OECD/Eurostat workshop

Application of advanced temporal disaggregation techniques to economic statistics

Tour Europe, La Defense, Paris

27 November 2003

RECENT DEVELOPMENTS IN THE FIELD OF BENCHMARKING AND TEMPORAL DISAGGREGATION

Tommaso Di Fonzo

University of Padova
difonzo@stat.unipd.it

OUTLINE OF THE PRESENTATION

1. **Working with log-transformed data**

2. **Dynamic extension: a first order dynamic model**

3. **Alternative model formulation**

4. **Dealing with systems of time series**

1. WORKING WITH LOG-TRANSFORMED DATA

- **Log-transformation is widely used in modelling economic series**

Non-additive transformation: $\ln y_t \neq \sum_{h=1}^{s} \ln y_{h,t}$

$$\Downarrow$$

cannot be directly applied in distribution problems

Pinheiro and Coimbra (1992), Salazar *et al.* (1994, 1997), Proietti (1999) and Aadland (2000), Mitchell *et al.* (2003)

Disaggregation of flows: $y_t = \sum_{h=1}^{s} y_{h,t}$

The disaggregated model

$$\ln y_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t} \qquad \boldsymbol{h=1,\ldots,s} \qquad \boldsymbol{t=1,\ldots,T}$$

$$z_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t} \qquad z_{h,t} \equiv \ln y_{h,t}$$

Expansion in Taylor series of $\ln y_{h,t}$ around

$$\bar{y}_t = \frac{1}{s}\sum_{h=1}^{s} y_{h,t} = \frac{y_t}{s}$$

$$\ln y_{h,t} = z_{h,t} \simeq \ln \bar{y}_t + \frac{1}{\bar{y}_t}\left(y_{h,t} - \bar{y}_t\right) = \ln y_t - \ln s + \frac{s y_{h,t}}{y_t} - 1$$

$$\sum_{h=1}^{s} z_{h,t} \simeq s \ln y_t - s \ln s + \frac{s\sum_{h=1}^{s} y_{h,t}}{y_t} - s = s \ln y_t - s \ln s$$

The (observable) aggregated model

$$z_t = \mathbf{x}_t^{'}\boldsymbol{\beta} + u_t \qquad \boldsymbol{t=1,\ldots,T} \qquad z_t = s\ln y_t - s\ln s$$

$$\Downarrow$$

$$\mathbf{z}_l = \mathbf{X}_l\boldsymbol{\beta} + \mathbf{u}_l \qquad \mathbf{V}_l = \mathbf{C}\mathbf{V}_h\mathbf{C}'$$

The BLU solution

$$\hat{\mathbf{z}}_h = \mathbf{X}_h\hat{\boldsymbol{\beta}} + \mathbf{V}_h\mathbf{C}'\mathbf{V}_l^{-1}\left(\mathbf{z}_l - \mathbf{X}_l\hat{\boldsymbol{\beta}}\right) \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}_l^{'}\mathbf{V}_l^{-1}\mathbf{X}_l\right)^{-1}\mathbf{X}_l^{'}\mathbf{V}_l^{-1}\mathbf{z}_l$$

$$\sum_{h=1}^{s}\hat{z}_{h,t} = z_t = s\ln y_t - s\ln s \qquad \boldsymbol{t=1,\ldots,T}$$

$$\hat{y}_{h,t} = \exp\left(\hat{z}_{h,t}\right)$$

Adjustment of the estimates

$$\sum_{h=1}^{s} \hat{y}_{h,t} \neq y_t \quad \Leftrightarrow \quad y_t - \sum_{h=1}^{s} \hat{y}_{h,t} = r_t \neq 0$$

- **Proietti (1999) suggests to 'smooth' discrepancies according to Denton (1971)**

$$\tilde{\mathbf{y}}_h = \hat{\mathbf{y}}_h + \left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{C}'\left[\mathbf{C}\left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{C}'\right]^{-1}\mathbf{r}_l$$

D: matrix (n,n) 'first differences'

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

SUMMARY VIEW

Type of		*Temporal aggregates*		*Adjustment needed for*
variable	*aggregation*	y_t	z_t	$\hat{y}_{h,t} = \exp\left(\hat{z}_{h,t}\right)$
Flow	Sum	$\sum_{h=1}^{s} y_{h,t}$	$s \ln y_t - s \ln s$	Yes
Index	Average	$\frac{1}{s}\sum_{h=1}^{s} y_{h,t}$	$\ln y_t$	Yes
Stock (eop)	Systematic sampling (eop)	$y_{s,t}$	$\ln y_t$	No
Stock (bop)	Systematic sampling (bop)	$y_{1,t}$	$\ln y_t$	No

* eop and bop stand for ‘end of period’ and ‘beginning of period’, respectively

THE *DELTALOG* MODEL

Measurement models formulated in terms of growth rates ⇔ logarithmic differences

$$\Delta \ln y_{h,t} = \ln \frac{y_{h,t}}{y_{h-1,t}} = \Delta z_{h,t}$$

$$z_{h,t} = \ln y_{h,t} \qquad\qquad y_{0,t} = y_{s,t-1}$$

$$\Delta z_{h,t} = \Delta \mathbf{x}_{h,t}^{'} \boldsymbol{\beta} + \varepsilon_{h,t}$$

Fernàndez (1981)

$$\hat{\mathbf{y}}_h = \exp\left(\hat{\mathbf{z}}_h\right)$$

$$\hat{\mathbf{z}}_h = \mathbf{X}_h \hat{\boldsymbol{\beta}} + \left(\mathbf{D}'\mathbf{D}\right)^{-1} \mathbf{C}' \left[\mathbf{C}\left(\mathbf{D}'\mathbf{D}\right)^{-1} \mathbf{C}'\right]^{-1} \left(\mathbf{z}_l - \mathbf{X}_l \hat{\boldsymbol{\beta}}\right)$$

$$\hat{\boldsymbol{\beta}} = \left\{\mathbf{X}_l^{'} \left[\mathbf{C}\left(\mathbf{D}'\mathbf{D}\right)^{-1} \mathbf{C}'\right]^{-1} \mathbf{X}_l\right\}^{-1} \mathbf{X}_l^{'} \left[\mathbf{C}\left(\mathbf{D}'\mathbf{D}\right)^{-1} \mathbf{C}'\right]^{-1} \mathbf{z}_l$$

