OECD/Eurostat workshop

Application of advanced temporal disaggregation techniques to economic statistics

Tour Europe, La Defense, Paris

27 November 2003

INTRODUCTION TO THE PRINCIPLES OF BENCHMARKING AND TEMPORAL DISAGGREGATION FOR ECONOMIC SERIES

Tommaso Di Fonzo

University of Padova
difonzo@stat.unipd.it

OUTLINE OF THE PRESENTATION

1. **Statement of the problem**

2. **Benchmarking: Denton's univariate procedure**

3. **Temporal disaggregation using related series: Chow and Lin's univariate approach**

4. **Benchmarking/Disaggregating a system of time series linked by an accounting constraint**

1. STATEMENT OF THE PROBLEM

- **Need for economic time series**

'High' frequency (quarterly, monthly)
'Long' time spans

for statisticians, econometricians and economic analysts

(*Action Plan on EMU Statistical Requirements*, september 2000)

Availability of complete, reliable (and timely) time series of the main economic aggregates

Quarterly National Accounts

Principal Economic Indicators

Unavailability of (part of) needed information

$\Downarrow$

hard or impossible to directly estimate

- **Use all the available information to get 'reasonable' estimates**

- **Quality and usefulness (*relevance*) depend on**

$\Rightarrow$ **Amount of available information**

$\Rightarrow$ **Way in which the various 'pieces' of information are combined**

TERMS OF THE PROBLEM-1

Benchmarking

Refers to the case where there are two sources of data for the same target variable, with different frequencies, and is concerned with correcting inconsistencies between the different estimates

Given monthly and annual estimates of retail trade series from different sources, reconcile the estimates
Make the annual total of SA series 'coherent' with the raw data

The dynamic profile of the target variable and of the unbenchmarked series should be similar. Warning: the step problem

TERMS OF THE PROBLEM-2

Temporal disaggregation of time series

Process of deriving high frequency data from low frequency data and, if available, related high frequency information

Given a 'low' frequency (yearly, quarterly) observed time series, estimate the 'high' frequency values (quarterly, monthly)

A temporal aggregation constraint must be fulfilled: the estimated figures must be 'coherent' with the temporally aggregated available data. Warning: model selection and choice of indicator series

FIELDS OF APPLICATION

Quarterly Accounts

give a quarterly breakdown of the figures in the annual accounts

Flash estimates

use the available information in the best possible way including, in the framework of a statistical model, the short-term available information and the low frequency data in a coherent way

Seasonal adjustment

Fulfil temporal (wrt raw data) and longitudinal (wrt accounting constraints, direct SA) coherence of SA series

ESA 1995 (12.04)

The statistical methods used for compiling quarterly accounts may differ quite considerably from those used for the annual accounts.

They can be classified in two major categories: direct procedure and indirect procedure.

Direct procedures are based on the availability at quarterly intervals, with appropriate simplifications, of the similar sources as used to compile the annual accounts.

On the other hand, indirect procedures are based on temporal disaggregation of the annual accounts data in accordance with mathematical and statistical methods using reference indicators that permit the extrapolation for the current year.

The choice between the different indirect procedures must above all take into account the minimisation of the forecast error for the current year, in order that the provisional annual estimates correspond as closely as possible to the final figures.

The choice between these approaches depends, among other things, on the information available at quarterly level.

The elements of a multivariate annual-quarterly indirect estimation/benchmarking problem

Gray shaded elements: unknown or unbenchmarked

Year T	Quarter h	$y_{1,t}$	$y_{01,T}$	$\cdots$	$y_{j,t}$	$y_{0j,T}$	$\cdots$	$y_{M,t}$	$y_{0M,T}$	z_t	$z_{0,T}$
	1	$y_{1,1}$		$\cdots$	$y_{j,1}$		$\cdots$	$y_{M,1}$		z_1	
	2	$y_{1,2}$		$\cdots$	$y_{j,2}$		$\cdots$	$y_{M,2}$		z_2	
	3	$y_{1,3}$		$\cdots$	$y_{j,3}$		$\cdots$	$y_{M,3}$		z_3	
	4	$y_{1,4}$		$\cdots$	$y_{j,4}$		$\cdots$	$y_{M,4}$		z_4	
1			$y_{01,1}$	$\cdots$		$y_{0j,1}$	$\cdots$		$y_{0M,1}$		$z_{0,1}$
	$\cdots$	$\cdots$			$\cdots$			$\cdots$		$\cdots$	
	1	$y_{1,4T-3}$		$\cdots$	$y_{j,4T-3}$		$\cdots$	$y_{M,4T-3}$		z_{4T-3}	
	2	$y_{1,4T-2}$		$\cdots$	$y_{j,4T-2}$		$\cdots$	$y_{M,4T-2}$		z_{4T-2}	
	3	$y_{1,4T-1}$		$\cdots$	$y_{j,4T-1}$		$\cdots$	$y_{M,4T-1}$		z_{4T-1}	
	4	$y_{1,4T}$		$\cdots$	$y_{j,4T}$		$\cdots$	$y_{M,4T}$		z_{4T}	
T			$y_{01,T}$	$\cdots$		$y_{0j,T}$	$\cdots$		$y_{0M,T}$		$z_{0,T}$
	$\cdots$	$\cdots$			$\cdots$			$\cdots$		$\cdots$	
	1	$y_{1,4N-3}$		$\cdots$	$y_{j,4N-3}$		$\cdots$	$y_{M,4N-3}$			
	2	$y_{1,4N-2}$		$\cdots$	$y_{j,4N-2}$		$\cdots$	$y_{M,4N-2}$		z_{4N-3}	
	3	$y_{1,4N-1}$		$\cdots$	$y_{j,4N-1}$		$\cdots$	$y_{M,4N-1}$		z_{4N-2}	
	4	$y_{1,4N}$		$\cdots$	$y_{j,4N}$		$\cdots$	$y_{M,4N}$		z_{4N-1}	
N			$y_{01,N}$	$\cdots$		$y_{0j,N}$	$\cdots$		$y_{0M,N}$	z_{4N}	$z_{0,N}$
	1	$y_{1,n-1}$		$\cdots$	$y_{j,n-1}$		$\cdots$	$y_{M,n-1}$		z_{n-1}	
	2	$y_{1,n}$		$\cdots$	$y_{j,n}$		$\cdots$	$y_{M,n}$		z_n	

