Lecture 5: Box-Jenkins methodology

Florian Pelgrin

University of Lausanne, École des HEC
Department of mathematics (IMEA-Nice)

Sept. 2011 - Dec. 2011

Road map

1. Introduction

- Present the practical and pragmatic approach of Box and Jenkins in order to build ARIMA models
  - Step 1 : Identification
  - Step 2 : Estimation (and selection)
  - Step 3 : Diagnostic checking
  - Step 4 : Model's use

- **Step 1** (identification) involves determining the order of the model required (p, d, and q) in order to capture the salient dynamic features of the data. This mainly leads to use graphical procedures (plotting the series, the `ACF` and `PACF`, etc).

- **Step 2** (estimation and selection) involves estimation of the parameters of the different models (using step 1) and proceeds to a first selection of models (using information criteria).

- **Step 3** (checking) involves determining whether the model(s) specified and estimated is adequate. Notably, one uses residual diagnostics.

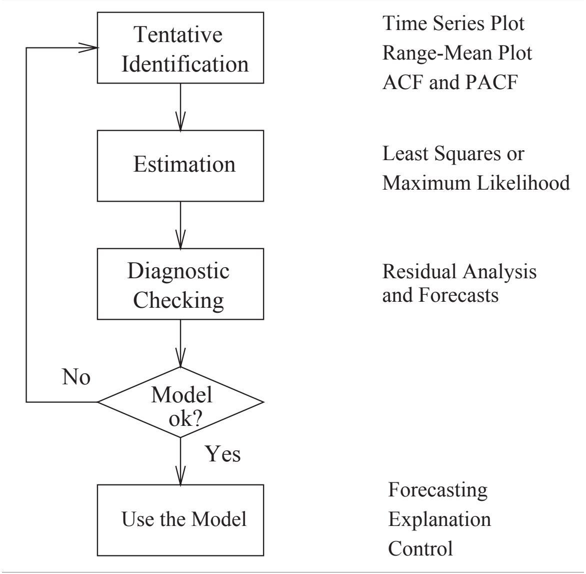


FIG.: Box-Jenkins methodology

2. Identification
2.1. Overview

Three objectives

1. Stationarity, non-stationarity ? What is the order of differentiation (d) ?
2. Seasonal component (ARIMA *versus* SARIMA model) ?
3. Identification of the ARMA order(p,q).

2.2. Identifying d

Two different approaches

- Graphical procedure : this involves plotting the data over time and the corresponding (partial) autocorrelation function.
  - If the ACF does not decrease to zero or at a very slow decay : this suggests non-stationarity (or long-memory effects).
  - Box and Jenkins (1976) recommend using the following differencing approach :
    1. Plot the autocorrelation function of the first-difference series
    2. Iterate the previous step until the ACF looks like the one of a stationary series
    3. Check the inverse autocorrelation function to avoid over-differencing.
- Test procedure : unit root tests (see Chapter 6)

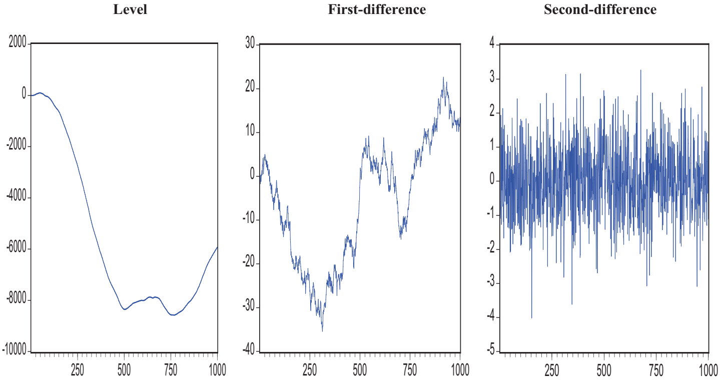


FIG.: Determining d...