Notice (flow variable)

$$\sum_{h=1}^{s} \ln \frac{y_{h,t}}{y_{h,t-1}} \simeq \Delta z_t = s \Delta \ln y_t$$

The (logarithmic) growth rate of the temporally aggregated variable can be seen as an approximation of the average of growth rates of the disaggregated variable (Aadland, 2000)

$$\sum_{h=1}^{s} \Delta \ln y_{h,t} = \mathbf{x}_t^{'} \boldsymbol{\beta} + \varepsilon_t \Leftrightarrow \quad s \Delta \ln y_t = \mathbf{x}_t^{'} \boldsymbol{\beta} + \varepsilon_t$$

In the other cases

$$\Delta \ln y_t = \mathbf{x}_t^{'} \boldsymbol{\beta} + \varepsilon_t$$

Temporal disaggregation and deltalog model

The variable of interest is disaggregated in such a way that its estimated growth rates be coherent (approximately for flows and indices, exactly for stock) with the observed low frequency growth rates

An example

Estimation of Italian monthly Industry Value Added (1970:01-2001:09)

Notice: No attention to specification issues, only to show how the technique works

$$\Delta \ln y_{h,t} = \beta_0 + \beta_1 \Delta \ln x_{h,t} + \varepsilon_{h,t}$$

t=1970q1,…,2001q3, h=1,2,3

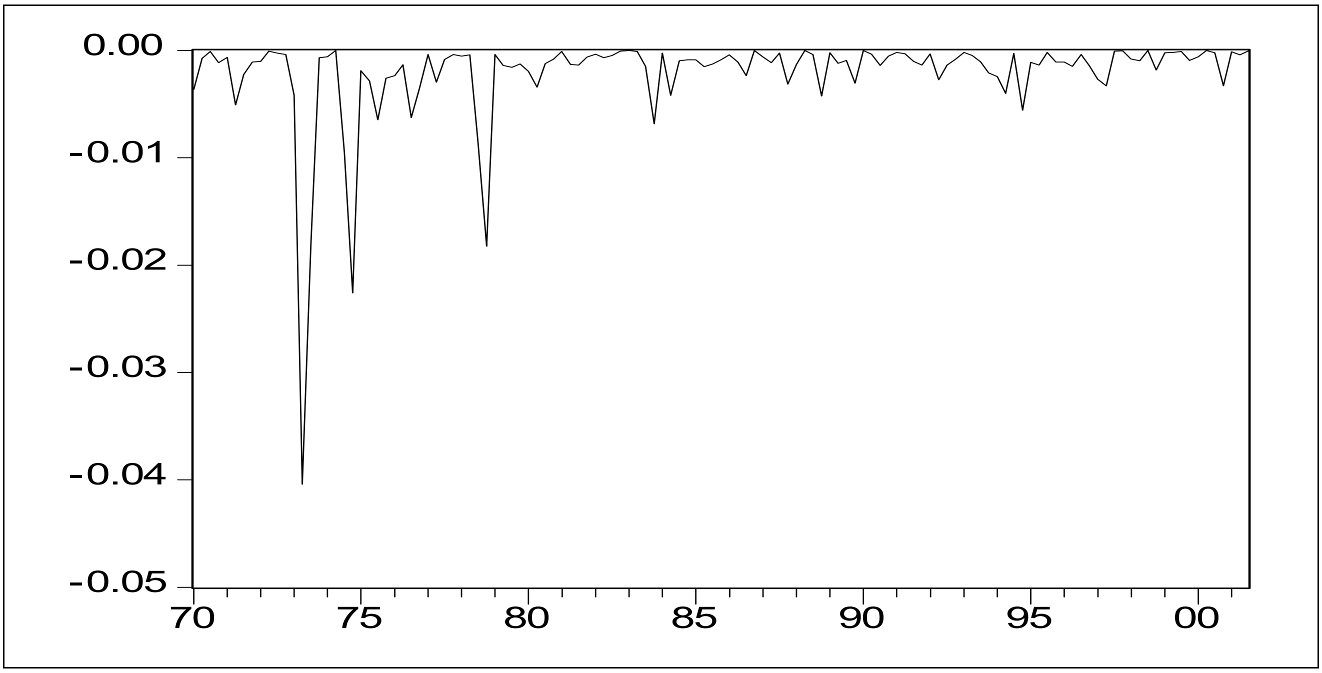
***Y*: V.A. Industry (in strict sense) (source: Istat)**

