WORKING PAPERS AND STUDIES

2003 EDITION

ISBN 92-894-5342-7
ISSN 1725-4825
Cat. No. KS-AN-03-014-EN-N

Workshop on Quarterly National Accounts

EUROPEAN COMMISSION

Europe Direct is a service to help you find answers to your questions about the European Union

New freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2003

ISBN 92-894-5342-7
ISSN 1725-4825

EUROSTAT - INSEE

Paris - Bercy

05/12/1994 - 06/12/1994

Workshop on

Quarterly National Accounts

Atelier sur les

Comptes Nationaux Trimestriels

Edited by
Roberto BARCELLAN and Gian Luigi MAZZI

Desktop publishing
Elisabeth ALMEIDA and Nelly DA SILVA

Special thanks
Stéphane GREGOIR for the organisation of the Workshop

Manuscript completed in April 2002

For further information, please contact

Roberto Barcellan - Gian Luigi Mazzi

Eurostat
5, rue Alphonse Weicker
L-2721 Luxembourg
Tel. (352) 4301 35802
Tel. (352) 4301 34351
Fax (352) 4301 33879

Roberto.Barcellan@cec.eu.int
Gianluigi.Mazzi@cec.eu.int

FOREWORD

Since the early 1990's, the importance of short term information has highly increased, in particular in the field of quarterly national accounts.

For the monitoring of the economy and for achievement of the objectives of the European Monetary Union, high-quality short term statistical instruments are in fact needed in order to provide the Community Institutions, Governments, Central Banks as well as economic and social operators with a set of comparable and reliable data.

Being aware of the importance of the short term information on national accounts, Insee and Eurostat organised a joint workshop in Paris-Bercy the 5 and 6 December 1994. At this workshop participated many specialists in the field of national accounts, and in particular of quarterly accounts, not only from the National Statistical Institutes but also from Central Banks and Universities.

The workshop was mainly dedicated to the different approaches for compiling quarterly accounts, based on econometric models, statistical surveys or indirect methods. The issue of revisions and validations of quarterly national accounts and of seasonal adjustment were also discussed. One session was dedicated to the users' point of view (most of the papers of the workshop are presented in this publication.)

The richness and the quality of the contributions of the Paris-Bercy workshop constituted an excellent basis for the Eurostat Handbook on Quarterly National Accounts drafted in the framework of the European System of Accounts (ESA95). This handbook is providing to Member States but also to Candidate Countries and other countries a precious and harmonised approach with a set of recommendations to be followed for the compilation of quarterly accounts.

Eurostat would like to thank INSEE for hosting this important workshop and all the participants, especially the authors of the papers for the excellent contributions.

Marco DE MARCH
Head of Unit
Eurostat - Unit B2
Economic Accounts and International Markets:
Production and Analyses

SUMMARY

Workshop on

Quarterly National Accounts

Atelier sur les

Comptes Nationaux Trimestriels

Program

INSEE-EUROSTAT WORKSHOP
5-6 December 1994

Monday 5 December 1994

8.30-9.00 Registration

9.00-9.15 ***Welcome*** address by Y. Franchet (Director General of EUROSTAT) and P. Mazodier (Directeur de la Direction des Etudes et des Synthèses Economiques – INSEE)

9.15-11.15 ***Revisions and validations of quarterly national accounts***

Chairman: Heinrich Lützel (Statistisches Bundesamt)

- "Revisions to Italian Quarterly National Accounts Aggregates: Some Empirical Results" – T. Di Fonzo (University of Brescia), S. Pisani, G. Savio (ISTAT)
- "Relations between Quarterly and Annual Accounts at INSEE" – V. Madelin (INSEE)
- "The Use of External Data Sources to Validate National Accounts Estimates" – K. Vernon (CSO)
- "Cyclical Patterns of the Spanish Quarterly National Accounts Series: A VAR Analysis" – A.M. Abad Garcia, E. Martin Quilis (INE)

11.30-13.00 ***User's point of view***

Chairman: Marco De March (EUROSTAT)

F. Carlucci (University of Rome)

A. Gubian (DARES)

P. Cuneo (BIPE)

M. Dramais (CEE-DGII)

14.30-16.45 ***New developments in the econometric-based approach***

Chairman: Enrico Giovannini (ISTAT)

Invited paper

- "Temporal Disaggregation of a System of Time Series when the Aggregate is Known: Optimal vs. Adjustment Methods" – Di Fonzo (University of Brescia) and "ECOTRIM: A Program for Temporal Disaggregation of Time Series" – R. Barcellan (University of Padova)
- "Estimating Missing Observations in ARIMA Models with the Kalman Filter" – V. Gomez (INE)

Contributed paper

- "Time Disaggregation of Economic Time Series: An Econometric Approach" – C. Lupi (ISTAT), G. Parigi (Bank of Italy)
- "Note on Temporal Disaggregation with Simple Dynamic Models" – S. Gregoir (INSEE)

17.00-18.30 Round table on the new SNA and the european coordination

Chairmen: S. Gregoir (INSEE) and G. L. Mazzi (EUROSTAT)

Tuesday 6 December 1994

8.30-10.15 ***Survey-based approach I***

Chairman: Jean-Etienne Chapron (INSEE)

- "The United Kingdom Approach to Quarterly National Accounts" – D. Caplan, S. Lambert (CSO)
- "Quarterly National Accounts of the Federal Statistical Office of Germany" – H. Lützel (Statistisches Bundesamt)
- "Special Aspects of Quarterly GDP Accounts in East Germany in the Initial Years after German Unification" – R. Hein (Statistisches Bundesamt)

10.30-12.30 ***Seasonal adjustment***

Chairman: Francesco Carlucci (University of Roma)

- "Comparison of Different Seasonal Adjustment Methods using Quarterly Data of National Accounts" – B. Fisher
- "Two Programs for Time Series Analysis, with an application to seasonal adjustment" – A. Maravall (European University Institute)

- "Trend-Cycle versus Seasonal Adjustment in Quarterly National Accounts" – A. Cristobal Cristobal, E.M. Quilis (INE)
- "Forecasting Changing Seasonal Components Using Periodic Correlation" – P.H. Franses, M. Ooms (Erasmus University)

13.45-15.45 ***Survey-based approach II***

Chairman: David Caplan (CSO)

- "Compilation of Quarterly National Accounts in Denmark" – T.M. Pedersen (Denmark Statistiks)
- "An Outline of the Swedish Quarterly National Accounts" – H. Runnquist, E. Karlsson (Statistics Sweden)
- "Data Flows in Quarterly National Accounting" – N. Van Stokrom (CBS)
- "Source and Use of Funds and Investments Statements in Quarterly Financial Accounts" – T. Grunspan (Bank of France)

16.00-18.00 ***Indirect methods***

Chairman: Tommaso Di Fonzo (University of Brescia)

Invited paper

- "GDP-Monthly Indicator" – E. Lääkäri (Statistics Finland)
- "Indicators of Monthly National Accounts" E. Salazar, R. Smith, S. Wright, M. Weale (Cambridge university – UK)
- "Quarterly Flash Estimates: A Proposal for Italy" – G. Bruno, E. Giovannini, C. Lupi (ISTAT)

Contributed paper

- "Using Data of Qualitative Business Survey for the Estimation of Quarterly National Accounts" – C. Coimbra (Lisboa)
- "Estimating Quarterly National Accounts for Southern Italy" – M. Carlucci, R. Zelli (University of Rome, "La Sapienza")

LIST OF AUTHORS

ABAD GARCIA Ana M.

BARCELLAN Roberto

BRUNO Giancarlo

CARLUCCI Margherita

COIMBRA Carlos

CRISTÓBAL CRISTÓBAL Alfredo

CUBADDA Gianluca

DI FONZO Tommaso

GIOVANNINI Enrico

GOMEZ Victor

GREGOIR Stéphane

HAIN Ralf

JANSSEN Ronald

KARLSSON Eddie

LÄÄKÄRI Erkki

LUPI Claudio

LÜTZEL Heinrich

MADELIN Virginie

MAZZI Gian Luigi

OOMENS Peter

PARIGI Giuseppe

PISANI Stefano

QUILIS Enrique Martín

RUNNQUIST Helena

SAVIO Giovanni

VAN STOKROM Nico

ZELLI Roberto

TABLE OF CONTENTS

Section 2

New Developments in the Econometrics-based Approach

Section 3

Survey-based Approach I

Survey-based Approach II

Section 4

Seasonal Adjustment

Section 5

Indirect Mehtods

SECTION 0 -

INTRODUCTION

Perspectives of Quarterly National Accounts in the new European System of Accounts

Gian Luigi MAZZI
EUROSTAT

This paper contains a synthesis of the main issues discussed during the workshop on Quarterly National Accounts that has been held in Paris on 3-4 December 1994. This workshop has been jointly organised by INSEE and EUROSTAT. This paper is intended to give an overlook on the main topics discussed in the papers as well as during the contributions in the round-table on the future of European Union Quarterly National Accounts. Moreover, the perspectives and the possible improvements of Quarterly National Accounts in the framework of the new European System of Accounts are also briefly discussed

1 Introduction

National Accounts are a continuously evolving system of economic identities aiming to supply a synthetic view of the economic reality. In its evolution extensions to social, regional, environmental and quarterly accounts have been introduced in the System of National Accounts.

Nevertheless, a significant difference can be observed between such extensions. Social, regional and quarterly accounts are already not only methodological but also practical issues in most Member States, so that data are already available. Environmental accounts still represent essentially a methodological issue, and only a limited number of quantitative applications are already been experimented.

From a macroeconomic point of view, Quarterly National Accounts (QNA) represent a major improvement. The possibility of compiling high frequency National Accounts figures - i.e. quarterly and monthly in the future (see Lääkäri, 2002) - is of special relevance when Quarterly data are fully integrated in the System of National Accounts, in order to obtain a better coverage and a complete description of the economic behaviour of a nation.

Even if Quarterly National Accounts can be considered a quite long-standing topic, only with the introduction of the new European System of Accounts their integration in the System of National Accounts has been explicitly stated as a crucial point. In the European Union Member States, QNA have been developed following different principles, mainly due to dissimilarity in the disaggregation level, coverage and compilation methods (see Lutzel 2002). In some cases QNA constitute an incomplete system with clear problems of integration with annual figures. The only Member State having a completely integrated system of QNA, at present, is the United Kingdom.

From the middle of ‘80s, Eurostat started an active methodological and empirical activity in this field of QNA, in order to obtain from all Member States integrated and reliable quarterly aggregates and in order to produce high quality European aggregates. Those data are intended to be really useful for analysts,

decision and policy makers as well as for monetary authorities.
Eurostat and National Statistical Institutes are more and more focusing on QNA mainly responding to the increasing role of quarterly data for short-term economic analysis and for the evaluation of the measures of economic policy.

This workshop has been jointly organised by INSEE and EUROSTAT and intends to provide useful and practical contributions on some key topics, such as:

- the Quarterly National Accounts compilation scheme in Member States,
- the methodology of calculation of QNA,
- the treatment of seasonality;
- flash estimates techniques.

2 The New European System of Accounts

The relevance of National Accounts was recognised early by the United Nations statistical service, which already in 1968 prepared a manual for the compilation of National Accounts. Eurostat has implemented and adapted those principles to the structure of European economies and in 1971 published by the European System of Accounts (ESA), later reviewed in 1979.

Both in the UN and the EU manuals, no attention was paid to the QNA. Nevertheless, in 1972 Eurostat decided to prepare a specific document on Quarterly Accounts, defining a simplified scheme for the compilation of high frequency National Accounts aggregates. Unfortunately, the document appeared to be too ambitious for the time, and it had no practical application.

At the end of the '80, the preparatory work for the New System of National Accounts started under the supervision of United Nations. The need for explicit treatment of Quarterly Accounts was addressed but in the final version of the manual SNA 93 there was no specific chapters on the compilation of QNA. From the Eurostat point of view, this represented an important drawback. Therefore, during the preparation of the European application of the SNA, it has been clearly stated that Quarterly National Accounts had be treated in a specific chapter.

This chapter included in the new ESA, although brief and not exhaustive, provided a description of the main specification of QNA: compilation rules, estimation techniques, treatment of seasonality, consistency between quarterly and annual accounts and revision policy. In particular it represented the first attempt for the improvement and harmonisation of QNA. At the same time, Eurostat launched a series of projects designed at the compilation of a handbook on Quarterly National Accounts, to the production of a special software for time disaggregation of economic series and to the implementation of a system of flash estimates of the most relevant quarterly aggregates.