Level

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.998	0.998	999.70	0.000
		2	0.997	-0.008	1997.1	0.000
		3	0.995	-0.008	2992.0	0.000
		4	0.993	-0.008	3984.5	0.000
		5	0.992	-0.008	4974.6	0.000
		6	0.990	-0.008	5962.0	0.000
		7	0.988	-0.008	6947.0	0.000
		8	0.986	-0.008	7929.2	0.000
		9	0.984	-0.008	8908.8	0.000
		10	0.982	-0.008	9885.7	0.000
		11	0.981	-0.008	10860.	0.000
		12	0.979	-0.008	11831.	0.000
		13	0.977	-0.008	12799.	0.000
		14	0.975	-0.008	13765.	0.000
		15	0.973	-0.008	14727.	0.000
		16	0.971	-0.008	15687.	0.000
		17	0.969	-0.008	16643.	0.000
		18	0.967	-0.008	17597.	0.000
		19	0.964	-0.008	18547.	0.000
		20	0.962	-0.008	19494.	0.000
		21	0.960	-0.008	20438.	0.000
		22	0.958	-0.008	21378.	0.000
		23	0.956	-0.008	22315.	0.000
		24	0.954	-0.008	23249.	0.000
		25	0.951	-0.008	24179.	0.000
		26	0.949	-0.008	25106.	0.000
		27	0.947	-0.008	26029.	0.000
		28	0.945	-0.008	26948.	0.000
		29	0.942	-0.008	27864.	0.000
		30	0.940	-0.008	28777.	0.000
		31	0.937	-0.008	29685.	0.000
		32	0.935	-0.007	30590.	0.000
		33	0.932	-0.007	31491.	0.000
		34	0.930	-0.007	32388.	0.000
		35	0.928	-0.007	33282.	0.000
		36	0.925	-0.007	34171.	0.000

First-difference

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.996	0.996	994.72	0.000
		2	0.993	0.038	1983.7	0.000
		3	0.990	0.020	2967.4	0.000
		4	0.987	0.002	3945.6	0.000
		5	0.983	-0.023	4918.2	0.000
		6	0.980	-0.005	5885.2	0.000
		7	0.976	-0.011	6846.3	0.000
		8	0.973	0.025	7802.0	0.000
		9	0.970	-0.028	8752.0	0.000
		10	0.966	-0.014	9696.1	0.000
		11	0.963	0.005	10634.	0.000
		12	0.959	-0.016	11567.	0.000
		13	0.955	-0.063	12492.	0.000
		14	0.951	-0.000	13411.	0.000
		15	0.947	-0.032	14323.	0.000
		16	0.943	-0.007	15227.	0.000
		17	0.939	0.007	16125.	0.000
		18	0.935	0.026	17016.	0.000
		19	0.931	-0.015	17900.	0.000
		20	0.927	-0.021	18778.	0.000
		21	0.922	-0.024	19648.	0.000
		22	0.918	0.010	20511.	0.000
		23	0.914	-0.008	21367.	0.000
		24	0.910	-0.004	22215.	0.000
		25	0.905	0.007	23057.	0.000
		26	0.901	-0.014	23892.	0.000
		27	0.897	-0.017	24719.	0.000
		28	0.892	-0.010	25539.	0.000
		29	0.888	-0.012	26351.	0.000
		30	0.883	-0.005	27155.	0.000
		31	0.878	-0.015	27952.	0.000
		32	0.874	0.038	28742.	0.000
		33	0.870	0.020	29525.	0.000
		34	0.865	-0.010	30301.	0.000
		35	0.861	0.054	31071.	0.000
		36	0.857	-0.001	31834.	0.000

