	16	-0.073	15.860	0.198
	17	0.042	16.681	0.214
	18	-0.030	17.115	0.250
	19	0.049	18.278	0.248
	20	-0.036	18.912	0.273
	21	-0.012	18.976	0.330
	22	-0.001	18.977	0.393
	23	0.109	24.683	0.171
	24	0.022	24.909	0.205

▷ *No coefficient is significant and also Q statistic is not significant for all lags.*

▷ *Thus we conclude that, with this test, we find no evidence of serial dependence in the residuals.*

▷ The mean of the residuals is not significantly different from zero and the variability of the residuals seems constant over time.

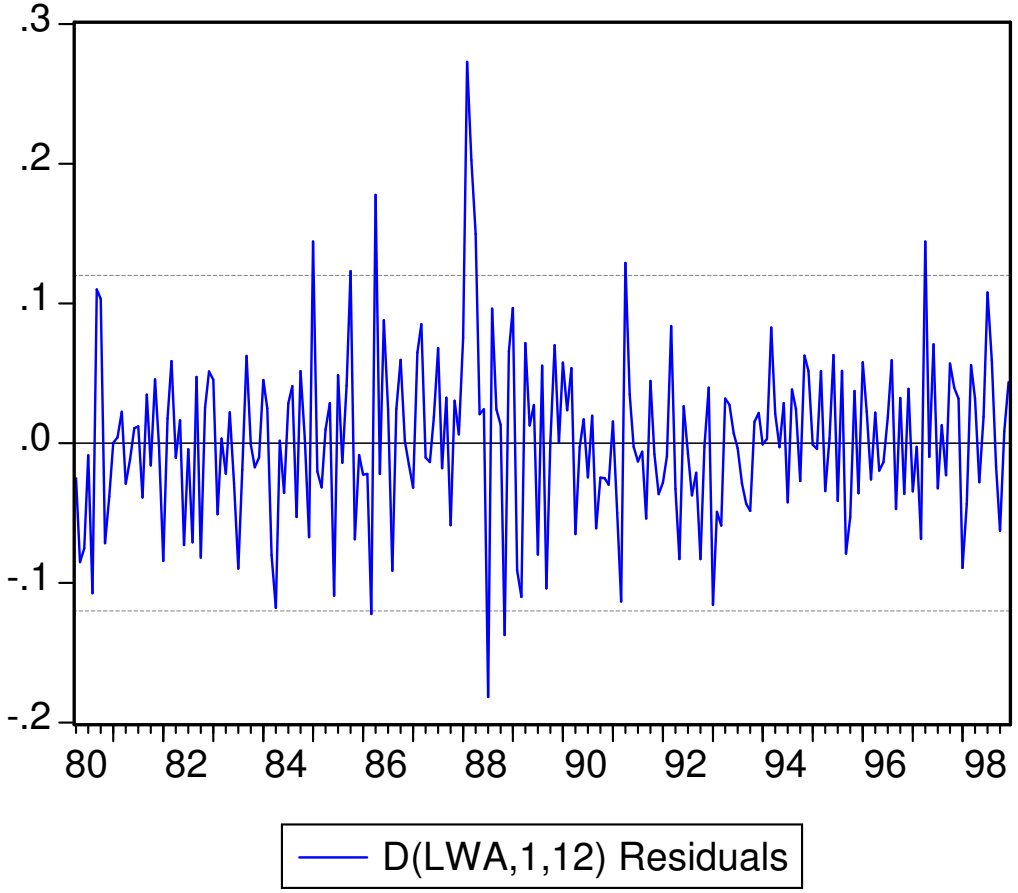


Series: Residuals	
Sample 1962M04 1999M12	
Observations 453	
Mean	0.001240
Median	-0.001309
Maximum	0.297403
Minimum	-0.309538
Std. Dev.	0.101856
Skewness	-0.124903
Kurtosis	3.189402
Jarque-Bera	1.854970
Probability	0.395547

▷ Is normally distributed? **Yes**

Model diagnosis - Examples

Example 94. *The figure gives the graph of the residuals of the the work related accidents series in logs estimated using an ARIMA*$(2,1,0) \times (0,1,2)_{12}$ *model. Again, some noticeable outliers are observed.*