TEMPORAL DISAGGREGATION PROCEDURES

Procedures without indicators

Smoothing procedures

Time series procedures

These methods should be used only when there are serious gaps in basic information

Procedures with indicators

Two-steps adjustment procedures

Optimal regression based procedures

HF: High Frequency (e.g., quarter, month)
LF: Low Frequency (e.g., year, quarter)

PROCEDURES WITHOUT INDICATORS-1

Smoothing procedures

(Lisman and Sandee, 1964; Boot *et al.*, 1967; Jacobs, 1994)

- **short term dynamics based on *a priori* assumption and purely mathematical techniques ensuring a sufficiently smoothed path**
- **extrapolation generally not allowed**
- **lack of information**
- **no real enlargement of the information base**

PROCEDURES WITHOUT INDICATORS-2

Time series procedures

(Al-Osh, 1989; Wei and Stram, 1990)

- **an ARIMA model for the LF series is used to get a 'reasonable' (in least squares sense) estimate of the HF data generating process**
- **allows for extrapolation**
- **measures of reliability of the estimates**
- **generally, no seasonal pattern can be recovered**
- **requires computational resources and an active role of the user**

PROCEDURES USING RELATED SERIES

The HF path is estimated on the basis of external HF information on logically and/or economically (not in a modelling, but in a 'measurement' sense!) related variables

Two steps adjustment methods

The process of estimation can be logically divided into two parts

1. **Preliminary estimation (from survey, according to some indirect estimation techniques,…) of the HF series of interest *not consistent with the LF counterpart***

2. **Adjustment (benchmarking) of the preliminary estimates in order to *fit* the known annual series**

Notice: In the multivariate case (system of time series) the second step includes the fulfilment of the contemporaneous accounting constraints

Preliminary estimation generally according to either

***direct way* (e.g., *sample survey*)**

or

***mathematical-statistical way* (e.g., by using a linear regression relationship between the LF target series and the LF related indicators)**

However obtained, preliminary estimates *do not generally satisfy the temporal aggregation constraints*

The LF discrepancy (residuals) must be distributed according to an adjustment procedure

The adjustment procedure (benchmarking) should operate by fitting the LF constraints and by altering the HF path of the preliminary estimates to the least extent possible

Optimal regression based procedures

Preliminary estimation and adjustment are simultaneously performed according to a statistically optimal procedure

The linkage between target series and indicators is formulated at HF level

A temporally aggregated (*LF and observable*) regression model is derived from the HF counterpart

Estimates are obtained as constrained and conditional forecasts of a generalized (autocorrelated disturbances) regression model

SOME DEFINITION

Distribution

LF (e.g., annual) data are either sums or averages of HF (e.g., quarterly) data (GDP, consumption, indexes and in general all flow variables and all average stock variables)

Interpolation

LF value equals by definition that of the last (or first) HF reference period (e.g., population at the end of the year, money stock, and all stock variables)

Extrapolation

Estimates of HF data are needed when the relevant LF value is not yet available

NOTATION

$y_{h,t}$: **variable of interest (unknown)**

$p_{h,t}$: **preliminary (unbenchmarked) estimate (known)**

$\mathbf{x}_{h,t}$: **vector (*k*,1) of related series (known)**

h=1,…,s, t=1,…,T $\Rightarrow$ ***n=sT***

s=3 $\Leftrightarrow$ quarterly to monthly
s=4 $\Leftrightarrow$ yearly to quarterly
s=12 $\Leftrightarrow$ yearly to monthly

Temporally aggregated series (known)

$$y_t = \sum_{h=1}^{s} c_h y_{h,t} \qquad t=1,\dots,T$$

Flows $\Rightarrow$ **c**=(1,1,…,1)′

Index $\Rightarrow$ **c**=(1,1,…,1)′/*s*

Stock $\Rightarrow$ **c**=(0,0,…,1)′

TEMPORAL AGGREGATION MATRICES

1. Distribution of flow variables

$$\mathbf{C}^{dist} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1 \end{bmatrix}$$

2. Estimation of index variables $\mathbf{C}^{ind} = \frac{1}{s}\mathbf{C}^{dist}$

3. Interpolation of stock variables

$$\mathbf{C}^{inter} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

4. Extrapolation:

$$\mathbf{C}^{extr} = \begin{bmatrix} & 0 & 0 \\ \mathbf{C}^{dist/ind/inter} & \vdots & \vdots \\ & 0 & 0 \end{bmatrix}$$

2. BENCHMARKING: DENTON'S UNIVARIATE PROCEDURE

References

- **Denton (1971)**
- **Helfand *et al.* (1977)**
- **Monsour and Trager (1979)**
- **Cholette (1984)**
- **Laniel and Fyfe (1990)**
- **Cholette and Dagum (1994)**
- **Durbin and Quenneville (1997)**
- **Chen, Cholette and Dagum (1997)**
- **Bloem *et al.* (2001)**
- **Quenneville *et al.* (2003)**