***X*: IPI (source: Bank of Italy)**

SA series

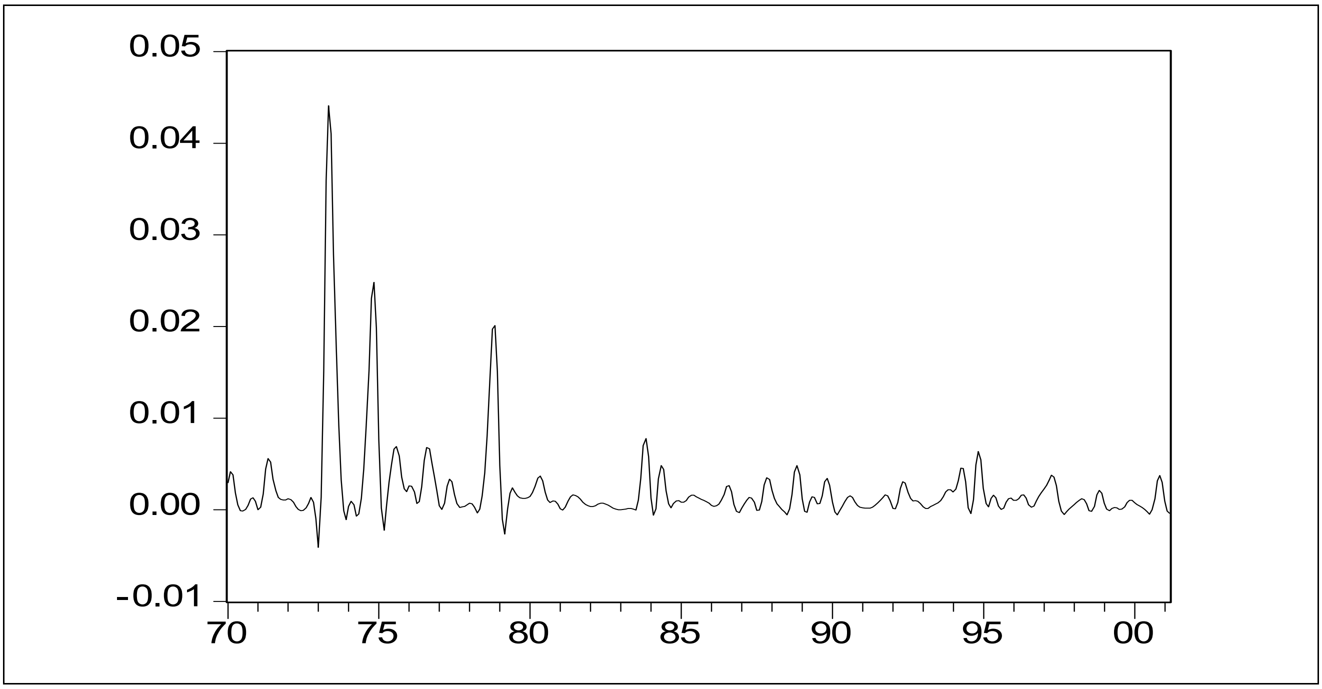
Quarterly discrepancies (%)

$$r_t = \frac{y_t - \hat{y}_t}{y_t} \times 100$$ ***t*=1970q1-2001q3**

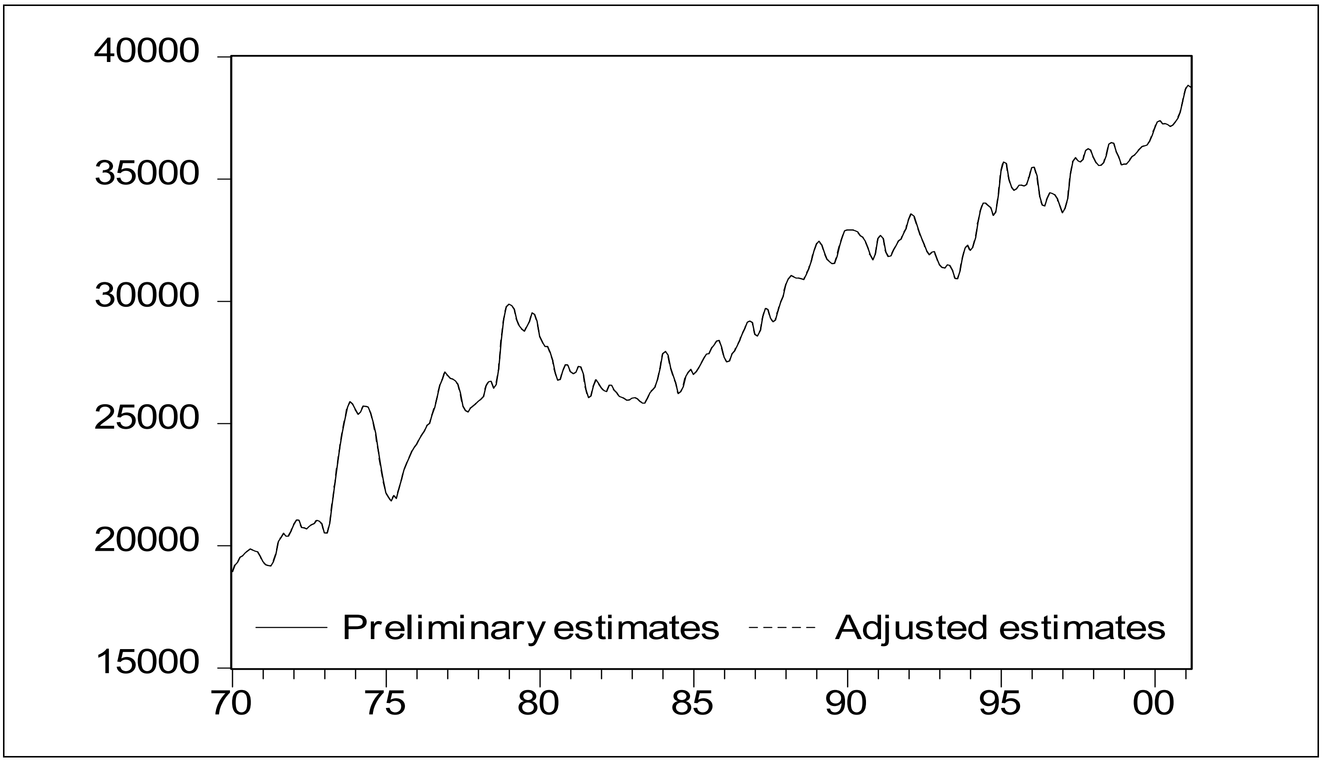


Monthly discrepancies (%)

$$r_{h,t} = \frac{\hat{y}_{h,t} - \tilde{y}_{h,t}}{\hat{y}_{h,t}} \times 100$$ ***h*=1,2,3** ***t*=1970q1-2001q3**



Preliminary ($\hat{y}_{h,t}$) and final ($\tilde{y}_{h,t}$) estimates ('technical' example)



2. DYNAMIC EXTENSION: A FIRST ORDER DYNAMIC MODEL

$$y_{h,t} = \phi y_{h-1,t} + \mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$$

$$|\phi| < 1 \qquad \varepsilon_{h,t} \sim WN\left(0, \sigma^2_{\varepsilon_h}\right) \qquad y_{0,t} \equiv y_{s,t-1}$$

$y_0 \equiv \eta$ **fixed (and unknown)**

Dynamic extension of the static approach by Chow and Lin

Gilbert (1977), Tserkezos (1991), Gregoir (1995), Salazar *et al.* (1998), Abeysinghe and Tay (2000), Santos Silva and Cardoso (2001), Di Fonzo (2002b, 2003a), Mitchell *et al.* (2003)