3 Handbook on quarterly national accounts

During the discussion of the ESA chapter on Quarterly National Accounts and in the occasion of the working group on National Accounts, many Member States explicitly requested Eurostat to provide a theoretical and practical manual for the compilation of Quarterly National Accounts.

The handbook on quarterly national accounts that Eurostat is setting up, will deal, as requested, with typical conceptual problems of QNA and will provide some guidelines on their treatment.

As a preliminary step, Eurostat set up a questionnaire in order to analyse the current situation of QNA in EU Member States and in main EU economic partners. The main outcome of this questionnaire should be an up-to-date overview of the current situation of QNA production. From this standpoint and together with the consideration of the accounting rules stated in the new ESA, the handbook will intend to provide:

- guidelines for the production of QNA in those Member States not yet compiling them;
- guidelines for the harmonisation of QNA compilation procedures for those Member States already producing them.

In this sense, the following aspects are going to be the core of the QNA handbook:

- description of the current procedures for QNA compilation in the Member States and main economic partners;
- integration of QNA in the context of the New European System of Accounts;
- reading of accounting rules in order to consider the shorter reference time horizon;
- identification of basic statistics and information sources available at quarterly frequency;
- theoretical background supporting existing methods used for compiling QNA;
- seasonal and calendar movements and their adjustment;
- consistency between quarterly and annual accounts.

Some well recognised experts in QNA will develop, together with Member States and Eurostat, the main theoretical aspects mentioned above, in order to provide a complete description and to supply assessments and suggestions for their treatment.

The final result stemming from the handbook will lead to the identification of a complete system set up by Eurostat in order to obtain integrated and harmonised QNA figures from all the Member States.

4 ECOTRIM: a software for temporal disaggregation of time series

Often economic time series are observed only at low frequency. Economic analysis more and more often requires high-frequency disaggregated data, therefore economic analysts as well as national Institutes of Statistics are confronted with the problem of temporal disaggregation for estimating such series.

ECOTRIM is a software written in GAUSS which supplies a set of mathematical and statistical techniques for carrying out temporal disaggregation. A detailed presentation of this programme is given in Barcellan and Di fonzo (2002).

The approach presented in ECOTRIM corresponds to the so-called “indirect approach” to the estimation of Quarterly National Accounts, which is already in use in some Member States, such as Italy, France and Spain (Madellin, Abad Garcia and Quilis, 2002). The high-frequency series are estimated using available information corresponding to low-frequency series and, sometimes, to some related series. Despite their large utilisation, such techniques are subject to some criticism deriving from a purely econometric point of view, as discussed in Lupi and Parigi (2002).

ECOTRIM has been developed on behalf of the European Commission – Eurostat - Directorate B *“Economic Statistics and Economic and Monetary Convergence”*.

One of the main advantages in the use of ECOTRIM is it offers the possibility to choose between a large set of techniques in order to select the most appropriate according to the available information.

In particular the program supplies a range of techniques concerning:

- temporal disaggregation of univariate time series by using or not using related series and fulfilling temporal aggregation constraints (the methods that ECOTRIM offers, follow the mathematical approach, the ARIMA model based approach, and the optimal, in the least square sense, approach);
- temporal disaggregation of multivariate time series for fulfilling both temporal and contemporaneous aggregation constraints (in this case too ECOTRIM proposes both adjustment and optimal techniques, in the least square sense);
- forecast of current year observations by using or not using available information from related series.

A data management section, a set of diagnostics and a graphic/display section allow users to carry out a complete analysis of the problem treated.

ECOTRIM can be used both in interactive mode and in batch mode (when many series have to be submitted to the disaggregation process).

ECOTRIM is going to be further developed by EUROSTAT, in order to respond to the different needs of operators treating sub-annual accounts. ECOTRIM is currently under development in order to incorporate some new interesting features in treating temporal disaggregation. With reference to recent

developments in the dynamic model approach literature, as well as the response to suggestions coming from National Institutes of Statistics, ECOTRIM will be enhanced with:

- dynamic model approaches (Gregoir, 2002);
- non-linear constraints;
- development of Kalman filter techniques (Gomez, 2002);
- specific accounting treatments;
- more detailed diagnostics sections;
- co-integration analysis aspects (Gregoir, 2002).

5 Flash estimates of quarterly national accounts

Flash estimates are an essential tool for producing timeline data for an integrated System of National Accounts. Such estimates are currently produced by some Member States such as Italy (see Bruno, Giovannini and others) and the Netherlands (see Jansen).

Eurostat contributes to the development of those techniques with a specific project for the production of flash estimates of main QNA aggregates at the European level - with a delay between 30 and 45 days after the end of the reference period. This delay has been chosen because it seems to be "reasonable" for flash estimates, as suggested by the experience of certain countries (UK, USA). Reducing such delay normally implies a considerable reduction in the available basic information.

The scope of this project is to ensure the construction of a coherent system of data for short-term economic analysis. Even if certain indicators (production indexes, prices, foreign trade, business surveys, etc.) are promptly available, they do not fulfil the requirements of homogeneity and consistency typical of an integrated System of National Accounts.

On the other hand, the first estimates of quarterly accounts are often released with a considerable delay, making them partly unsuitable for the users needs. In particular, the business cycle analysis and evaluation of the economic and monetary policy measures require shorter delays.

Given these premises, an efficient way of estimating those aggregates should take into account the sub-annual information coming from different sources and should find out an efficient way to incorporate it into the accounting framework. In this way the need for prompt information, with the need for a consistent information system can be reconciled.

For the accomplishment of this objective, time-series analysis techniques (ARIMA models, transfer functions, VAR) are applied to the quarterly aggregates and certain sub-annual statistics to be used as indicators. This methodology has been preferred to structural econometric modelling, in order to overcome the identification and estimation problems connected to structural models.

Using this approach, considerations coming from the economic theory are generally disregarded, while attention is focused on the statistical relation between quarterly accounts variables and a set of basic statistics (Giovannini, Lupi, Bruno, Cubadda, 2002). Moreover, this procedure is consistent with the actual production process of QNA.

It should be clear that flash estimates are not simply an early release of quarterly accounts, but a rather different "product". In fact, it has been widely used as qualitative data, with the characteristics of combining qualitative information (e.g. Business and Consumer Surveys) with time-series techniques. In addition, the level of disaggregation is normally different from the one normally used to compile QNA.

The first step of this procedure is the selection, among available data, of those indicators showing the following features:

- stable relationship with the corresponding quarterly aggregate to be estimated;
- good forecasting performance;
- early availability.

After this first step, the relationship between the indicator and the aggregate to be estimated can be identified and evaluated in order to obtain forecasting models. Those two steps are interrelated, and can lead to a revision of the model.

Some other issues may be involved in the work. First of all, the treatment of seasonality: the use of raw data and applying seasonal adjustment on the results or the direct use of seasonally adjusted data do not necessarily gives the same results. This is a rather important point, given the importance of seasonally adjusted figures. Second, the extent of the revisions is likely to influence the estimation, so far as it relies on the past history of the aggregates.

6 Conclusions

The new system of European National Accounts will constitute a major improvement for all National Accounts and a real enhancement for Quarterly National Accounts.

Eurostat and Member States are requested to take an active role in this process by improving both methodological background and organisational framework for Quarterly Accounts. The preparation of a handbook of Quarterly Accounts will provide the guidelines in order to ensure the production of reliable, harmonized and comparable Quarterly National Accounts for the EU countries and for European aggregates. Furthermore, the enlargement process of the European Union will bring in this challenge also to the Accession countries.

Therefore, the compilation of Quarterly National Accounts and the reduction of the delay for data availability require important methodological developments. In this sense this workshop has represented an important forum to bring together National Statistical Offices, academics and users. The contributions presented during this workshop gave a high-quality input for developments.

References

Garcia A. M. A. and Quilis E. M. , "Cyclical patterns of the Spanish quarterly national accounts: a VAR analysis" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Barcellan R., " Ecotrim: A program for temporal disaggregation of time series" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002

Bruno G., Cubadda G., Giovannini E., Lupi C. , "The flash Estimate of the Italian Real Gross Product" in in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Di Fonzo T. , "Temporal disaggregation of a system of time series when the aggregate is known: Optimal VS. Adjustment methods" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002

Gregoir S., "Propositions pour une desagregation temporelle basée sur des modèles dynamiques simples" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Gomez V., "Estimating observations in ARIMA models with the Kalman filter" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Lääkäri E., "The monthly GDP indicators" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Lupi C. and Parigi G., "Temporal disaggregation of economic time series: Some econometric issues" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Lützel H., " Quarterly National Accounts of the Federal Statistical" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Madelin V., "La relation entre les comptes annuels et les comptes trimestriels à l'INSEE" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

Quilis E. M., "Cyclical patterns of the Spanish quarterly national accounts: a VAR analysis" in proceedings of the workshop of Quarterly National Accounts edited by Barcellan R. and Mazzi G.L. – Luxembourg 2002.

SECTION 1 -
REVISIONS AND VALIDATIONS OF QUARTERLY NATIONAL ACCOUNTS

Revisions to Italian Quarterly National Accounts Aggregates: Some empirical results

Tommaso DI FONZO ; Dipartimento di Scienze Statistiche, Università di Padova
Stefano PISANI; Dipartimento Contabilità Nazionale e Analisi Economica, Istat
Giovanni SAVIO; Dipartimento Contabilità Nazionale e Analisi Economica, Istat

1 Introduction[1]

In common with most statistical agencies in other countries, in Italy Istat periodically revises quarterly national accounts estimates: almost all published variables are first released on a preliminary basis to satisfy the users' need for timely information, and are then revised at a later period to incorporate information that was not available at the time of the preliminary release.

This gives rise to a revision process characterized by a preliminary (or provisional) estimate and by a subsequent series of current revisions, routinely produced, that follow closely upon one another. From time to time, additional, more unusual revisions also occur which take place at infrequent and irregular intervals. These `extraordinary' revisions can include changes in concepts and definitions of the aggregates and/or information coming from decennial censuses and improved estimation procedures.

Revisions to economic time series have been a continuing interest since the early studies by Zellner (1958) and Morgenstern (1963). In fact, the analysis of revisions provides a basis both for assessing the accuracy of provisional in relations to final estimates and for improving the methods of estimation used to compile provisional estimates. It cannot, of course, tell us anything about the reliability of the final estimates: a particular figure might never be revised and yet still be very unreliable. Assessment of the reliability of the final estimates must rely upon direct evidence about the data sources and the estimation technique on which they are based (Hibbert, 1981).

A history of systematic bias in the provisional estimates might lead either to a direct adjustment to the series in question so as to offset the presumed continuing bias (Rao, Srinath and Quenneville, 1989), or to more fundamental changes such as the collection of additional data or advances in the estimation technique.

While the features, nature and size of the revision process in the Italian annual national accounts estimates can be found in Trivellato (1986, 1987), in this paper we examine quarterly rather annual data of GDP and its selected components.

A comprehensive analysis of the revisions would encompass the full set of published data, in both current and constant prices, including both raw and seasonally adjusted data. Growth rates as well as levels would be analyzed and a complete documentation about the revisions themselves and about the reasons why they were necessary would be collected and investigated. However, such an analysis would require considerable time and resources.

1 The views expressed in this paper are those of the authors and not necessarily of any institution o r organization. We would like to thank Carolina Testa for her help in the collection and organization of the data s et as well as for first, preliminary empirical analysis.

The scope of the present paper is therefore more limited: its main objectives are a first insight into the revision process which characterizes the Italian quarterly national accounts, an examination of its distinctive features and, finally, a discussion of some issues concerning the behaviour of the preliminary estimates.

The magnitude of the revisions produced by Istat since the second half of eighties is discussed, presenting an overview of how the initial estimates differ from their final figures in both levels and growth rates. Then, according to some recent contributions (Patterson and Heravi, 1991, Pisani and Savio, 1994), we perform an econometric analysis of the revision process. Such an analysis is organized in three steps.

Firstly, we compare the short-medium term components of different versions of data. Secondly, we apply the concepts of integration and cointegration to verify the stationarity of the revision process. Lastly, in order to examine whether preliminary estimates of the main quarterly aggregates are efficient forecasts of the final figure, the hypotheses of unbiasedness, weak efficiency and orthogonality are tested.

