October 2008

NOTES ON PROGRAMS TRAMO AND SEATS©

TRAMO PART

Time Series Regression with ARIMA Noise,

Missing Observations and Outliers

Agustín Maravall

Bank of Spain

Observed time series:

$$z = (z_{t_1}, z_{t_2}, \ldots, z_{t_m})$$

$$1 = t_1 < t_2 < \ldots < t_m = T$$

- There may be missing observations.

- Perhaps z is the log of the original observations

MODEL

$$z_t = y_t' \beta + x_t$$

$$y_t \equiv \left[\begin{array}{l} \text{Matrix with n} \\ \\ \text{regression variables} \end{array} \right]$$

$$= \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ y_{12} & y_{22} & \cdots & y_{n2} \\ \vdots & \vdots & & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{nT} \end{bmatrix}$$

$$\uparrow \qquad\qquad\qquad \uparrow$$

First reg. var; n^{th} reg. var

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \text{Regression parameters}$$

x_t = ARIMA "noise" (possibly nonstationary)

Hence, in

$$z_t = y_t' \beta + x_t$$

$\downarrow$ deterministic component

$\downarrow$ stochastic component

Examples of "deterministic" component:

$$y_t' \beta = a + bt$$

$$\text{or} \quad = \sum \text{dummy variables}$$

$$\text{or} \quad = \text{Trading} - \text{day}$$

...

<u>ARIMA model for the noise x_t</u>

(the most important part of z_t)

$B \equiv$ Backward operator

$$B^j z_t = z_{t-j}$$

$$\phi(B)\delta(B)x_t = \theta(B)a_t + c \qquad (1)$$

$a_t \sim$ niid $(0, V_a)$

$\equiv$ white-noise innovations

$\phi(B)$: Stationary AR polynomial in B

$\delta(B)$: Non-stationary AR polynomial in B (unit roots)

$\theta(B)$: Invertible MA polynomial in B.

For seasonal series, the polynomials typically have a "multiplicative" structure.

Let

s= no. of observations per year (frequency of observation)

TRAMO mostly aimed at the following frequencies:

s = 12: montly, (s = MQ)

6: bimonthly,

4: quarterly,

3: 4-month observ.,

2: semester data,

1: annual obs.

Multiplicative structure:

$$\delta(B) = (1-B)^{d}\ (1-B^{s})^{d_s} =$$

$$= \nabla^{d}\ \nabla_{s}^{d_s}$$

$\equiv$ Differencing operator (Nonstationary unit roots)

(d_s = BD)

$$\phi(B) = \phi_p(B)\ \Phi_{p_s}(B^{s}) =$$

$$(1+\phi_1 B + \ldots + \phi_p B^{p})\ (1+\Phi_1 B^{s} + \Phi_2 B^{2s})$$

$\uparrow$ Stationary regular AR polynomial in B

$\uparrow$ Stationary seasonal AR pol. in B^s

$$\theta(B) = \theta_q(B)\ \Theta_{q_s}(B^{s}) =$$

$$= (1+\theta_1 B + \ldots + \theta_q B^{q})\quad (1+\Theta_1 B^{s} + \Theta_2 B^{2s})$$

$\uparrow$ invertible regular MA pol. in B

$\uparrow$ invertible seasonal MA pol. in B^s

(In program: q_s = BQ

p_s = BP)

Recommended limits:

$d, p, q \quad \leq \quad 3$

$d_s, p_s, q_s \quad \leq \quad 2$ (1 if TRAMO is used with SEATS)

Hence, FULL MODEL:

$$\underbrace{\phi(B)\,\Phi(B^s)\,\nabla^d\,\nabla_s^{d_s}}_{\text{full AR polynomial}}\,x_t = \underbrace{\theta(B)\,\Theta(B^s)}_{\text{full MA polynomial}}\,a_t + c$$

Note:

To simplify notation, sometimes we shall simply write

$$\phi(B)\,x_t = \theta(B)\,a_t$$

where: ↓ ↓

full AR pol. full MA pol.

x_t is assumed INVERTIBLE

(for observed series, a sensible assumption).

Recall: if

$$\phi(B)\, x_t = \theta(B)\, a_t$$

is ARIMA model for x_t, x_t is invertible when $\theta(B)^{-1}$ converges (i.e., when roots of $\theta(B) = 0$ are larger than 1 in moduli).

Therefore, if (c = a constant)

$$\theta(B) = c\,(b_1 - B)(b_2 - B)\ldots(b_Q - B)$$

$$(b_1, b_2, \ldots, b_Q \equiv \text{roots in B of } \theta(B) = 0),$$

the roots have to satisfy:

$$|b_j| > 1. \qquad \forall_j$$

Equivalently, in terms of the inverse roots

$$r_j = \frac{1}{b_j},$$

$$\theta(B) = (1 - r_1 B)\ldots(1 - r_Q B),$$

invertibility implies:

$$|r_j| < 1, \quad \forall j$$

In TRAMO, assumption is enforced. Thus, if

$$|r_j| \rightarrow 1 \quad ,$$

in the limit, it is set to

$$|r_j| = .99 \qquad \text{(XL in program)}$$

- No numerical problems

- Good protection against overdifferencing.

Assume model is

$$x_t = a_t \qquad \text{(w.n.)}$$

If we overdifference

$$\nabla x_t = (1 - B)\, a_t \quad ,$$

TRAMO will yield something close to

$$\nabla x_t = [1 - (XL)\, B]\, a_t \; .$$

For ex., for our series length ($\leq$ 600), the model

$$\nabla x_t = (1 - .99\, B)\, a_t \qquad (2)$$

causes no problem and is indistinguishable from

$$x_t = \mu + a_t$$

- No loss of d. of freedom if (2) is used.

- (2) could be rewritten as

$$x_t = \mu^{(t)} + a_t$$

with

$\mu^{(t)} =$ a very slowly evolving mean (an "adaptive" feature).

In the short-run, (2) is likely to perform better.

REGRESSION VARIABLES:

- Input by user
- generated by program:

  Trading Day (several specifications)

  Easter

  Intervention variables

  Outliers

WHAT TRAMO DOES:

- Exact ML estimation of the regression-ARIMA model

  (Conditional/unconditional LS procedures are also available)
- Diagnostics
- Detection and correction of outliers:

  AO: Additive Outlier

  TC: Transitory Change

  LS: Level Shift

  IO: Innovational outlier
- Optimal interpolation of Missing Observations

  (and associated SE)
- Computes optimal Forecasts (and associated SE)

  "Optimal" ≡ Minimum MSE

Some important features

- Pretesting:

  * Log/Level

  * TD

  * EE

- Automatic Model Identification (AMI)

joint with

- Automatic Outlier Detection and Correction (AODC)

AUTOMATIC MODEL IDENTIFICATION

Two steps:

1) Obtains the degree of differencing

(max. orders: $\nabla^2 \nabla_s$)

2) Obtains the multiplicative stationary ARMA model

$0 \leq (p, q) \leq 3$

$0 \leq (p_s, q_s) \leq 2$

* BIC criterion

* Favors "Balanced"models

(order of full AR = order of full MA)

Procedure is combined with AUTOMATIC OUTLIER DETECTION AND CORRECTION.

SUMMARY OF THE METHODOLOGY

1) Kalman filtering with no M.O. and no regression parameters.

a) Stationary series

$$\phi(B)\, z_t = \theta(B)\, a_t\,, \qquad a_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

We use the following state space form representation of ARMA model (Akaike, 1974):

$$r = \max\{p, q+1\}$$

$$x_t = \begin{pmatrix} z_t \\ z_{t+1,t} \\ \vdots \\ z_{t+r-1,t} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \dots & -\phi_1 \end{pmatrix} x_{t-1} + \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \psi_{r-1} \end{pmatrix} a_t \qquad \text{(A)}$$

$$z_t = (1, 0, \dots, 0)\, x_t$$

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \sum_{i=0}^{\infty} \psi_i B^i\,, \text{ (Psi-weights)}$$

$$\phi_i = 0, \quad i > p, \quad p = dg\{\phi\}, \quad q = dg\{\theta\}.$$

When the series is stationary, TRAMO uses the algorithm of Morf, Sidhu and Kailath (1974), with an improvement based on Mélard (1984).

$$z = (z_t, \ldots, z_N)' \quad \text{Var}(z) = \sigma^2\,\Omega = \sigma^2 LL',$$

$L \equiv$ lower triangular. (Cholesky factorization of a p.d. covariance matrix)

Likelihood:

$$\begin{aligned}\ell(z) &= (2\pi\,\sigma^2)^{-N/2}\,|\,\Omega\,|^{-1/2}\,\exp\left\{-\frac{1}{2\sigma^2}\,z'\,\Omega^{-1}\,z\right\}\\ &= \ell(z_N \mid z_{N-1}, \ldots, z_1)\ldots\ell(z_2 \mid z_1)\,\ell(z_1)\\ &= (2\pi\,\sigma^2)^{-N/2}\,|\,\Pi_{t=1}^N\,\sigma_{t,t-1}^2\,|^{-1/2}\,\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^N e_t^2\right\},\end{aligned}$$

where

$$e_t = (z_t - \hat{z}_{t,t-1})/\sigma_{t,t-1} \qquad \leftarrow \text{standardized innovations}$$

$$\hat{z}_{t,t-1} = E(z_t \mid z_{t-1}, \ldots, z_1)$$

$$\sigma_{t,t-1}^2 = E(z_t - \hat{z}_{t,t-1})^2 / \sigma^2.$$

The Kalman Filter recursively computes e_t and $\sigma_{t,t-1}^2$ for $t = 1, \ldots, N$, with starting conditions derived from the marginal distribution, namely

$$\hat{z}_{1,0} = E(z_1) = 0 \quad \text{and} \quad \sigma_{1,0}^2 = \text{Var}(z_1)/\sigma^2.$$

σ^2 can be concentrated out of the likelihood using

$$\hat{\sigma}^2 = e' e / N, \qquad e = (e_1, \ldots, e_N)'$$

and maximizing $\ell(z)$ is equivalente to minimizing the nonlinear sum of squares

$$S = |L|^{1/N} e'e |L|^{1/N},$$

where

$$|L| = \Pi_{t=1}^{N} \sigma_{t,t-1}; \quad e_t = (z_t - \hat{z}_{t,t-1}) / \sigma_{t,t-1}.$$

It can be shown that

$$e = L^{-1} z,$$

⇑ residuals computed by TRAMO

where

$$\text{Var}(z) = \sigma^2 LL' = \sigma^2 \Omega$$

Thus e is white noise (0, σ^2)

The unconditional least squares method minimizes $S = e'e$.

b) Nonstationary series

If there are no M.O., we can difference the series to achieve stationarity, and proceed as before. However, we shall be interested in general in using a representation for the levels of the series.

$$\phi(B)\,\delta(B)\,z_t = \theta(B)\,a_t\ , \quad a_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

$$\delta(B) = 1 + \delta_1 B + \ldots + \delta_d B^d:$$

polynomial with Unit Roots ($\therefore$ Total Differencing)

The state space form is similar to that of the stationary case but uses the original series. Hence, it is as in the stationary case with $\phi(B)$, $\psi(B)$, and p replaced by

$$\phi^*(B) = \phi(B)\,\delta(B)\ ,$$

$$\psi^*(B) = \theta(B)\,/\,\phi(B)\ ,$$

$$p^* = p + d\ ,$$

and x_t replaced by z_t in expression (A) .

Further, define

$$z = (z_1, \ldots, z_d,\ z_{d+1}, \ldots, z_N)'$$

$$x_t = \delta(B)\,z_t \equiv \text{differenced series}$$

$$x = (x_{d+1}, \ldots, x_N)'$$

We define the likelihood as the conditional likelihood (in the levels)

$$\ell\,(z_N,\ldots,z_{d+1} \mid z_d,\ldots z_1)\ .$$

It can be obtained through the prediction error decomposition

$$\ell\,(z_N,\ldots,z_{d+1} \mid z_d,\ldots z_1) = \ell\,(z_N \mid z_{N-1},\ldots, z_1)$$

$$\text{x}\ldots\text{x}\ \ell\,(z_{d+1} \mid z_d,\ldots z_1)\ ,$$

using the Kalman Filter.

Under some straightforward assumptions concerning the starting conditions (z_1 ,…, z_d), it is shown in Gómez and Maravall (1994, JASA) that

$$\ell(x) = \ell\,(z_N,\ldots,z_{d+1} \mid z_d,\ldots z_1)$$

⇑ Box-Jenkins likelihood ⇑ conditional likelihood

Estimation of ARMA Models

Three estimation methods for ARMA models are available:

- Conditional least squares method (CLS)
- Unconditional least squares method (ULS)
- Exact maximum likelihood method (EML)

Fast recursive procedures for likelihood evaluation, based on the Kalman filter, make it advisable to use the exact maximum method. This method (EML) implies maximization of a highly nonlinear function of the parameters.

There is a need for good starting values for the optimization method used to estimate the model parameters. One good option is the Hannan-Rissanen's method. It is computationally cheap, because it uses only linear regressions, and it provides estimates that are close to the exact maximum likelihood estimates.

Because it is

- simple,

- intuitive,

I proceed to summarize the HR method.

Basic idea behind H-R method

Assume invertible ARMA (1,1) model:

$$z_t + \phi\, z_{t-1} = a_t + \theta\, a_{t-1} \quad , \quad |\,\theta\,| < 1 \ .$$

or

$$(1 + \phi\, B)\, z_t = (1 + \theta\, B)\, a_t \quad .$$

Since model is invertible, we can express it as

$$(1 + \theta\, B)^{-1}\, (1 + \phi\, B)\, z_t = a_t \quad ,$$

where

$$\frac{1}{1 + \theta\, B} = 1 - \theta\, B + \theta^2\, B^2 - \theta^3\, B^3 + ...$$

Given that $\theta^j \rightarrow 0$ as $j \rightarrow \infty$, we can truncate the model, and write it as:

$$(1 + \pi_1\, B + \pi_2\, B^2 + ... + \pi_k\, B^k)\;\; z_t = a_t \quad .$$

Hence we can approximate the model with a long AR.

LS estimation of the AR yields consistent estimators of the

residuals a_t.

With these estimated residuals we can build the series $\hat{a}_{t-1}$.

Replacing a_{t-1} by $\hat{a}_{t-1}$ in ARMA, we can express it as

$$z_t + \phi\, z_{t-1} - \theta\, \hat{a}_{t-1} = a_t \quad ,$$

or

$$z_t = \beta_1\, y_{1t} + \beta_2\, y_{2t} + a_t \quad ,$$

where

$$\beta_1 = -\phi \quad , \qquad y_{1t} = z_{t-1} \quad ,$$
$$\beta_2 = \theta \quad , \qquad y_{2t} = \hat{a}_{t-1} \quad .$$

β_1 and β_2 (and hence ϕ and θ) can be consistently estimated by OLS.

This provides consistent estimators of ϕ and θ.

Modified Hannan-Rissanen method:

We wish to estimate the ϕ - and θ - parameters in an invertible ARMA (p, q) model

$$\phi(B)\, z_t = \theta(B)\, a_t \ ,$$

where

$$z_t = \delta(B)\, x_t \ ,$$

$$\text{order}\, [\phi(B)] = p \ ,$$

$$\text{order}\, [\theta(B)] = q \ ,$$

1) We compute

$$n = \max \left\{ \ln^2 (N - d),\ 2 \max \{p, q\} \right\} \ ,$$

and fit the (long) AR:

$\Pi_n(B)\, z_t = u_t$, ("Pi-weights" representation).

We obtain $\hat{\Pi}(B)$ and compute the residuals ($t \geq 1$)

$$\hat{u}_t = \sum_{j=0}^{n} \hat{\Pi}_j\, z_{t-j} \quad , \quad \hat{\Pi}_0 = 1 \ ,$$

where $z_t = 0$ if $t \leq 0$

$\hat{u}_t$ is a consistent estimator of a_t .

2) The ϕ and θ parameters are estimated using $\hat{u}_{t-j}$ as a replacement for a_{t-j} in the model.