The adjusted (benchmarked) estimates are obtained according to a *movement preservation principle*

on levels

⇓

Additive (First Differences) variant

minimize wrt to $y_{h,t}$ the quadratic loss function

$$\sum_{\tau=2}^{n}\left[\left(y_\tau - y_{\tau-1}\right)-\left(p_\tau - p_{\tau-1}\right)\right]^2 \equiv \sum_{\tau=2}^{n}\left[\left(y_\tau - p_\tau\right)-\left(y_{\tau-1} - p_{\tau-1}\right)\right]^2$$

on proportionate levels

⇓

Proportional (First Differences) variant

minimize wrt to $y_{h,t}$ the quadratic loss function

$$\sum_{\tau=2}^{n}\left(\frac{y_\tau - p_\tau}{p_\tau}-\frac{y_{\tau-1} - p_{\tau-1}}{p_{\tau-1}}\right)^2 \equiv \sum_{t=2}^{n}\left(\frac{y_\tau}{p_\tau}-\frac{y_{\tau-1}}{p_{\tau-1}}\right)^2$$

in both cases: subject to the aggregation constraint $\mathbf{C}\mathbf{y}_h = \mathbf{y}_l$

In general, using matrix notation, the objective function (with M (*n*x*n*) p.d.)

$$(\mathbf{y}_h - \mathbf{p})'\mathbf{M}(\mathbf{y}_h - \mathbf{p}) = \mathbf{y}_h'\mathbf{M}\mathbf{y}_h - 2\mathbf{p}'\mathbf{M}\mathbf{y}_h + \mathbf{p}'\mathbf{M}\mathbf{p}$$

must be optimized while fulfilling the constraint $\mathbf{C}\mathbf{y}_h = \mathbf{y}_l$

The benchmarked estimates are given by

$$\tilde{\mathbf{y}}_h = \mathbf{p} + \mathbf{M}^{-1}\mathbf{C}'\left(\mathbf{C}\mathbf{M}^{-1}\mathbf{C}'\right)^{-1}\left(\mathbf{y}_l - \mathbf{C}\mathbf{p}\right)$$

AFD variant

(approximate solution, assuming $y_0 - p_0 = 0$)

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

PFD variant

(approximate solution, assuming $y_0 / p_0 = 1$)

$$\mathbf{M} = \hat{\mathbf{p}}^{-1}\mathbf{D'D}\hat{\mathbf{p}}^{-1} \qquad \hat{\mathbf{p}} = \begin{bmatrix} p_1 & 0 & 0 & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p_n \end{bmatrix}$$

3. TEMPORAL DISAGGREGATION USING RELATED SERIES

The standard (linear and static) approach

References

- **Chow and Lin (1971, 1976)**
- **Bournay e Laroque (1979)**
- **Fernández (1981)**
- **Litterman (1983)**
- **Di Fonzo (1987, 2002b)**
- **Wei and Stram (1990)**
- **Chan (1993)**
- **Eurostat (1999)**

STANDARD REFERENCE (STATIC) MODEL

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

$\mathbf{x}_{h,t}$ $(k,1)$ $\qquad$ $\boldsymbol{\beta}$ $(k,1)$ $\qquad$ $u_{h,t} \perp \mathbf{x}_{h,t}$

$$\mathbf{y}_h = \mathbf{X}_h\boldsymbol{\beta} + \mathbf{u}_h$$

$$E\left(\mathbf{u}_h\mathbf{u}_h' \mid \mathbf{X}_h\right) = \mathbf{V}_h$$

$$\mathbf{C} = \mathbf{I}_T \otimes \mathbf{c}' \Rightarrow \quad \mathbf{y}_l = \mathbf{C}\mathbf{y}_h$$

BLU ESTIMATION OF $\mathbf{y}_h$

Premultiply by C

$$\Downarrow$$

$$\mathbf{C}\mathbf{y}_h = \mathbf{C}\mathbf{X}_h\boldsymbol{\beta} + \mathbf{C}\mathbf{u}_h$$

$$\Downarrow$$

$$\mathbf{y}_l = \mathbf{X}_l\boldsymbol{\beta} + \mathbf{u}_l$$

$$\mathbf{y}_l = \mathbf{C}\mathbf{y}_h \qquad \mathbf{X}_l = \mathbf{C}\mathbf{X}_h$$

$$E\left(\mathbf{u}_l\mathbf{u}_l' \mid \mathbf{X}_h\right) = \mathbf{V}_l = \mathbf{C}\mathbf{V}_h\mathbf{C}'$$

Classical solution (Chow and Lin, 1971)

$$\hat{\mathbf{y}}_h = \mathbf{X}_h\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\mathbf{u}}_l$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_l'\mathbf{V}_l^{-1}\mathbf{X}_l\right)^{-1}\mathbf{X}_l'\mathbf{V}_l^{-1}\mathbf{y}_l$$

GLS estimator of β in the aggregated (observable) model

$$\mathbf{y}_l = \mathbf{X}_l\boldsymbol{\beta} + \mathbf{u}_l$$

$$\mathbf{L} = \mathbf{V}_h\mathbf{C}'\mathbf{V}_l^{-1} \qquad \hat{\mathbf{u}}_l = \mathbf{y}_l - \mathbf{X}_l\hat{\boldsymbol{\beta}}$$