Example 94. *The figure gives* ACF *of the residuals.*

Correlogram of Residuals

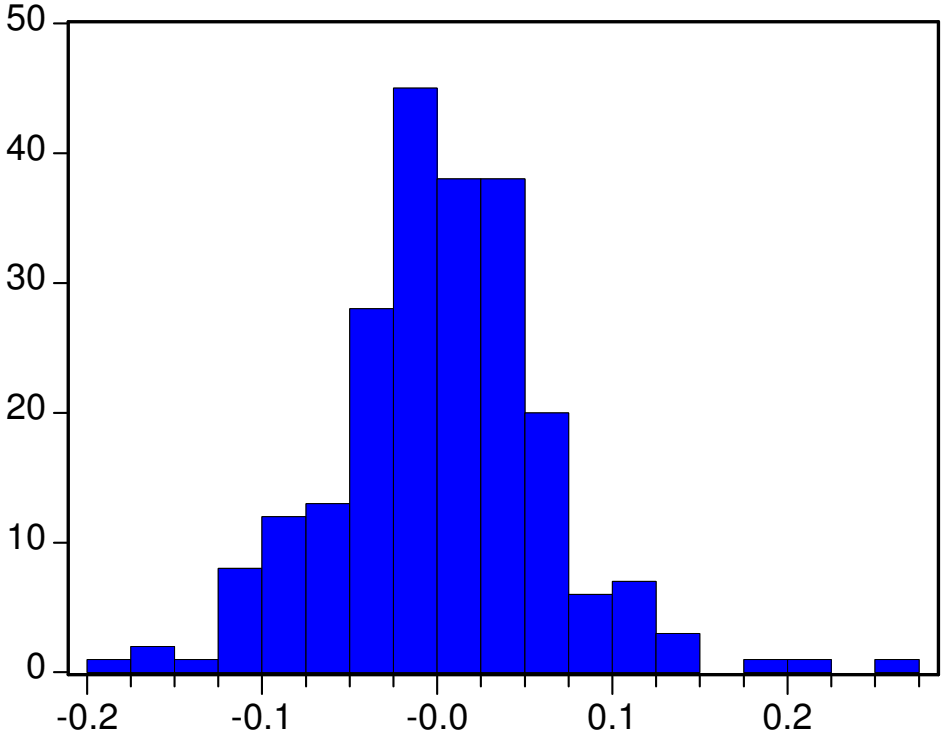
Date: 02/18/08 Time: 15:21
Sample: 1980M04 1998M12
Included observations: 225
Q-statistic probabilities adjusted for 4 ARMA te

Autocorrelation		AC	Q-Stat	Prob
	1	-0.013	0.0356	
	2	-0.024	0.1663	
	3	-0.029	0.3625	
	4	0.029	0.5575	
	5	-0.039	0.9122	0.340
	6	0.055	1.6196	0.445
	7	-0.091	3.5709	0.312
	8	0.109	6.3584	0.174
	9	0.109	9.1827	0.102
	10	-0.041	9.5743	0.144
	11	0.030	9.7879	0.201
	12	-0.005	9.7949	0.280
	13	-0.063	10.765	0.292
	14	0.173	17.973	0.055
	15	0.001	17.973	0.082
	16	-0.128	21.969	0.038
	17	0.076	23.387	0.037
	18	-0.061	24.299	0.042
	19	-0.048	24.867	0.052
	20	0.059	25.741	0.058
	21	0.024	25.885	0.077
	22	0.096	28.200	0.059
	23	0.112	31.386	0.037
	24	-0.002	31.387	0.050

▷ *No coefficient is significant but the* Q *statistic is significant for lags* ≥ 16.

▷ *Thus we conclude that, with this test, we reject the serial independence of the residuals.*

Example 94. *These atypical observations may leads to the rejection of any test of normality.*



Series: RESID	
Sample 1979M01 2000M12	
Observations 225	
Mean	0.001720
Median	0.001828
Maximum	0.262545
Minimum	-0.191480
Std. Dev.	0.062037
Skewness	0.306754
Kurtosis	4.752281
Jarque-Bera	32.31450
Probability	0.000000

```
type: {'AO'  'LS'}
date: {'11-88'  '02-88'}
```

▷ The presence of noticeable outliers suggests studying these effects before trying more complex models for this series.

Model stability test

▷ If the model is suitable, the prediction errors for one step ahead are normal random variables of zero mean, and constant variance.

▷ As a result, if we have estimated the model with data up to $t = T$ and next we generate predictions. Let $\widehat{a}_{T+1}, ..., \widehat{a}_{T+h}$ denote the one step ahead prediction errors, the variable

$$Q = \sum_{j=1}^{h} \widehat{a}_{T+j}^2 / \sigma^2$$

is a χ^2 with h degrees of freedom.