From:

$$z_t + \sum_{j=1}^{p} \phi_j \, z_{t-j} - \sum_{j=1}^{q} \theta_j \, a_{t-j} = a_t \quad ,$$

we minimize the SS:

$$\sum_{t=m}^{N-d} \left\{ \sum_{j=0}^{p} \phi_j \, z_{t-j} - \sum_{j=1}^{q} \theta_j \, \hat{a}_{t-j} \right\}^2 \quad ,$$

where m = max {p + 1, q + 1}, ϕ_0 =1, and $\hat{a}_{t-j} = \hat{u}_{t-j}$.

(If q > 0, another regression is used to correct for bias.)

HR is used in intermediate steps. Final estimation can be done through CLS, ULS, or EML methods.

Example

Suppose that we want to estimate the parameters of the MA (1) model

$$z_t = a_t + \theta a_{t-1}, \qquad t = 1,\dots, N,$$

where a_t is ditributed $N(0, \sigma^2)$, and let $z = (z_1,\dots, z_N)'$ be the observed series, with Var (z) = $\sigma^2 \Sigma$. Then, the log-likelihood is, up to a constant,

$$\lambda(z) = -\frac{1}{2}\left\{ N \ln(\sigma^2) + \ln|\Sigma| + z'\Sigma z/\sigma^2 \right\}.$$

The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = z'\Sigma z/N$$

and the σ^2-maximized log-likelihood $\ell(z)$ is, up to a constant,

$$\ell(z) = -\frac{1}{2}\left\{ N \ln(\hat{\sigma}^2) + \ln|\Sigma| \right\}.$$

Then,

- the conditional least squares method minimizes $z'\Sigma z$, under the assumption that $a_0 = 0$. That is, it minimizes $\sum_{t=1}^{N} a_t^2$, where the a_t are recursively obtained from $a_t = z_t - \theta\, a_{t-1}$, $a_0 = 0$.

- the unconditional least squares method minimizes $z'\Sigma z$, but estimating by maximum likelihood the initial condition a_0. Box and Jenkins used the so-called "backcasting procedure". Modern recursive methods, like the Kalman filter or the innovations algorithm of Brockwell and Davis (1992), yield

$$\hat{a}_{t+1} = z_{t+1} - \frac{\theta}{D_t}\,\hat{a}_t$$

$$D_{t+1} = 1 + \theta^2 - \frac{\theta^2}{D_t} \ ,$$

  with the initial conditions $\hat{a}_1 = z_1$ and $D_1 = 1 + \theta^2$. Then, the method minimizes $z'\Sigma z = \sum_{t=1}^{N} \hat{a}_t^2$. It can be shown that $D_t \to 1$ as $t \to \infty$.

- the exact maximum likelihood method minimizes $z'\Sigma z\ |\Sigma|^{1/N}$, where, with the above notation, $z'\Sigma z = \sum_{t=1}^{N} \hat{a}_t^2$ and $|\Sigma| = \Pi_{t=1}^{N} D_t$.

- HR would, in essence, use LS to obtain

$$z_t + \hat{\phi}_1 z_{t-1} + \ldots + \hat{\phi}_p\ z_{t-p} = \hat{a}_t$$

  with p long enough to “whitten” the series. The θ-parameters would be estimated by LS in

$$z_t = \theta\ w_t + a_t$$

$$\text{with } w_t = \hat{a}_{t-1}$$

DIAGNOSTIC CHECKING

- Main diagnostics are RESIDUAL - based tests, where we use the estimated residuals $\hat{a}_t$ to test the hypothesis

$$H_0 : [\hat{a}_t] \sim \text{niid}\ (0, V_a).$$

- Also important: out-of-sample Forecast test:

Do the out-of-sample forecast errors behave in agreement with the model?

1. Residual-based tests

Once we have estimated the model, we construct the residual series. If exact maximum likelihood has been used, the residuals are defined by

$$\hat{a}_t = \frac{z_t - z_{t|t-1}}{\sigma_{t|t-1}} \ ,$$

where $z_{t|t-1}$ is the one-step ahead forecast and $\sigma_{t|t-1}$ is its root mean squared error. These are given recursively by the Kalman filter.

Number of Residuals in TRAMO

NZ : number of observations ;

δ : degree of differencing needed to achieve stationarity;

(Ex: $\nabla\nabla_{12} \rightarrow \delta = 13$)

K : # of outliers + TD ε EE variables + additional regression variables

m : number of missing observations

Number of residuals produced by the KF that should be niid:

NA = NZ – m – (d. of f. lost) ,

where degrees of freedom lost = δ + k.

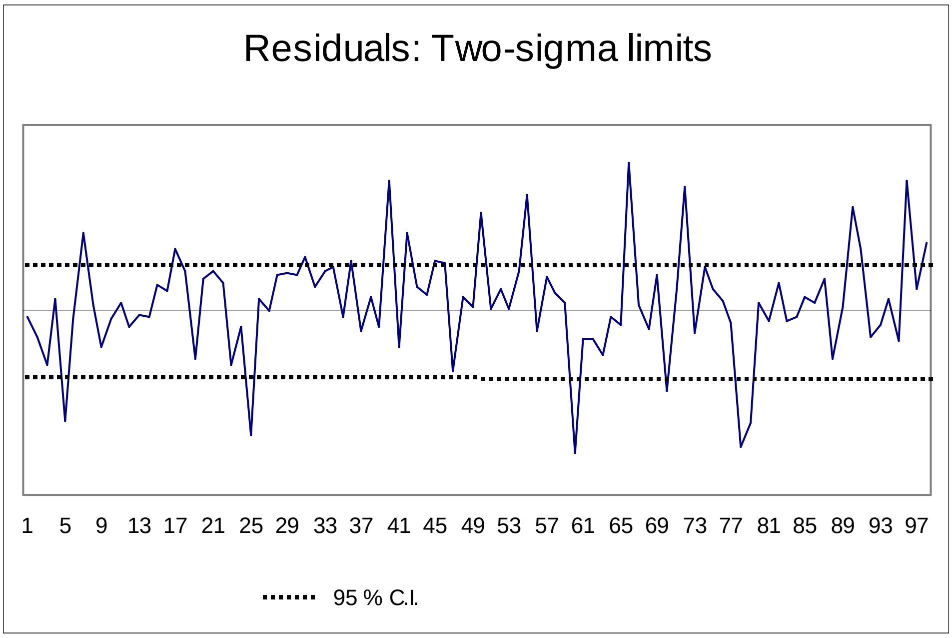
<u>Remark</u>: to facilitate filtering, SEATS will estimate the residuals associated with the d. of f. lost and the missing values.

After the residuals have been obtained, we can

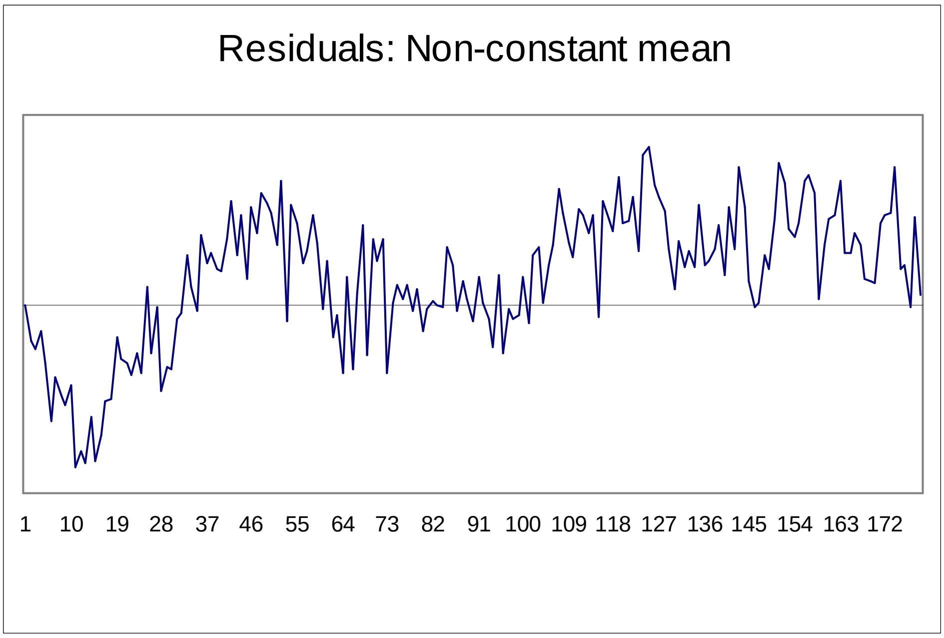
- **Inspect the graph of** $[\hat{a}_t]$

  Examples of problems that can be detected through inspection of the graph of $[\hat{a}_t]$:

  1. Abnormal number of outliers (> say 5%) $\rightarrow$ Nonnormality



2. Non-constant mean (an indication of non-stationary residuals).



<u>Note</u>:

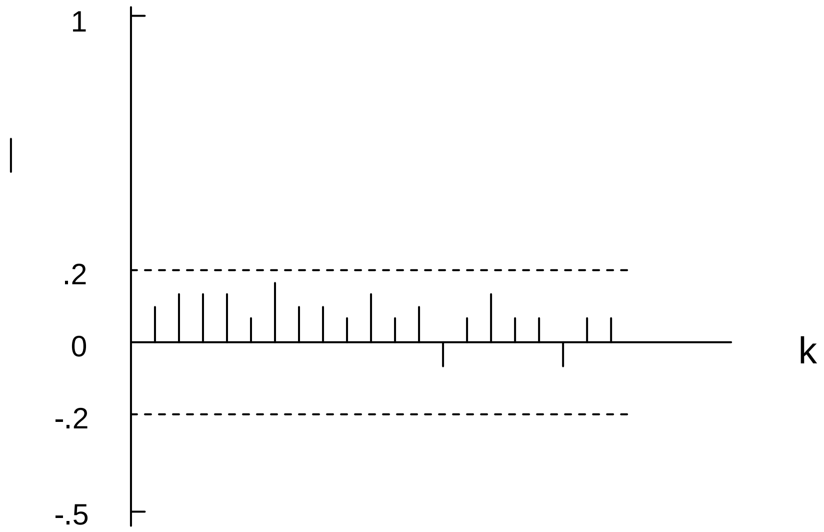
If d is increased, it is a safe procedure to increase q, in order to protect against overdifferencing.

Ex:

We fit

$$\nabla x_t = (1 + \theta B)\, a_t + \mu$$

and the ACF of $\hat{a}_t$ shows:



This may be O.K.,

but the persistence of positive values can also point to Nonstationarity (lack of “fast” convergence).

Thus one may try d = 2.

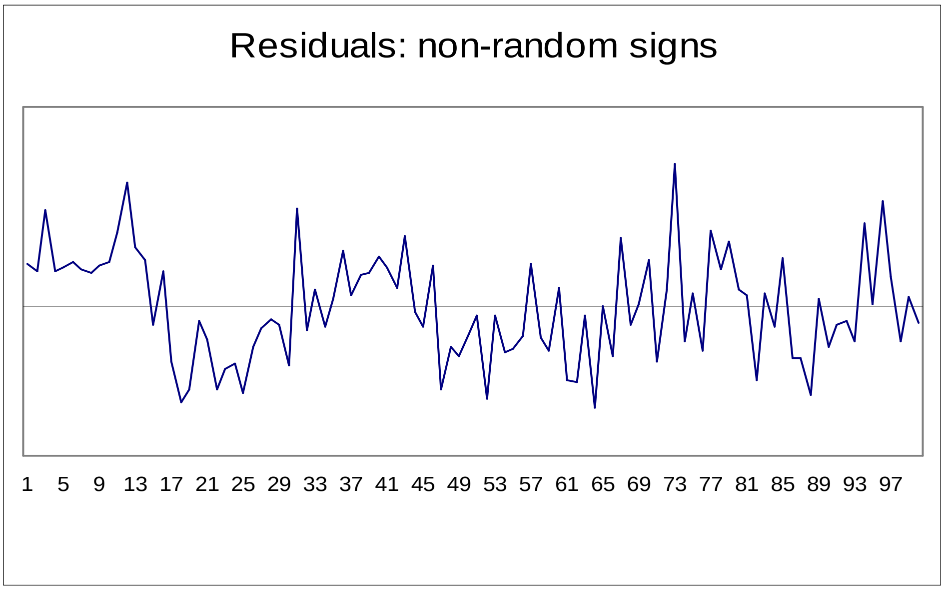
Then, increase q to 2, and fit

$$\nabla^2 x_t = (1 + \theta_1 B + \theta_2 B^2)\, a_t \quad .$$

- If overdifferenced: one of the roots of $\theta(B)$ will be close to (1-B).

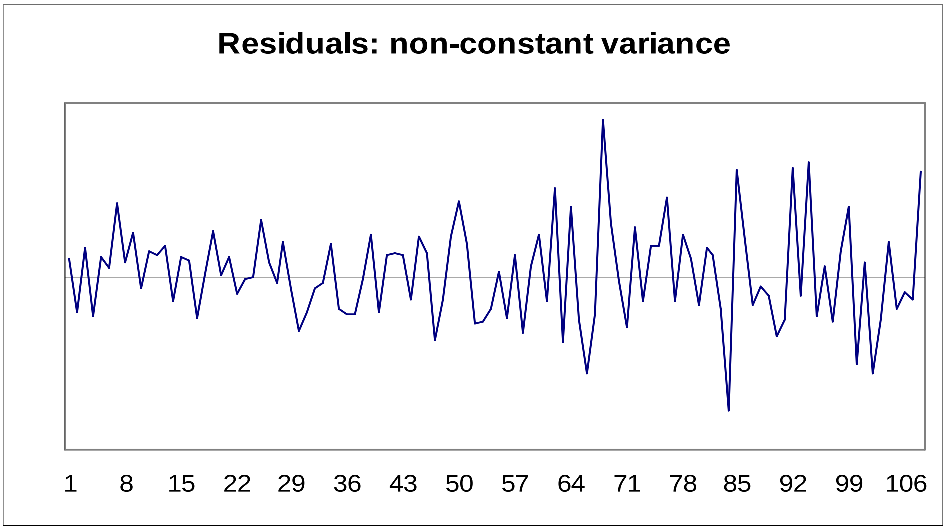
- If there is no need for q = 2, then $t(\hat{\theta}_2)$ is not significant, and q can then be reset to 1.