- **Suited to deal with different (interesting) relationships between *Y* and *X***
- **If *Y* and *X* are (non stationary and) cointegrated variables**
  $\Rightarrow$ **reparameterization of an ECM**

Matrix formulation of the model

$$\phi \neq 0 \qquad \Rightarrow \qquad \mathbf{D}_\phi \mathbf{y}_h = \mathbf{X}_h \boldsymbol{\beta} + \mathbf{q}\eta + \boldsymbol{\varepsilon}_h = \mathbf{Z}_h \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_h$$

$$\mathbf{D}_\phi = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix},$$

$$\mathbf{q} = (\phi, 0, 0, \ldots, 0)' \qquad \mathbf{Z}_h = [\mathbf{X}_h \,|\, \mathbf{q}] \qquad \boldsymbol{\gamma} = [\boldsymbol{\beta}' \,|\, \eta]'$$

$$\mathbf{D}_\phi^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \phi & 1 & 0 & \cdots & 0 & 0 \\ \phi^2 & \phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & \phi & 1 \end{bmatrix}$$

$$\mathbf{y}_h = \mathbf{D}_\phi^{-1} \mathbf{Z}_h \boldsymbol{\gamma} + \mathbf{D}_\phi^{-1} \boldsymbol{\varepsilon}_h = \mathbf{Z}_{h,\phi} \boldsymbol{\gamma} + \mathbf{u}_h$$

$$E\left(\mathbf{u}_h\mathbf{u}_h'\right)=\sigma_\varepsilon^2\mathbf{V}_h$$

$$\mathbf{V}_h=\mathbf{D}_\phi^{-1}\left(\mathbf{D}_\phi^{-1}\right)'=\begin{bmatrix}1 & \phi & \phi^2 & \cdots & \phi^{n-2} & \phi^{n-1}\\ \phi & 1 & \phi & \cdots & \phi^{n-3} & \phi^{n-2}\\ \phi^2 & \phi & 1 & \cdots & \phi^{n-4} & \phi^{n-3}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & \phi & 1\end{bmatrix}$$

The aggregated model

$$\mathbf{y}_l=\mathbf{C}\mathbf{Z}_{h,\phi}\boldsymbol{\gamma}+\mathbf{C}\mathbf{u}_h \qquad \mathbf{Z}_{l,\phi}=\mathbf{C}\mathbf{Z}_{h,\phi} \qquad \mathbf{V}_{l,\phi}=\mathbf{C}\mathbf{V}_{h,\phi}\mathbf{C}'$$

Parameter estimation

$$\ell\left(\phi,\boldsymbol{\beta}\right)=$$

$$-\frac{T}{2}\ln 2\pi-\frac{1}{2}\ln\left|\mathbf{V}_{l,\phi}\right|-\frac{1}{2}\left(\mathbf{y}_{l,\phi}-\mathbf{Z}_{l,\phi}\boldsymbol{\beta}\right)'\mathbf{V}_{l,\phi}^{-1}\left(\mathbf{y}_{l,\phi}-\mathbf{Z}_{l,\phi}\boldsymbol{\beta}\right)$$

Observations

- **wrt Gregoir (1995) and Salazar *et al.* (1998, 2003), the problem of estimating figures for the first *s* periods is solved in a simpler and straightforward manner**

- **Estimates are obtained according to the 'classical' Chow and Lin's approach**

$$\hat{\mathbf{y}}_h = \mathbf{Z}_{h,\hat{\phi}}\hat{\boldsymbol{\gamma}}_{\hat{\phi}} + \mathbf{V}_{h,\hat{\phi}}\mathbf{C}'\mathbf{V}_{l,\hat{\phi}}^{-1}\left[\mathbf{y}_l - \mathbf{Z}_{l,\hat{\phi}}\hat{\boldsymbol{\gamma}}_{\hat{\phi}}\right]$$

$$\hat{\boldsymbol{\gamma}}_{\hat{\phi}} = \left(\mathbf{Z}_{l,\hat{\phi}}^{'}\mathbf{V}_{l,\hat{\phi}}^{-1}\mathbf{Z}_{l,\hat{\phi}}\right)^{-1}\mathbf{Z}_{l,\hat{\phi}}^{'}\mathbf{V}_{l,\hat{\phi}}^{-1}\mathbf{y}_l$$

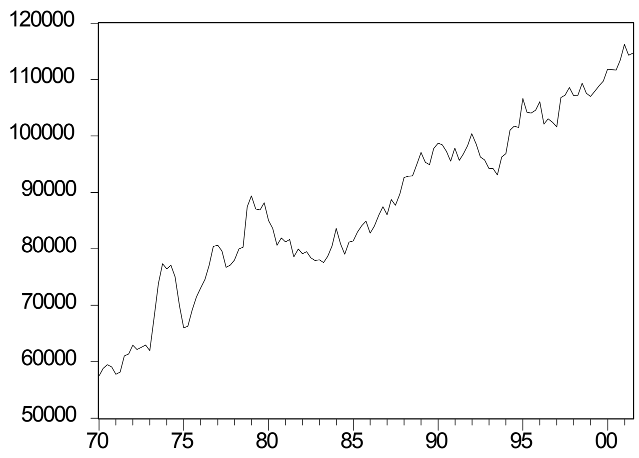
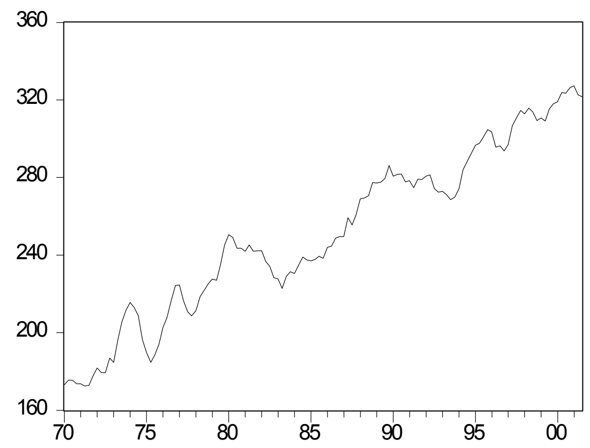
$$E\left(\hat{\mathbf{y}}_h - \mathbf{y}\right)\left(\hat{\mathbf{y}}_h - \mathbf{y}\right)' =$$