The paper is organized as follows. In section 2 we give an outline of the Italian revision process and show the links between the various annual and quarterly estimates. Moreover, we discuss the information content of different versions of an aggregate at the same time to establish homogeneous comparisons. Section 3 is devoted to a descriptive analysis of the accuracy of provisional estimates at current prices, both in levels and in growth rates. In section 4 the revision process is analyzed from a different, though complementary, point of view. The above mentioned econometric approach is worked out, to better understand the salient temporal features of the revision process particularly with respect to the effects of the extraordinary revision published in 1987. Brief conclusions and indications for future research work on this subject find place in section 5.

2 Outline of the Italian revision process

The current publication plan of the Italian quarterly national accounts may be characterized in the following way (Giovannini, 1993, p. 18).

The first estimate for any given quarter is released within about 110 days of the quarter's end.

This provisional estimate is subsequently revised every 110 days both in the same year and during the following two years.

Finally, the estimates are revised on the second half of April of the following three years.

We denote by 'version *t.q*' the quarterly time series published 110 days after the end of quarter *q* of year *t*. Thus, for example (see table 1), the first preliminary estimate of GDP at current prices for the first quarter of 1992 belongs to the 1992.1 version, the first revised estimate for the same quarter belongs to the 1992.2 version and so on.

So far the above mentioned preliminary estimate has been revised 9 times, the last one on the second half of October 1994. Other four estimates of the 1992.1 figure are planned in future, and precisely along with the 1994.3, 1994.4, 1995.4 and, finally, 1996.4 versions. Normally these revisions complete the process, and no further changes take place.

The revision process is then an extensive one: after the first, preliminary estimate for each quarterly aggregate, from ten to thirteen further estimates continue to be published during a period variable between five and six years. More precisely, the release timing provides for fourteen vintages of the first quarter estimate, thirteen of the second quarter, twelve of the third quarter and, finally, eleven vintages of the fourth quarter data.

As the quarterly national accounts estimation procedure is of indirect type, in the sense that quarterly series are obtained by temporal disaggregating annual data (Istat, 1985, Di Fonzo, 1987), the quarterly revision process is strictly related to the annual one. The complete publication plan for a quarterly national accounts aggregate at year *t* is shown in table 2, from which we can get a first impression of the relationships that exist between the various quarterly estimates with respect to the relevant annual figure.

Three types of current annual estimates and revisions for year *t* can be recognized:

P_t: first preliminary annual estimate, released in year *t+1*;

Tab. 1: Preliminary and revised estimates of Italian GDP (1992.1-1994.2, billions of current lire, seasonal ly adjusted data

Version	Date of reference									
	1992				1993				1994	
	1	2	3	4	1	2	3	4	1	2
1992.1	371674									
1992.2	371158	374278								
1992.3	371374	375319	377308							
1992.4	373264	377065	378234	378627						
1993.1	373123	376049	377223	380795	383414					
1993.2	373602	377238	377158	379192	382336	388362				
1993.3	372778	377120	377533	379759	382140	388674	389840			
1993.4	372493	373606	376198	379326	382434	389035	392026	396619		
1994.1	373089	376257	375786	379190	382829	389020	391125	397140	400035	
1994.2	372986	375976	375747	379614	383047	388899	391039	397130	401973	400874

Tab. 2: Current publication plan of the Italian quarterly national accounts aggregates for year *t*

Version	Quarters of the year t				Annual data
	1	2	3	4	
t.1	$N^1_{t,1}$				n.a.
t.2	$N^2_{t,1}$	$N^1_{t,2}$			n.a.
t.3	$N^3_{t,1}$	$N^2_{t,2}$	$N^1_{t,3}$		n.a.
t.4	$P^1_{t,1}$	$P^1_{t,2}$	$P^1_{t,3}$	$P^1_{t,4}$	P_t
t+1.1	$P^2_{t,1}$	$P^2_{t,2}$	$P^2_{t,3}$	$P^2_{t,4}$	P_t
t+1.2	$P^3_{t,1}$	$P^3_{t,2}$	$P^3_{t,3}$	$P^3_{t,4}$	P_t
t+1.3	$p^4_{t,1}$	$P^4_{t,2}$	$P^4_{t,3}$	$P^4_{t,4}$	P_t
t+1.4	$R^1_{t,1}$	$R^1_{t,2}$	$R^1_{t,3}$	$R^1_{t,4}$	R_t
t+2.1	$R^2_{t,1}$	$R^2_{t,2}$	$R^2_{t,3}$	$R^2_{t,4}$	R_t
t+2.2	$R^3_{t,1}$	$R^3_{t,2}$	$R^3_{t,3}$	$R^3_{t,4}$	R_t
t+2.3	$R^4_{t,1}$	$R^4_{t,2}$	$R^4_{t,3}$	$R^4_{t,4}$	R_t
t+2.4	$D^1_{t,1}$	$D^1_{t,2}$	$D^1_{t,3}$	$D^1_{t,4}$	D_t
t+3.1					
t+3.2					
t+3.3					
t+3.4	$D^2_{t,1}$	$D^2_{t,2}$	$D^2_{t,3}$	$D^2_{t,4}$	D_t
t+4.1					
t+4.2					
t+4.3					
t+4.4	$D^3_{t,1}$	$D^3_{t,2}$	$D^3_{t,3}$	$D^3_{t,4}$	D_t

R_t: first revised annual estimate, released in year *t+2*;

D_t: second revised annual estimate, released in year *t+3*.

From table 2 four groups of current quarterly estimates clearly emerge:

$N^{v}_{t,j}$: Within year, not constrained $\begin{cases} v=1,2,3 & j=1 \\ v=1,2 & j=2 \\ v=1 & j=3 \end{cases}$

$P^{v}_{t,j}$: Constrained to P_t, =1,2,3,4 j=1,2,3,4;

$R^{v}_{t,j}$: Constrained to R_t, =1,2,3,4 j=1,2,3,4;

$D^{v}_{t,j}$: Constrained to D_t, =1,2,3,4, j=1,2,3,4;

Where *v* and *j* denote the within year vintage and the quarter, respectively. The following accounting relationships hold for quarterly and annual estimates of flow variables (we consider a release time spanning from *t=2* to *t=n+1*):

$$\sum_{j=1}^{4} P^{v}_{t,j} = P_t \qquad \begin{matrix} v=1,2,3,4 \\ t=1,\mathrm{K},n \end{matrix}$$

$$\sum_{j=1}^{4} R^{v}_{t,j} = R_t \qquad \begin{matrix} v=1,2,3,4 \\ t=1,\mathrm{K},n-1 \end{matrix}$$

$$\sum_{j=1}^{4} D^{v}_{t,j} = D_t \qquad \begin{matrix} v=1,2,3 \\ t=1,\mathrm{K},n-2 \end{matrix}$$

According to this current revision scheme, the final part of each version will be revised several times. As far as the last published series (1994.2 version) is concerned, a complete overview of the future planned revisions is given in table 3.

This picture is an idealized one, however, since it abstracts from the additional, more unusual revisions which also occur.

In Appendix we present the actual publication schedule for the Italian quarterly national accounts series from 1985 to the present time [2]. It is clear that the current revision process has been upset by several additional revisions, whose characteristics have been summarized in table 4.

We call these extraordinary revisions *historical* or *occasional*[3]. They take place at infrequent and irregular intervals, that affect the historical estimates for more than five years. The added information for historical revisions most usually comes from decennial censuses, other kinds of enlargements of the information basis and improved estimation procedures. The occasion of an historical revision is also used to implement conceptual and definitional changes to the national accounts framework and/or changes in the base year for constant prices evaluation. Historical revisions usually are major events which have wide-ranging effects through the detailed components of the accounts, and through time.

Tab. 3: Future planned revisions for the last published quarterly national accounts

Period	Actual type of estimate	Number of revisions	Year of final release
90.1-90.4	D^2	1	1995
91.1-91.4	D^1	2	1996
92.1-92.4	R^3	4	1997
93.1-93.4	P^3	8	1998
94.1	N^2	12	1999
94.2	N^1	12	1999

2 Quarterly national accounts publication actually started in 1976, but was discontinued in 1982. In 1985 Istat adopted a new estimation procedure (Istat, 1985), still valid nowadays. The analysis will be enlarg ed to the early estimates in a next paper.

3 A different characterization of the occasional revisions, unsuitable for the Italian case, is due t o Smith (1977).

Among the revisions reported in table 4, the one published in 1987 (1986.4 version) distinguishes itself for its peculiarity and, as we shall see, for considerable effects on the main aggregates. As Smith (1977) pointed out, it is not possible to make general statements about the characteristics of historical revisions, since each one is rather unique. However the 1987 revision is the only one that has been concerned with changes in the conceptual framework of the Italian quarterly national accounts (Istat, 1993).

It should be noted that the within year revisions are slightly different from the subsequent revisions. Estimates $N^{v}_{j,t}$ are extrapolations based on related indicators which are available on a quarterly basis: besides the different (less) amount of information on which they are based, unlike the other estimates they do not satisfy any accounting constraint, because the annual benchmark is not known at the release time. Therefore the within year revisions depend only on the revisions to these related indicators, while for estimates other than the primary source for revisions is new information on the annual benchmark.

Another important point about the revision process is that the presence of occasional revisions that are superimposed upon the current revision process results in *mixed* revisions (jointly 'current' and 'extraordinary'), whose nature is not homogeneous with the current revision sequence.

All these facts have to be kept in mind when comparing the various estimates, to properly appreciate the information content offered by them.

3 The accuracy of provisional estimates

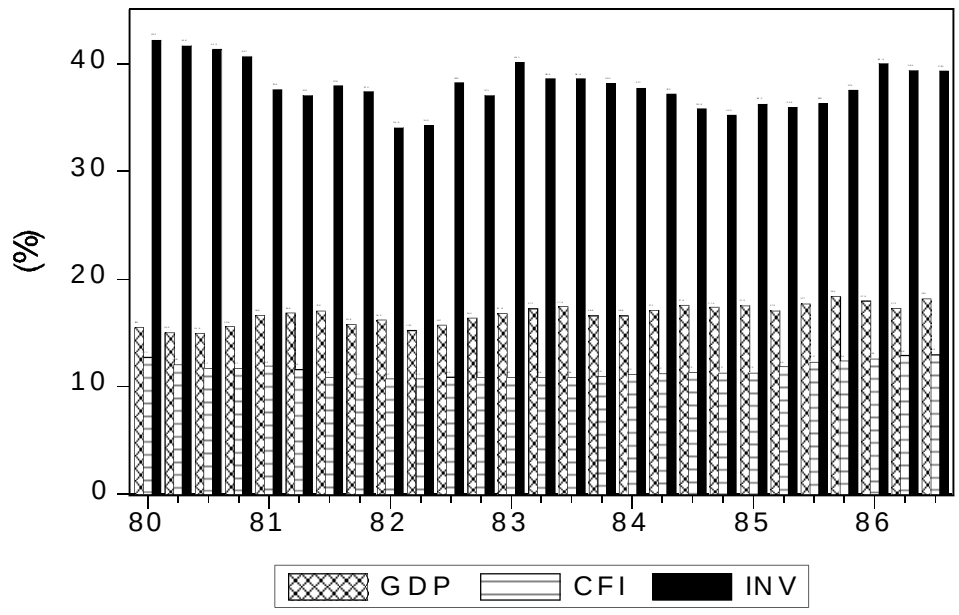
The descriptive analyses of the current revision process are carried out on the following current prices, seasonally adjusted aggregates:

GDP:	Gross Domestic Product;
CFI:	Final Domestic Consumption;
INV:	Gross Fixed Capital Formation;
IMP:	Imports of Goods and Services;
EXP:	Exports of Goods and Services.

The data set includes all published versions from 1984.4 to 1994.2 (see the Appendix).