3. Signs are not random (lack of independece)



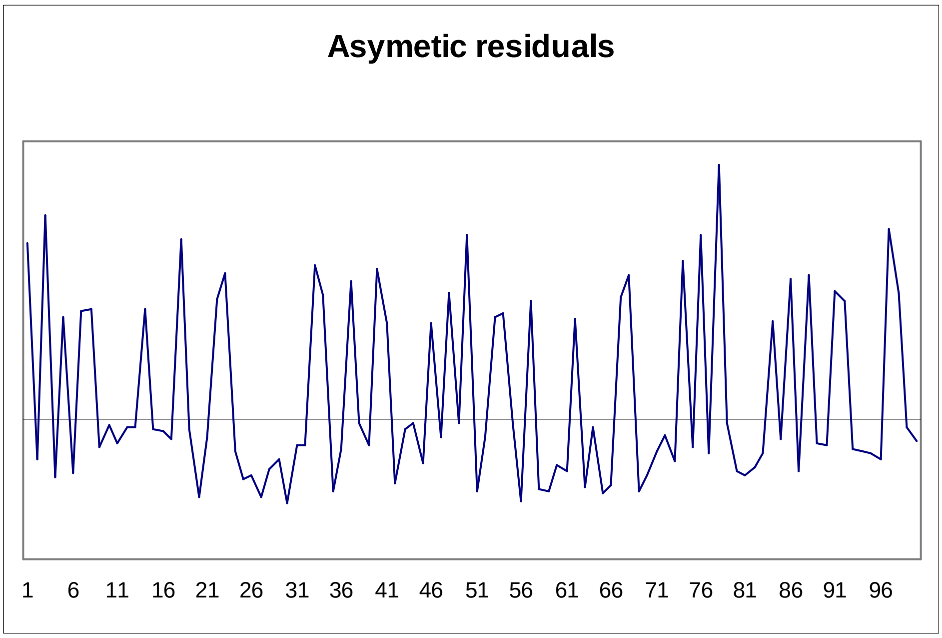
$$\left.\begin{array}{l} p(a_t > 0 \mid a_{t-1} > 0) \\ p(a_t > 0 \mid a_{t-1} < 0) \end{array}\right\} \neq p(a_t > 0)$$

4. Heteroscedastic residual variance (∴ Nonlinear model)



Variance in last half > variance in first half

5. Asymetric residuals ($\therefore$ Nonnormal)



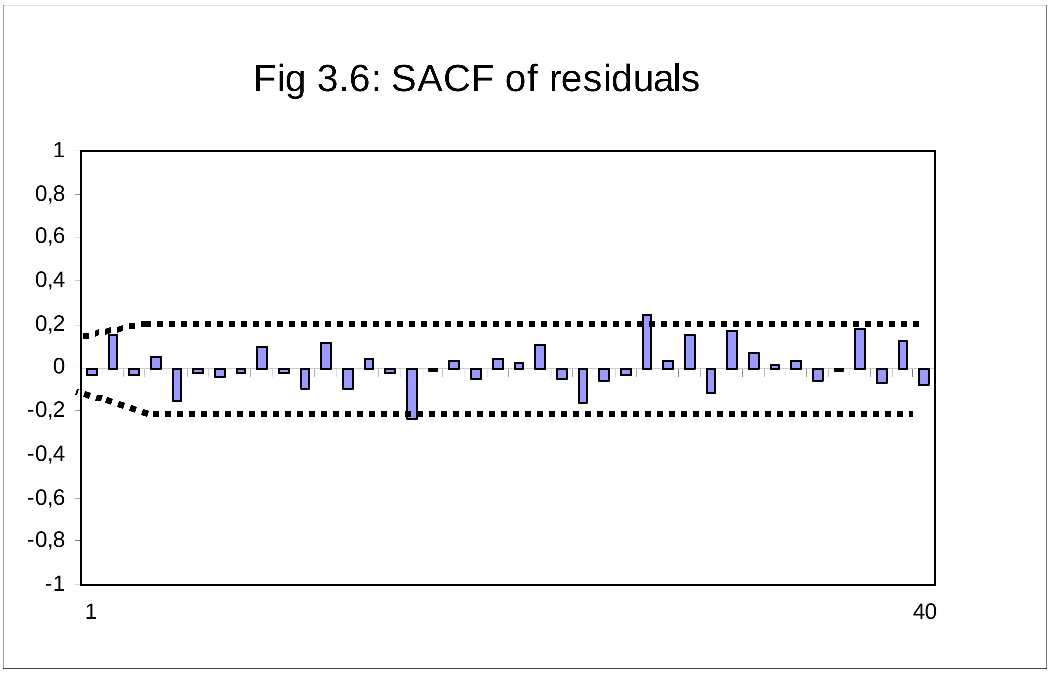
few but large $a_t > 0$;

many but small $a_t < 0$.

Remark: on occasion, problem is associated with a few years at the beginning $\rightarrow$ Remove them?

- **Examine the sample autocorrelation function (SACF) of** $[\hat{a}_t]$

$$\text{SACF}\,(\hat{a}_t) = \hat{\rho}_k\,(\hat{a}_t) \quad , \quad k = 1, 2, 3, \ldots$$



- The individual ACs $\hat{\rho}_k\,(\hat{a}_t)$ should not be significant.

For k not too small (1, 2,…) ,

If $\hat{a}_t$ is white noise (and N not too small)

$$\text{SE}\,[\hat{\rho}_k\,(\hat{a}_t)] \approx \frac{1}{\sqrt{N}}$$

Ex: significance at 95% level (roughly):

$$|\,\hat{\rho}_k\,(\hat{a}_t)\,| > \frac{2}{\sqrt{N}}$$

Note:

At 95% level, approx. 1 every 20, on average, should be expected to be significant, even when series is w.n.

Note:

For k = 1, 2,…

SE of $\hat{\rho}'s$ are smaller (see Box and Jenkins, 1970)

- The SACF ($\hat{a}_t$) should not display systematic behavior

  * in the levels

  * in the signs

Examples of problems that can be detected through inspection of the SACF $(\hat{a}_t)$:

1. Too many $\hat{\rho}_k$ $(\hat{a}_t)$ are significant

(if 95% level: more than 5%

if 99% level: more than 1%…)

$\rightarrow [\hat{a}_t]$ is not uncorrelated (w.n.)

A word of caution:

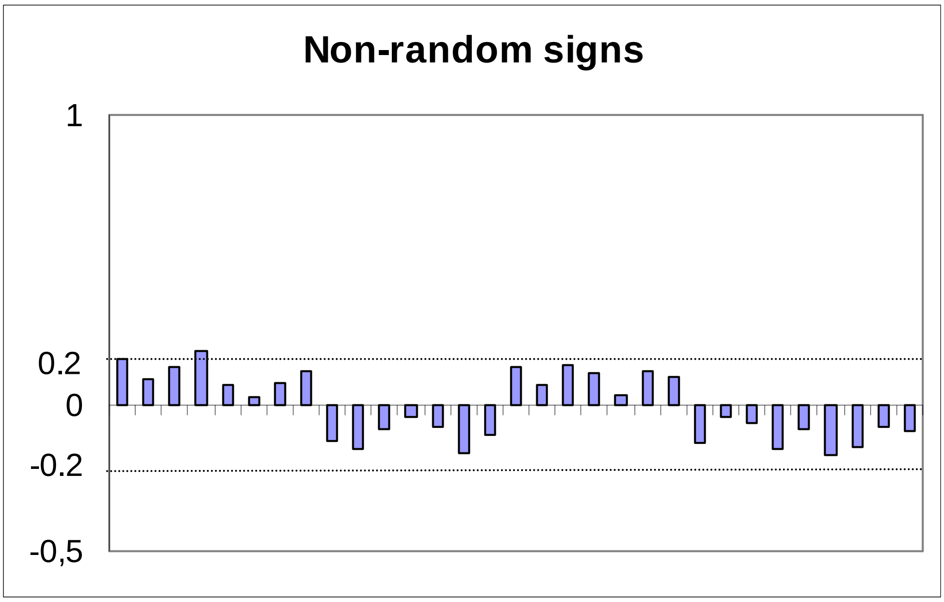
Concentrate on "meaningful" lags:

(1, 2, …, 12, …, 24)

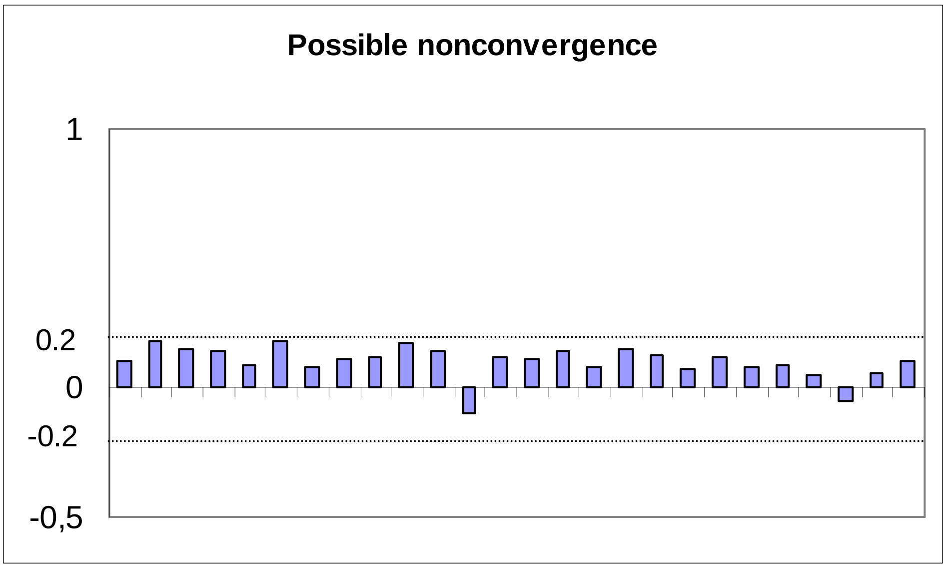
Do not pay much attention to, say,

$\hat{\rho}_{17} = .16$ (SE $=$ .07)

2. Non-random signs



3. Possible Nonconvergence



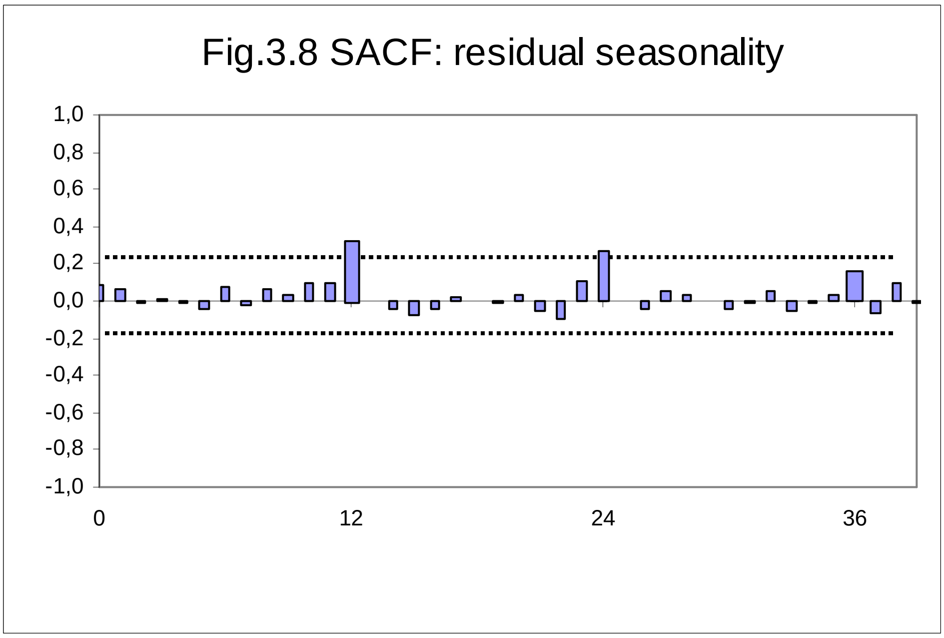
→ NS

IMA (1,1) with $\theta \approx -1$ may have similar SACF → Increase ∇ (and q) ?

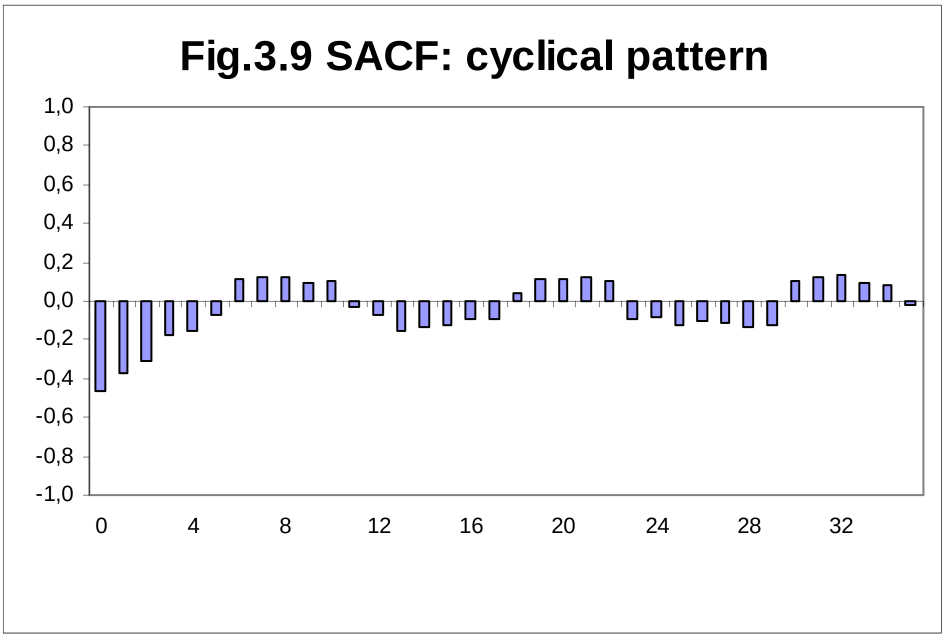
4. <u>Non-convergence for seasonal lags</u>

$\hat{\rho}_{12}, \quad \hat{\rho}_{24}, \quad \hat{\rho}_{36} \quad > 0$

$\rightarrow$ Need for seasonal differencing ∇_s? (or IMA $(1,1)_s$)



5. <u>Some hidden pattern</u>



$\rightarrow$ Cyclical behaviour ?

AR (2)

ARMA (2, 2) ?

WARNING (already mentioned)

$$\left.\begin{array}{c}\hat{\rho}_1(\hat{a}_t)\\ \vdots \\ \hat{\rho}_k(\hat{a}_t)\end{array}\right\}\text{ all depend on } [x_1,...,x_N].$$

This dependence may induce "spurious" correlation between

$$\hat{\rho}_j(\hat{a}_t) \text{ and } \hat{\rho}_i(\hat{a}_t)$$

$\rightarrow$ Appearance of systematic pattern for large lags

THUS: DO NOT OVERREAD SACF

Concentrate mostly on:

* Low-order ACs.
* Seasonal ACs

Additional diagnostics

- MEAN (residuals) = 0

Standard t-test on $\hat{\mu}[\hat{a}_t]$.

- Lack of residual AUTOCORRELATION

The Ljung-Box-Pierce (LBP) statistics Q(h) provides a "portmanteau" test for overall significance of residual AC.

$$Q(h) = N(N+2) \sum_{j=1}^{h} \hat{\rho}_a^2(j) / (N-j)$$

where $\hat{\rho}_a^2$ (j) is the j-lag SAC of the residuals

If $[\hat{a}_t]$ is residual from fitted ARMA (p, q), asympt.:

$$Q(h) \sim X^2(h-(p+q))$$

For ex.:

Monthly data (s=12) $\rightarrow$ Use h = 24.

Quarterly data (s=4) $\rightarrow$ Use h = 12 or 16.

Annual data (s=1) $\rightarrow$ Use h = 8.

If model is Airline Model (default)

$$\nabla \nabla_s x_t = (1+\theta_1 B)(1+\theta_s B^s) a_t$$

2 parameters (p + q = 2)

$\because Q(h) \sim X^2(h-2)$

Critical values (95%): Approx. ,

s = 12, h = 24 → < 34

s = 4, h = 16 → < 24

- To concentrate only on SEASONAL AC in residuals, use

$$Q_s = N\,(N+2)\sum_{j=1}^{3}\left[\hat{\rho}_{js}(\hat{a}_t)\right]^2 / (N - js)$$

(s = # obs./year)

Pierce, 76:

approx. $\sim X_2^2$

(95% : < 6)

This is approximate test for RESIDUAL SEASONALITY.

Remark:

The Q_s test only makes sense for residual seasonality if

$$\rho_s > 0$$

When $\rho_s < 0$, autocorrelation is associated with a two-year cycle.