▷ Since σ^2 is estimated by means of $\widehat{\sigma}^2$, the variable

$$\frac{Q}{h} = \frac{\sum_{j=1}^{h} \widehat{a}_{T+j}^2 / h}{\widehat{\sigma}^2} \tag{212}$$

is an F with h and $T - p - q$ degrees of freedom, with T being the number of initial data points and $p + q$ the number of estimated parameters.

▷ Hence, if Q/h is significantly large, this suggests that the model is not suitable.

Predictions

Punctual predictions

▷ Predictions of the estimated model can be carried using the estimated parameters as if they were the true ones.

▷ Moreover, if the model is correct the estimated parameters, which are those that minimize the one step ahead prediction errors, are also those that minimize the prediction error to any horizon.

▷ Nevertheless, if the model is not correct this property is not necessarily true.

▷ To illustrate this aspect, let us assume that a series has been generated by an AR(2) but it has been estimated, erroneously, as an AR(1). If we use least squares the parameter is estimated with:

$$\min \sum (z_t - \alpha z_{t-1})^2$$

and the estimator $\widehat{\alpha}$ obtained minimizes the quadratic one step ahead prediction error.

Punctual predictions

▷ The result is $\widehat{\alpha} = r_1$, with r_1 being the first order autocorrelation coefficient of the data.

▷ Let us assume that we are interested in a two-step ahead prediction. We could state that if the data have been generated by an AR(1) the optimum predictor is, the conditional expectation, $\phi^2 z_{t-1}$ and use $\widehat{\alpha}^2 z_{t-1}$.

▷ Alternatively, we could directly obtain the β coefficient of an AR(1) predictor that minimizes the quadratic two-step ahead prediction error, minimizing:

$$\min \sum (z_t - \beta z_{t-2})^2$$

and we find that $\widehat{\beta} = r_2$, the second order autocorrelation coefficient.

▷ If the data had been generated with an AR(1) since the theoretical values would verify $\rho_2 = \rho_1^2 = \phi^2$, the predictor of z_{t+2} from z_{t-1} assuming an AR(1), $\widehat{\alpha}^2 z_{t-2}$, would coincide approximately for samples with $r_2 z_{t-2}$.

Punctual predictions

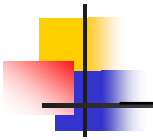
▷ Nevertheless, since the true model is the AR(2), both predictors will be quite different and the second will have a smaller mean square prediction error.

▷ Indeed, as the true model is AR(2), r_1 estimates the theoretical value of the first autocorrelation coefficient of this process, which is $\phi_1/(1-\phi_2)$, and r_2 will estimate $\phi_1^2/(1-\phi_2)+\phi_2$. The prediction $r_2 z_{t-2}$ is closer to the optimal than to the $\widehat{\alpha}^2 z_{t-2}$.

▷ To summarize, if the model is correct, the prediction to any horizon is obtained using the parameters estimated to one horizon.

▷ Nevertheless, if the model is not correct, we can improve the predictions by estimating the parameters for each horizon.

▷ This idea has been researched by some authors estimating the parameters at different horizons. If the model is correct, the parameters obtained will be approximately equal, but if the model is incorrect, we will find that it is necessary to use different parameters for different horizons.



Predictions

Prediction intervals

▷ The prediction intervals that we studied in section 7 were calculated assuming known parameters and only taking into account the uncertainty due to the future innovations.

▷ In practice, when the parameters of the model are estimated with the data, there are four types of uncertainty in the prediction, associated with the lack of knowledge about:

1. Future innovations.
2. The distribution of the innovations.
3. The true values of the parameters.
4. The model that generated the data.

Prediction intervals

▷ The first source of uncertainty is inevitable and does not depend on the sample size. As the future values of the series depend on future unknown innovations, we will have uncertainty which grows with the prediction horizon.

▷ The importance of the three remaining sources of uncertainty depend on the size of the sample. In general the effect of these uncertainties is small for large samples (long series), but can be significant in smaller samples (fewer than 50 data points).

▷ If we have a long series we can run a reliable test to see whether the distribution of the innovations is normal and, if we reject it, we can estimate and use a closed to real distribution of the innovations starting from the residuals.

▷ With short series, the power of a normality test is low, thus there is always greater uncertainty with respect to the distribution of the innovations. A possible solution is to use **bootstrap techniques**.

Prediction intervals

▷ The third source of uncertainty: the effect of the estimation of the parameters also increases with short series, since the estimation error diminishes with the sample size.

▷ Finally, with short series it is difficult to choose between similar models since the confidence intervals of the parameters are wide.