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

$$= \left(\mathbf{I}_n - \mathbf{LC}\right)\mathbf{V}_h + \left(\mathbf{X}_h - \mathbf{LX}_l\right)\left(\mathbf{X}_l'\mathbf{V}_l^{-1}\mathbf{X}_l\right)^{-1}\left(\mathbf{X}_h - \mathbf{LX}_l\right)'$$

THE PROBLEM: THE CHOICE OF $\mathbf{V}_h$

- **How identify $\mathbf{V}_h$ from a 'reasonable' estimate of $\mathbf{V}_l$?**

- **In general, $\mathbf{V}_h$ cannot be uniquely identified from the relationship $\mathbf{V}_l = \mathbf{C}\mathbf{V}_h\mathbf{C}'$**

DGP for $u_{h,t} \mid \mathbf{X}_h$ (Eurostat, 1999)

- Chow and Lin (1971): $u_{h,t} \mid \mathbf{X}_h \sim AR(1)$
- Fernàndez (1981): $u_{h,t} \mid \mathbf{X}_h \sim \text{Random Walk}$
- Litterman (1983): $u_{h,t} \mid \mathbf{X}_h \sim ARIMA(1,1,0)$
- Wei and Stram (1990): $u_{h,t} \mid \mathbf{X}_h \sim ARIMA(p,d,q)$

Computational simplicity of estimate, availability of *ad hoc* procedures (Ecotrim)

$\Downarrow$

Chow and Lin's AR(1) variant is the most used (Istat, INSEE, INE, Eurostat)

<u>Problems</u>

- **Uncorrect specification (bad quality of the LF regression, residuals generated by integrated processes)**
- **Modelling using raw/sa related indicators**

The usefulness of this procedure largely depends on the validity of the assumed regression model

4. BENCHMARKING/DISAGGREGATING A SYSTEM OF TIME SERIES LINKED BY AN ACCOUNTING CONSTRAINT

References

- **Rossi (1982)**
- **Cholette (1988)**
- **Taillon (1988)**
- **Di Fonzo (1990)**
- **Chen and Dagum (1997)**
- **Eurostat (1999)**
- **Di Fonzo and Marini (2003)**

THE PRACTICAL PROBLEM

- **Missing data problem**

  Estimate $y_{h,j}$, $j = 1,\ldots,M$

  to be solved in an indirect estimation framework

The available data

- *M* **temporally aggregated (say, annual) time series**

  $y_{l,j}$, $j = 1,\ldots,M$

- **a contemporaneously aggregated high-frequency (say, quarterly) time series**

  $\mathbf{z}$

The elements of a multivariate annual-quarterly indirect estimation/benchmarking problem (again!)

Year T	Quarter h	$y_{1,t}$	$y_{01,T}$	$\cdots$	$y_{j,t}$	$y_{0j,T}$	$\cdots$	$y_{M,t}$	$y_{0M,T}$	z_t	$z_{0,T}$
	1	$y_{1,1}$		$\cdots$	$y_{j,1}$		$\cdots$	$y_{M,1}$		z_1	
	2	$y_{1,2}$		$\cdots$	$y_{j,2}$		$\cdots$	$y_{M,2}$		z_2	
	3	$y_{1,3}$		$\cdots$	$y_{j,3}$		$\cdots$	$y_{M,3}$		z_3	
	4	$y_{1,4}$		$\cdots$	$y_{j,4}$		$\cdots$	$y_{M,4}$		z_4	
1			$y_{01,1}$	$\cdots$		$y_{0j,1}$	$\cdots$		$y_{0M,1}$		$z_{0,1}$
	$\cdots$	$\cdots$			$\cdots$			$\cdots$		$\cdots$	
	1	$y_{1,4T-3}$		$\cdots$	$y_{j,4T-3}$		$\cdots$	$y_{M,4T-3}$		z_{4T-3}	
	2	$y_{1,4T-2}$		$\cdots$	$y_{j,4T-2}$		$\cdots$	$y_{M,4T-2}$		z_{4T-2}	
	3	$y_{1,4T-1}$		$\cdots$	$y_{j,4T-1}$		$\cdots$	$y_{M,4T-1}$		z_{4T-1}	
	4	$y_{1,4T}$		$\cdots$	$y_{j,4T}$		$\cdots$	$y_{M,4T}$		z_{4T}	
T			$y_{01,T}$	$\cdots$		$y_{0j,T}$	$\cdots$		$y_{0M,T}$		$z_{0,T}$
	$\cdots$	$\cdots$			$\cdots$			$\cdots$		$\cdots$	
	1	$y_{1,4N-3}$		$\cdots$	$y_{j,4N-3}$		$\cdots$	$y_{M,4N-3}$			
	2	$y_{1,4N-2}$		$\cdots$	$y_{j,4N-2}$		$\cdots$	$y_{M,4N-2}$		z_{4N-3}	
	3	$y_{1,4N-1}$		$\cdots$	$y_{j,4N-1}$		$\cdots$	$y_{M,4N-1}$		z_{4N-2}	
	4	$y_{1,4N}$		$\cdots$	$y_{j,4N}$		$\cdots$	$y_{M,4N}$		z_{4N-1}	
N			$y_{01,N}$	$\cdots$		$y_{0j,N}$	$\cdots$		$y_{0M,N}$	z_{4N}	$z_{0,N}$
	1	$y_{1,n-1}$		$\cdots$	$y_{j,n-1}$		$\cdots$	$y_{M,n-1}$		z_{n-1}	
	2	$y_{1,n}$		$\cdots$	$y_{j,n}$		$\cdots$	$y_{M,n}$		z_n	