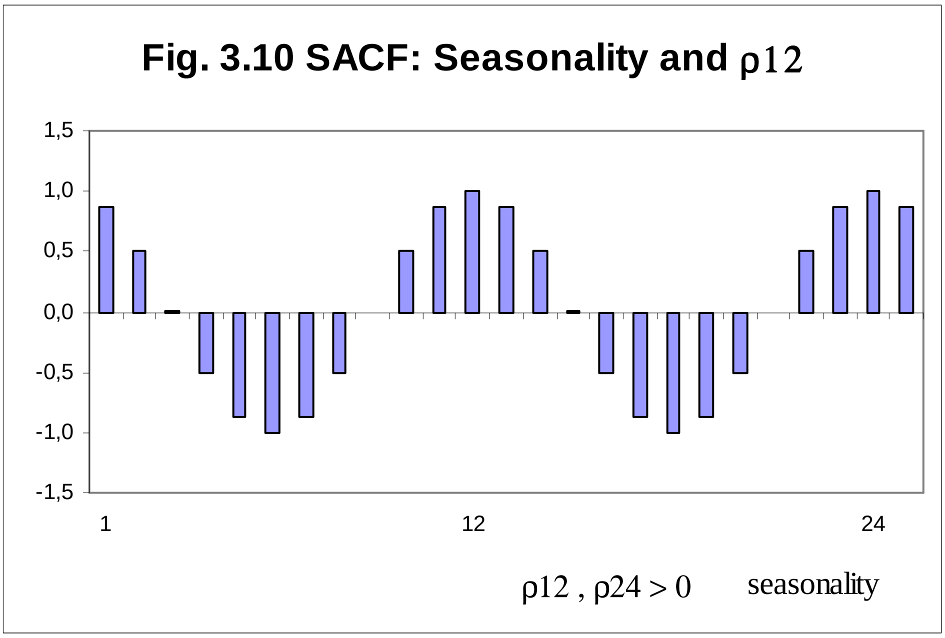


Fig. 3.10 SACF: Seasonality and ρ12
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
1
12
24
ρ12 , ρ24 > 0
seasonality

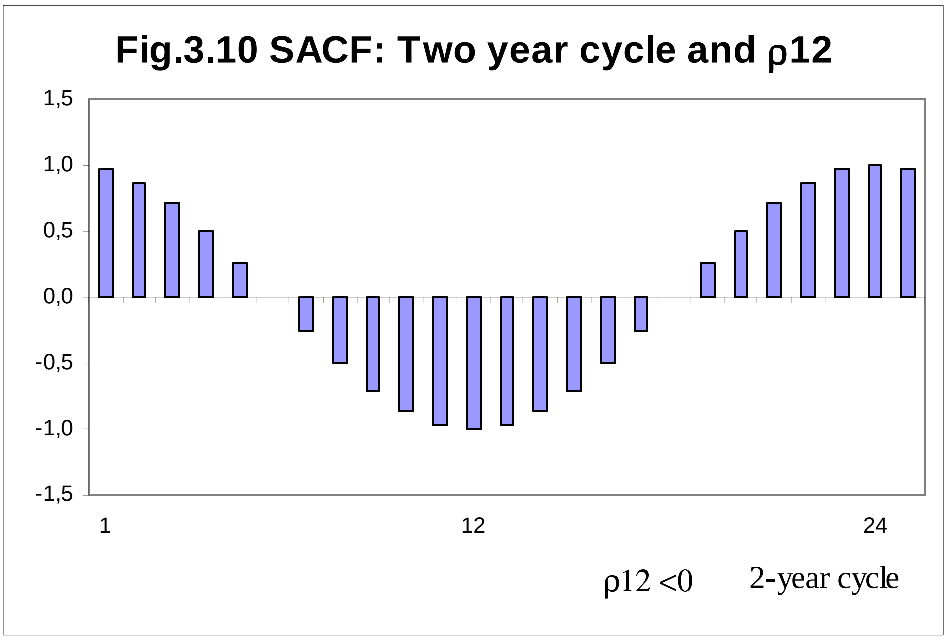


Fig.3.10 SACF: Two year cycle and ρ12
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
1
12
24
ρ12 <0
2-year cycle

(In the SEATS summary output, when $\rho_s < 0$, and hence Q_s is not associated with seasonality, the value of Q_s is made zero.)

- NORMALITY of residuals

a) <u>Skewness</u>: $m_3 = \dfrac{\sum \hat{a}_t^3}{N\hat{\sigma}_a^3}$

If Normal: Asympt.: $m_3 \sim N\left(0, SE(m_3)\right)$

$$SE(m_3) = \left(\frac{6}{N}\right)^{1/2}$$

Hence:

$$H_0 : m_3 = 0 \rightarrow |m_3| < 2\,SE(m_3)$$

b) <u>Kurtosis</u>: $m_4 = \dfrac{\sum \hat{a}_t^4}{N\hat{\sigma}_a^4}$

If Normal: Asympt. : $m_4 \sim N\left(3, SE\ (m_4)\right)$

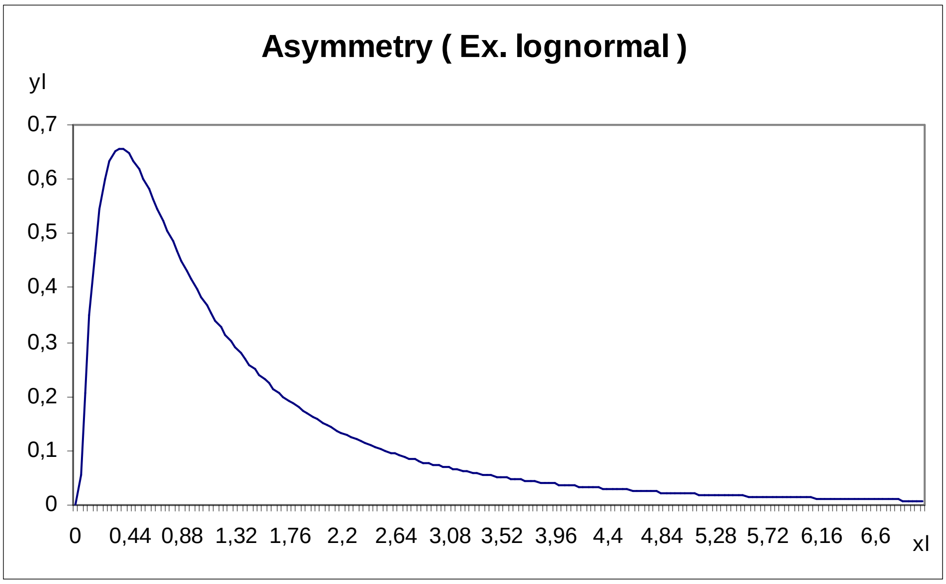
$$SE\ (m_4) = \left(\frac{24}{N}\right)^{1/2}$$

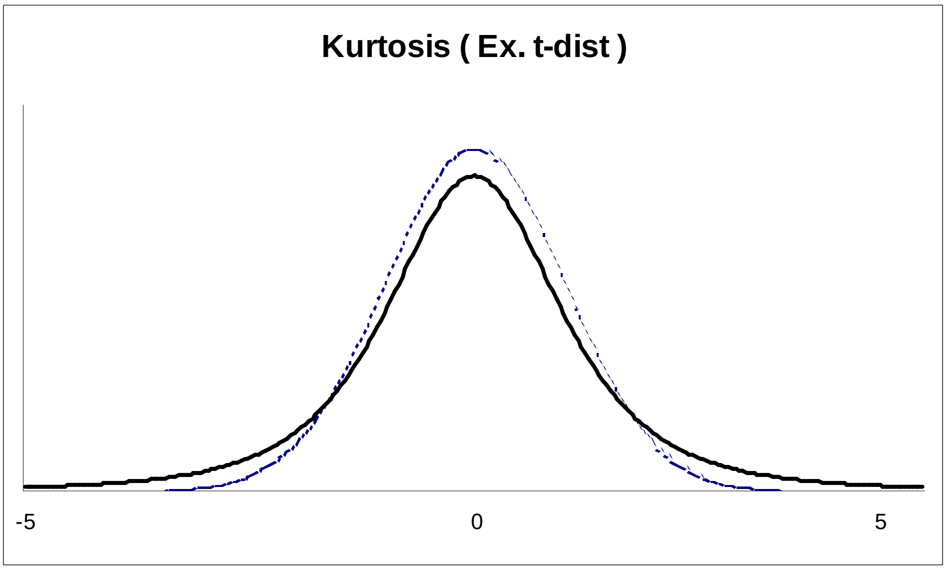
Hence

$$H_0 : m_4 = 3 \rightarrow |m_4 - 3| < 2\,SE(m_4)$$

Skewness → Asymmetry

Kurtosis → Thick tails





Remark:

Excess skewness is more damaging.

Point estimators are in general robust w.r. to excess kurtosis.

c) Joint test: (Bowman-Shenton)

$$M = \frac{N}{6} m_3^2 + \frac{N}{24} (m_4 - 3)^2$$

Under $H_0 : a_t \sim N$

asympt. : $M \sim X_2^2$

Hence, Normality rejected if M>6

- LINEARITY of the process

a) If x_t = linear stochastic process (= linear combination of Normal residuals) then (Maravall, JBES, 1983)

$$\rho_k\left(x_t^2\right) = \left[\rho_k\left(x_t\right)\right]^2, \qquad k = 1, 2, \ldots$$

Given that $\left|\rho_k\left(x_t\right)\right| < 1$,

if, for some k, $\left|\rho_k\left(x_t^2\right)\right| > \left|\rho_k\left(x_t\right)\right|$

$\rightarrow$ indication of NL

Previous test is applied to residuals, in which case, linearity implies $\rho\left(a_t\right) = \rho\left(a_t^2\right) = 0$.

For ex. : for monthly data

$\hat{\rho}_{12}(\hat{a}_t) = -.1$

$\hat{\rho}_{12}(\hat{a}_t^{\,2}) = .6$

$\rightarrow$ evidence of seasonal NL

b) Portmanteau-test: (McLeod)

LBP: $Q_h(\hat{a}_t)$ and

$Q_h(\hat{a}_t^{\,2})$ have similar asympt. $dist^n$

$Q_h\ (\hat{a}_t^2) \sim X_h^2$

$\therefore$ LBP – test can be applied in the same way to $\hat{a}_t$ and to $\hat{a}_t^{\,2}$

If $Q_h(\hat{a}_t)$ < critical X^2 value (d.f. : h-(p+q))

while

$Q_h(\hat{a}_t^{\,2})$ > critical X^2 value (d.f. : h)

$\rightarrow$ indication of NL

(also, non-normality)

Remark:

Test on a_t^2 works well when

H_{ALT} : a_t = ARCH, GARCH, …, Bilinear, …

These nonlinear structures are likely to have a small effect on point estimation.

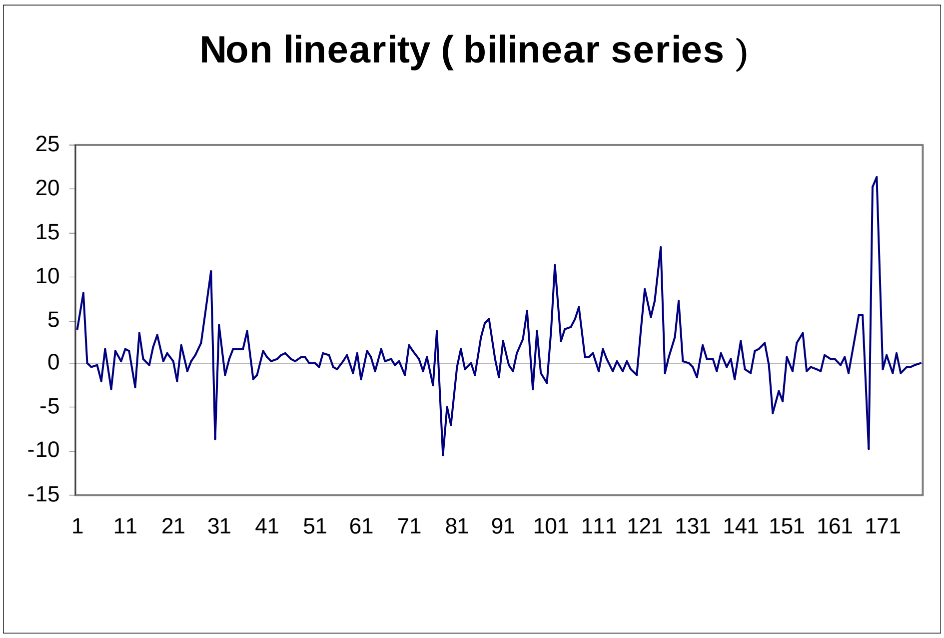
They indicate that $Var(a_t)$ changes with time

(the ARIMA estimation $\rightarrow$ "an average").

We learn more about the series, but our point estimates of

- ϕ, θ - param.
- Forecasts
- Seasonal component
- etc.,

will remain acceptable. The non-linearity affects mostly the SE of the point estimates.



- RANDOMNESS of residuals:

"Run" test

(a run: a sequence of + + + … or of - - -).

Asympt. ~ N (0, 1)

hence < 2 in abs value.

2. Out-of-sample Forecast test

Needs a separate run of TRAMO.

Setting

NBACK = - K (K > 0)

TRAMO fits model for first

(N – K) observations.

Then, with parameters fixed, it

- computes sequentially 1–period–ahead forecasts for the K periods removed;
- obtains the forecast errors incurred.

Let

N_r = number of residual

= N – d – # of M.O. – dof lost because of estimation of regression variables and outliers.

V_a : in-sample residual variance (N_r – K residuals).

V_0 : out-of-sample forecast-error variance (K forecast errors).

Under

H_o : Model is valid for out – of – sample period,

$$\frac{V_0}{V_a} \sim F(K, N_r - K)$$

Hint:

s = 12 → Use K = 12 (short series), 18, 24 (long series).

$$\left.\begin{array}{l} s = 4 \\ s = 1 \end{array}\right\} \rightarrow \text{Use} \quad K = 8$$

Test makes more sense as a one-sided test (if out-of-sample variance is smaller, not much reason to reject model).

Remark: At present, RSA parameter cannot be used with NBACK < 0.

If out-of-sample testing is desired for the automatic procedure, the parameters have to be entered 1 by 1.

Ex. :	INIC = 3	IDIF = 3	(AMI)
	IATIP = 1		(AODC)
	LAM = -1		(LOG – LEVEL)
	ITRAD = -1		(TD)
	….		

The automatic procedure is applied in this case to the first $(N_r - K)$ observations. The model obtained is used for the out-of-sample forecasts.

Hint:

Main criteria for choosing among models:

- Out-of-sample forecast test (often, indecisive)
- BIC $\rightarrow$ minimum;
- MSE $(\hat{a}_t) \rightarrow$ minimum. But care with overparametrization;
- Q $(\hat{a}_t) \rightarrow \langle$ critical value;
- (Joint Normality test; Skewness).

Additional considerations:

- Parsimony (few parameters);
- Balanced models (order of total AR = order of total MA)

If TRAMO is to be used with SEATS, avoid Q>P if possible, or P>>Q (chances that the model has no admissible decomposition increase).

SEATS will provide additional criteria for choosing among models.

Remark:

TRAMO allows the user to fix some (or all) model parameters.

Before fixing parameters, the **autocorrelation matrix of the parameter estimates** should be considered.

For example, assume we are estimating an IMA (2, 2) and the t-values are,

$t\left(\hat{\theta}_1\right) = .97$

$t\left(\hat{\theta}_2\right) = 2.01$

and that we decide to set, in a second round, $\theta_1 = 0$.

If $\rho\left(\hat{\theta}_1, \hat{\theta}_2\right) =$ high and positive, then setting $\theta_1 = 0$ may well yield $t\left(\hat{\theta}_2\right) < 1.96$, and no parameter is then significant.

But now, the Q-statistics may show some autocorrelation in the residuals.

If, on the contrary,

$$\rho\left(\hat{\theta}_1, \hat{\theta}_2\right) < 0,$$

then, setting $\theta_1 = 0$ will make $\hat{\theta}_2$ more significant.

Having passed the diagnostics, we can proceed to **inference**.

A particularly important application is FORECASTING.

Forecasting ARIMA Models

Two kinds of forecasts can be considered.

- Conditional forecasts or infinite memory forecasts.
- Finite memory forecasts.

a) **Infinite memory forecasts** are easier to compute. They are based on the assumption that the semi-infinite sample $\{z_j : j \leq t\}$ is available to forecast the observation z_{t+1}, although in practice only a finite series $z = (z_1, ..., z_t)'$ is available.

The procedure to obtain these forecasts is the following.