In order to obtain a meaningful picture of the quarterly revision process, it is first necessary to establish which comparisons is appropriate to carry out. For each of the five series examined we consider five sets of

Fig. 1: Percentage relative revisions induced by the 1987 historical revision: GDP, CFI and INV



Tab. 4: Historical revisions of annual and quarterly national accounts

Version	Type of revision	Reference period	
		annual data	quarterly data
1984.4	New estimation procedure of quarterly national accounts (Istat, 1985)	--	1970.1-1984.4
1986.1	Completion of 1985 revision (Istat, 1987)	--	1970.1-1986.1
1986.4	Use of census data. New estimates of employment. Estimates of hidden economy. Change of the base year. New procedure of constant price evaluation. Input-Output table for 1982 (Istat, 1990, 1993)	1980-1986	1980.1-1986.4
1987.4	Small firms survey 1985-1986	1983-1987	1983.1-1987.4
1988.4	Retrospective estimation starting from 1970 (Istat, 1992)	1970-1981 1983-1988	1970.1-1988.4
1990.4	Input-Output tables for 1985 and 1988. Small firms survey for various years (Istat, 1991). Change of the base year	1970-1990	1980.1-1991.3 (*)
1992.2	Revision of the estimation procedure of quarterly national accounts. Completion of 1991 revision	--	1970.1-1992.2

estimates, $N^1_{t,j}$, $P^1_{t,j}$, $R^1_{t,j}$, $D^1_{t,j}$, $F_{t,j}$, respectively, where by $F_{t,j}$ we denote the series of "final" estimates, for which the current revision process can be regarded as finished[4].

The comparison between $N^1_{t,j}$ and $P^1_{t,j}$ tell us how much of the revision is carried out in the first stage, while the revision carried out in the second and third stage can be appreciated by comparing $P^1_{t,j}$ with $R^1_{t,j}$ and $R^1_{t,j}$ with $D^1_{t,j}$, respectively[5]. The comparison between $N^1_{t,j}$ and $F_{t,j}$ permits to evaluate the overall effect of the revision process.

Given the particularly strong effect on GDP, CFI and INV of the revision published in 1987 (see fig. 1), it seems appropriate to keep complete and 'pure' comparisons distinct, where in the latter case we omit the 'mixed' comparisons related to the above mentioned historical revision (see table 5 for an overview of the data availability and of the nature of comparisons).

A graphical support to this choice is given by fig. 2, in which GDP percentage relative errors (see below) of $N^1_{t,j}$ as compared with $P^1_{t,j}$ are presented. The 1987 historical revision results in abnormally large relative errors in the first three quarters of 1986 [6]. As far as IMP and EXP are concerned, the effects of the 1987 historical revision are less marked (see fig. 3 and fig. 4)[7].

We provide summary statistics for different comparisons between preliminary and revised series. According with Biggeri and Trivellato (1991), relative errors are analyzed to evaluate the accuracy of preliminary estimates of the levels and errors to evaluate the accuracy of preliminary growth rates.

4 In other words, the series $F_{t,j}$ corresponds to the 1994.2 version up to the 1989.4 figure.

5 We consider only current revisions related to different annual benchmarks. However, an analysis of the `intermediate' revisions would be very useful to appreciate the effects of the revisions in the related series used in the estimation procedure.

6 Similar results were found for CFI and INV too.

7 However, the descriptive analysis has been carried out on the same set of observations for all aggregates.

Fig. 2: Percentage relative errors of estimates $N^{1}_{j,t}$ as compared with $P^{1}_{t,j}$ GDP

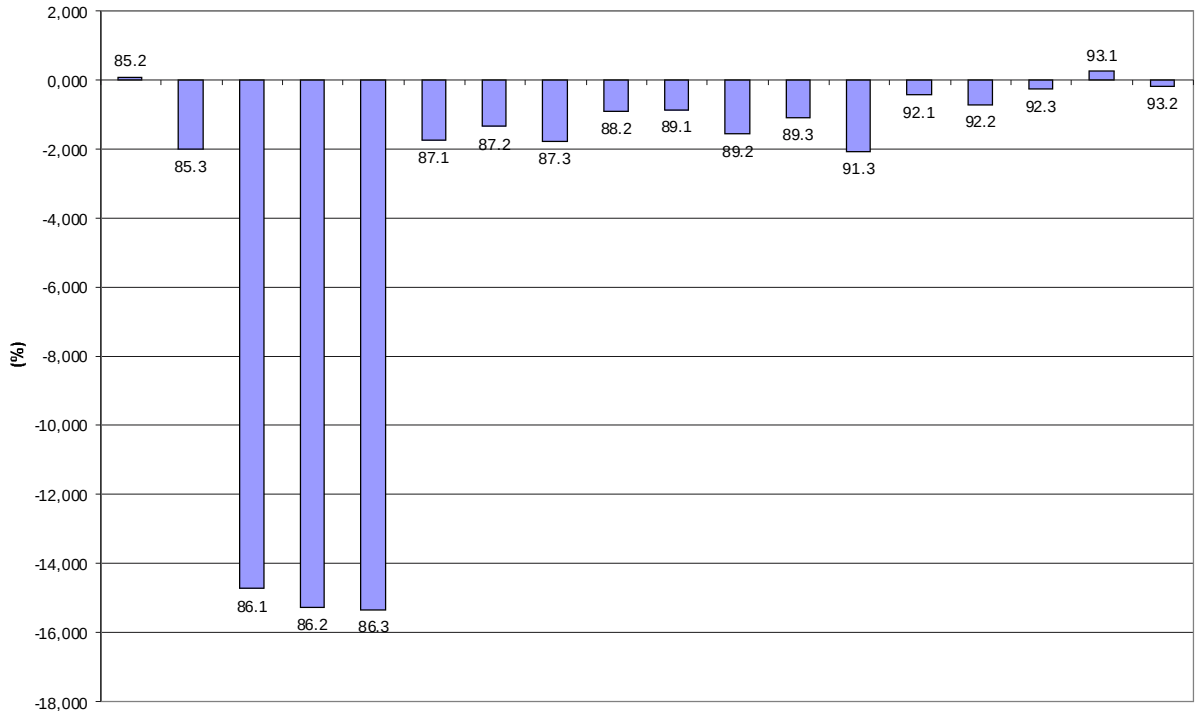
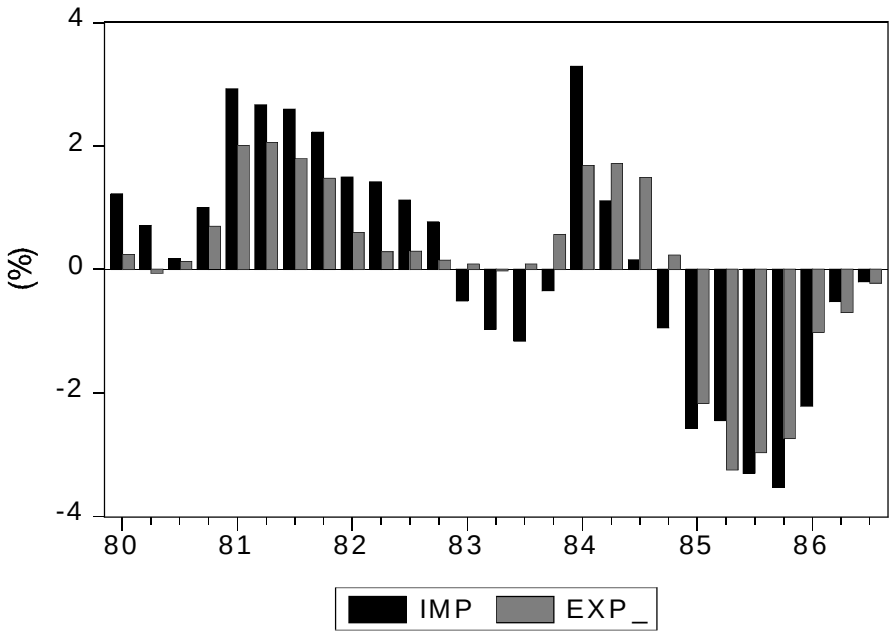


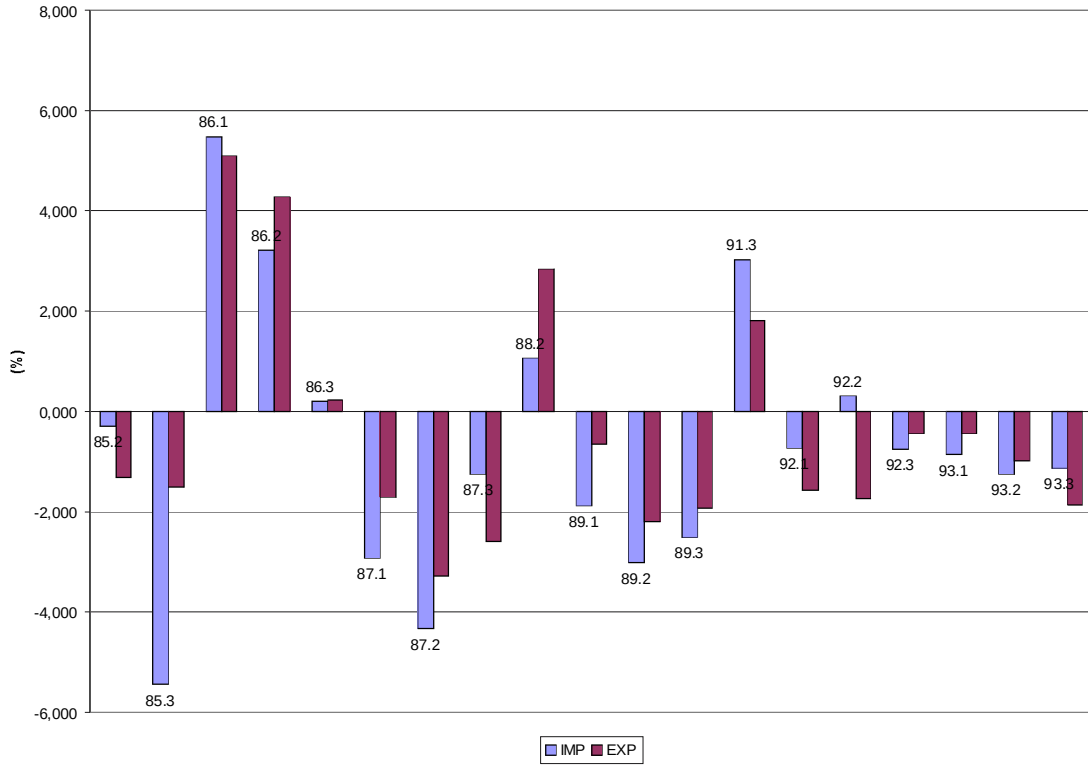
Fig. 3: Percentage relative revisions induced by the 1987 historical revision: IMP and EXP



Tab. 5: Data availability for the analysis of the current revision process of the Italian quarterly nationa l accounts

Year and quarter		Current estimates					Comparisons			
		N^1	P^1	R^1	D^1	F	$N^1 - P^1$	$P^1 - R^1$	$R^1 - D^1$	$N^1 - F$
1983	1	•	•	x	x	x	•	•	p	•
	2	•	•	x	x	x	•	•	p	•
	4	•	•	x	x	x	•	•	p	•
	3	•	•	x	x	x	•	•	p	•
1984	1	•	x	x	x	x	•	p	m	•
	2	•	x	x	x	x	•	p	m	•
	3	•	x	x	x	x	•	p	m	•
	4	•	x	x	x	x	•	p	m	•
1985	1	•	x	x	x	x	•	m	p	•
	2	x	x	x	x	x	p	m	p	m
	3	x	x	x	x	x	p	m	p	m
	4	•	x	x	x	x	•	m	p	•
1986	1	x	x	x	x	x	m	p	p	m
	2	x	x	x	x	x	m	p	p	m
	3	x	x	x	x	x	m	p	p	m
	4	•	x	x	x	x	•	p	p	•
1987	1	x	x	x	x	x	p	p	p	p
	2	x	x	x	x	x	p	p	p	p
	3	x	x	x	x	x	p	p	p	p
	4	•	x	x	x	x	•	p	p	•
1988	1	•			•		•	p	•	•
	2				•		p	p	•	p
	3	•			•		•	p	•	•
	4	•			•		•	p	•	•
1989	1	x	x	•	x	x	p	•	•	p
	2	x	x	•	x	x	p	•	•	p
	3	x	x	•	x	x	p	•	•	p
	4	•	x	•	x	x	•	•	•	•
1990	1	x	•	x	x	•	•	•	p	•
	2	x	•	x	x	•	•	•	p	•
	3	x	•	x	x	•	•	•	p	•
	4	•	•	x	x	•	•	•	p	•
1991	1	•	x	x	x	•	•	p	p	•
	2	•	x	x	x	•	•	p	p	•
	3	x	x	x	x	•	p	p	p	•
	4	•	x	x	x	•	•	p	p	•
1992	1	x	x	x	•	•	p	p	•	•
	2	x	x	x	•	•	p	p	•	•
	3	x	x	x	•	•	p	p	•	•
	4	•	x	x	•	•	•	p	•	•

Fig. 4: Percentage relative errors of estimates $N_{j,t}^{1}$ as compared with $P_{j,t}^{1}$: IMP and EXP



More precisely, let v_t and e_t the error and the relative error, respectively, in a provisional estimate (p_t) with respect to a revised estimate (r_t):

$$v_t = p_t - r_t \qquad e_t = \frac{p_t - r_t}{r_t}.$$

In order to analyze the preliminary estimates of levels we use the following indices:

1. mean relative error, $\bar{e}$;
2. mean absolute relative error, $\bar{e}'$;
3. relative error standard deviation, s_e;
4. square root of the mean quadratic relative error, d_e;
5. bias component of the mean quadratic relative error, U_e^b.