Further(!) notation

M: **number of series to be disaggregated**

n: **number of HF periods**

N: **number of LF periods**

s: **temporal aggregation order (number of infra-annual periods)**

n > sN $\Rightarrow$ constrained extrapolation

The aggregation constraints

Contemporaneous $$\sum_{j=1}^{M} \mathbf{y}_{h,j} = \mathbf{z}$$

Temporal $$\mathbf{C}\mathbf{y}_{h,j} = \mathbf{y}_{l,j} \qquad j = 1, \ldots, M$$

$$\mathbf{C} = \left[\mathbf{I}_N \otimes \mathbf{c}' \vdots \quad \mathbf{0} \right]$$

$$\left(\mathbf{1}_M^{'} \otimes \mathbf{I}_n\right)\mathbf{y}_h = \mathbf{z} \qquad\qquad \left(\mathbf{I}_M \otimes \mathbf{C}\right)\mathbf{y}_h = \mathbf{y}_l$$

$$\Downarrow$$

$$\mathbf{H}\mathbf{y}_h = \mathbf{y}_a$$

H: $\left[\left(n + NM\right) \times nM\right]$ **aggregation matrix**

$$\mathbf{H} = \begin{bmatrix} \mathbf{1}_M^{'} \otimes \mathbf{I}_n \\ \mathbf{I}_M \otimes \mathbf{C} \end{bmatrix} \qquad\qquad \mathbf{y}_a = \begin{bmatrix} \mathbf{z} \\ \mathbf{y}_l \end{bmatrix}$$

Contemporaneous aggregation of temporally aggregated series

$$\sum_{j=1}^{M} y_{lj,T} = \sum_{h=1}^{s} z_{s(T-1)+h} = z_{0,T} \qquad T = 1,\ldots,N,$$

$$\Downarrow$$

$$\left(\mathbf{1}_M^{'} \otimes \mathbf{I}_N\right)\mathbf{y}_l = \mathbf{C}\mathbf{z}$$

Matrix H has not full row rank

N **aggregated observations are redundant,**

only *r=n+N(M-1)* 'free'

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_w \\ \cdots \\ \mathbf{H}_M \end{bmatrix}$$

$$\mathbf{H}_w = \begin{bmatrix} \mathbf{1}'_{M-1} \otimes \mathbf{I}_n & \mathbf{I}_n \\ \mathbf{I}_{M-1} \otimes \mathbf{C} & \mathbf{0} \end{bmatrix} \qquad \mathbf{H}_M = \left[\mathbf{0} \;\vdots\; \mathbf{C} \right]$$

$$\mathbf{W} = \left[\mathbf{C} \;\vdots\; -\left(\mathbf{1}'_{M-1} \otimes \mathbf{I}_N \right) \right] \qquad \mathbf{R} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{W} \end{bmatrix}$$

$$\Downarrow$$

$$\mathbf{H}_M = \mathbf{W}\mathbf{H}_w \qquad \mathbf{H} = \mathbf{R}\mathbf{H}_w$$

These results will turn out useful in handling the disaggregation formulae

Two distinct situations

- M **preliminary HF time series available**

$$\mathbf{p}_j, \qquad j = 1, \ldots, M$$

$$\sum_{j=1}^{M} \mathbf{p}_j \neq \mathbf{z}$$

and/or $\mathbf{p}_j$ **doesn't comply with** $\mathbf{y}_{lj}$

- **use of a set of HF related indicators** $\mathbf{X}_j$ **to obtain indirect estimates**

Distinction not necessarily as strict as it seems: preliminary HF series could have been individually obtained using related indicators

Benchmarking (two steps)

1. **proportional adjustment (Eurostat, 1999)**

2. **raking (iterative proportional fitting, Laniel and Fyfe, 1990)**

3. **Denton's multivariate benchmarking (*movement preservation principle*)**

$$\min_{\mathbf{y}_h} (\mathbf{y}_h - \mathbf{p})'\mathbf{M}(\mathbf{y}_h - \mathbf{p})$$

subject to $\mathbf{H}\mathbf{y}_h = \mathbf{y}_a$

$$\Downarrow$$

$$\tilde{\mathbf{y}}_h = \mathbf{p} + \mathbf{M}^{-1}\mathbf{H}'\left(\mathbf{H}\mathbf{M}^{-1}\mathbf{H}'\right)^{-}(\mathbf{y}_a - \mathbf{H}\mathbf{p})$$

$\left(\mathbf{H}\mathbf{M}^{-1}\mathbf{H}'\right)^{-}$ **: MP generalized inverse**

Multivariate Denton's benchmarking: formulae not involving inversion of singular matrices

$$\left(\mathbf{H}\mathbf{M}^{-1}\mathbf{H}'\right)^{-} = \mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{M}_w^{-1}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'$$

$$\mathbf{M}_w = \mathbf{H}_w\mathbf{M}^{-1}\mathbf{H}_w'$$

$$\Downarrow$$

$$\tilde{\mathbf{y}}_h = \mathbf{p} + \mathbf{M}^{-1}\mathbf{H}_w'\mathbf{M}_w^{-1}\left(\mathbf{y}_w - \mathbf{H}_w\mathbf{p}\right)$$

$$\mathbf{y}_w = \mathbf{H}_w\mathbf{y}$$

Multivariate Denton's benchmarking: the choice of M

$\mathbf{I}_M \otimes (\mathbf{D}'\mathbf{D})$	Denton additive first difference
$\mathbf{I}_M \otimes (\mathbf{D}'\mathbf{D}'\mathbf{D}\mathbf{D})$	Denton additive second difference
$\mathbf{P}^{-1}(\mathbf{I}_M \otimes \mathbf{D}'\mathbf{D})\mathbf{P}^{-1}$	Denton proportional first difference
$\mathbf{P}^{-1}(\mathbf{I}_M \otimes \mathbf{D}'\mathbf{D}'\mathbf{D}\mathbf{D})\mathbf{P}^{-1}$	Denton proportional second difference
$(\mathbf{S} \otimes \mathbf{I}_n)^{-1}$	Rossi (1982) multivariate adjustment