$$\left(\mathbf{I}_n - \mathbf{L}_{\hat{\phi}}\mathbf{C}\right)\mathbf{V}_{h,\hat{\phi}} + \left(\mathbf{X}_h - \mathbf{L}_{\hat{\phi}}\mathbf{X}_l\right)\left(\mathbf{X}_l^{'}\mathbf{V}_{l,\hat{\phi}}^{-1}\mathbf{X}_l\right)^{-1}\left(\mathbf{X}_h - \mathbf{L}_{\hat{\phi}}\mathbf{X}_l\right)'$$

$$\mathbf{L}_{\hat{\phi}} = \mathbf{V}_{h,\hat{\phi}}\mathbf{C}'\mathbf{V}_{l,\hat{\phi}}^{-1}$$

The technique at work: a complete example
(estimation of Italian monthly industrial V.A.)

The quarterly series

Value added industry	Industrial Production Index
	

Non stationary

Residual-based cointegration test does not reject the cointegration hypothesis

Quarterly regression in levels: not significant intercept

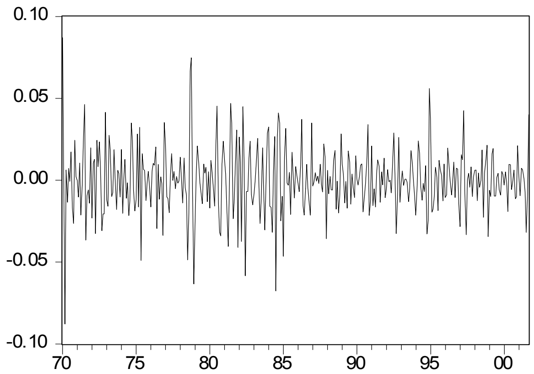
Parameters' estimates

Variant	α	β	η	ϕ
1: Lev.	185.342 (1.197)	60.178 (33.068)	18524.1 (15.397)	0.823 (23.793)
2: Lev.	---	60.973 (197.2)	18653.7 (15.543)	0.826 (24.080)
3: Log.	0.982 (42.127)	0.157 (29.712)	9.832 (225.5)	0.837 (25.300)
3: Log.*	0.966 (41.583)	0.154 (29.309)	---	0.840 (25.226)

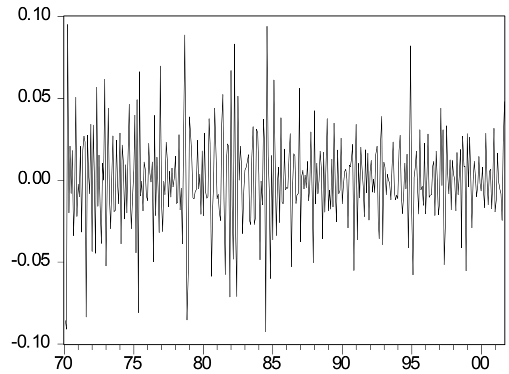
In parentheses are reported *t*-statistics.
* According to Salazar *et al.* (1997)

Discrepancies between estimates according to 2 and 3

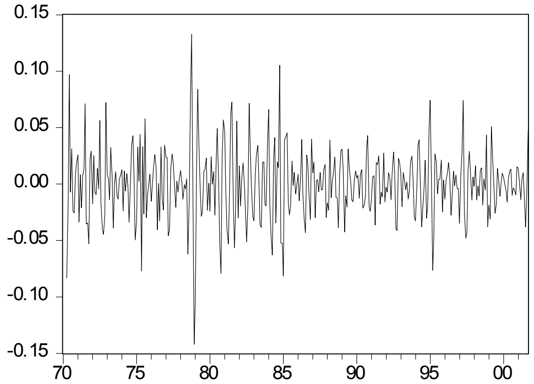
Levels (%)