Let:

$\hat{z}_{N+j|N}$ = Forecast of z_{N+j} made at N =

$$E(z_{N+j} | z_N, z_{N-1}, z_{N-2}, ...) = \text{(or, in compact form)}$$
$$= E_N z_{N+j} \quad .$$

We use as example ARIMA (1, 1, 2)

$$(1 + \phi B) \nabla z_t = (1 + \theta_1 B + \theta_2 B^2) a_t .$$

Denote:

$$\phi(B) = 1+\phi_1 B+\phi_2 B^2 = (1+\phi B)(1-B)$$

$$= \text{Total AR (including U.R.)}$$

Model:

$$z_N+\phi_1 z_{N-1}+\phi_2 z_{N-2} = a_N+\theta_1 a_{N-1}+\theta_2 a_{N-2},$$

or:

$$z_{N+1} = -\phi_1 z_N-\phi_2 z_{N-1}+a_{N+1}+\theta_1 a_N+\theta_2 a_{N-1}$$

Given that,

$E_N\ a_{N+j} = 0, \qquad j>0,$

(future innovations cannot be forecasted)

$E_N\ a_{N+j} = \hat{a}_{N+j} \qquad j \le 0,$

(past innovations are approximately equal to the sequence of residuals, or of the 1-period-ahead forecast errors already made,)

then $E_N\ z_{N+1}$ can be expressed as

$$\hat{z}_{N+1|N} = -\phi_1 z_N-\phi_2 z_{N-1}+\theta_1 \hat{a}_N+\theta_2 \hat{a}_{N-1} \quad ,$$

↑______↑ ↑______↑

observed available residuals

Similarly, $E_N\ z_{N+2}$ is given by

$$\hat{z}_{N+2|N} = -\phi_1 \hat{z}_{N+1|N} - \phi_2 z_N + \theta_2 \hat{a}_N \quad ,$$

and, for $j > 2$ ($= q$, order of the MA part of the model)

$$\hat{z}_{N+j|N} = -\phi_1 \hat{z}_{N+j-1|N} - \phi_2 \hat{z}_{N+j-2|N}$$

or the finite difference equation

$$\phi(B)\, \hat{z}_{N+j|N} = 0$$

with B operating on j.

This difference equation provides the forecasts for horizons larger than q periods ("Eventual" Forecast Function).

Therefore, forecasts can be computed recursiverly.

Note:

For an ARIMA model

$$\phi(B)x_t = \phi(B)a_t \quad ,$$

where $\phi(B)$ may include unit roots,

* the eventual ACF follows

$$\phi(B)\rho_k = 0 \quad ,$$

with B operating on k (k > q) ;

* the eventual forecast function (FF) follows

$$\phi(B)\hat{x}_{T+j|T} = 0 \quad ,$$

with B operating on j (j > q) ;

Therefore, the eventual ACF and FF display the same dynamic behavior
(we forecast the correlations, so to speak).

b) **Finite Memory Forecasts**

If $\hat{z}_{N+k|N}$ denotes the finite memory forecast of z_{N+k} made at N, then

$$\hat{z}_{N+k|N} = \Sigma_{k,N}\, \Sigma_N^{-1} z,$$

where z is the finite series, $\sum_{k,N} = Cov(z_{N+k}, z)$ and $\sum_N = Var(z)$.

It can be obtained recursively with, for example, the Kalman filter.

Given an ARIMA (p, d, q) model, the Eventual Forecast Function is verified by both infinite and finite forecasts

$$\hat{z}_{N+k|N} = -\phi_1 \hat{z}_{N+k-1|N} - \ldots - \phi_{p+d} \hat{z}_{N+k-p-d|N}, \qquad k > q,$$

where $1 + \phi_1 B + \ldots + \phi_{p+d} B^{p+d}$ is the autoregressive polynomial, which includes the unit roots, and it is understood that

$\hat{z}_{N-j} = z_{N-j}$ if $j \geq 0$.

Therefore, the Eventual Forecast Function is driven by the roots of the AR polynomial.

Confidence intervals

Let $\hat{z}_{N+k|N}$ be the forecast of z_{N+k} based on the finite series $z = (z_1, ..., z_N)'$ and let $\sigma^2_{N+k|N} = \text{Mse}\left(\hat{z}_{N+k|N}\right)$. Then, assuming normality, the usual confidence intervals are obtained as

$$\left(\hat{z}_{N+k|N} - c\sigma_{N+k|N}, \hat{z}_{N+k|N} + c\sigma_{N+k|N}\right),$$

where, for example, $c = 1.96 \approx 2$ (95% confidence level).

If the series has been transformed by taking logs, confidence intervals are obtained as

$$[\exp\left(\hat{z}_{N+k|N} - c\sigma_{N+k|N}\right), \exp\left(\hat{z}_{N+k|N} + c\sigma_{N+k|N}\right)]$$

Remarks on Forecasting

a) TRAMO uses exact projection forecasts (on finite sample) $\rightarrow$ KF.

SEATS uses projections on semi-infinite sample $\rightarrow$ BJ.

If sample size (N) is not too small, both forecasts are very similar.

b) ARIMA forecasts are useful as short-term forecasts.

For monthly or quarterly series, maximum horizon $\approx$ 1 (2?) year.

Good short-term behaviour due to flexibility and adaptive nature.

Long-term extrapolation of this flexibility induces long-term forecast instability (see Maravall, 2000).

A Remark on Forecast Errors

Consider model:

$$\phi(B)\, x_t = \theta(B)\, a_t \quad ,$$

and express it in terms of the Psi-weights

(= “pure” MA representation)

$$x_t = \left[\frac{\theta(B)}{\phi(B)}\right] a_t$$

$$x_t = \psi(B)\, a_t =$$
$$= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots$$

Consider error in forecast of x_{N+k} made at period N:

$$e_{N+k|N} = x_{N+k} - \hat{x}_{N+k|N} \quad .$$

At the end of Part I of these notes, we saw that

$$e_{N+k|N} = a_{N+k} + \psi_1 a_{N+k-1} + \ldots + \psi_{k-1} a_{N+1} \quad .$$

Denote by $V_e(j)$ the variance of the j-th periods-ahead forecast

$$V_e(j) = \mathrm{Var}\,(e_{N+j\,|\,N}) \quad .$$

Then,

$$V_e(1) = V_a$$

$$V_e(2) = (1+\psi_1^2)\ V_a$$

$$V_e(3) = (1+\psi_1^2+\psi_2^2)\ V_a$$

and so on.

If model is stationary (ARMA), as $j \rightarrow \infty$

$$V_e\ (j) = \left(1+\sum_{i=1}^{j-1}\psi_i^2\right) V_a \qquad \rightarrow \qquad V_x < \infty \quad .$$

If model is non-stationary (ARIMA), as $j \rightarrow \infty$

$V_e\ (j)$ increases without bound, because the sum $\sum_{i=1}^{j-1}\psi_i^2$ does not converge.

EXTENSIONS OF THE ARIMA MODEL

1. Missing observations

2. Regression Variables
    2.1. Calendar Effects.
        a) Trading Day effect
        b) Easter Effect
        c) Leap-Year Effect
        d) Holidays
    2.2. Intervention variables generated by the program
    2.3. Regression variables entered by user.

3. Outliers.

REGARIMA model:

ARIMA model + those extensions.

MISSING OBSERVATIONS

Kalman filtering with missing observations and no regression parameters.

There are two ways to treat M.O. in TRAMO:

a) *Skipping approach*:

It uses the original series and computes the likelihood "jumping" over the periods for which there is no observation.

b) *Additive outlier approach*: It uses the differenced series

(previously the missing observations are replaced with arbitrary values; Gómez, Maravall, and Peña, 1999).

a) **Skipping (SK) approach**. If the observed series is $z = \left(z_{t_1},...,z_{t_M}\right)'$, where $0 < t_1 < ... < t_M$, the likelihood is defined as

$$l(z) = l\left(z_{t_M},...,z_{t_{d-k+1}} \mid z_{d-k},...,z_1\right),$$

where we have assumed that there are k missing observations among the first d periods.

If k > 0 and there are some M.O. among these first values $\{z_1,...,z_d\}$, they are treated as fixed parameters and estimated along with the other regression parameters of the model.

Starting conditions for the K.F. are obtained similarly to when there are no M.O. The K.F. is used like in Jones (1980) for prediction and likelihood evaluation (computation of the likelihood "skips" the periods for which observations are missing). The state-space representation for the nonstationary levels is employed.

For interpolation, a simplified version of the Kalman fixed-point smoother (FPS) is used. See the paper of Gómez and Maravall (1994).

b) **Additive outlier (AO) approach**. The holes corresponding to the missing values are filled by linear interpolation of the last observation before and the first observation after the (perhaps a sequence of) missing value(s). Then, an additive outlier is specified for each missing observation.

Having assigned tentative values to the M.O., the differenced series can be used for likelihood evaluation.

It can be shown that the skipping approach is equivalent to the additive outlier approach if a correction is made in the determinental term of the likelihood of the latter.

(See Gómez, Maravall, and Peña, J. Econometrics 1999).

Remark:

1. When there are not many M.O., it is simpler to use the Additive Outlier approach.
2. In the AO approach, however, given that the M.O. are treated as A.O. through regression, the total number of M.O. is limited by the maximum number of regression variables allowed by the program (around 30).
3. If the number of M.O. exceeds this limitation, the SK approach (with the KF and FPS) has to be used.
4. AMI (+AODC) requires the AO approach.

Thus, if there are many M.O. the automatic features of TRAMO cannot be directly applied.

The following two-step procedure often works well.

Automatic modelling with many M.O:

If there are more M.O. than those allowed by the AO approach,

as is typically the case when interpolating higher frequency data

(ex.: estimate monthly values from quarterly or annual observations),

to use the automatic modelling features, one can proceed as follows:

* Series with seasonality

a) Use SK approach first using the default (Airline) model (perhaps with $\theta_1 = -.4$, $\theta_s = -.6$).

b) On the interpolated series, perform AMI + AODC. This produces a new model.

c) With the new model use the SK approach to obtain new interpolations.

d) Go back to b) and iterate (2 or 3 iterations usually are enough).

* Series with no seasonality

Replace the default model by an IMA (1,1) –perhaps with mean– model, and proceed as before.

REGRESSION VARIABLES

Kalman filtering with regression parameters

a) **Stationary series**.

Suppose the observed series $z = (z_1,...,z_N)'$ follows the model.

$$z = Y\beta + x,$$

where $x = (x_1,...,x_N)'$ and

$$\phi(B)\, x_t = \theta(B)\, a_t \quad , \qquad a_t \overset{i.i.d}{\sim} N(0, \sigma^2) \quad .$$

Let $Var(x) = \sigma^2\Omega = \sigma^2 LL'$, $L \equiv$ lower triangular.

Premultiplying $(z = Y\beta + x)$ by L^{-1},

$$L^{-1}z = L^{-1}Y\beta + L^{-1}x \qquad (1)$$

Given that $L^{-1}x$ (= a_t) is white noise, previous equation is a standard regression model for which OLS are efficient.

The QR algorithm applied to $L^{-1}Y$ yields an orthogonal matrix Q such that.

$Q'L^{-1}Y = \begin{pmatrix} R \\ 0 \end{pmatrix} \Rightarrow$ upper triangular with nonzero elements in the main diagonal.

Partitioning $Q' = \begin{pmatrix} Q_1' \\ Q_2' \end{pmatrix}$ conforming to $Q'L^{-1}Y = \begin{pmatrix} R \\ 0 \end{pmatrix}$, it is obtained that

$$\begin{aligned} Q_1'L^{-1}\ z &= R\beta + Q_1'L^{-1}\ x \\ Q_2'L^{-1}\ z &= \qquad Q_2'L^{-1}\ x \end{aligned}$$

From this, $\hat{\beta} = R^{-1}\ Q_1'L^{-1}\ z$ and $\hat{\sigma}^2 = z'\left(L^{-1}\right)' Q_2 Q_2' L^{-1}\ z/(N-k)$,

where k is the number of elements in β.

(Q $\Rightarrow$ Housholder transformation.)

If in model $z = Y\beta + x$, the K.F. is applied to z and to the columns of Y, one obtains $L^{-1}\,z$ and $L^{-1}Y$. Therefore, aplication of the Kalman filter transforms the GLS model

$$z = Y\beta + x$$

into the OLS model

$$L^{-1}\ z = L^{-1}Y\beta + L^{-1}x$$

b) **Nonstationary series** $\Rightarrow$ Similar

See the papers of Gómez and Maravall (1994), and Gómez, Maravall and Peña (1999) for the case when there are M.O.

Parameter estimation. In the regression model

$$z = Y\beta + x,$$

β and σ^2 can be concentrated out of the likelihood. The parameters of the ARMA model are estimated using Gauss-Marquardt's method.

The function to minimize is

$$S = |L|^{1/N} z'\left(L^{-1}\right)' Q_2 Q_2' L^{-1} z |L|^{1/N} = r'r$$

Some ARMA parameters can be fixed.

In summary, in model:

$$z_t = y_t' \beta + x_t$$

$$\phi(B)\, x_t = \theta(B)\, a_t$$

- Regression parameters β are estimated by GLS (conditional on the ϕ and θ ARIMA parameters).
- New estimators of ϕ and θ are then obtained by EML (conditional of the previous $\hat{\beta}$ estimators).
- New estimators of $\hat{\beta}$ are obtained by GLS (conditional on the new ϕ and θ ARIMA parameters).
- Iterate until convergence.

(Notice that the KF directly transforms the GLS problem into a simple OLS one).

CALENDAR EFFECTS

These are important cases of regression variables

– linked to the calendar of each year, and

– generated by the program.

Trading Day Variables

Traditionally, six variables have been used to model the trading day effect. These are: (no. of Mondays) – (no. of Sundays), …., (no. of Saturdays) – (no. of Sundays).

The motivation for using these variables is that it is desirable that the sum of the effects of each day of the week cancel out. Mathematically, this can be expressed by the requirement that the trading day coefficients $\beta_j, \quad j = 1,...,7$, verify

$$\sum_{j=1}^{7} \beta_j = 0$$

which implies $\beta_7 = -\sum_{j=1}^{6} \beta_j$.

There is the possibility of considering a more parsimonious modeling of the trading day effect by using one variable instead of six. In this case, the days of the week are first divided into two categories: working days and non-working days. Then the variable is defined as (no. of (M, T, W, Th, F)) – (no. of Sat, Sund) x 5/2).

Again, the motivation is that it is desirable that the trading day coefficients $\beta_j, \quad j = 1,...,7$, verify

$$\sum_{j=1}^{7} \beta_j = 0$$

Since $\beta_1 = \beta_2 = ... = \beta_5$ and $\beta_6 = \beta_7$, we have $5\beta_1 = -2\beta_6$.

The set of TD variables may also include the

<u>Leap-Year Effect</u>

Leap-year variable	=	0 for months that are not february.
	=	-.25 for februaries with 28 days.
	=	.75 for februaries with 29 days (Leap-year).

(Over a four-year period the effect cancels out).

The choice of the TD specification is made by the user (parameter ITRAD).

If workdays have roughly similar effects (ex.: some IPI series) and the series is not long, the parsimonious specification (working/non working day) that uses 1 parameter is often preferable (gain in d.of f.).

For some series (ex.: currency series) Mondays have very different effects from tuesdays, and so on. In this case, the 6 variable specification is preferable.

<u>Holidays</u>

A file with the holidays can be treated by TRAMO in two ways:

– As part of the TD variable (parameter IUSER = -2). In this case:

  holiday = an additional nonworking day.

– As a separate regression effect. (If coefficients are close, better to incorporate it as part of the TD).

Easter Variable

The variable models a constant change in the level of daily activity during the d days before Easter. The value of d can be supplied by the user (by default IDUR = 6).

The variable has zeros for all months different from March and April.

The value assigned to March is equal to $p_M – m_M$, where p_M is the proportion of the d days that fall on that month and m_M is the mean value of the proportions of the d days that fall on March over a long period of time.

The value assigned to April is $p_A – m_A$, where p_A and m_A are defined analogously.

Usually, a value of $m_M = m_A = 1/2$ is a good approximation.

Since $p_A – m_A = 1 – p_M – (1 – m_M) = - (p_M – m_M)$, the sum of the effects of both months, March and April, cancel out, a desirable feature.

Remarks:

- The duration of the effect is controlled by the parameter IDUR = d days. Its value typically ranges between 3 and 8.