In turns, to evaluate the accuracy of preliminary growth rates we consider the following indices:

1. mean error, $\bar{v}$;
2. mean absolute error, $\bar{v}'$;
3. error standard deviation, s_v
4. square root of the mean quadratic error, d_v;
5. bias component of the mean quadratic error, U_v^M

In other words, we calculate :

Errors	Relative errors
$\bar{v} = \frac{1}{n}\sum_{t=1}^{n} v_t$	$\bar{e} = \frac{1}{n}\sum_{t=1}^{n} e_t$
$\bar{v}' = \frac{1}{n}\sum_{t=1}^{n} \lvert v_t \rvert$	$\bar{e}' = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
$s_v = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(v_t - \bar{v}\right)^2}$	$s_e = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(e_t - \bar{e}\right)^2}$
$d_v = \sqrt{\frac{1}{n}\sum_{t=1}^{n} v_t^2}$	$d_e = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
$U_v^M = \frac{\left(\bar{p} - \bar{r}\right)^2}{d_v^2}$	$U_e^b = \frac{\bar{e}^2}{d_e^2}$

Clearly, $\bar{v}$ and $\bar{e}$ are not measures of the accuracy of provisional estimates. However, together with $\bar{v}'$ and $\bar{e}'$, respectively, they permit to evaluate if, and eventually how much, the errors are always (or almost always) in the same direction.

From table 6 it emerges that, as far as the aggregates in question and given the small number of quarters token into account, on average the preliminary $N^1_{t,j}$ vintage underestimates the final figure (by about 2% for INV, 1.5% for GDP, 1.4% for both CFI and IMP, and by a smaller amount for EXP). This trend is very systematic for GDP, CFI and INV, as the comparison between $\bar{e}$ and $\bar{e}'$ clearly shows, while the corrections to IMP and EXP are not always in the same direction.

It should be noted that the weight of the first revision (from $N^1_{t,j}$ to $P^1_{t,j}$) (i) is relatively more marked ($N^1_{t,j}$ underestimates $P^1_{t,j}$ by about 1% for GDP) and (ii) is generally characterized by a considerable systematicity (for GDP U^b_e is equal to .6352).

The systematicity in underestimating $F_{t,j}$ for GDP, CFI and INV leads, on average, to an underestimation of quarterly growth rates (about 0.9% for INV and 0.7% for GDP, see table 7).

Even in this case (i) the effect of the first revision stage is larger (from .24 to .61 percentage points on average) as compared with the following two, and (ii) the intermediate revisions and the overall one of IMP and EXP have different characteristics from those of GDP, CFI and INV.

A more marked, and systematic, underestimation characterizes the early estimates of annual growth rates (see table 8) for all aggregates, perhaps with the only exception of EXP, for which the overall revision process does not show a marked upward tendency.

It has to be noted that both quarterly and annual growth rates have been calculated on the basis of the same version of data. This means that we compare levels at different stages of revision. For example, the current first quarter growth rates of an aggregate at year *t* are given by

$$n^1_{t,1} = \frac{N^1_{t,1} - P^1_{t-1,4}}{P^1_{t-1,4}};$$

$$p^1_{t,1} = \frac{P^1_{t,1} - R^1_{t-1,4}}{R^1_{t-1,4}};$$

$$r^1_{t,1} = \frac{R^1_{t,1} - D^1_{t-1,4}}{D^1_{t-1,4}}.$$

In all three cases not homogeneous figures are used, in the sense that they don't belong to the same annual benchmark revision stage. [8]

The annual growth rates are *always* calculated on the basis of estimates pertaining to different stages of the annual revision process. Indeed,

$$na^1_{t,j} = \frac{N^1_{t,j} - P^1_{t-1,j}}{P^1_{t-1,j}}, \; j = 1,2,3;$$

$$pa^1_{t,1} = \frac{P^1_{t,j} - R^1_{t-1,j}}{R^1_{t-1,j}}, \; j = 1, \mathrm{K}, 4;$$

$$ra^1_{t,1} = \frac{P^1_{t,j} - D^1_{t-1,j}}{D^1_{t-1,j}}, \; j = 1,2,3,4.$$

As Patterson and Heravi (1992) pointed out, in order to isolate the factors influencing the revisions the growth rates should be calculated on the same vintage of data. Each numerator of the above fractions can actually be viewed as the sum of a 'pure difference' effect and of a 'revision' effect. More precisely, denoting by $y^v_{t,j}$ the -th vintage of an aggregate at quarter j of year t and by $y^{v+4}_{t,j}$ the subsequent $(v+4)$ -th vintage of the same value, it is

$$y^v_{t,j} - y^{v+4}_{t-1,j} = \left(y^v_{t,j} - y^v_{t-1,j} \right) - \left(y^{v+4}_{t-1,j} - y^v_{t-1,j} \right),$$

where $y^v_{t,j} - y^v_{t-1,j}$ measures the pure difference effect and $y^{v+4}_{t-1,j} - y^v_{t-1,j}$ the revision effect. If, as it often happens, on average $y^v_{t-1,j} < y^{v+4}_{t-1,j}$, the annual growth rates tend to be underestimated by the early estimates.

However, the relatively scarcity of data due to the irregular sequence of publication and to the intervention of the historical revision published in 1987, as well as the peculiarity and the relatively brief history of the Italian quarterly national accounts,

8 This problem is less noticed for $d^1_{t,1} = \frac{D^1_{t,1} - D^2_{t-1,4}}{D^2_{t-1,4}}$, because both $D^1_{t,j}$ and $D^2_{t,j}$ satisfy the same accounting constraint with D_t.

induced us to operate on the same version rather than on the same vintage.

On the whole, the empirical evidence shows that the current preliminary estimates have a tendency to underestimation, in a more marked way for GDP, CFI and INV, less for IMP and EXP. Moreover, the revision process is somewhat 'convergent', in the sense that it has a sufficiently stable order so that $N^1 \leq P^1 \leq R^1 \leq F$.
As one could expect, these results are in agreement with those found for the Italian annual revision process[9] (Trivellato, 1986, Biggeri and Trivellato, 1991). They also confirm a generalized phenomenon to many countries (Glejser and Schavey, 1974, Smith, 1977, Young, 1993).

4 An econometric approach to the analysis of the revision process

As it has been shown in the preceding analyses, a noticeable process of revision of Italian quarterly national accounts took place during the 1980s, mainly caused by the need to quantify appropriately the structural changes experienced over the last two decades.

To give an adequate assessment of the innovations induced by successive revisions, within the Department of National Accounts and Economic Analysis of Istat the need was felt to define techniques which would permit a complete appraisal of the changes occurring to various quarterly and annual aggregates, thus rendering their economic interpretation possible.

For this purpose it was decided to use econometric techniques, which are considered most appropriate for a quantitative assessment of the phenomena under study. Particularly, in Pisani and Savio (1994) the following approach was proposed.

Firstly, the main statistical features of the cyclical component of the various preliminary versions of a series and that of the final one are compared. Secondly, integration and cointegration properties of different versions are considered. Lastly, viewing preliminary versions as forecasts of the final version, the issue of rationality is considered with regard to the aspects of unbiasedness and (weak) efficiency.

It should be noted that the analyses carried out in this order provide different, but exhaustive, indications about the revision process. For example, cyclical components could show different patterns, but the series should be cointegrated if the associated revision is stationary. Furthermore, in this case the early version could well be an unbiased though inefficient forecast of the final version.

This approach seems to be very useful in terms of its efficacy in the analysis of the revisions involving considerable changes in the series, in particular of those defined here as historical or occasional. The analysis will then focus mainly on the constant prices 1986.3, 1988.4 and 1990.3 versions, respectively.
These versions are compared with the final one [10], conventionally defined as 1994.2. All the analyses of this section are carried out using logarithmic transformations of the rebased (at 1985 prices) variables [11].

4.1 The analysis of cyclical components

The first step of the analysis consists in comparing the transitory (or cyclical) component of each preliminary version with that of the final version. In Pisani and Savio (1994) such an analysis was carried out using, as a method of decomposition of the series, the estimation of adequate structural models (see e.g. Harvey, 1989).

9 Though not yet concluded, the analyses on quarterly constant prices aggregates generally confirm th ese results.

10 When comparing different versions, the maximum available time span is used.

11 The rebasing has been done for each component of the analyzed variables as they are currently publi shed by Istat. Methodological as well as practical aspects related to this issue (Rushbrook, 1978) will be deeply analyzed in future work.

Tab. 6: Indices of accuracy of provisional estimates of levels <

Comparison	n	$\bar{e}$	$\bar{e}'$	s_e	d_e	U_e^b
GDP						
$N^1 - P^1$	16	-0.0095	.0099	.0072	.0119	.6352
$P^1 - R^1$	24	-.0019	.0045	.0053	.0057	.1180
$R^1 - D^1$	24	-.0015	.0053	.0073	.0075	.0395
$N^1 - F$	7	-.0147	.0147	.0036	.0151	.9417
CFI						
$N^1 - P^1$	16	-.0053	.0072	.0079	.0095	.3088
$P^1 - R^1$	24	-.0017	.0047	.0055	.0057	.0859
$R^1 - D^1$	24	-.0018	.0034	.0045	.0048	.1365
$N^1 - F$	7	-.0136	.0136	.0044	.0143	.9057
INV						
$N^1 - P^1$	16	-.0043	.0215	.0240	.0243	.0314
$P^1 - R^1$	24	-.0047	.0093	.0106	.0116	.1654
$R^1 - D^1$	24	.0029	.0074	.0103	.0107	.0738
$N^1 - F$	7	-.0215	.0215	.0081	.0230	.8763
IMP						
$N^1 - P^1$	16	-.0137	.0192	.0197	.0240	.3266
$P^1 - R^1$	24	-.0013	.0083	.0106	.0107	.0155
$R^1 - D^1$	24	.0002	.0080	.0128	.0128	.0002
$N^1 - F$	7	-.0137	.0209	.0178	.0225	.3728
EXP						
$N^1 - P^1$	16	-.0110	.0168	.0150	.0186	.3497
$P^1 - R^1$	24	.0013	.0104	.0140	.0141	.0080
$R^1 - D^1$	24	.0011	.0087	.0123	.0124	.0079
$N^1 - F$	7	-.0040	.0224	.0271	.0274	.0218

Tab. 7: Indices of accuracy of provisional estimates of quarterly growth rates

Comparison	n	$\bar{v}$	$\bar{v}'$	s_v	d_v	U_v^M
GDP						
$N^1 - P^1$	16	-0.0049	.0059	.0057	.0076	.4255
$P^1 - R^1$	24	-.0015	.0041	.0049	.0052	.0894
$R^1 - D^1$	24	.0011	.0044	.0061	.0062	.0297
$N^1 - F$	7	-.0069	.0069	.0043	.0082	.7196
CFI						
$N^1 - P^1$	16	-.0024	.0030	.0033	.0041	.3416
$P^1 - R^1$	24	-.0005	.0031	.0038	.0039	.0194
$R^1 - D^1$	24	.0002	.0018	.0023	.0023	.0088
$N^1 - F$	7	-.0026	.0036	.0038	.0046	.3111
INV						
$N^1 - P^1$	16	-.0036	.0096	.0109	.0115	.0967
$P^1 - R^1$	24	-.0015	.0074	.0086	.0088	.0302
$R^1 - D^1$	24	.0002	.0054	.0073	.0073	.0004
$N^1 - F$	7	-.0094	.0102	.0058	.0110	.7223
IMP						
$N^1 - P^1$	16	-.0061	.0109	.0126	.0140	.1917
$P^1 - R^1$	24	.0002	.0103	.0144	.0144	.0003
$R^1 - D^1$	24	-.0018	.0106	.0182	.0183	.0101
$N^1 - F$	7	.0032	.0187	.0216	.0218	.0218
EXP						
$N^1 - P^1$	16	-.0056	.0103	.0120	.0132	.1771
$P^1 - R^1$	24	.0005	.0137	.0195	.0195	.0007
$R^1 - D^1$	24	-.0021	.0122	.0191	.0193	.0120
$N^1 - F$	7	.0062	.0209	.0242	.0250	.0606