D: $(n \times n)$ matrix performing first differences

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

The BLUE approach (single step)

$$\mathbf{y}_{h,j} = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{u}_j \qquad j = 1, \ldots, M$$

$$E(\mathbf{u}_j) = \mathbf{0} \qquad E(\mathbf{u}_i \mathbf{u}_j') = \mathbf{V}_{ij} \quad i, j = 1, \ldots, M$$

$\mathbf{X}_j$**:** $(n \times K_j)$ **matrices of related series**

$$\begin{bmatrix} \mathbf{y}_{h,1} \\ \mathbf{y}_{h,2} \\ \vdots \\ \mathbf{y}_{h,M} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_M \end{bmatrix}$$

$$\mathbf{y}_h = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad E(\mathbf{u}) = \mathbf{0} \quad E(\mathbf{u}\mathbf{u}') = \mathbf{V} = \{\mathbf{V}_{ij}\}$$

Observed, aggregated regression model

$$\mathbf{y}_a = \mathbf{X}_a \boldsymbol{\beta} + \mathbf{u}_a$$

$$\mathbf{X}_a = \mathbf{H}\mathbf{X} \qquad \mathbf{u}_a = \mathbf{H}\mathbf{u}$$

$E(\mathbf{u}_a \mathbf{u}_a') = \mathbf{V}_a = \mathbf{H}\mathbf{V}\mathbf{H}'$ **singular**

The solution (Di Fonzo, 1990)

Multivariate extension of the optimal univariate approach (Chow and Lin, 1971)

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{L}\left(\mathbf{y}_a - \mathbf{X}_a\hat{\boldsymbol{\beta}}\right)$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_a^{'}\mathbf{V}_a^{-}\mathbf{X}_a\right)^{-1}\mathbf{X}_a^{'}\mathbf{V}_a^{-}\mathbf{y}_a$$

$$\mathbf{L} = \mathbf{V}\mathbf{H}'\mathbf{V}_a^{-}$$

$$E\left[(\hat{\mathbf{y}} - \mathbf{y})(\hat{\mathbf{y}} - \mathbf{y})'\right] = (\mathbf{I}_n - \mathbf{L}\mathbf{H})\mathbf{V} + (\mathbf{X} - \mathbf{L}\mathbf{X}_a)\left(\mathbf{X}_a^{'}\mathbf{V}_a^{-}\mathbf{X}_a\right)^{-1}(\mathbf{X} - \mathbf{L}\mathbf{X}_a)'$$

NB: Disaggregation formulae involving MP generalized inverse

The BLUE in terms of *r* 'free' observations

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{L}_w\left(\mathbf{y}_w - \mathbf{X}_w\hat{\boldsymbol{\beta}}\right)$$

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_w^{'}\mathbf{V}_w^{-1}\mathbf{X}_w\right)^{-1}\mathbf{X}_w^{'}\mathbf{V}_w^{-1}\mathbf{y}_w$$

$$\mathbf{L}_w = \mathbf{V}\mathbf{H}_w^{'}\mathbf{V}_w^{-1}$$

NB: The results are invariant with respect to the choice of a particular sub-vector of $\mathbf{y}_a$, provided it has dimension (Nx1) (Chen and Dagum, 1997, express a different view)

Extrapolation

$n > MN$

The relevant temporally aggregated values are not yet available

Estimate $\mathbf{y}_e = \begin{bmatrix} \mathbf{y}_{e1}^{'} & \cdots & \mathbf{y}_{ej}^{'} & \cdots & \mathbf{y}_{eM}^{'} \end{bmatrix}'$

Constrained extrapolation

$\mathbf{z}_e \qquad \mathbf{H}_e\mathbf{y}_e = \mathbf{z}_e \qquad \mathbf{H}_e = \mathbf{1}_M^{'} \otimes \mathbf{I}_R$

$$\hat{\mathbf{y}}^* = \mathbf{X}^*\hat{\boldsymbol{\beta}}^* + \mathbf{L}^*\left(\mathbf{y}_a^* - \mathbf{X}_a^*\hat{\boldsymbol{\beta}}^*\right)$$

$$\hat{\boldsymbol{\beta}}^* = \left[\left(\mathbf{X}_a^*\right)'\left(\mathbf{V}_a^*\right)^{-}\mathbf{X}_a^*\right]^{-1}\left(\mathbf{X}_a^*\right)'\left(\mathbf{V}_a^*\right)^{-}\mathbf{y}_a^*$$

Pure extrapolation: $\mathbf{z}_e$ not available

$$\hat{\mathbf{y}}_e = \mathbf{X}_e\hat{\boldsymbol{\beta}} + \boldsymbol{\Omega}\mathbf{H}'\mathbf{V}_a^{-}\left(\mathbf{y}_a - \mathbf{X}_a\hat{\boldsymbol{\beta}}\right)$$

SOME OPERATIONAL OBSERVATIONS
(Moauro and Savio, 2001)

Though the behaviour of NSIs could sometimes appear conservative enough in the introduction of new techniques, (…) their introduction in a routine process (such as that of the estimation of quarterly national accounts) requires that at least the following requirements are fulfilled:

1) the techniques should be flexible enough to allow for a variety of time series to be treated easily, rapidly and without too much intervention by the producer;

2) they should be accepted by the international specialized community;

3) the techniques should give reliable and meaningful results;

4) the statistical procedures involved should be run in an accessible and well known, possibly user friendly, and well sounded software program, interfacing with other relevant instruments typically used by data producers (i.e. seasonal adjustment, forecasting, identification of regression models,…).