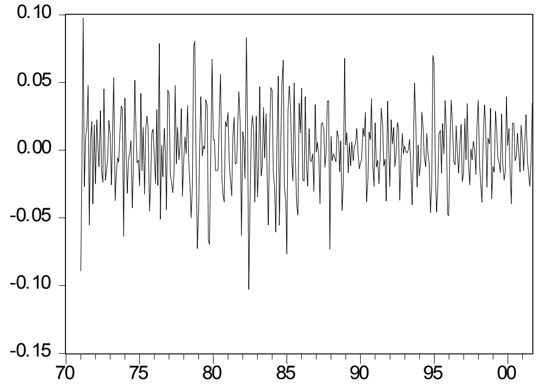
Monthly % changes



Quarterly % changes

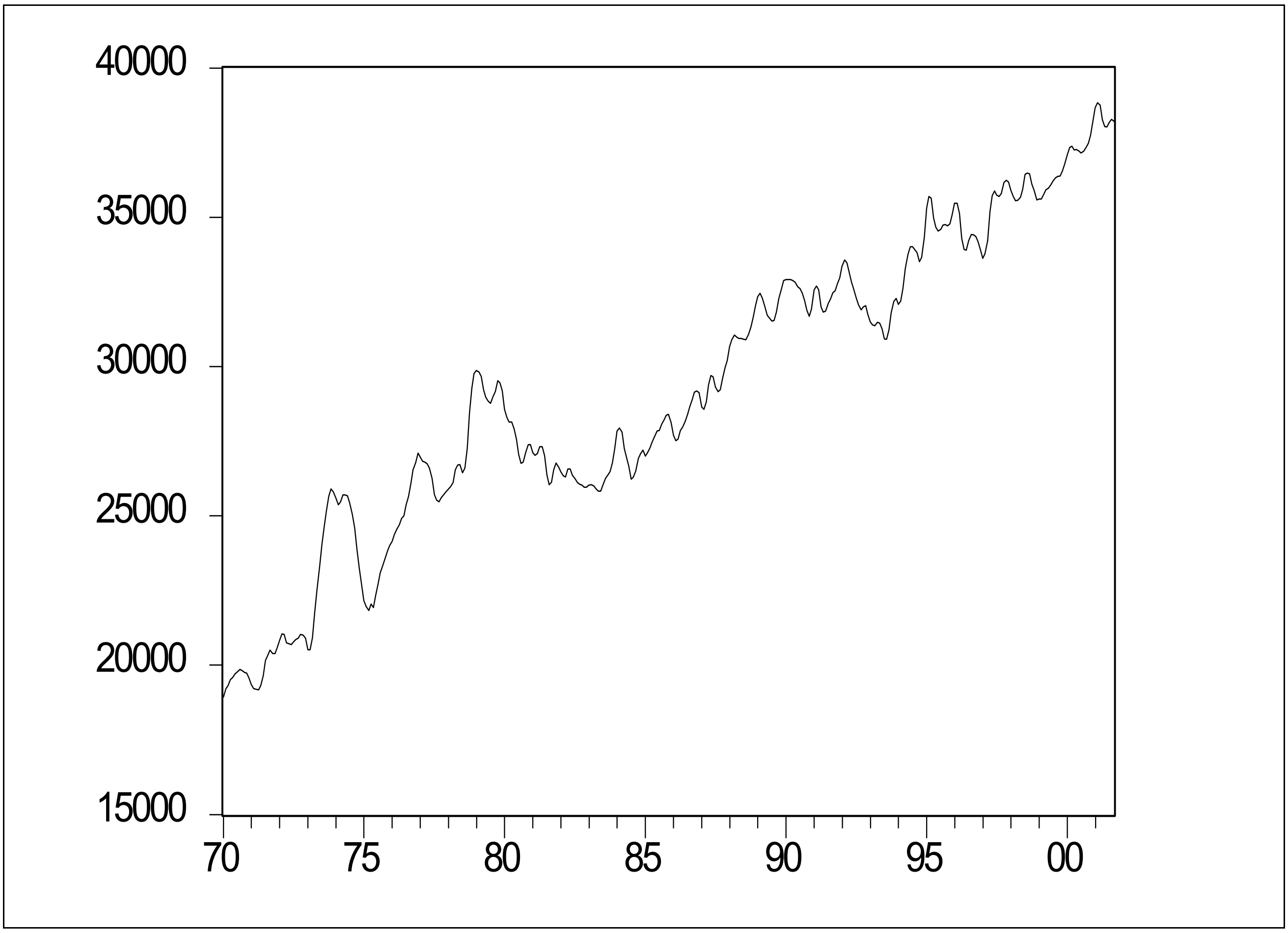


Annual % changes



Monthly industrial value added

Estimated series according to a first order dynamic model in logs



DYNCHOW

A Gauss programme for temporal disaggregating a time series (Eurostat copyright)

Generalities

- **10 different high-frequency models are considered**

- **Both original and log-transformed data can be modelled**

- **Flexible enough to deal with several cases often encountered in practical work**

- **Documented (Di Fonzo, 2003b)**

Details

- **Four different types of aggregation**

- **The high-frequency model can be expressed either in levels or in logs of the variable of interest**

- ***Dynchow* uses the right approximation for each type of aggregation, producing high-frequency estimated values coherent with the known (original, not transformed) aggregated series**

- **Estimation is performed *via* a non linear optimization routine , either CML or CO (sensible safe of computing time wrt scanning)**

Running *Dynchow*

{ydisag,arpar,coeff,diaggls,uroot}=
dynchow(y0,X,hfm,lnt,s,aggr,ap,em,op,nlap,nstep);

INPUT VARIABLES FOR DYNCHOW

y0: time series to be disaggregated (m x 1)
X: high-frequency related series (n x k) - WITHOUT CONSTANT!
hfm: high-frequency model
lnt: log-transformation
s: aggregation order
aggr: type of aggregation
ap: estimation approach (1: Santos Silva and Cardoso, 2: Salazar et al.)
em: estimation method (1: Estimated GLS, 2: Maximum Likelihood)
op: optimization procedure (1: Non linear, 2: Scanning)
nlap: application module used (1: CML, 2: CO).
Needed only if op==1. Forced to 2 if em==1
nstep: number of steps for the scanning procedure
(11<= nstep <=501). Needed only if op==2.

OUTPUT OF DYNCHOW

ydisag: estimated disaggregated series + log-estimates + s.e. (n x 3)
arpar: estimated autoregressive parameter + s.e. + t-stat + p-value (1 x 4)
coeff: estimated regression coefficients + s.e + t-stat + p-values (k x 4)
diaggls: regression diagnostics (10 x 2)
uroot: unit roots tests (6 x 8; produced only if m > 20)

DYNCHOW: **input parameters for the specification of the model**

hfm	High frequency model		
1	First order dynamic model with constant	$(1-\phi L)y_{h,t} = \alpha + \mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
2	First order dynamic model without constant	$(1-\phi L)y_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
3	First differences regression model with drift	$(1-L)y_{h,t} = \alpha + (1-L)\mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
4	First differences regression model without drift	$(1-L)y_{h,t} = (1-L)\mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
5	First differences in *y* with drift	$(1-L)y_{h,t} = \alpha + \mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
6	First differences in *y* without drift	$(1-L)y_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
7	Chow and Lin: Static regression model with constant and AR(1) disturbances	$y_{h,t} = \alpha + \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t}$ $(1-\phi L)u_{h,t} = \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
8	Chow and Lin: Static regression model without constant and with AR(1) disturbances	$y_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t}$ $(1-\phi L)u_{h,t} = \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
9	Litterman: Static regression model with constant and ARIMA(1,1,0) disturbances	$y_{h,t} = \alpha + \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t}$ $(1-\phi L)(1-L)u_{h,t} = \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$
10	Litterman: Static regression model without constant and with ARIMA(1,1,0) disturbances	$y_{h,t} = \mathbf{x}'_{h,t}\boldsymbol{\beta} + u_{h,t}$ $(1-\phi L)(1-L)u_{h,t} = \varepsilon_{h,t}$	$t=1,\ldots,T \quad h=1,\ldots,s$