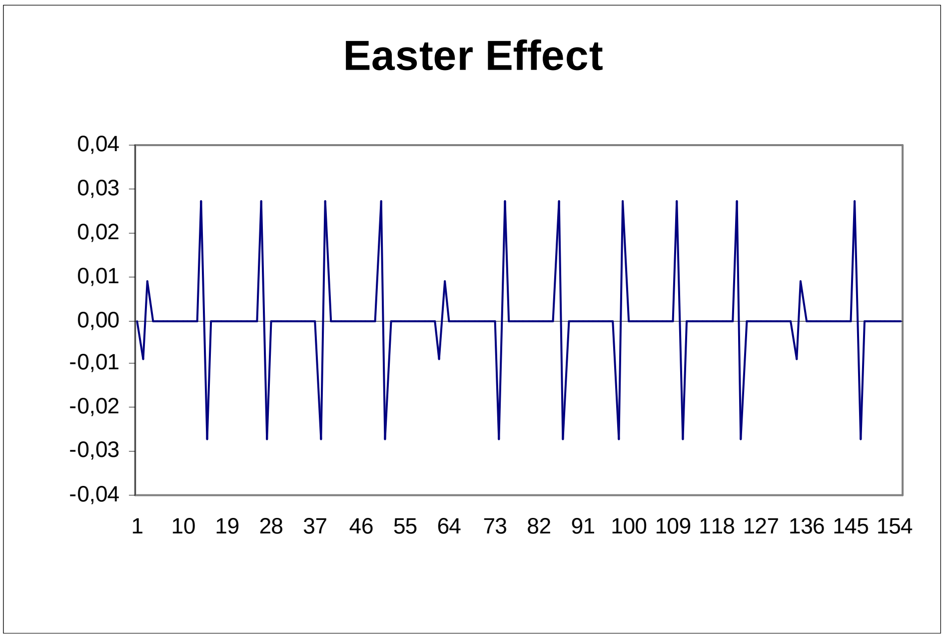
- TRAMO has a facility for pretesting for the presence of TD, LY, and EE (later).

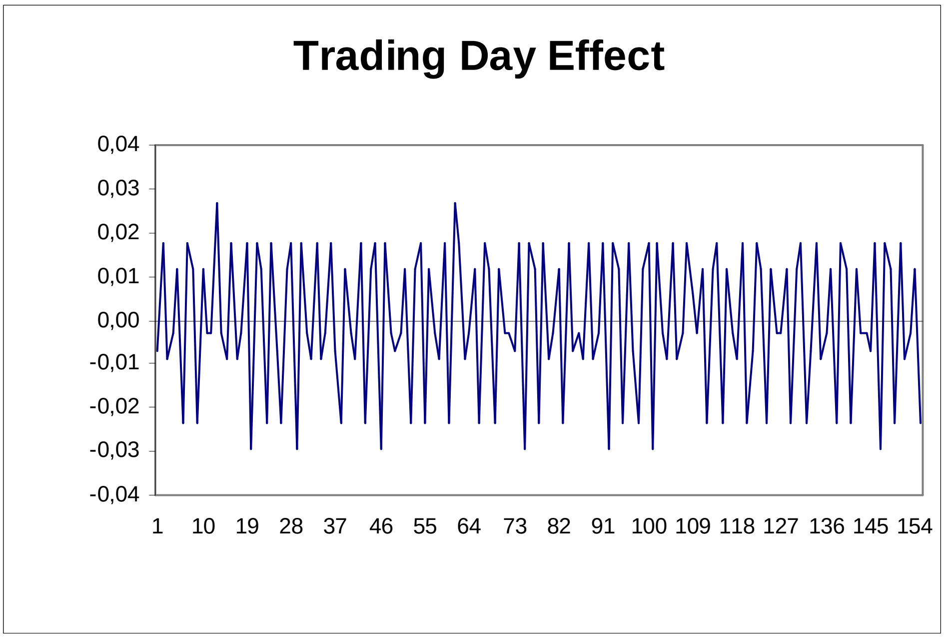
- TD effects typically characterize series reflecting economic activity.

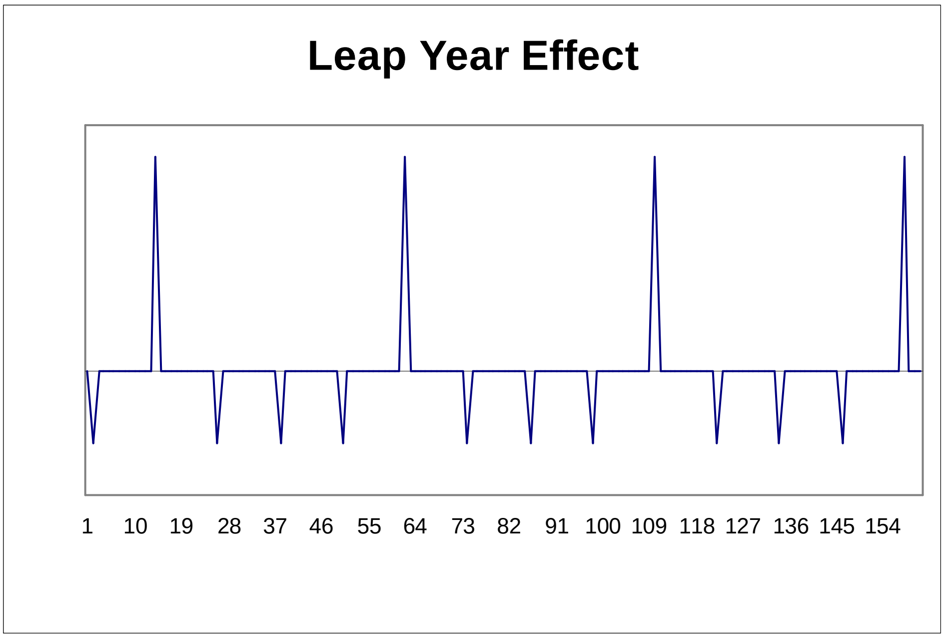
- If TD is significant, most often adding the holidays variable improves significantly the results.

- If forecasts of the series are desired, the holiday regression variable has to be extended over the forecasting horizon (2 years when TRAMO is used with SEATS).

- Given that the holidays vary among countries (for example, in Spain it may be appropriate to separate national festivities from the festivities of the autonomous regions), it is strongly recommend that each country builds its own (set of) holiday regression variable(s), extended over the forecast period.


Easter Effect
0,04
0,03
0,02
0,01
0,00
-0,01
-0,02
-0,03
-0,04
1 10 19 28 37 46 55 64 73 82 91 100 109 118 127 136 145 154

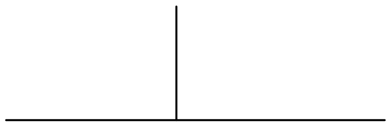

Trading Day Effect
0,04
0,03
0,02
0,01
0,00
-0,01
-0,02
-0,03
-0,04
1 10 19 28 37 46 55 64 73 82 91 100 109 118 127 136 145 154


Leap Year Effect
1
10
19
28
37
46
55
64
73
82
91
100
109
118
127
136
145
154

INTERVENTION VARIABLES

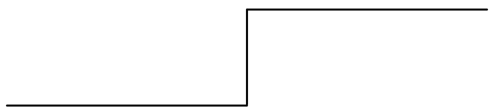
Intervention variables are regression variables used to model certain abnormal effects, such as strikes, major changes in economic policy, natural disasters, etc. (see Box and Tiao, 1975). Examples of intervention variables are

– impulses

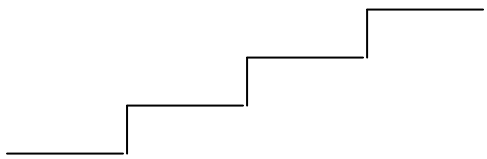


or

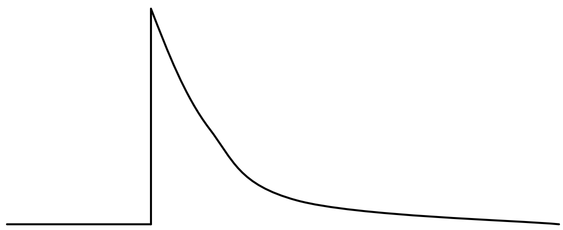
– level shifts



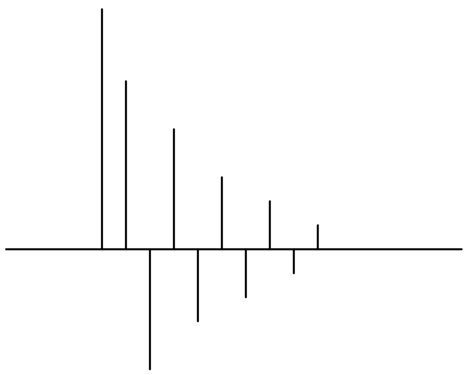
or



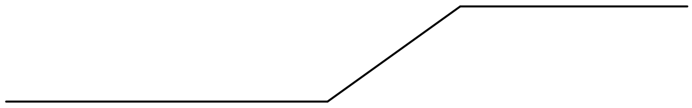
– Temporary changes



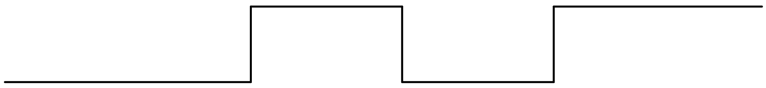
or



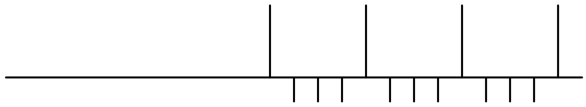
– ramps



– alternance of regimes



– seasonal outlier



These paterns (and many more) can be parametrized in TRAMO as intervention variables.

An intervention variable is modelled as any possible sequence of ones and zeros, on which some operators may be applied. In particular, these operators are of the type:

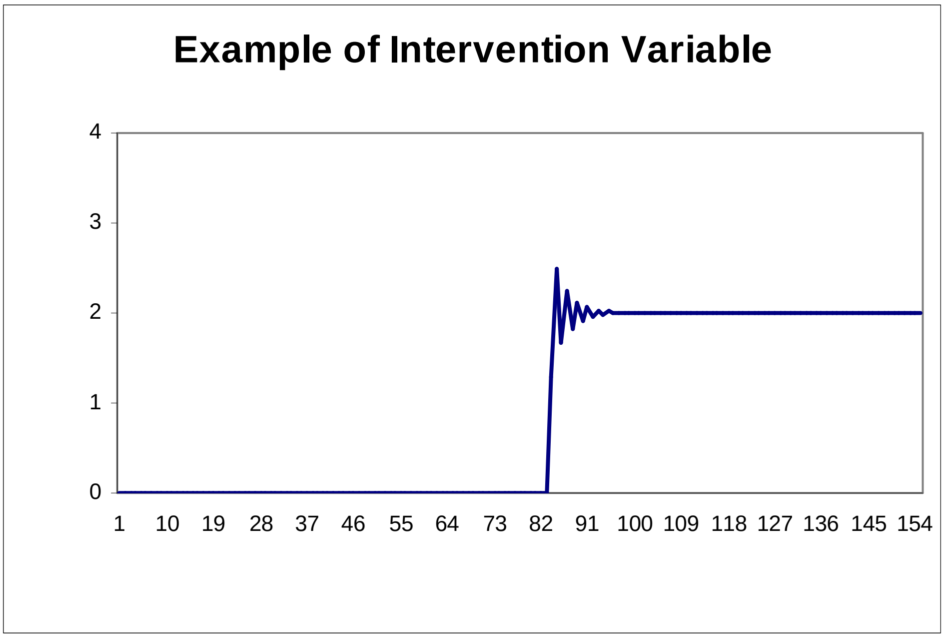
a) $1/(1 - \delta B)$, where $0 < \delta \leq 1$;

b) $1/(1- \delta_S B^S)$, where $0 < \delta_S \leq 1$;

c) $1/[(1-B)(1-B^S)]$

A rich set of intervention variables can be generated in this way.

Thus, for example, the combination of a) with $\delta = 1$ and with $\delta = -.7$ would produce a level shift that is reached after a sequence of (damped) overshooting/undershooting.



Intervention variables often are:

– variables reflecting some special event:

1) with known date of occurrence;

2) whose effect has an "a priori" (pehaps roughly known) dynamic pattern (permanent or transitory …)

Remark:

All outliers included in Automatic Outlier Detection,

AO, TC, LS

can be directly imposed as intervention variables.

OTHER REGRESSION VARIABLES

Regression variables can also be entered by the user.

– weather, temperature,

– (socio) economic variables related to the series (explanatory variables, leading indicators,…)

Example:

Monet. Aggreg. = f (political elections);

Exports = f (foreign income and exchange rate)

Sales of air conditioners = f (temperature).

At present, Transfer Function (or Distributed Lag) models (with lagged values of regression variables) can be estimated with TRAMO as:

$$z_t = \sum_i \omega_i(B) r_{it} + y_t' \beta + x_t$$

ARIMA

Other reg. var. (TD, EE, Outliers, Interv. var.,…)

Set of explanatory variables entered by user (with a distributed lag).

The effect of r_{it} is spread over time:

Ex:

$$\omega_1(B) = \omega_0^{(1)} + \omega_1^{(1)} B + \omega_{12}^{(1)} B^{12}$$

$\rightarrow$ instantaneous effect + 1-month lagged effect + 1-year lagged effect.

$$\omega_2(B) = \omega_1^{(2)} B + \omega_2^{(2)} B^2 + \omega_3^{(2)} B^3$$

$\rightarrow$ Leading indicator with the effect lagged 1 month, and spread over 3 months.

Remark:

If available series is:

$[z_1, ..., z_N]$

and forecasts are seeked for

$t = N+1, ..., N+FH$ $\leftarrow$ FH: Forecast horizon,

then:

all regression variables have to extend from

t = 1 to t = N + FH

(parameter ILONG)

If the regression variable is stochastic (for ex., an economic variable entered by user), future values of the variable (to cover the forecasting period) can be obtained by running TRAMO first (perhaps in an automatic manner) on the explanatory variable.

OUTLIERS

In addition to possible gross errors, such as a misplaced point or an erroneous dating, time series data are often subject to the influence of some special, often nonrepetitive, events; for example: implementation of a new regulation, major political or economic changes, modifications in the way the variable is measured, ocurrence of a strike or a natural disaster, etc.

Consequently, discordant observations and various types of abrupt changes are often present in time series data.

TRAMO uses an improved Chen and Liu-type procedure for outlier detection and correction (see Gómez and Maravall, 2001).