(Partial) Bibliography on Benchmarking and Temporal disaggregation of economic series

Aadland D.M. (2000), Distribution and interpolation using transformed data, *Journal of Applied Statistics*, 27: 141-156.

Abeysinghe T. (2000), Modeling variables of different frequencies, *International Journal of Forecasting*, 16: 117-119.

Abeysinghe T. and C. Lee (2000), Dynamic regressions with variables observed at different frequencies (*mimeo*).

Al-Osh M. (1989), A dynamic linear model approach for disaggregating time series data, *Journal of Forecasting*, 8: 85-96

Astolfi R., D. Ladiray, G.L. Mazzi, F. Sartori and R. Soares (2000), *A monthly indicator of GDP for the Euro-zone*, Eurostat (*mimeo*).

Barcellan, R. (2002), *ECOTRIM: a program for temporal disaggregation of time series*, Workshop on Quarterly National Accounts, Eurostat, Theme 2 Economy and finance: 79-95.

Barcellan R. and T. Di Fonzo (1994), Disaggregazione temporale e modelli ARIMA, Società Italiana di Statistica, *Atti della XXXVII Riunione Scientifica*, 2: 355-362.

Bloem A., R.J. Dippelsman and N.Ø. Mæhle (2001), *Quarterly National Accounts Manual. Concepts, data sources, and compilation*, Washington DC, International Monetary Fund.

Boot J.C.G., Feibes W. and J.H.C. Lisman (1967), Further methods of derivation of quarterly figures from annual data, *Applied Statistics*, 16, 1: 65-75.

Bournay J. and G. Laroque (1979), Réflexions sur la méthode d'élaboration des comptes trimestriels, *Annales de l'INSEE*, 36: 3-30.

Cabrer B. and J.M. Pavia (1999), Estimating M (>1) quarterly time series in fulfilling annual and quarterly constraints, *International Advances in Economic Research*, 5, 3: 339-349.

Chan (1993), Disaggregation of annual time-series data to quarterly figures: a comparative study, *Journal of Forecasting*, 12: 677-688.

Chen Z.G., Cholette P.A. and E.B. Dagum (1997), A nonparametric method for benchmarking survey data via signal extraction, *Journal of the American Statistical Association*, 92, 440: 1563-1571.

Chen Z.G. and E.B. Dagum (1997), A recursive method for predicting variables with temporal and contemporaneous constraints, American Statistical Association, *Proceedings of the Business and Economic Statistics Section*: 229-233.

Cholette P.A. (1984), Adjusting sub-annual series to yearly benchmarks, *Survey Methodology*, 10, 1: 35-49.

Cholette P.A. (1988), *Benchmarking systems of socio-economic time series*, Working Paper, Time Series Research and Analysis Division, Statistics Canada.

Cholette P.A. and E.B. Dagum (1994), Benchmarking time series with autocorrelated survey errors, *International Statistical Review*, 62, 3: 365-377.

Chow G. and A.L. Lin (1971), Best linear unbiased interpolation, distribution and extrapolation of time series by related series, *The Review of Economics and Statistics*, 53, 4: 372-375.

Chow G. and A.L. Lin (1976), Best linear unbiased estimation of missing observations in an economic time series, *Journal of the American Statistical Association*, 71, 355: 719-721.

Cuche N.A. and M.K. Hess (2000), Estimating monthly GDP in a general Kalman filter framework: evidence from Switzerland, *Economic & Financial Modelling*, Winter 2000: 153-193.

Denton F.T. (1971), Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization, *Journal of the American Statistical Association*, 66: 99-102.

Di Fonzo T. (1987), *La stima indiretta di serie economiche trimestrali*, Padova, Cleup.

Di Fonzo T. (1990), The estimation of M disaggregated time series when contemporaneous and temporal aggregates are known, *The Review of Economics and Statistics*, 72, 1: 178-182

Di Fonzo T. (2002a), *Temporal disaggregation of a system of time series when the aggregate is known. Optimal vs. adjustment methods*, Workshop on Quarterly National Accounts, Eurostat, Theme 2 Economy and finance: 63-77.

Di Fonzo T. (2002b), *Temporal disaggregation of economic time series: towards a dynamic extension*, Università degli Studi di Padova, Dipartimento di Scienze Statistiche, Working Paper 2002.17.

Di Fonzo T. (2003a), Temporal disaggregation using related series: log-transformation and dynamic extension, *RISEC*, 50, 3: 371-400.

Di Fonzo T. (2003b), *Construction of higher frequency data when only lower ones are available. The routine DYNCHOW*, September 12, 2003 (mimeo).

Di Fonzo T. and M. Marini (2003), *Benchmarking systems of seasonally adjusted time series according to Denton's movement preservation principle*, Università degli Studi di Padova, Dipartimento di Scienze Statistiche, Working Paper 2003.09.

Durbin J. and B. Quenneville (1997), Benchmarking by state space models, *International Statistical Review*, 65,1: 23-48.

Eurostat (1999), *Handbook of quarterly national accounts*, Luxembourg, European Commission.

Fernández R.B. (1981), A methodological note on the estimation of time series, *The Review of Economics and Statistics*, 63: 471-478.

Gilbert, C.L. (1977) Regression using mixed annual and quarterly data, *Journal of Econometrics*, 5: 221-239.