Second-difference

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.065	-0.065	4.2762	0.039
		2	-0.004	-0.008	4.2911	0.117
		3	0.000	-0.001	4.2912	0.232
		4	0.027	0.027	5.0225	0.285
		5	0.003	0.007	5.0338	0.412
		6	0.022	0.023	5.5160	0.480
		7	-0.026	-0.023	6.1867	0.518
		8	0.016	0.012	6.4449	0.598
		9	-0.012	-0.011	6.5839	0.680
		10	-0.018	-0.021	6.9093	0.734
		11	0.045	0.044	8.9591	0.626
		12	0.060	0.065	12.544	0.403
		13	-0.014	-0.004	12.741	0.468
		14	0.025	0.025	13.396	0.496
		15	0.029	0.031	14.239	0.507
		16	-0.009	-0.009	14.321	0.575
		17	-0.001	-0.004	14.321	0.644
		18	-0.005	-0.007	14.345	0.706
		19	-0.002	-0.003	14.349	0.763
		20	0.021	0.018	14.793	0.788
		21	-0.006	-0.000	14.834	0.831
		22	0.010	0.011	14.936	0.865
		23	-0.013	-0.018	15.103	0.891
		24	-0.005	-0.010	15.132	0.917
		25	0.015	0.014	15.372	0.932
		26	0.027	0.022	16.125	0.933
		27	0.004	0.007	16.141	0.950
		28	-0.013	-0.012	16.327	0.961
		29	-0.010	-0.012	16.432	0.970
		30	0.033	0.031	17.582	0.965
		31	-0.056	-0.055	20.793	0.917
		32	0.027	0.019	21.527	0.919
		33	0.015	0.019	21.758	0.933
		34	-0.071	-0.072	26.988	0.798
		35	-0.006	-0.010	27.020	0.831
		36	-0.026	-0.030	27.707	0.837

FIG.: Determining d...

2.3. Seasonality

- A stochastic process is said to be a seasonal (or periodic) time series with periodicity s if Z_t and Z_{t+ks} have the same distribution.
- Such seasonal series are common in business, economics and finance
  - Business : The series of monthly sales of a department store, etc ;
  - Economics : Disaggregate price series, unemployment, components of GDP, monetary aggregates, stocks, etc ;
  - Finance : Intraday data, month-effect, day-effect, etc.
- Seasonal variations can constitute a large part of total variation in certain (macroeconomic) time series.

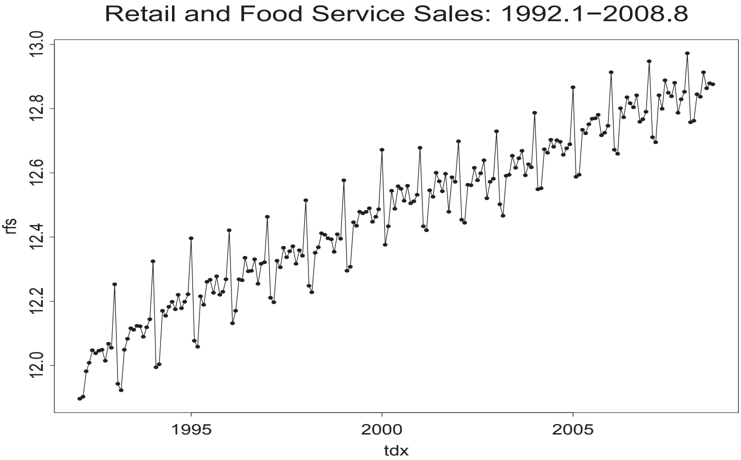


FIG.: Determining *d*...