Notice: the user should avoid to put deterministic variables as either a constant or a linear deterministic trend in vector $\mathbf{x}_{h,t}$, because they are explicitly (the constant) or implicitly (the linear trend) treated by the program

Major drawbacks of *Dynchow*

- **Research programme, not well suited for practices of massive routine indirect estimation of series**

- **Runs under Gauss (it is not a system-independent exe file)**

3. ALTERNATIVE MODEL FORMULATION

Two starting points

- **How to deal with seasonal time series (e.g., what can we do when the seasonal patterns of the related indicators do not fit the one – expected - of the series of interest)?**

- **Enlarge the (static and) dynamic formulation in (simple and) feasible ways**

Points to be stressed

- ***Annual aggregation of quarterly (or monthly) data destroys seasonality***

- ***What about quarterly aggregation of monthly data?***

Further issues

- **Links with ECM representation:**

meaningful and feasible in order to developing disaggregation formulae?

Constrained ECM representation of the simple first order dynamic model

$$y_t = \alpha_0 + \gamma_0 x_t + \alpha_1 y_{t-1} + \varepsilon_t$$

$$\Downarrow$$

$$\Delta y_t = \gamma_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + \varepsilon_t$$

$$\beta_0 = \frac{\alpha_0}{1-\alpha_1} \qquad \beta_1 = \frac{\gamma_0}{1-\alpha_1}$$

- **Exploit the state-space framework and Kalman filter (Harvey and Pierse, 1984, Harvey, 1989, Cuche and Hess, 2000)**

4. DEALING WITH SYSTEMS OF TIME SERIES

Extension 1: VAR-based benchmarking

Extension 2: BLUE – feasible multivariate AR(1)

Extension 3: Multivariate dynamic Chow-Lin

Extension 1: VAR-based benchmarking

Guerrero and Nieto (1999)

Benchmarking model exploiting the autoregressive features of a set of preliminary series

$$\mathbf{p}_t = \left(p_{1,t}, p_{2,t}, \ldots, p_{j,t}, \ldots, p_{M,t}\right)'$$

$$\mathbf{y}_t = \left(y_{1,t}, y_{2,t}, \ldots, y_{j,t}, \ldots, y_{M,t}\right)'$$

$\mathbf{p}_t$ **and** $\mathbf{y}_t$ **share the same autocorrelation structure** $\Rightarrow$ **VAR(*p*)**

$$\mathbf{\Pi}(L)(\mathbf{y}_t - \mathbf{p}_t) = \boldsymbol{\varepsilon}_t \qquad t = 1, \ldots, n$$

$$\mathbf{\Pi}(\mathbf{y}_g - \mathbf{p}_g) = \boldsymbol{\varepsilon}_g$$

$$E(\boldsymbol{\varepsilon}_g \mid \mathbf{p}_g) = \mathbf{0} \qquad E(\boldsymbol{\varepsilon}_g \boldsymbol{\varepsilon}_g' \mid \mathbf{p}_g) = \mathbf{P} \otimes \mathbf{\Sigma}$$

$$\hat{\mathbf{y}}_g = \mathbf{p}_g + \mathbf{A}(\mathbf{y}_{a,g} - \mathbf{C}_g \mathbf{p}_g)$$

$$\mathbf{A} = \mathbf{\Pi}^{-1}(\mathbf{P} \otimes \mathbf{\Sigma})\mathbf{\Pi}^{-1'}\mathbf{C}_g'[\mathbf{C}_g\mathbf{\Pi}^{-1}(\mathbf{P} \otimes \mathbf{\Sigma})\mathbf{\Pi}^{-1'}\mathbf{C}_g']^{-}$$

$$Cov(\hat{\mathbf{y}}_g - \mathbf{y}_g \mid \mathbf{p}_g) = (\mathbf{I}_{Mn} - \mathbf{A}\mathbf{C}_g)\mathbf{\Pi}^{-1}(\mathbf{P} \otimes \mathbf{\Sigma})\mathbf{\Pi}^{-1'}$$

Multivariate Denton's benchmarking procedure with a data-based weighting scheme

$$\mathbf{M} = \mathbf{\Pi}'(\mathbf{P} \otimes \mathbf{\Sigma})^{-1}\mathbf{\Pi}$$

NB: slight differences in notation (data organized by time and not by variable)

A two-stage operational procedure

First stage

- **Estimate a VAR(*p*) for $\mathbf{p}_t$**

- **Get an estimate of Π ($\tilde{\Pi}$) assuming**

$$\mathbf{P} \otimes \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{Mn}$$

- **'Tentative' disaggregated series**

$$\tilde{\mathbf{y}}_g$$

Second stage

- **Test whiteness of the residual series**

$$\hat{\Pi}(\tilde{\mathbf{y}}_g - \mathbf{p}_g) = \tilde{\boldsymbol{\varepsilon}}$$

to verify $\mathbf{P} \otimes \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{Mn}$

- **If not, fit a VAR(*p*) model to $\tilde{\varepsilon}_t$ and derive the implied estimate of $\mathbf{P} \otimes \boldsymbol{\Sigma}$**

- **Estimate $\hat{\mathbf{y}}_g$**

Empirically validate the compatibility between $\mathbf{y}_g$ and $\mathbf{p}_g$

A discrepancy measure: Wald statistic

$$DM = (\mathbf{y}_{a,g} - \mathbf{C}_g\mathbf{p}_g)'[\mathbf{C}_g\mathbf{\Pi}^{-1}(\mathbf{P}\otimes\mathbf{\Sigma})\mathbf{\Pi}^{-1'}\mathbf{C}_g^{'})]^{-}(\mathbf{y}_{a,g} - \mathbf{C}_g\mathbf{p}_g)$$

Asymptotically χ_g^2

$$g = rank[\mathbf{C}_g\mathbf{\Pi}^{-1}(\mathbf{P}\otimes\mathbf{\Sigma})\mathbf{\Pi}^{-1'}\mathbf{C}_g^{'})]$$