Ginsburgh V.A. (1973), A further note on the derivation of quarterly figures consistent with annual data, *Applied Statistics*, 22: 368-374.

Gregoir S. (1995), *Propositions pour une désagrégation temporelle basée sur des modèles dynamiques simples*, INSEE (*mimeo*).

Gudmundsson G. (1999), Disaggregation of annual flow data with multiplicative trends, *Journal of Forecasting*, 18:33-37.

Guerrero V.M. (1990), Temporal disaggregation of time series: an ARIMA-based approach, *International Statistical Review*, 58: 29-46.

Guerrero V.M. and J. Martinez (1995), A recursive ARIMA-based procedure for disaggregating a time series variable using concurrent data, *TEST*, 2: 359-376.

Guerrero, V.M. and F.H. Nieto (1999), Temporal and contemporaneous disaggregation of multiple economic time series, *TEST*, 8, 2: 459-489.

Harvey A.C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge, Cambridge University Press.

Harvey A.C. and R.G. Pierse (1984), Estimating missing observations in economic time series, *Journal of the American Statistical Association*, 79, 385: 125-131.

Helfand S.D., Monsour N.J. and M.L. Trager M.L. (1977), Historical revision of current business survey estimates, American Statistical Association, *Proceedings of the Business and Economic Statistics Section*: 246-250.

Hodgess E.M. and W.W.S. Wei (2000), Temporal disaggregation of stationary bivariate time series, *Linear Algebra and its Applications*, 321: 175-196.

Hotta L.K. and K.L. Vasconcellos (1999), Aggregation and disaggregation of structural time series models, *Journal of Time Series Analysis*, 20: 155-171.

Jacobs J. (1994), *'Dividing by 4': a feasible quarterly forecasting method?*, Department of Economics, University of Groningen (*mimeo*).

Laniel N. and K. Fyfe (1990), Benchmarking of economic time series, *Survey Methodology*, 16, 2: 271-277.

Lisman J.H.C. and J. Sandee (1964), Derivation of quarterly figures from annual data, *Applied Statistics*, 13, 2: 87-90.

Litterman R.B. (1983), A random walk, Markov model for the distribution of time series, *Journal of Business and Economic Statistics*, 1: 169-173.

Mitchell J., R.J. Smith, , M.R. Weale, S. Wright and E.L. Salazar (2003), *An indicator of monthly GDP and an early estimate of quarterly GDP growth* (mimeo).

Moauro G. and G. Savio (2001), *Disaggregation of time series using common components models* (mimeo).

Monsour N.J. and M.L. Trager (1979), Revision and benchmarking of business time series, American Statistical Association, *Proceedings of the Business and Economic Statistics Section*: 333-337.

Pinheiro M. and C. Coimbra (1992), *Distribution and Extrapolation of Time Series by Related Series Using Logarithms and Smoothing Penalties*, Gabinete de Estudios Económicos, Instituto Nacional de Estatistica, Working Paper.

Powell A. (1969), Aitken estimator as a tool in allocating predetermined aggregates, *Journal of the American Statistical Association*, 64: 913-922.

Proietti T. (1999), Distribution and interpolation revisited: a structural approach, *Statistica*, 58: 411-432.

Quenneville B., Huot G., Cholette P., Chiu K. and T. Di Fonzo (2003), *Adjustment of seasonally adjusted series to annual totals*, paper presented at the Statistics Canada Symposium, Ottawa, October 28-31, 2003.

Rossi, N. (1982), "A note on the estimation of disaggregate time series when the aggregate is known", *The Review of Economics and Statistics*, 64, 4: 695-696.

Salazar E.L., R.J. Smith and M. Weale (1997), *Interpolation using a Dynamic Regression Model: Specification and Monte Carlo Properties*, NIESR Discussion Paper n. 126.

Salazar E.L., R.J. Smith, M. Weale and S. Wright (1994), *Indicators of monthly national accounts*, presented at the I.N.S.E.E.-EUROSTAT 'Quarterly National Accounts Workshop', Paris-Bercy, 5-6 December, 1994 (*mimeo*).

Salazar E.L., R.J. Smith, M. Weale and S. Wright (1998), *A monthly indicator of UK GDP* (*mimeo*).

Santos Silva J.M.C. and F.N. Cardoso (2001), The Chow-Lin method using dynamic models, *Economic Modelling*, 18: 269-280.

Solomou, S. and Weale, M. (1993), "Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920-38", *Journal of the Royal Statistical Society*, A, 156, 1: 89-105.

Taillon J. (1988), Dessaisonalisation d'un tableau avec contraintes linéaires: Application aux flux financiers canadiens, *The Canadian Journal of Statistics*, 16: 97-107.

Thomas J.J. and K.F. Wallis (1971), Seasonal variation in regression analysis, *Journal of the Royal Statistical Society*, A, 134, 1: 57-72.

Tserkezos D.E. (1991), A distributed lag model for quarterly disaggregation of the annual personal disposable income of the Greek economy, *Economic Modelling*, 8: 528-536.

Tserkezos D.E. (1993), Quarterly disaggregation of the annualy known Greek gross industrial product using related series, *Rivista Internazionale di Scienze Economiche e Commerciali*, 40: 457-474.

Weale, M. (1992), "Estimation of data measured with error and subject to linear restrictions", *Journal of Applied Econometrics*, 7: 167-174.

Wei W.W.S. and D.O. Stram (1990), Disaggregation of time series models, *Journal of the Royal Statistical Society*, 52:453-467.

Wilcox J.A. (1983), Disaggregating data using related series, *Journal of Business & Economic Statistics*, 1, 3: 187-191.