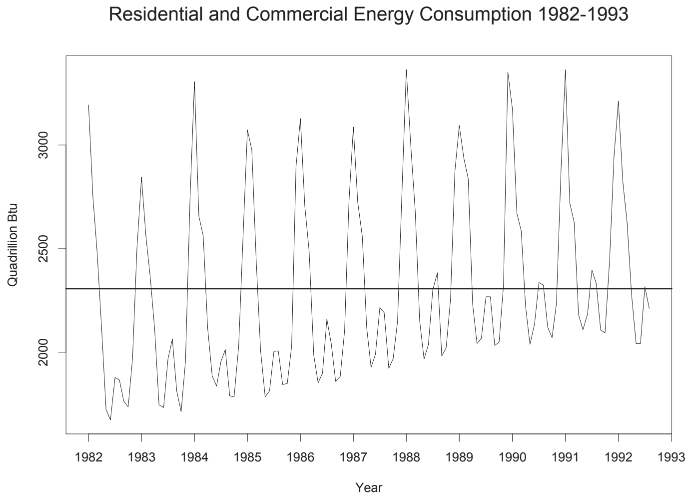


FIG.: Determining *d*...

- Seasonality can be assessed using graphical procedures :
  - Plot the series ;
  - The `ACF` ;
  - The spectral density
- Which treatment(s) for seasonality ?
  - Working with seasonally-adjusted series ?
  - Joint modeling of the seasonal and nonseasonal component ?

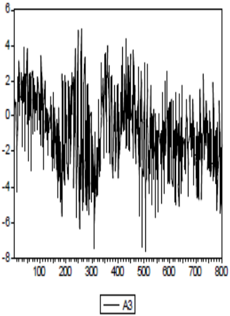
- Using seasonally-adjusted series does not mean that the seasonal pattern is completely removed. This is particularly true if the entire span of data is not used.

- Seasonally adjusting (or pre-filtering) a time series (e.g., the well-known Census X-12 makes extensive use of such filters) can distort sone of its important properties and may complicate further analysis.

- For a joint modeling, two approaches :

  - Models with seasonal dummy variables ;
  - SARIMA models (see Appendix).

Example: Seasonality



Sample: 1 800
Included observations: 800

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.597	0.597	286.55	0.000
		2	0.351	-0.009	385.51	0.000
		3	0.176	-0.046	410.54	0.000
		4	0.120	0.054	422.10	0.000
		5	0.120	0.063	433.81	0.000
		6	0.093	-0.018	440.81	0.000
		7	0.099	0.049	448.69	0.000
		8	0.082	0.003	454.11	0.000
		9	0.122	0.089	466.11	0.000
		10	0.280	0.261	529.74	0.000
		11	0.474	0.312	712.41	0.000
		12	0.786	0.639	1215.8	0.000
		13	0.463	-0.477	1390.9	0.000
		14	0.279	0.047	1454.4	0.000
		15	0.137	-0.050	1469.7	0.000
		16	0.088	-0.028	1476.1	0.000
		17	0.093	0.004	1483.2	0.000
		18	0.065	0.001	1486.6	0.000
		19	0.076	0.047	1491.4	0.000
		20	0.044	-0.057	1493.0	0.000
		21	0.072	0.009	1497.3	0.000
		22	0.210	-0.023	1533.7	0.000

FIG.: Identifying seasonality...

2.4. Identifying p and q

- The model order (p, q) can be determined by using graphical plots of the ACF and PACF.
- Main characteristics of ARMA(p,q) models :

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after q	Tails off
PACF	Cuts off after p	Tails off	Tails off

Different shapes of the autocorrelation function :

- **Exponential decay to zero** : Autoregressive model (use the partial autocorrelation plot to identify the order p)
- **Damped oscillations decaying (exponentially) to zero** : Autoregressive model
- **One or more spikes, the rest is essentially zero** : Moving average model (order q identified by where autocorrelation plot becomes zero)
- **Exponential decay starting after a few lags** : Mixed autoregressive and moving average model

Different shapes of the autocorrelation function (cont'd) :

- **No significant autocorrelations (zero or close to zero)** : White noise
- **High values at fixed intervals** : Include seasonal autoregressive terms
- **No decay to zero or very slow decay** : Non-stationarity or long-memory effects...

- In practice, identifying p and q using the `ACF` and `PACF` involves a trial and error approach...with more or less subjectivity in interpreting these functions : "real data" rarely exhibit simple patterns !