Tab. 8: Indices of accuracy of provisional estimates of annual growth rates

Comparison	n	$\bar{v}$	$\bar{v}'$	s_v	d_v	U_v^M
GDP						
$N^1 - P^1$	16	-0.0073	.0079	.0055	.0092	.6378
$P^1 - R^1$	24	-.0005	.0060	.0078	.0078	.0046
$R^1 - D^1$	24	-.0004	.0051	.0068	.0068	.0030
$N^1 - F$	7	-.0057	.0065	.0076	.0095	.3651
CFI						
$N^1 - P^1$	16	-.0037	.0065	.0064	.0074	.2527
$P^1 - R^1$	24	.0001	.0043	.0055	.0055	.0001
$R^1 - D^1$	24	-.0006	.0030	.0035	.0036	.0292
$N^1 - F$	7	-.0051	.0070	.0075	.0090	.3192
INV						
$N^1 - P^1$	16	-.0013	.0189	.0217	.0218	.0033
$P^1 - R^1$	24	-.0088	.0145	.0157	.0180	.2377
$R^1 - D^1$	24	.0018	.0076	.0111	.0113	.0247
$N^1 - F$	7	-.0247	.0265	.0180	.0305	.6535
IMP						
$N^1 - P^1$	16	-.0167	.0201	.0175	.0242	.4757
$P^1 - R^1$	24	-.0009	.0119	.0167	.0167	.0031
$R^1 - D^1$	24	.0006	.0101	.0161	.0161	.0014
$N^1 - F$	7	-.0236	.0239	.0182	.0298	.6257
EXP						
$N^1 - P^1$	16	-.0120	.0204	.0187	.0223	.2925
$P^1 - R^1$	24	.0003	.0096	.0120	.0120	.0006
$R^1 - D^1$	24	.0013	.0122	.0188	.0189	.0046
$N^1 - F$	7	-.0098	.0250	.0280	.0297	.1082

Here we use a different method; it consists in the application of the Hodrick and Prescott (1980) filter for determining the trend component of a series [12].

The main characteristic of such a filter is that it emphasizes the medium and high-frequency movements in the data, those that are usually associated with business cycle. It should be noted that the filter is simple to use and highly operational, the definition of the trend this procedure adopts is quite intuitive, it is able to render stationary series that are integrated up to the fourth order and, finally, the estimated trend component is endowed with the same features of integration and cointegration as the original series (Lupi, 1992). Being statistical in nature, it represents a particular way of viewing the data which can be very productive.

Figure 5 reports the cyclical component of these aggregates for the various versions considered in the analysis. In the seventies the cyclical components of the various versions are very close, while in the eighties they are more irregular and, consequently, show more marked differences.Nonetheless, the graphs show that the main turning points, as well as the phase and the period of the cycles, are substantially similar, independently of the aggregate and the version considered. This result parallels those obtained in Pisani and Savio (1994) for GDP.

The contemporaneous cross correlations between the early and final version (see table 9) give some quantitative insights on the degree of correspondence between the cyclical components. The correlations are very high and less than 0.9 in just three cases, in correspondence with the first version considered for INV, IMP and EXP.

4.2 Unit root analysis and cointegration tests

It was recently stressed that the comparison between provisional and final versions of a given series may conveniently be carried out by means of reference to the concept of cointegration (see Patterson and Heravi, 1991, Pisani and Savio, 1994).

In fact, an important aspect to be verified during the analysis of the revision process is the stationarity of the revisions resulting from different versions and the existence of a cointegration relationship between the various provisional versions and the final one.

Such an analysis is found to be particularly relevant both for the activity of the official statistical agencies and for the users of data. In fact, if we accept the hypothesis that the residuals from a cointegrating regression between, say, a preliminary version and the final one of a series are I(1) rather than I(0), then an I(1) variable, or combination of variables, must have been omitted[13]. Thus, an important aspect to be considered is the issue of whether preliminary versions of data are cointegrated with the final realization. Testing for cointegration requires a preliminary analysis of the degree of integration of the series examined. To this end, the conventional test proposed by Dickey and Fuller (Fuller, 1976, Dickey and Fuller, 1979, 1981) is used, the results of which are given in table 10.

Given the strong upwards trend in most of the data series, the functional form used in the test is that with drift and deterministic trend [14]. The number of differences used in the tests was chosen so that it suffices the regression passing an LM test for autocorrelation of residuals up to lag 8.

---

12 Briefly, the filter proposed by Hodrick and Prescott permits the extraction from the series y_t of a trend component τ_t defined as the solution of the following minimum problem

$$\min_{\tau_t} \sum_{t=1}^{n} (y_t - \tau_t)^2 + \mu \sum_{t=2}^{n-1} \left[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}) \right]^2, \mu > 0.$$

Fluctuations are defined from the trend component, $(y_t - \tau_t)$.

13 On the other hand, if the hypothesis of cointegration was accepted, the associated revision (derive d from the joint restriction on the corresponding cointegrating regression that the constant is zero and the c ointegrating coefficient is unity) could be non-stationary. We shall return to this aspect later, when we will d iscuss about the rationality tests.

Tab. 9: Estimated contemporaneous cross-correlations between the cyclical component of the final version and those of different preliminary versions

Version	GDP	CFI	INV	IMP	EXP
1986.3	0.944	0.909	0.898	0.851	0.807
1988.4	0.989	0.968	0.962	0.973	0.968
1990.3	0.974	0.957	0.937	0.971	0.964

The null hypothesis of integration of order 1 is strongly accepted for all the series considered, excluding the first and the final version of imports, for which the test statistic is only marginally below the 5% critical value.

In these two cases, however, as in the others, the null hypothesis is rejected by an analogous procedure applied to the first differences of the series.

Now, we can consider the tests of cointegration between the final version and each preliminary estimate.
Table 11 reports the results obtained using Johansen's maximum likelihood procedure (see Johansen, 1988, 1989) for the choice of the cointegration rank between couples of variables.

The order of the VAR was carried out on the basis of the results provided by a number of criteria: Akaike, Hannan and Quinn, and Schwartz [15].

Following recent results obtained by Reimers (1993), in case of conflict the results obtained from the Hannan and Quinn's criterion was preferred [16]. In only five cases the resulting VAR order was modified, increasing it until the hypotheses of whiteness and normality of residuals were satisfied. The tests considered were those of Ljung and Box (1978) for serial correlation, Jarque and Bera (1981) and Mardia (1980) for the normality hypothesis.

Tab. 10: Dickey-Fuller unit root tests for log-level and first differences of different versions*

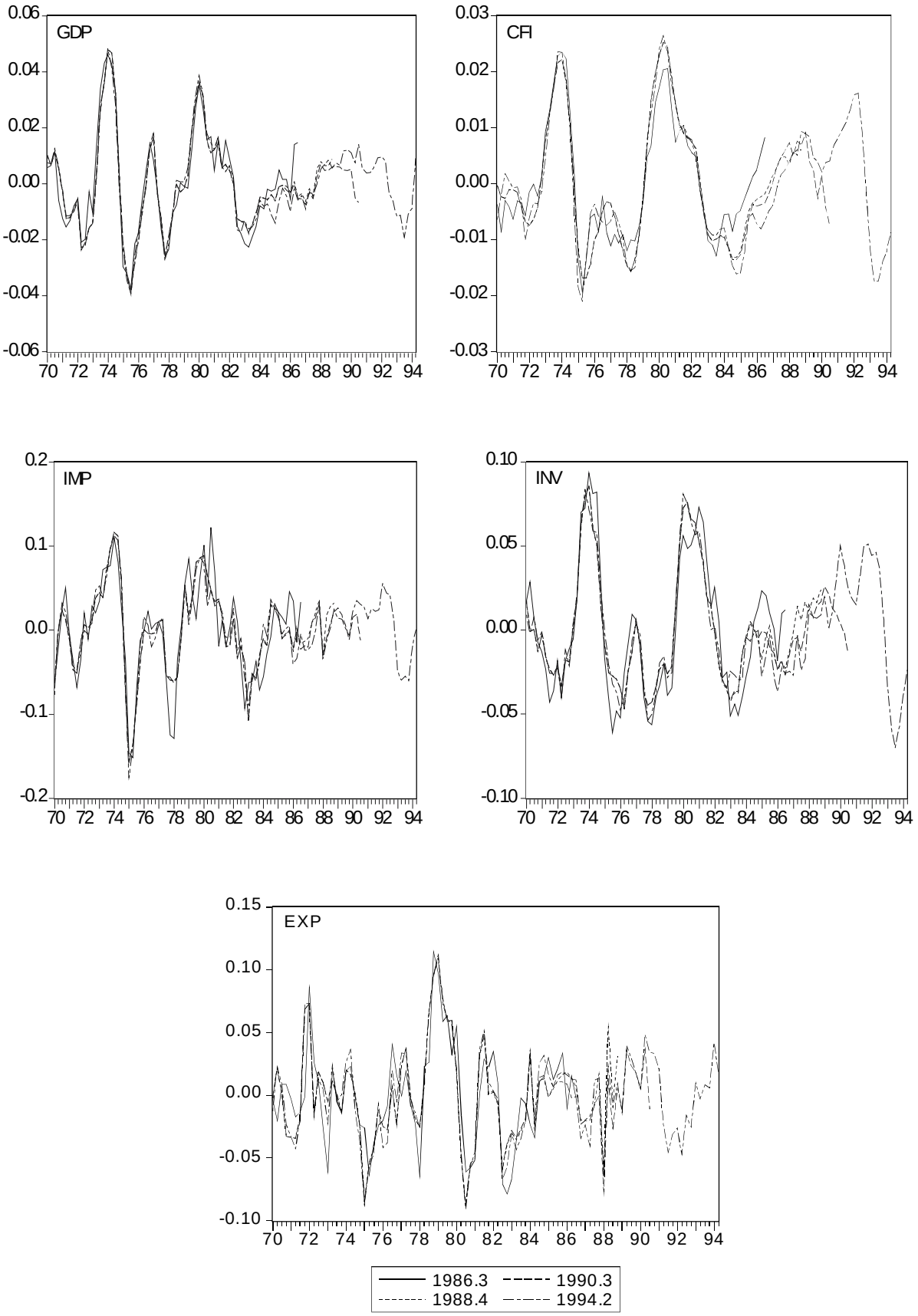
Series	1994.2	1990.3	1988.4	1986.3
GDP	-2.5054 (1)	-3.2447 (1)	-3.1015 (1)	-2.5279 (1)
Δ GDP	-5.5959 (0)	-4.7364 (0)	-4.5282 (0)	-5.4797 (0)
CFI	-1.0140 (2)	-1.5690 (0)	-2.0729 (2)	-2.4538 (1)
Δ CFI	-4.9364 (1)	-5.0858 (2)	-4.2561 (1)	-4.6179 (0)
INV	-2.5354 (1)	-2.3180 (1)	-2.2314 (3)	-2.9060 (1)
Δ INV	-5.8423 (0)	-4.9659 (0)	-4.6528 (2)	-5.2346 (0)
IMP	-3.5218 (1)	-1.9009 (0)	-2.2500 (0)	-3.6149 (1)
Δ IMP	-7.6797 (0)	-6.8469 (0)	-6.6661 (0)	-5.5993 (3)
EXP	-3.0644 (0)	-2.8386 (0)	-2.7403 (0)	-2.3008 (0)
Δ EXP	-10.8241 (0)	-10.1938 (0)	-10.0698 (0)	-8.5313 (0)

14 The results obtained in the specifications without trend are qualitatively analogous to those shown in the paper and are available from the authors on request.