The effect of an outlier is modelled by

$$z_t^* = \omega \upsilon(B)\, I_t(T) + z_t$$

where $\upsilon(B)$ is a quotient of polynomials in B, that models the type of outlier (its effect over time), I_t (T) is an indicator variable signaling the time (T) of ocurrence of the outlier,

$$I_t(T) = \begin{cases} 1 & \text{if} \quad t = T \\ 0 & \text{otherwise} \end{cases}$$

ω represents the impact of the outlier at time T ("instantaneous impact") and z_t is the outlier-free series, which follows the model (when no other regression variables are present).

$$\phi(B)\delta(B)z_t = \theta(B)a_t$$

In automatic detection and correction, by default, 3 types of outliers can be considered:

ADDITIVE OUTLIER: $AO : \upsilon(B) = 1$

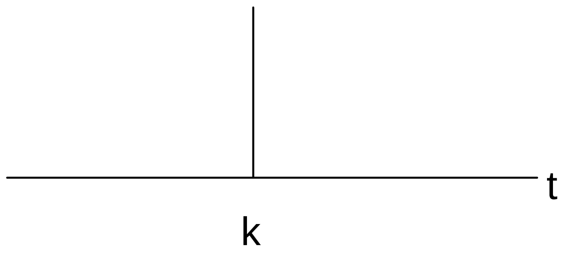
TRANSITORY CHANGE: $TC : \upsilon(B) = 1/(1 - \delta B)$, $(\delta = .7$ default value)

LEVEL SHIFT: $LS : \upsilon(B) = 1/(1 - B)$

One can also include a 4th type:

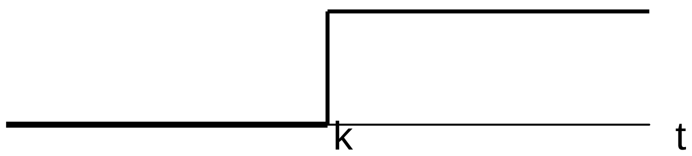
INNOVATION OUTLIER: $IO : \upsilon(B) = \theta(B)/[\phi(B)\delta(B)] \Rightarrow$ shock in the innovations a_t

1) Additive Outlier:



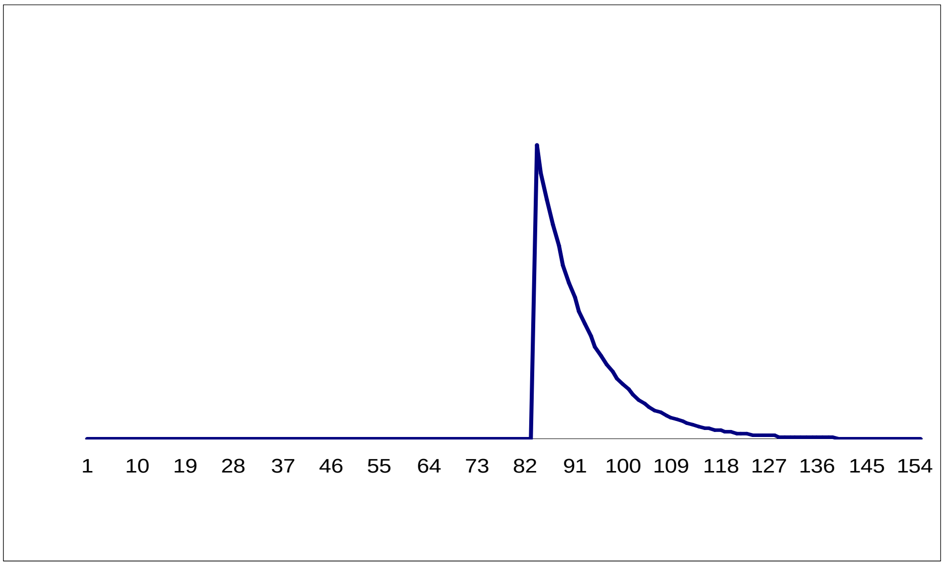
$$x_t = \ldots + \beta d_{kt} + \ldots \quad ; \qquad d_{kt} = \begin{cases} 1 & t = k \\ 0 & \text{otherwise} \end{cases}$$

2) Level Shift = LS ("step function")



$$x_t = \ldots + \beta \frac{1}{\nabla} d_{kt} + \ldots ; \qquad d_{kt} \text{ as before}$$

3) Transitory Change = TC



$$x_t = \ldots + \beta \frac{1}{1-\delta B} d_{kt} + \ldots$$

d_{kt} as before

$0 < \delta < 1$
||
.7 default

4) Innovational outlier = IO $(a_t + \beta)$

$$x_t = \ldots + \beta \frac{\theta(B)}{\phi(B)\,D} d_{kt} + \ldots$$

For D = $\nabla\nabla_{12}$ (for example) $\Rightarrow$ IO becomes unbounded and may dominate the series.

IO's are of questionable interest. I do not recommend their use.

An important "conceptual"diference in TRAMO:

- If we have a "priori" information on a special event likely to influence the series, we model it as an

  INTERVENTION VARIABLE .

  It can have the form of an outlier (for ex., an LS),

  but it is directly entered as a

  REGRESSION VARIABLE

  (It may or may not be significant)

- On the contrary:

  OUTLIER CORRECTION is a tool to achieve n.i.i.d. residuals.

  - Remove non-normality
  - Remove spurious effects on ACF, distorsion of parameter estimates and of forecasts.

  Their location and type is "a priori" unknown.

  However, they can often be given expost explanation.

  (for ex.: change of base of an index, currency devaluation, economic integration, …)

  In this case, better to respecify it as an intervention variable (removing it from AODC).

TRAMO philosophy:

- Minimize the number of outliers needed.

  Yet many series are in need of outlier correction.

  (They distort identification, parameter estimates, forecasts, estimates of the components,…)

  Rough measure:

  Between 50-60% of series need outlier correction.

  Maximum number of outlier/series:

  it depends on series,

  and also on series length:

  – shorter series may require a larger proportion of outliers.

  Rough measure:

  max. # : < 4-5 % of observations.

  optimal range : (0 -2%)

  ↓

  monthly data: 1 outlier / 4 years (+)

  quarterly data: 1 outlier / 12 years (+)

- When a series is routinely treated (for ex, every month), once an appropriate model has been identified, several of its features should be fixed for the next periods. One of these features is the type and location of the outliers. To that effect, outliers can also be entered directly.

  Setting IUSER = 2, NSER = 2 ,

  and entering the position and type:

67	AO	135	LS
↑	↑		
Observ. number	Type of outlier		

  the model will estimate an AO for t = 67, and an LS outlier for t = 135.

- USEFUL HINTS for series "treated with care":

a) If TRAMO finds no outlier for the default critical level (VA), it may be worth it to decrease VA until first outlier(s) appears. Then:

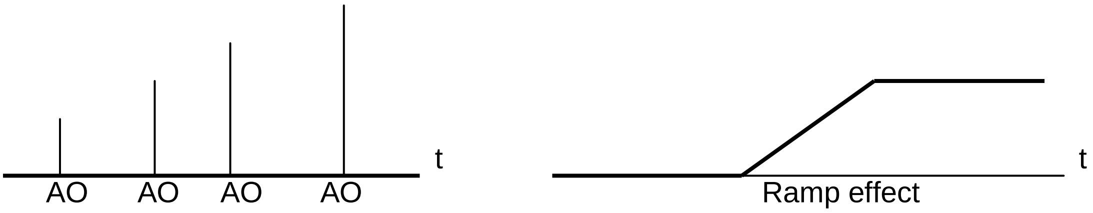
   - Look at the significance ("masking" effects are tricky)
   - Is there an expost explanation?
   - Is the fitting significantly improved?

If there is no clear justification for keeping the outlier, drop it.

b) Perhaps outliers can be reparametrized in a more parsimonious way (using intervention-variable specifications)

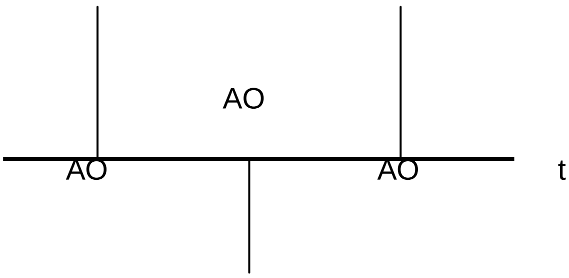
Obvious examples:

1)



# of param: 4 ➔ 1

2)

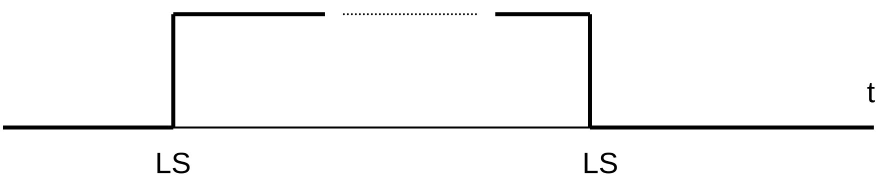


of similar magnitude

Regression var: (0, 1, 0, -1, 0, 1, 0)

# of param: 3 ➔ 1

3)



Regression var: (0, 1, 0)

# of param: 2 ➔ 1

- Gain in degrees of freedom
- Gain in efficiency
- Likely gain in forecasting

<u>ONE POSSIBLE ADDITION:</u>

<u>SEASONAL OUTLIERS</u>

AO, LS, TC ➔ Do not affect seasonal component.

Yet, on ocassion, series may indicate:

- Seasonality, for some particular month(s), has become less (or more) volatile (Typical ex.: IPI in Spain, Italy ➔ August)
- Perhaps this is indicated by large # of AO for that month.

There can also be institutional reasons for abnormal changes in the seasonal pattern.

Examples:

Germany and effect of the evolution of the December pay on retail trade series. (The peaks for December disappear.)

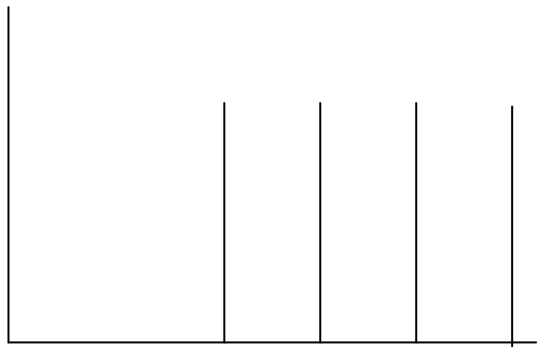
Italy: "cuaresma" and marriages. (The drop for Easter disappears.)

A convenient outlier :

SEASONAL LEVEL SHIFT

SLS outlier (monthly data)

$$x_t = \begin{pmatrix} \text{Other} \\ \text{reg.} \\ \text{effects} \end{pmatrix} + \underbrace{\beta \frac{1}{\nabla_{12}} d_{kt}}_{\text{sls}} + \text{ARIMA} \qquad d_{kt} = \begin{cases} 1 & t = k \\ 0 & \text{otherwise} \end{cases}$$

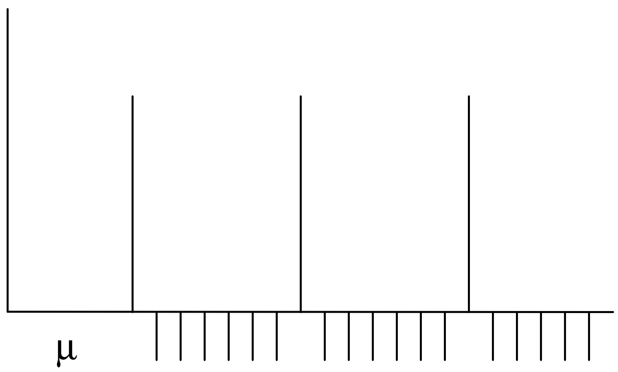


Mean $(\mu) > 0$ $\rightarrow$ Seasonal effect needs to be centered,

$\hookrightarrow$ Mean assigned to trend

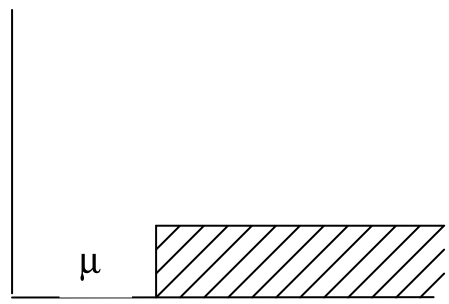
SLS is split into two effects,

a) One that is centered



Purely seasonal effect

b) the mean



Constant mean (to trend)

Alternatively, one can prefer the pure seasonal effect without the (small) effect on the mean. Use then the regression variable:

$$\left[\frac{11}{12}, \underbrace{\frac{-1}{12}, \frac{-1}{12}, \ldots \frac{-1}{12}}_{11\,\text{Times}}, \frac{11}{12}, \underbrace{\frac{-1}{12}, \ldots, \frac{-1}{12}}_{11\,\text{Times}}, \ldots \right]$$

Specification of the dynamic effect should depend on “a priori” information:

Example:

1. Salary increase = a new special bonus in April.

   Obvious choice: SLS (with mean effect)

2. Removal of Christmas bonus payment, that will now be spread uniformly over the year.

   Obvious choice: the purely seasonal version (no mean effect).

- At present, Seasonal Outlier is not included in Automatic Outlier Detection (needs to be entered as regression variable)

- Centering needs to be done in SEATS

- There can be more months involved

  German example:

$$+\beta\frac{1}{\nabla_{12}}(d_{kt}+d_{k-1,t})+...$$

K = December 94

One needs to enter

ISEQ = 1, DELTAS = 1,… ,

and then the position and number of months per year affected by the outlier

119 2

↑

At observation 119 (say, a December) a SLS will affect all future Dec. and Januaries

Of course, there are several alternative ways of modelling seasonal (and nonseasonal) outliers (see Kaiser and Maravall, 2001).

AODC has to select just a few basic outliers.

Detection and correction for the effect of an outlier

Assume, first, that we know the location and type of outlier.

Let $z^* = \left(z_1^*,...,z_N^*\right)'$ and $z = \left(z_1,...,z_N\right)'$ be the observed and the outlier-free series, respectively. Put

$$Y = \left(\upsilon_1(B)\, I_1(T),...,\upsilon_N(B)\, I_N(T)\right)'$$

Then,

$$z^* = Y\omega + z$$

where, to simplify, we suppose that z follows a stationary ARMA model. Let $\mathrm{Var}\,(z) = \sigma^2\Omega = \sigma^2 LL'$, with L lower triangular.

Then,

$L^{-1}z^* = L^{-1}Y\omega + L^{-1}z$

is an OLS regression model. If

$$r^* = L^{-1}z^*,\ X = L^{-1}Y \text{ and } r = L^{-1}z,$$

we can write

$$r^* = X\omega + r, \qquad r \sim N\left(0, \sigma^2 I_N\right),$$

If model parameters are known

Let $\tilde{\omega} = (X'X)^{-1}X'r^*$ and $\tau = (X'X)^{1/2}\tilde{\omega}/\sigma$.

Then, $\tau \sim N(0,1)$

Normally, the parameters of the model are unknown and one has to estimate them. Then, we have

$$\hat{\omega} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{r}^*, \qquad \hat{\tau} = (\hat{X}'\hat{X})^{1/2}\hat{\omega}/\hat{\sigma},$$

where $\hat{\tau}$ is asymptotically equivalent to τ.

Assume now that we know the location (t=T), but not the type of outlier.

For

t=T, we compute: $\hat{\omega}_{AO}(T), \hat{\omega}_{TC}(T), \hat{\omega}_{LS}(T)$

and $\lambda_T = \max\{|\hat{\tau}_{AO}(T)|, |\hat{\tau}_{TC}(T)|, |\hat{\tau}_{LS}(T)|\}$

Use $\lambda_T > C$ to test for significance.

C: An "a priori" set critical value (parameter VA in program).

If we don't know the timing of the outlier, we compute λ_t for t=1,…,N and use

$$\lambda = \max_t \lambda_t = \left|\hat{\tau}_{tp}(T)\right|$$

If $\lambda > C$, there is an outlier of type tp (AO, TC, LS) at T.

We correct for this outlier, and start the process again to see if there is another outlier.

Outliers are removed one by one, until we obtain $\lambda_T < C$.

Now we proceed to joint estimation of the multiple outliers:

$$z_t^* = z_t + \sum_{i=1}^{k} \omega_i \upsilon_i(B) I_{t_j}(t_j)$$

We have to perform multiple regressions to avoid (as much as possible) masking effects.

The standard deviation of the residuals is estimated using the MAD estimator

$$\hat{\sigma} = 1.483 \times \text{median}\left\{\left|r_t^* - \tilde{r}^*\right|\right\}$$

where $\tilde{r}^* \equiv$ median of $r^* = L^{-1}z^*$ (it is an estimator that is robust with respect to the possible presence of outliers.)

Algorithm for Automatic Outlier Detection and Correction

Initialization

If there are any regression variables in the model, including the mean, the regression coefficients are estimated by OLS and the series is corrected for their effect.

Stage I: Detection and estimation of outliers one by one

I.1) The ARIMA parameters are estimated, using the Hannan-Rissanen's method, and the series corrected for all regression effects present at the time, including the outliers so far detected.

I.2) Considering the estimates of the ARIMA parameters obtained in I.1 as fixed, the regression coefficients are estimated by GLS and their *t* statistics are computed. To this end, the fast algorithm of Morf, Sidhu and Kailath (1974) is used, followed by the *QR* algorithm. New estimated residuals are obtained.

I.3) With the estimated residuals obtained in I.2, the robust MAD estimator of the standard deviation of the residuals is computed.