- The practical rule is :

  1. To choose an upper bound for p, say $p_{\max}$, and q, say $q_{\max}$;
  2. Estimate all models with $0 \leq p \leq p_{max}$ and $0 \leq q \leq q_{max}$;
  3. Use information criteria (or other procedures) to discriminate among the competing models.

3. Estimation and information criteria
3.1. Estimation

This step involves the estimation of the parameters of the models identified (specified) in Step 1 (with or without constraints). This can be done using (see Chapter 3, part 5)

- (Nonlinear) Linear least squares method ;
- Maximum likelihood estimation
- (Generalized) Method of moments
- Etc

3.2. Information criteria
(a) Overview

- The goodness-of-fit of the model can be assessed with the residual variance

$$\hat{\sigma}^2(k) = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2$$

where T is the number of observations used for the estimation, k is the total number of parameters estimated (e.g., $k = p + q + 2$), $\hat{\epsilon}_t$ is the adjusted residual at time t.

  - The residual sum of squares is inversely proportional to the number of degrees of freedom : "large" models with many parameters are often chosen when the sample size is large ;
  - Choosing over-parameterized (profligate) models tends to fit to data specific features, which would not be replicated out-of-sample ⇒ the models may appear to fit the data very well (with a large R^2) but would give inaccurate forecasts !

- Broadly speaking, an information criterion measures the goodness of fit of the model by assigning an informational cost to the number of parameters to estimate (making clear the idea that it is always possible to more or less trade off p versus q in selecting the ARMA orders)

- Information criteria embody two factors

  1. A term which is a function of the residual sum of squares ;
  2. A term which penalizes for the loss of degrees of freedom from adding extra parameters.

- Using an additional lag leads to two opposite effects

  1. $\downarrow$ residual sum of squares ;
  2. $\uparrow$ value of the penalty term

  $\Rightarrow$ IC $\downarrow$ if and only if $|\Delta SSR| > |\Delta PT|$

- The object is thus to choose the number of parameters which **minimizes** the value of the information criteria.

(b) Akaike information criterion

Definition

The Akaike information criterion (AIC) is defined to be

$$\texttt{AIC}(k) = \log\left(\hat{\sigma}^2(k)\right) + \frac{2}{T}(k)$$

where k is the total number of parameters estimated.

Remarks :

- AIC may give more than one minimum (p, q) ;
- AIC depends on the normality assumption ;
- AIC is not consistent : it will deliver on average too large a model (even with $T \to \infty$)—AIC tends to over-parameterize.
- AIC is generally efficient (small average variation in selected model orders from different samples within a given population).

(c) Schwarz bayesian information criterion

Definition

The Schwarz bayesian information criterion (SBIC) is defined to be

$$\texttt{SBIC}(p,q) = \log\left(\hat{\sigma}^2(k)\right) + \frac{k}{T}\log(T)$$

where k is the total number of parameters estimated.

Remarks :

- SBIC embodies a much stiffer penalty term than AIC, i.e. SBIC penalizes larger models more than AIC and thereby tends to select lower-order models than the AIC.
- SBIC is strongly consistent in selecting p and q : if the data are truly generated by an ARMA(p,q) process, the SBIC picks the true model with probability 1 as the sample size approaches infinity.
- SBIC is less efficient than AIC.

(d) Hannan-Quinn information criterion

Definition

The Hannan-Quinn information criterion (HQIC) is defined to be

$$\texttt{HQIC}(p,q) = \log\left(\hat{\sigma}^2(k)\right) + \frac{2k}{T}\log\left(\log(T)\right)$$

where k is the total number of parameters estimated.

Remarks :

- HQIC contains a penalty term which is somewhere between AIC and SBIC.
- HQIC is strongly consistent in selecting p and q.