▷ For example, if the data have been generated with $(1 - .7B - .15B^2)z_t = a_t$ it will be difficult to choose with a sample of $T = 50$ between an AR(1) and an AR(2), since the estimation error of the parameters is of order $1/\sqrt{T} = 1/\sqrt{50} = .14$, a similar size to that of the parameter, hence the usual test will indicate that this parameter is not significant.

▷ With small or medium sample sizes there are usually several models which are compatible with the observed series, and we cannot ignore the fact that the selected model may be wrong. A possible solution is the **model average procedures**

Predictions

Prediction intervals for large samples

▷ If the sample size is large, the main uncertainty is that which is due to innovations and the others sources:

2. The distribution of the innovations.
3. The true values of the parameters.
4. The model that generated the data.

can be ruled out.

▷ Assuming normality, we can construct confidence intervals in large samples for the prediction of 95% taking the estimators as parameters and calculating the interval as in the case in which the parameters are known. For example, for 95% the interval is:

$$z_{T+k} \in \widehat{z}_T(k) \pm 1,96\ \widehat{\sigma}\ \left(1 + \widehat{\psi}_1^2 + ... + \widehat{\psi}_{k-1}^2\right)^{1/2}.$$

Prediction intervals for large samples

Example 95. *The figure shows the predictions generated for the work related accident series in the hypothesis of large samples and assuming normal distributions. Notice that the confidence intervals grow as the prediction horizon increases.*

Predictions

Bootstrap prediction intervals

▷ If the innovations are not normal, or if the sample size is not large and we do not wish to ignore the uncertainty due to the estimation of parameters, we can use bootstrap techniques to generate the predictions.

▷ A simple bootstrap procedure that takes into account the uncertainty due to the estimation of the parameters is the following (see, v.g. Thombs and Shucany (1990); Alonso, Peña and Romo (2002) and Pascual, Romo and Ruiz (2004)):

Outline of the resampling procedure:

$$(X_1,\ldots,X_N) \Rrightarrow \widehat{AR(p)} \Rrightarrow \begin{cases} X_1^{*(1)},\ldots,X_N^{*(1)} & \Rrightarrow & \widehat{AR(p)}^{*(1)} \\ \vdots & & \vdots \\ X_1^{*(B)},\ldots,X_N^{*(B)} & \Rrightarrow & \widehat{AR(p)}^{*(B)} \end{cases}$$

Bootstrap prediction intervals

1. Given $\boldsymbol{X}$, we obtain some estimates of the autoregressive parameters: $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_p)$.

2. We calculate the residuals: $\widehat{\varepsilon}_t = X_t - \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i}$, for $t = p+1, p+2, \ldots, N$.

3. We obtain the empirical distribution function of the centered residuals $\widetilde{\varepsilon}_t = \widehat{\varepsilon}_t - (N-p)^{-1} \sum_{t=p+1}^{N} \widehat{\varepsilon}_t$ by:

$$F_N^{\widetilde{\varepsilon}}(x) = \frac{1}{N-p} \sum_{t=p+1}^{N} I\left(\widetilde{\varepsilon}_t \leq x\right).$$

4. We obtain $N-p$ i.i.d. observations from $F_N^{\widetilde{\varepsilon}}$ denoted by $(\varepsilon_{p+1}^{*(b)}, \ldots, \varepsilon_N^{*(b)})$.

Bootstrap prediction intervals

5. We fix the first p values $(X_1^{*(b)}, \ldots, X_p^{*(b)})$ and we obtain the remaining $N-p$ observations $(X_{p+1}^{*(b)}, \ldots, X_N^{*(b)})$ by:

$$X_t^{*(b)} = \sum_{i=1}^{p} \widehat{\phi}_i X_{t-i}^{*(b)} + \varepsilon_t^{*(b)}.$$

6. Given $\boldsymbol{X}^{*(b)} = (X_1^{*(b)}, \ldots, X_n^{*(b)})$, we can calculate the bootstrap $\widehat{\boldsymbol{\phi}}^{*(b)} = (\widehat{\phi}_1^{*(b)}, \ldots, \widehat{\phi}_p^{*(b)})$ or some other statistics of interest.

A fundamental remark on the above resampling procedure:

▷ This resampling method must be modified if the goal is to construct prediction intervals since it does not replicate the conditional distribution of the future observations X_{T+h} given the observed data $\boldsymbol{X}$.