I.4) If $x = (x_{d+1},...,x_N)'$, where *d* is the degree of the differencing polynomial, denotes the differenced series, the statistics

$$\hat{\tau}_{AO}(t),\ \hat{\tau}_{LS}(t) \text{ and } \hat{\tau}_{TC}(t)$$

are computed for $t = d+1,...,N$. To this end, the residuals computed in I.2 and the MAD obtained in I.3 are used.

Let, for each $t = d+1,..,N$,

$$\lambda_t = \max\{|\hat{\tau}_{AO}(t)|, |\hat{\tau}_{TC}(t)|, |\hat{\tau}_{LS}(t)|\}$$

If

$$\lambda = \max_t \lambda_t = |\hat{\tau}_{tp}(T)| > C$$

where C is the pre-selected critical value, then there is a possible outlier of type tp at T. The subindex tp can be AO, TC, or LS. If no outlier has been found the first time the algorithm passes through this point, then it is stopped. The series is free of outlier effects. If no outlier has been found, but it is not the first time that the algorithm passes through this point, then go to II.1. If, on the contrary, an outlier has been found, then correct the series for all regression effects, using the estimates obtained in I.2 and the last outlier coefficient estimate obtained while computing λ, and go back to I.1 to iterate.

Stage II: Multiple Regression

II.1) Using the estimates of the multiple regression and their t statistics obtained the last time the algorithm passed through I.2, check whether there are any outliers with a t statistic $|\hat{\tau}_i| < C$, where C is the same critical value than in I.4. If there aren't any, stop. If, on the contrary, there are some, then remove the one with the lowest absolute t-value and go back to I.2 to iterate.

Remarks

1. The procedure used to incorporate or reject outliers is similar to the stepwise regression procedure for selecting the “best” regression equation.

2. One of the advantages of the previous algorithm is that all estimations of the ARIMA parameters are performed using linear regressions and are, therefore, computationally cheap and fast. Another advantage consists of using “exact” instead of approximate residuals. Finally, the use of multiple regressions avoids the detection of spurious outliers or the lack of detection of outliers due to masking effects.

3. In I.2, the Kalman filter is applied to the data and the columns of the regression matrix. This has the effect of moving from a GLS to an OLS model.

4. In I.3, “exact” residuals are used. However, the regression matrix corresponding to the outlier, each one a vector, is (conditionally) filtered, using the filter implied by the inverse of the model for the series.

5. In TRAMO, the critical value C is the parameter VA. By default, its value is set as a function of the number of observations in the series (NZ).

min. value: $VA = 3$ when $NZ \leq 50$

max. value: $VA = 4$ when $NZ \geq 450$

For $50 \leq NZ \leq 400$,

$$VA = 3 + .0025\,(NZ - 50)$$

AUTOMATIC MODEL IDENTIFICATION

Identification of an ARIMA model in the Box-Jenkins tradition is simple, but depends, in an important way, on the analyst.

Broadly:

a) The unit roots are determined by the differencing required to produce a SACF that clearly converges ("… fast enough").

b) The orders of the stationary polynomials are infered from several statistics (all are sample values):

   ACF : Autocorrelation function,

   PACF : Partial ACF

   EACF : Extended ACF

   IACF : Inverse ACF

   ……

   with the analyst and the program iterating, according to the graphs.

In TRAMO,

Automatic Model Identification

(AMI)

works very differently. It also deals with a variety of added features:

* Missing values
* Outliers
* Calendar effects

………..

Brief summary of the procedure followed in TRAMO.

Pretest for the Log-level Specification

TRAMO contains two tests. (In the present version, only the second one is used.)

The fist test consists first of a trimmed range-mean regression. If the slope is positive, logs are selected; if it is negative, levels are used. When the slope is close to zero, a selection is made based on the BIC criterion when applied to the default model using both specifications.

The second test consists of direct comparison of the BICs of the default model in levels and in logs (with a proper correction). A parameter, FCT, permits to bias the test to favor logs or to favor levels.

Pretest for Trading Day and Easter Effects

The pretests are made with regressions using the default model for the noise and, if the model is subsequently changed, the test is redone.

Thus, the output file of TRAMO may say at the beginning "Trading day is not significant", yet the model finally estimated may contain TD variables (or viceversa).

A Remark on the use of the default (Airline) model

Pretesting and the starting point of AMI depend heavily on the Airline model

$$\nabla \nabla_{12} z_t = (1+\theta_1 B)(1+\theta_{12} B^{12}) a_t + \mu$$

(monthly series)

Two important reasons:

– Many studies have shown it is appropiate for large number of real series (40-60%)

  In what concerns the seasonal structure, the multiplicative

  IMA_{12} (1,1) structure is overwhelmingly used.

– The Airline model approximates well many other models.

  * Of course, this will be true for models of the type

$$\nabla \nabla_{12} x_t = \theta(B) a_t + \mu$$

  * But also models with

    - very weak or no trend
    - deterministic trend

    or

    - no seasonality
    - deterministic seasonality

    will be relatively well approximated

For example:

<u>Deterministic trend / no trend</u>

Model

$$\nabla_{12}\ x_t = (1+\theta_{12}\ B^{12})\ a_t + \mu\, t + c \quad ,$$

or model

$$\nabla_{12}\ x_t = (1+\theta_{12}\ B^{12})\ a_t \quad ,$$

become, after differencing,

$$\nabla\nabla_{12}\ x_t = (1-B)\ (1+\theta_{12}\ B^{12})\ a_t + \mu$$

($\mu = 0$ for the second model).

This Non-invertible model, with a U.R. shared by AR and MA polynomials, presents serious numerical problems.

However, it will be well approximated by

$$\nabla\nabla_{12}\ x_t = (1-.99\ B)\ (1+\theta_{12}\ B^{12})\ a_t + \mu \quad ,$$

which is an invertible model and presents no problem.

Deterministic seasonality / no seasonality

If model is

$$\nabla x_t = (1+\theta B)\, a_t + \mu$$

or

$$\nabla x_t = (1+\theta B)\, a_t + \sum_{i=1}^{12} \beta_i\, d_{it} \quad ,$$

where d_{it} = monthly dummy variables, after seasonal differencing, they both become

$$\nabla\nabla_{12}\, x_t = (1+\theta B)\,(1-B^{12})\, a_t \quad .$$

As before, it will be well approximated by

$$\nabla\nabla_{12}\, x_t = (1+\theta B)\,(1-.99\, B^{12})\, a_t \quad .$$

In TRAMO-SEATS, when estimating ARIMA parameters,

- if modulus of an AR root approaches 1, it is fixed equal to 1;
- if modulus of an MA root approaches 1, it is fixed equal to a value smaller than but very close to 1 (parameter XL, by default = .99).

(MODERATE) OVERDIFFERENCING CAUSES LITTLE DAMAGE.

Unit roots are estimated well due to the following general result:

Result:

Assume general model

$\phi(B)z_t = \theta(B)a_t$,

where $\phi(B)$ and $\theta(B)$ may contain nonstationary or noninvertible (unit) roots.

Let $\hat{\beta}$ be the MLE of a ϕ or θ parameter.

Then, for

- parameters associated with $\begin{Bmatrix} \text{stationary} \\ \text{invertible} \end{Bmatrix}$ roots,

$\hat{\beta} \rightarrow \beta$: order of convergence: $\dfrac{1}{\sqrt{N}}$

- parameters associated with $\begin{Bmatrix} \text{non stationary} \\ \text{non invertible} \end{Bmatrix}$ roots,

$\hat{\beta} \rightarrow \beta$: order of convergence: $\dfrac{1}{N}$

("Superconsistency" properties of parameters associated with Unit Roots. True for AR and MA roots).

Implication:

Unit AR or MA roots are very well captured.

Hence, for example,

- if the series is $z_t = a_t$, (white noise)

and we fit an Airline model, $\hat{\theta}_1$ and $\hat{\theta}_{12}$ will approach -1. TRAMO fixes them to be close to -1, but invertible, so that we would end up with something like

$$\nabla \nabla_{12} z_t = (1-.99B)\left(1-.97B^{12}\right)a_t$$

Given that there are no deterministic components, near cancellation of

∇ with (1-.99B), and

∇_{12} with (1-.97 B^{12})

will reproduce the series a_t nearly untouched.

- If the series contains

deterministic trend: $\mu t + c$,

or

deterministic seasonal: $\sum_{i=1}^{11} \beta_i d_{it}$,

these will be reproduced (nearly untouched) with the same Airline model.

- If z_t is the IMA (1, 1) model with mean:

  $$\nabla z_t = (1 + \theta B)\, a_t + \mu$$

  and the data is quarterly, fitting an Airline model would yield something like

  $$\nabla \nabla_4 z_t = (1 + \theta B)\,(1 - .99B^4)\, a_t \ ,$$

  and the IMA (1,1) would be (nearly) reproduced,

  and so on (see Maravall, 1998).

Hence, because of its encompassing properties, the Airline model is excellent:

- as a standard example for ARIMA models,
- as a starting point in the modelling process,
- as a reference benchmark,
- as the model to use for pretesting.

Automatic Model Identification in the Presence of Outliers

The algorithm iterates between the following two stages

1. Automatic outlier detection and correction

2. Automatic model identification

The first model used is the default model. The automatic model identification is performed with the series corrected for the outliers and other regression effects present at the time. If the model changes, the automatic detection and correction of outliers is performed again from the beginning.

We already looked at AODC; now we summarize AMI of the ARIMA model.

<u>Automatic ARIMA model identification</u>

Suppose the series $\{x_t\}$ follows the model

$$\phi(B)\delta(B)x_t = \theta(B)a_t, \qquad a_t \overset{i.i.d.}{\sim} N(0,\sigma^2).$$

TRAMO proceeds in two steps:

1) First, identify $\delta(B)$ (unit roots)

2) Second, identify the ARMA model, i.e, $\phi(B)$ and $\theta(B)$.

1) <u>Identification of the Nonstationary polynomial δ (B)</u>

To determine the appropriate differencing of the series, we discard unit root testing.

Problems with Unit Root Testing:

– When regular and seasonal U.R. may be present, available tests have low power.

For example, in

$$\nabla z_t = (1 - .8B) a_t \ , \qquad (1)$$

$$\nabla_{12} z_t = (1 - .8B^{12}) a_t \ ,$$

U.R. would most likely be rejected due to the large MA root.

[AR approx. to (1): (1-.2B-…) $z_t = a_t$]

– Besides, in AMI + AODC, one may try thousands of models, where:

next try depends on previous results.

→ Serious DATA MINING problem: the size of the test is a function of prior rejections and acceptances.

→ No way of knowing the true size of the test.

We follow an alternative approach:

Decide “a priori”, instead of a fictitious size, the following value:

How large the modulus of a root should be in order to accept it as 1 (unit root)?

For AR and MA roots the criterion is different (roughly: unit AR roots are O.K.; unit MA roots should be avoided.)

For AR root:

mod >.95 $\Rightarrow$ root made unit root.

For MA root:

mod >.99 $\Rightarrow$ mod. of root made .99.

So as to have invertibility.

With .99 no numerical problems appear.

We use some very useful results (Tiao, Tsay) on superconvengence of unit AR roots.

Example:

Let true model be the AR (3):

$$(1-.5B)\nabla^2 z_t = a_t$$

Because of superconsistency of U.R. estimators:

* If we estimate simply an AR (1) $\rightarrow$ the first U.R. (∇) is likely to be captured.

* If we estimate, again, an AR (1) on the previous residuals, $\rightarrow$ the second U.R. (∇) is likely to be captured.

* Alternatively, if we start by estimating an AR (2), both U.R. (∇^2) are likely to be captured.

* Further increases in p (the order of the AR), or further AR(1) fits to residuals will not point to a U.R.

The previous results extend in a straightforward manner to SEASONAL U.R.

TRAMO uses these results.

First, the model AR (2) AR_s (1) with mean,

$$\left(1+\phi_1 B+\phi_2 B^2\right)\left(1+\phi_s B^s\right)\left(z_t-\mu\right)=a_t \quad ,$$

is estimated by HR.

As already mentioned, if the modulus of an MA root is relatively large, the bias in the estimator of the AR parameter can be large, and the U.R. can be missed.

Therefore, after detecting U.R. with AR fits, TRAMO uses ARMA (1,1) fits to detect U.R. that might not have been captured because of ignoring possibly large MA roots.

Hence, after the pure AR fit, TRAMO fits models of the form ARMA (1,1) $ARMA_s$ (1,1) with mean

$$\left(1+\phi B\right)\left(1+\Phi B^s\right)\left(z_t-\mu\right)=\left(1+\theta B\right)\left(1+\Theta B^s\right)a_t \quad .$$

The residuals of the last estimated model are used for a pre-test to specify a mean or not.

(Unit roots are identified 1 by 1.)

2) <u>Identification of the stationary ARMA model: $\phi(B)$ and $\theta(B)$</u>

The program selects the orders (p, q), where $p = dg\{\phi(B)\}$ and $q = dg\{\theta(B)\}$, corresponding to the lowest $BIC_{p,q}$, where

$$BIC_{p,q} = \ln\left(\hat{\sigma}^2_{p,q}\right) + (p+q)\frac{\ln(N-d)}{N-d}.$$

It searches for models of the form

$$\phi_{p_r}(B)\phi_{p_s}\left(B^s\right)x_t = \theta_{q_r}(B)\theta_{q_s}\left(B^s\right)a_t$$

over the range

$0 \le p_r, q_r \le 3,\qquad 0 \le p_s, q_s \le 2$ (1 if used with SEATS)

This is done sequentially (for fixed regular polynomials, the seasonal ones are obtained, and viceversa). The search favors parsimony and “balanced” models (similar AR and MA orders). See Gómez and Maravall (2001).

Remarks:

- Because all possible combinations are checked, the number of models fitted under the (AMI + AODC) procedure typically runs into the thousands. The complete procedure takes a fraction of a second.

- When analyzing series with care, TRAMO may suggest a few models (perhaps 2 or 3) all of which could be reasonably acceptable. When used with SEATS, looking for, among these models, the one that provides a better decomposition may provide additional tools for the choice (see Kaiser-Maravall, 2001; Maravall-Sánchez, 2000; Maravall, 2002).

An Application

The performance of TRAMO applied in an entirely automatic manner with the full automatic model identification and outlier detection and correction procedure has been analysed in some detail at Eurostat on 13227 series (EUROSTAT, 1996b) and by Fischer and Planas (2000). The series were short-term indicators (all activities) for all 15 European Union member states, the European aggregates, USA, and Japan. We summarize some of the Eurostat findings.

– 60% of the series follow a multiplicative model; for 40% the components are additive.

– 50% of the series can be modelled with the default model (i.e., an Airline-type model).

– For 87% of the series the fit obtained was good, for an additional 11% the fit was reasonably acceptable, and only for 2% the fit was poor. In fact, the authors concluded “Automatic modelization works much better than expected!”.

  Fischer and Planas find even more favorable results (and the work is not yet finished).

– 15% of the series showed a stationary behaviour; for 2% of all series the order of the regular differencing was larger than 1.

– 50% of the series did not require outlier adjustment; 27% presented one outlier; 23% presented two or more outliers.

– Split into four groups of indicators, the results are the following.

	Production	Other Indicators	Foreign Trade	Business Surveys	TOTAL
Nº. of Series	2627	2577	6976	1047	13227
Mult/addit model (%)	63/37	59/41[a)]	75/25	7/93[b)]	64/36
Airline model (0,1,1) x $(0,1,1)_s$ (%)	71	37	49	31	50
Differencing order < 1/>1 (%)	8/0.4	5/7	19/0.3	32/0	15/2
AR or MA regular polynomial > 2 (%)	5	11	6	3	7
Ljung Box: good/bad (%)	88/1	77/7	90/1	92/0.6	87/2
No seasonality found (%)	1	9	7	47	9
Outliers none/one/ more (%)	44/22/34	43/23/34	54/30/26	54/25/21	50/27/23

a) Splitted up into indicators, the percentage of the multiplicative model is employment: 29% / turnover: 77% / new orders: 76%.

b) Since the balance can be negative, the model for most of the series is additive.