(e) Practical implementation

- General methodology
  1. Set upper bounds, $p_{\max}$ and q_{max}, for the AR and MA order, respectively ;
  2. Fit all possible ARMA(p,q) models for $p \leq p_{\max}$ and $q \leq q_{\max}$ using a **common sample size** ;
  3. The best models (possibly more than one !) satisfy :

$$\min_{p \leq p_{\max}, q \leq q_{\max}} \mathtt{AIC}(p,q)$$
$$\min_{p \leq p_{\max}, q \leq q_{\max}} \mathtt{SBIC}(p,q)$$
$$\min_{p \leq p_{\max}, q \leq q_{\max}} \mathtt{HQIC}(p,q).$$

Remark : Finite sample corrections are available !

- Final model selection should not be based exclusively on any of these information criteria : goodness-of-fit is important but typically not the only relevant information criterion for choosing a model (especially, if the main goal is to generate forecasts...)

- Different information criteria may recommend different models.

- Even if one decides to only use one of these criteria, several models that are close to the minimum information criterion value might be considered (adjacent models)...

- In other words, all "reasonable" models should remain as candidates for the final selection and may be assessed by further diagnostic checks.

4. Diagnostic checking
4.1. Overview

- The main objective is to check the adequacy of the model(s) selected in the previous step.
- Notably all of the relevant information from the data should be extracted by the model. For instance, the part of the data unexplained by the model (residuals) should be small and no systematic or predictable patterns should be left in the residuals (i.e., weak white noise).
- Diagnostic testing in the Box-Jenkins methodology essentially involves the statistical properties of the error terms (normality assumption, weak white noise assumption).

4.2. Residual diagnostics
(a) Procedures

- (ϵ_t) is a (weak) white noise process, i.e. the residuals of an estimated model should exhibit white noise-like behavior : departure from this assumption means that some information can still be exploited in the modeling...

- Two methods
  - Graphical procedure
    1. Plot of the residuals : do they exhibit systematic patterns ?
    2. Use the SACF and SPACF (as in the identification step) of the residuals : Do they have significant elements ?
  - Testing procedure : autocorrelation tests.

(b) Autocorrelation tests

- One problem with checking the significance of individual (partial) autocorrelation is that each element might be individually insignificant, but all (or a subset) of the elements may be jointly significant.

- The Box-Pierce Q-statistic (or portmanteau test) tests the joint hypothesis that the first K autocorrelations of the adjusted error terms are jointly zero :

$$H_0 : \rho_{\hat{\epsilon}}(1) = \rho_{\hat{\epsilon}}(2) = \cdots = \rho_{\hat{\epsilon}}(K) = 0.$$

- The test statistic is given by :

$$Q = T \sum_{k=1}^{K} \hat{\rho}_{\hat{\epsilon}}^{2}(k)$$

where $\hat{\rho}_{\hat{\epsilon}}^{2}(k)$ is the k-th order sample autocorrelation of the estimated residuals, T is the sample size, and K is chosen sufficiently large.

- The Q-test has an asymptotic chi-square (χ^2) distribution with $K - p - q$ degrees of freedom. The null hypothesis of uncorrelated (estimated) residuals is rejected if the Q exceeds the tabulated critical value (for a chosen significance level).
- The Q-test is only asymptotically valid and may perform poorly for small and medium size samples.
- Refinements have been introduced in the literature : Ljung-Box test, McLeod-Li test, Monti's test, etc.

(c) Discussion

- Diagnostic testing based on autocorrelation tests could only reveal a model that is underparameterised ("too small") and would not reveal a model that is overparameterised ("too large").

- Autocorrelation of residuals can give rise to common factors especially in overfitted models. This makes estimation difficult and the statistical tests ill-behaved. For example, if the true data generating process is an ARMA(p,q) and one deliberately then fits an ARMA(p+1,q+1) there will be a common factor (all of the parameters in the latter model can be identified).

- Other residual procedures include the tests for the normality assumption (especially for maximum likelihood estimation), etc.