15 See Lütkepohl (1991) for details.

16 In fact, Reimers shows by Monte Carlo simulations that Hannan and Quinn's test performs better in t erms of corrected percentages of selection than the others mentioned in the text.

Fig. 5: Cyclical components of some versions derived from the Hodrick-Prescott filter



According to the preferred functional form, a linear trend is included in both I(0) and I(1) components (see Johansen, 1992a, 1992b) for reasons analogous to those which make preferable, in an univariate framework, the model with trend as opposed to that with only the drift, when the series is characterized by a clear tendency. In these cases, the tests performed in the model without trend would lead one to accept the hypothesis of I(1) even if the variable was in fact trend-stationary.

The results summarized in table 11 show that, while for IMP and EXP the hypothesis of cointegration is strongly accepted for all the preliminary versions, for the other couples of variables it is not possible to define, at a level of significance of 5%, a common (stochastic) trend.

In particular, for INV the hypothesis of cointegration is accepted for all the versions considered, excluding that of 1986.3, for which the results of the tests based on the maximum eigenvalue (column 5) and on the trace (column 6) are contradictory. Such an ambiguity might originate in the low power of the test when the cointegrating relation is close to non-stationarity (see Johansen and Juselius, 1992).

The hypothesis of cointegration is accepted for all versions of CFI except the last one (1986.3), while for GDP we just find *i=1* for 1990.3[17]. Combining these results with those highlighted in the previous paragraph, we should affirm that the revision process might have led to substantial changes in the long-term components of most series, while leaving their cyclical components virtually unaffected. This may be due to the fact that either the revision has no cycle, or the cycle has the same characteristics as that present in the early versions.

Such results indicate that occasional revisions may involve significant differences in log-term components of data series. It should be stressed that the results obtained from 'non-updated' macroeconometric models *vis-à-vis* changes in national accounts - such as those related to rebasing, methodologies adopted or reference sources-might therefore be decidedly misleading.

4.3 Efficiency and unbiasedness of the revision process

Harvey *et al.* (1983) and Mankiw *et al.* (1984), amongst others, have noted that preliminary versions can be considered as forecasts of the final version, or as preliminary measurements of the final version subject to a measurement error. In the former sense, the one considered in this work, a number of tests widely used, for example, to verify the hypothesis of rational expectations (unbiasedness, serial correlation, efficiency and orthogonality) may be valid as an aid for a probabilistic interpretation of the differences between preliminary versions of data and the final one. In this sense, a common starting point is given by the following equation:

$$y_t^m = \alpha + \beta y_t^v + u_t, \qquad (1)$$

Where y_t^m and y_t^v represent the preliminary and the final version of *y* respectively. Equation (1) is traditionally interpreted as the basis for the unbiasedness test, which verifies the hypothesis $H_0 : (\alpha, \beta) = (0,1)$.

However Wallis (1989) has pointed out that the test is in fact stricter: the mantained hypothesis should be identified with an efficient, though not necessarily fully efficient, predictor. In other words, the test actually verifies the hypothesis of efficiency (in weak sense) of expectations. In fact, efficiency requires that the revision $\left(y_t^m - y_t^v\right)$ is unforeseeable from information at time *t*. Since part of this information includes (beyond the constant) the provisional version y_t^v, it follows that a finding of $\alpha \neq 0$ and/or $\beta \neq 1$ in (1) suggests inefficient use of information.

Furthermore, unbiasedness requires that the mean revision is zero (Holden and Peel, 1990), which could occur even if the preliminary version was an inefficient predictor.

17 The differences in terms of cointegration properties might originate in the sharp re-evaluation of the service sector due to the revision process, in particular for those activities included in the branch of bu siness services. This may have determined more marked changes in the intermediate consumptions component *vis-à-vis* final consumption.

Tab. 11: Johansen maximum likelihood cointegration analysis: tests of the cointegration rank betwe en the final and different preliminary versions*

Series	Version	N. Of lags in the VAR	i	$-n\ln(1-\hat{\lambda}_i)$	$-n\sum\ln(1-\hat{\lambda}_i)$	N. Of coint. vectors
GDP	1990.3	2	1 2	19.02 9.49	28.50 9.49	1
	1988.4	2	1 2	16.27 9.29	25.56 9.29	0
	1986.3	2	1 2	10.29 3.92	14.21 3.92	0
CFI	1990.3	5	1 2	28.04 6.92	34.95 6.92	1
	1988.4	2	1 2	22.55 5.90	28.04 5.90	1
	1986.3	2	1 2	12.44 6.60	19.04 6.60	0
INV	1990.3	2	1 2	40.31 5.97	46.28 5.97	1
	1988.4	2	1 2	30.06 6.99	30.06 6.99	1
	1986.3	2	1 2	14.02 9.59	23.61 9.59	0-1
IMP	1990.3	2	1 2	26.90 4.11	31 4.11	1
	1988.4	2	1 2	24.75 4.81	29.56 4.81	1
	1986.3	5	1 2	26.74 8.79	35.53 8.79	1
EXP	1990.3	2	1 2	19.77 8.20	26.01 8.20	1
	1988.4	1	1 2	45.83 7.71	53.55 7.71	1
	1986.3	1	1 2	44.44 5.92	50.37 5.92	1

* 5% critical values are 18.96 and 12.25 for the $\hat{\lambda}_{MAX}$ test, respectively for $i = 1$ and $i = 2$.
For the trace test 5%critical values are 25.32 and 12.25.

Weak efficiency is a sufficient, but not a necessary condition for unbiasedness [18]. As a direct test of bias we can consider estimating a regression of the revision on a constant:

$$y_t^m - y_t^v = \alpha_1 + \varepsilon_t \qquad (2)$$

If the constant is statistically different from zero, we reject the hypothesis of unbiasedness. Note that weak efficiency and unbiasedness are particular cases of the orthogonality test. The orthogonality condition requires preliminary revisions be unpredictable using the information available at the time of the forecast. Two points must be stressed:

- Brown and Maital (1981) have shown that the hypotheses of unbiasedness and weak efficiency might well be valid if the residuals in (1) and (2) are serially correlated. In fact, as the forecast is ade h steps ahead, the residuals may be characterized by an MA(h-1) process. Following Hansen and Hodrick (1980), they propose to adjust the covariance matrix given by the OLS regression obtaining correct standard errors for the coefficients. This is the estimation method here used.
- There is a difference of approach between these tests and those for cointegration. The former ones assume stationary variables and moving average disturbances, while the latter assume non-stationary variables and disturbances which are $NID(0, \sigma^2)$.

  As Patterson and Heravi (1991) note, this difference does not appear to have been overcome yet in the literature. The results of running the regressions given by (1) an (2) obtained are given in table 12. The test statistic for the null hypothesis is shown as a $\chi^2(2)$ and a $\chi^2(1)$ for the weak efficiency and unbiasedness tests respectively.

The null hypothesis of weak efficiency is surprisingly accepted only for INV and the 1988.4 version, while unbiasedness is found for CFI and INV for all versions excluding 1986.3. On the contrary, EXP is always, though marginally, unbiased. The values given by α_t provide measures of the mean bias.

Compared with the version 1986.3, the aggregate which underwent sharpest growth was INV (+30.0%), followed by GDP (+12.1%) and the other variables the quotas of which - in real terms - grew between 5% and 6%. In most of these cases, such re-evaluations caused changes in the level of the series, while not substantially altering their cyclical and long-term components.

5 Conclusions

This paper has documented and discussed the revisions to the current prices, seasonally adjusted expenditure components of Italian gross domestic product published over the range 1985 - 1994, and, for the same aggregates, to various versions of constant prices, seasonally adjusted data.

A number of conclusions have been drawn, concerning the relative sizes of the absolute revisions to different components, the biases in the preliminary estimates and the evolutionary patterns being followed by the revision process. We observe that:

- virtually all important components of GDP experienced significant revisions;
- GDP, CFI and INV have marked downward biases in their preliminary estimates while IMP and EXP are touched relatively less by the revision process;
- these results are generally confirmed by an econometric comparison of a number of constant prices versions with the last published figures;
- the short-medium term dynamics of three representative constant prices versions remained substantially unaffected by the revision process;
- the permanent component has been strongly affected by the 1987 historical revision; the weak efficiency and unbiasedness tests suggest that after that date the revision process behaves in a more regular way.

Obviously, the present paper is only a beginning effort in the analysis of quarterly national accounts revisions, and many other important issues remain to be investigated:

18 There is no inconsistency, however, between a finding of rejection of weak efficiency and an accept ance of the hypothesis of unbiasedness.

- are the seasonally adjusted estimates revised more or less, on average, than the unadjusted estimates, and are the biases stronger or weaker?
- is there any monotonous pattern in the sequence of successive revisions that could be used in reducing the average size of corrections?
- what do the revision process tell us about the behaviour of growth rates? Have the revisions to the levels left the growth rates substantially unaffected?
- what's about the other national accounts aggregates?

Another related issue involves the use of preliminary estimates in econometric models. Such a problem has received much attention in literature for the obvious implications on the reliability of the conclusions that an applied economic researcher can draw (see, among others, Grimm and Hirsch, 1983, Howrey, 1984, Trivellato and Retto re, 1986, Nijman, 1989).

Tab. 12: Efficiency and unbiasedness tests*

Series	Version	$\hat{\alpha}$	$\hat{\beta}$	$x^2(2)$	$\hat{\alpha}_1$	$x^2(1)$
GDP	1990.3	0.2309 (0.0322)	0.9813 (0.0027)	102.0085 (0.00)	0.0047 (0.0021)	5.0379 (0.02)
	1988.4	0.2205 (0.0221)	.9822 (0.0018)	133.8009 (0.00)	0.0053 (0.0021)	6.2547 (0.01)
	1986.3	-3.5386 (0.2686)	1.3073 (0.026)	1992.6460 (0.00)	0.1219 (0.0212)	33.1358 (0.00)
CFI	1990.3	0.1435 (0.0461)	0.9878 (0.0039)	10.9096 (0.00)	-0.0006 (0.0014)	0.1641 (0.69)
	1988.4	0.0808 (0.0280)	0.9932 (0.0023)	8.7086 (0.01)	0.0004 (0.0009)	0.1380 (0.71)
	1986.3	-4.2053 (0.0924)	1.3641 (0.0080)	2231.6148 (0.00)	0.0605 (0.0256)	5.5802 (0.02)
INV	1990.3	-0.1550 (0.0400)	1.0147 (0.0038)	19.5871 (0.00)	0.0010 (0.0009)	1.2729 (0.26)
	1988.4	-0.0907 (0.0812)	1.0086 (0.0013)	2.3498 (0.31)	0.0005 (0.0005)	0.9906 (0.32)
	1986.3	0.9025 (0.8746)	0.9417 (0.0844)	983.5052 (0.00)	0.3035 (0.0100)	880.6516 (0.00)
IMP	1990.3	-0.1750 (0.1938)	1.0140 (0.0184)	22.6953 (0.00)	-0.02719 (0.0070)	15.2612 (0.00)
	1988.4	-0.4248 (0.1360)	1.0378 (0.0013)	73.2108 (0.00)	-0.0259 (0.0075)	12.0318 (0.00)
	1986.3	2.5406 (0.0823)	0.7611 (0.0079)	2042.6419 (0.00)	0.0487 (0.0313)	2.4225 (0.12)
EXP	1990.3	0.1868 (0.0457)	0.9827 (0.0044)	45.9330 (0.00)	0.0058 (0.0034)	2.8680 (0.09)
	1988.4	0.2351 (0.0350)	0.9779 (0.0034)	90.4598 (0.00)	0.0057 (0.0040)	2.0602 (0.15)
	1986.3	1.8608 (0.0571)	0.8250 (0.0056)	3101.9622 (0.00)	0.0575 (0.0287)	4.0067 (0.005)

* 5% Columns from 3 to 5 refer to equation (1), those from 6 to 7 to equation (2). Standard errors are reported in parentheses below the coefficients. In square brackets marginal significance levels of the x^2 test are specified. $x^2(2)$ Test verifies the hypothesis $H_0:(\alpha,\beta) = (0,1)$ in (1), $x^2(1)$ verifies the hypothesis $H_0:\alpha_1 = 0$ in (2).