If *DM* rejects $\Rightarrow$ other preliminary series have to be found

The outlined procedure as implemented in the routine *MULTDIS*

- p_g **is computed using univariate AR(1) Chow and Lin**

- **The order of VAR is found according to AIC up to lag order 4 (should be appropriate for quarterly seasonally adjusted series)**

Extension 2: BLUE - feasible AR(1)

$$u_{j,\tau} = \rho_j u_{j,\tau-1} + \varepsilon_{j,\tau} \qquad \tau = 1,\dots,n \quad j = 1,\dots,M$$

$$\mathbf{P}_j = \begin{bmatrix} 1 & 0 & & 0 & 0 \\ -\rho_j & 1 & \cdots & 0 & 0 \\ 0 & -\rho_j & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\rho_j & 1 \end{bmatrix}$$

$$\mathbf{P}_j \mathbf{u}_j = \boldsymbol{\varepsilon}_j \qquad \mathbf{u}_j = \mathbf{P}_j^{-1} \boldsymbol{\varepsilon}_j,$$

$$\mathbf{P}_j^{-1} = \begin{bmatrix} 1 & 0 & & 0 & 0 \\ \rho_j & 1 & \cdots & 0 & 0 \\ \rho_j^2 & \rho_j & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_j^{n-1} & \rho_j^{n-2} & \cdots & \rho_j & 1 \end{bmatrix}.$$

$$E\left(\mathbf{u}_i \mathbf{u}_j'\right) = \sigma_{ij} \left(\mathbf{P}_j' \mathbf{P}_i\right)^{-1}$$

A procedure to jointly estimate the *M* time series

Adaption of Cabrer and Pavía (1999)

1. **Get estimates $\tilde{\rho}_j$ and $\tilde{\mathbf{y}}_j$ using univariate AR(1) Chow and Lin**

2. **Calculate $\tilde{\mathbf{u}}_j = \tilde{\mathbf{y}}_j - \mathbf{X}_j \tilde{\boldsymbol{\beta}}_j^{OLS}$**

3. **ADF test on $\tilde{\mathbf{u}}_j$: equal ρ_j to unity where the null is accepted**

4. **Calculate $\tilde{\varepsilon}_{j,t} = \tilde{u}_{j,t} - \tilde{\rho}_j \tilde{u}_{j,t-1}$ and estimate**
$$\tilde{\sigma}_{ij} = \frac{1}{n} \sum_{t=1}^{n} \tilde{\varepsilon}_{i,t} \tilde{\varepsilon}_{j,t}$$

5. **Estimate y using the BLUE formulae**

Extension 3: Multivariate Dynamic Chow-Lin

$$y_{j,\tau} = \phi_j y_{j,\tau-1} + \mathbf{x}'_{j,\tau}\boldsymbol{\beta}_j + \varepsilon_{j,\tau}$$

$$\tau = 1,\ldots,n\text{, } j = 1,\ldots,M$$

$$\mathbf{D}_{\phi,j}\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{q}_j\eta_j + \boldsymbol{\varepsilon}_j \qquad j = 1,\ldots,M$$

$$\mathbf{D}_{\phi,j} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\phi_j & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\phi_j & 1 \end{bmatrix} \qquad \mathbf{q}_j = \left(\phi_j, 0, \ldots, 0\right)'$$

$$\mathbf{y}_j = \mathbf{D}_{\phi,j}^{-1}\mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{D}_{\phi,j}^{-1}\mathbf{q}_j\eta_j + \mathbf{D}_{\phi,j}^{-1}\boldsymbol{\varepsilon}_j$$

$$\mathbf{u}_j = \mathbf{D}_{\phi,j}^{-1}\boldsymbol{\varepsilon}_j \sim \textbf{AR(1)}$$

$$\mathbf{V}_{ij} = E\left(\mathbf{u}_i\mathbf{u}'_j\right) = \sigma_{ij}\left(\mathbf{D}'_{\phi,j}\mathbf{D}_{\phi,i}\right)^{-1} \qquad i, j = 1,\ldots,M$$

A procedure to jointly estimate the *M* time series

Adaption of multivariate AR(1)

1. **Get estimates $\breve{\phi}_j$ and $\breve{\mathbf{y}}_j$ using the univariate dynamic generalization of Chow and Lin (**DYNCHOW**)**

2. **Calculate $\breve{\mathbf{u}}_j = \breve{\mathbf{y}}_j - \mathbf{Z}_{\phi,j}\breve{\boldsymbol{\gamma}}_j^{OLS}$, where**
$\mathbf{Z}_{\phi,j} = \mathbf{D}_{\phi,j}^{-1}\left[\mathbf{X}_j \vdots \mathbf{q}_j\right]$, $\boldsymbol{\gamma}_j = \left[\boldsymbol{\beta}'_j, \eta_j\right]'$

3. **Calculate $\breve{\varepsilon}_{j,t} = \breve{u}_{j,t} - \breve{\rho}_j \breve{u}_{j,t-1}$ and estimate**
$$\breve{\sigma}_{ij} = \frac{1}{n}\sum_{t=1}^{n}\breve{\varepsilon}_{i,t}\breve{\varepsilon}_{j,t}$$

4. **Estimate y using the BLUE formulae**

The routine *MULTDIS*

MULTivariate DISaggregation of time series
(Eurostat copyright)

- **Written in GAUSS language**

- **Takes back the work done for the multivariate case in ECOTRIM**

- **Multivariate AR(1)**

- **Preliminary version of Guerrero and Nieto**

Running MULTDIS

{ydisag, coeff, diagmult} =
MULTDIS (ya, x, pmind, z, hfm, s, taggr, caggr, out);

INPUT

ya **: matrix of series to be disaggregated**

x **: high-frequency related series**
WITHOUT CONSTANT!

pmind **: reference index of x with respect to ya**

z **: contemporaneous high-frequency constraint**

hfm **: high-frequency multivariate model**
1: white noise 2: random walk
3: AR(1) 4: Guerrero and Nieto

s **: aggregation order**

taggr **: type of temporal aggregation (1: sum, 2: average, 3: sample of the last, 4: sample of the first)**

caggr **: type of contemporaneous aggregation**
(1: sum, 2: average)

out **: type of output (1: full, 0: light)**