Bootstrap prediction intervals

Outline of the resampling procedure:

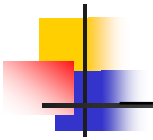
$$(X_1,\ldots,X_N) \Rightarrow \widehat{AR(p)} \Rightarrow \begin{cases} X_1^{*(1)},\ldots,X_N^{*(1)} & \Rightarrow \quad \widehat{AR(p)}^{*(1)} \\ \vdots & \vdots \\ X_1^{*(B)},\ldots,X_N^{*(B)} & \Rightarrow \quad \widehat{AR(p)}^{*(B)} \end{cases}$$

$$\begin{array}{lll} \Rightarrow & (X_{T-p},\ldots,X_T) & X_{T+1}^{*(1)},\ldots,X_{T+h}^{*(1)} \\ & \vdots & \vdots \\ \Rightarrow & \underbrace{(X_{T-p},\ldots,X_T)}_{Observed} & \underbrace{X_{T+1}^{*(B)},\ldots,X_{T+h}^{*(B)}}_{futures} \end{array}$$

7. Compute future bootstrap observations by the recursion:

$$X_{T+h}^* - \bar{X} = -\sum_{j=1}^{p} \widehat{\phi}_j^* (X_{T+h-j}^* - \bar{X}) + \varepsilon_t^*,$$

where $h > 0$, and $X_t^* = X_t$, for $t \leq T$.



Bootstrap prediction intervals

▷ This approach can be generalized to take into account as well the uncertainty in the selection of the model.

▷ If we can find a distribution for the possible orders of the model, we can generate realizations of the process using different orders and in each one calculate the prediction errors.

▷ In this way we obtain a distribution generated from prediction errors that we can use to construct prediction intervals.

▷ See:

- Alonso, Peña and Romo (2004): A model average approach.

- Alonso, Peña and Romo (2002): A nonparametric approach.

Predictions

Prediction by model average

▷ Let us assume that k models have been estimated for a series and that we have the BIC values for each model.

▷ The BIC values are, except for constants, $-2logP(M_i|D)$, where $P(M_i|D)$ is the probability of model M_i given the data D. Then,

$$P(M_i|D) = ce^{-\frac{1}{2}BIC_i},$$

where c is constant.

▷ With these results, we can transform the BIC values into a posteriori probabilities of different models by means of

$$P(M_i|D) = \frac{e^{-\frac{1}{2}BIC_i}}{\sum_{j=1}^{k} e^{-\frac{1}{2}BIC_j}}.$$

Prediction by model average

▷ The probability distribution of a new observation is then a mixed distribution

$$\sum P(M_i|D)f(z|M_i) \tag{213}$$

where $f(z|M_i)$ is the density function of the new observation in accordance with the model M_i.

▷ For example, for one period ahead if we generate predictions with each model and let $\widehat{z}_T^{(i)}(1)$ be the prediction for the period $T+1$ with model M_i, the expectation of the distribution (213) is:

$$\widehat{z}_T(1) = \sum_{i=1}^{k} \widehat{z}_T^{(i)}(1)P(M_i|D)$$

which is the result of combining the predictions of all the models to obtain a single aggregate prediction.

Prediction by model average

▷ This way of calculating predictions is known as **Bayesian Model Average**.

▷ In general this prediction is more accurate, on average, than that generated by a single model.

▷ Moreover, it allows us to construct more realistic prediction intervals than those obtained by ignoring the uncertainty of the parameters and the model.

▷ Letting $\widehat{\sigma}_i^2$ denote the variance of the innovations of the model M_i, which coincides with the one step ahead prediction error using this model, the variance of the combination of models is:

$$var(\widehat{z}_T(1)) = \sum_{i=1}^{k} \widehat{\sigma}_i^2 P(M_i|D) + \sum_{i=1}^{k} (\widehat{z}_T^{(i)}(1) - \widehat{z}_T(1))^2 P(M_i|D)$$

which allows a construction of more realistic prediction intervals.

Example 96. *We are going to see how to combine the predictions for the vehicle registration series. Let us assume that the three possible models are those indicated in table:*

Model	BIC_{Eviews}	$\Pr(M\|Data)$
ARIMA(0,1,1)$\times(0,1,1)_{12}$	-1.398	0.335
ARIMA(0,1,1)$\times(1,1,1)_{12}$	-1.466	0.347
ARIMA(0,1,1)$\times(2,1,0)_{12}$	-1.292	0.318

▷ We can calculate the a posteriori probabilities of each model by means of

$$\Pr(M_1|D) = \frac{\exp(-\frac{1}{2} \times -1.398)}{\exp(-\frac{1}{2}\times) + \exp(-\frac{1}{2} \times -1.466) + \exp(-\frac{1}{2} \times -1.292)} = 0.335$$

and analogously for other models.

▷ In this case the uncertainty with respect to the model is big and the predictions obtained by means of the combination are not similar to those of the best model.