References

Biggeri L. and U. Trivellato (1991) , "The evaluation of errors in national accounts data: provisional and revised estimates", Survey Methodology, 17, 1, pp. 99-119.

Brown B.W. and S. Maital (1981) , "What do economists know? An empirical study of experts' expectations", Econometrica, 49, 2, pp. 491-504.

Dickey D.A. and W.A. Fuller (1979) , "Distribution of the estimators for autoregressive time series with a unit root", Journal of the American Statistical Association, 74, pp. 427-431.

Dickey D.A. and W.A. Fuller (1981) , "Likelihood ratio statistics for autoregressive time series with unit root", Econometrica, 49, pp. 1057-1072.

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

Fuller W.A. (1976), Introduction to statistical time series, New York, Wiley.

Giovannini E. (1993), "L'informazione statistica per le decisioni: alcune linee generali e un esempio illustrativo", Quaderni di Ricerca. Interventi e Relazioni, Istat, Sistema Statistico Nazionale.

Glejser H. and P. Schavey (1974) , "An analysis of revisions on national accounts data for 40 countries", The Review of Income and Wealth, 20, pp. 317-332.

Grimm B.T. and A.A. Hirsch (1983) , "The impact of the 1976 NIPA benchmark revision on the structure and predictive accuracy of the BEA quarterly econometric model", in M.F. Foss (ed.), The U.S. National Income and Product Accounts. Selected topics, National Bureau of Economic Research, Studies in Income and Wealth, 47, pp. 333-382.

Hansen L.P. and R.J. Hodrick (1980) , "Forward exchange rates as optimal predictors of future spot rates: an econometric analysis", Journal of Political Economy, 88, pp. 829-853.

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

Harvey A.C., C.R. McKenzie, D.P.C. Blake and M.J. Desai (1983), "Irregular data revisions", in A. Zellner (ed.), Applied time series analysis of economic data, Washington D.C., US Bureau of the Census, pp. 329-347.

Hibbert J. (1981), "Revisions to quarterly estimates of gross domestic product", Economic Trends, 331, pp. 82-87.

Hodrick R. and E. Prescott (1980), "Post-war U.S. business cycles: an empirical investigation", Carnegie Mellon University, (mimeo).

Holden K. and D.A. Peel (1990), "On testing for unbiasedness and efficiency of forecasts", The Manchester School, 58, pp. 120-127.

Howrey E.P. (1984), "Data revision, reconstruction, and prediction: an application to inventory investment", The Review of Economics and Statistics, 66, 3, pp. 386-393.

Istat (1985), "I conti economici trimestrali. Anni 1970-1984", Supplemento al Bollettino Mensile di Statistica, 12.

Istat (1987), "Miglioramenti apportati ai conti economici trimestrali. Serie con base dei prezzi 1970"', Collana d'Informazione, n. 4.

Istat (1990), "Nuova contabilità nazionale", Annali di Statistica, vol. 9.

Istat (1991), "Conti economici delle imprese con meno di 10 addetti. Anni 1986 e 1988", Collana d'informazione, 45.

Istat (1992), "I conti economici trimestrali con base 1980", Note e Relazioni, 1.

Istat (1993), "The underground economy in Italian economic accounts", Annali di Statistica, serie X, vol. 2.

Jarque, C.M. and A.K. Bera (1980) , "Efficient tests for normality, homoscedasticity and serisl correlation", Economics Letters, 6, pp. 255-259.

Johansen S. (1988), "Statistical analysis of cointegration vectors", Journal of Economics Dynamics and Control, 12, pp. 231-254.

Johansen S. (1989), Likelihood based inference on cointegration. Theory and applications, Venezia, Cafoscarina.

Johansen S. (1992a), "Determination of the cointegration rank in the presence of a linear trend", Oxford Bulletin of Economics and Statistics, 54, pp. 383-397.

Johansen S. (1992b), The role of the constant term in cointegration analysis of nonstationary variables, Institute of Mathematical Statistics, University of Copenhagen, Preprint 92-1.

Johansen S. and K. Juselius (1992) , "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK", Journal of Econometrics, 53, pp. 211-244.

Ljung G.M. and G.E.P. Box (1978) , "On a measure of lack of fit in time series models", Biometrika, 65, pp. 297-303.

Lupi C. (1992), "Measures of persistence, trends and cycles in I(1) time series: a rather skeptical viewpoint", Applied Economics Discussion Paper, 137, Institute of Economics and Statistics, University of Oxford.

Lütkepohl H. (1991) , Introduction to multiple time series analysis, Berlin, Springer-Verlag.

MacKinnon J.G. (1991), "Critical values for cointegration tests", in R.F. Engle and C.W.J. Granger (eds.), Long-run economic relationships: readings in cointegration, Oxford, Oxford University Press, pp. 267-276.

Mankiw N.G., D.E. Runkle and M.D. Shapiro (1984), "Are preliminary announcements of the money stock rational forecasts?", Journal of Monetary Economics, 14, pp. 15-27.

Mardia K.V. (1980), "Tests for univariate and multivariate normality", in Handbook of Statistics, vol. I, Amsterdam, North-Holland, pp. 279-320.

Morgenstern O. (1963) , On the accuracy of economic observations, Princeton, Princeton University Press.

Nijman T. (1989), "A natural approach to optimal forecasting in case of preliminary observations", (mimeo).

Patterson K.D. (1992) , "Revisions to the components of the trade balance for the United Kingdom", Oxford Bulletin of Economics and Statistics, 54, 3, pp. 103-120.

Patterson K.D. and S.M. Heravi (1991) , "Data revisions and the expenditure components of GDP", The Economic Journal, 101, 407, pp. 887-901.

Patterson K.D. and S.M. Heravi (1992) , "Efficient forecasts or measurement errors? Some evidence for revisions to United Kingdom GDP growth rates", The Manchester School, 60, 3, pp. 249-263.

Pisani S. and G. Savio (1994) , "Le revisioni delle serie storiche trimestrali di Contabilità Nazionale", in Istat, Avanzamenti metodologici e statistiche ufficiali, Roma, pp. 203-229.

Rao J.N.K., K.P. Srinath and B. Quenneville (1989), "Estimation of level and change using current preliminary data", in D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh (eds.), Panel surveys, New York, Wiley, pp. 457-479.

Reimers H.E. (1993) , "Lag order determination in cointegrated VAR systems with application to small German macro-models", paper presented at the Econometric Society European Meeting '93, Uppsala, August (mimeo).

Rushbrook J.A. (1978), "Rebasing the national accounts - why and how and the effects", Economic Trends, 293, pp.105-110.

Smith P. (1977), An analysis of the revisions to the Canadian National Accounts, Statistics Canada, Econometric Research Staff Working Paper.

Trivellato U. (ed.) (1986), Errori nei dati preliminari, previsioni e politiche economiche, Padova, Cleup.

Trivellato U. (ed.) (1987), Attendibilità e tempestività delle stime di contabilità nazionale, Padova, Cleup.

Trivellato U. and E. Rettore (1986), "Preliminary data errors and their impact on the forecast error of simultaneous-equation models", Journal of Business & Economic Statistics, 4, pp. 445-453.

Wallis K.F. (1989), "Macroeconomic forecasting: a survey", The Economic Journal, 99, pp. 28-64.

Young A.H. (1993), "Reliability and accuracy of the quarterly estimates of GDP", Survey of Current Business, 73, 10, pp. 29-43.

Zellner A. (1958), "A statistical analysis of provisional estimates of gross national product and its components, of selected national income components, and of personal savings", Journal of the American Statistical Association, 53, 281, pp. 54-65.

APPENDIX

Actual Publication Schedule of Italian Quarterly National Acounts Aggregates

Year and quarter of reference

Vers.:	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
1984.4	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x										
1985.1	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	.									
1985.2																x x									
1985.3																x x x									
1985.4												x x x x	x x x x	x x x x	x x x x	x x x x									
1986.1	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x								
1986.2															x x x x	x x x x	x x								
1986.3															x x x x	x x x x	x x x								
1986.4	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x								
1987.1	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .						x x x x	x x x x	x							
1987.2	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .						x x x x	x x x x	x x							
1987.3	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .						x x x x	x x x x	x x x							
1987.4	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .				x x x x	x x x x	x x x x	x x x x	x x x x							
1988.1	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	.						
1988.2	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .							x x x x	x x x x	x x						
1988.3	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . .						
1988.4	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x						
1989.1																				x					
1989.2																		x x x x	x x x x	x x					
1989.3																		x x x x	x x x x	x x x					
1989.4																x x x x	x x x x	x x x x	x x x x	x x x x					
1990.1																			x x x x	x x x x	x				
1990.2																			x x x x	x x x x	x x				
1990.3																			x x x x	x x x x	x x x				
1990.4	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .				
1991.1	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	.			
1991.2	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. .			
1991.3	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x			
1991.4	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .								x x x x	x x x x	x x x x	x x x x	x x x			
1992.1	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .	. . . .											x x x x	x x x x	x		
1992.2	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x x x	x x		
1992.3																					x x x x	x x x x	x x x		
1992.4																			x x x x	x x x x	x x x x	x x x x	x x x x		
1993.1																						x x x x	x x x x	x	
1993.2																						x x x x	x x x x	x x	
1993.3																						x x x x	x x x x	x x x	
1993.4																				x x x x	x x x x	x x x x	x x x x	x x x x	
1994.1																							x x x x	x x x x	x
1994.2																							x x x x	x x x x	x x
	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94

x: new estimate

.: no published estimate

La relation entre les comptes annuels et les comptes trimestriels à l'INSEE

Virginie MADELIN
INSEE

1 Introduction

L'élaboration de la comptabilité nationale française se caractérise par le fait que les méthodes de construction des comptes trimestriels et des comptes annuels sont fondamentalement différentes, mais qu'il existe néanmoins de fortes interactions et une cohérence temporelle entre les deux systèmes. Les comptes annuels sont bâtis à partir d'une approche exhaustive qui exploite toutes les informations statistiques et comptables disponibles et travaille à un niveau très fin de la nomenclature (600 et 90 postes). Les comptes trimestriels privilégient les enchaînements temporels entre les grandeurs macro-économiques en s'appuyant sur une information plus agrégée. Ils reposent sur une méthode économétrique.

Les comptables trimestriels n'interviennent dans le processus d'élaboration des comptes annuels que lors de la première évaluation d'une année en mars de l'année suivante. L'information dont disposent alors les comptables annuels est confrontée à celle qui a été traitée et structurée en termes de comptabilité nationale par les comptables trimestriels. L'approche plus conjoncturelle des comptables trimestriels s'enrichit du détail de l'analyse des comptables annuels pour aboutir à une évaluation commune, à quelque détails près, de l'année écoulée. Elle contribue ainsi à remettre en question certaines méthodes ou pratiques d'évaluation mais aussi l'interprétation des comptes nationaux. Lors de cette période de concertation, des problèmes récurrents sont rencontrés. Comme les comptables trimestriels travaillent à un niveau relativement agrégé (12 branches, 13 produits), des déformations sensibles apparaissent entre la structure de leur TES (Tableaux Entrées-Sorties) et celle du TES agrégé des comptables annuels. De même, l'absence d'indicateur infra-annuel pour certaines branches de l'activité entraîne des erreurs dans la construction des comptes trimestriels. Enfin, les comptables annuels privilégient les comptes aux prix de l'année précédente, alors que les comptables trimestriels raisonnent aux prix de l'année de base (1980), ce qui entraîne, lorsqu'on s'éloigne de l'année de base, des distorsions dans les estimations.

Les comptes trimestriels n'interviennent plus dans le processus d'élaboration des comptes d'une année lors des exercices de révision suivants effectués durant trois années pour prendre en compte l'enrichissement de l'information disponible. Ils deviennent des clients des comptables annuels et introduisent dans leur base de donnée les évaluations révisées des trois dernières années.

Cette orthogonalité entre les deux méthodes d'élaboration assure qu'il n'existe pas de biais auto-entretenu dans l'approche économétrique des comptes trimestriels.

Le système français de comptabilité nationale comprend deux volets articulés, les comptes annuels d'une part, les comptes trimestriels d'autre part. Les premiers visent à donner une information

macro-économique annuelle avec le maximum de précision et de détail. Les seconds ont pour objectif de préciser et d'anticiper à un rythme infra-annuel cette information. Bien que travaillant à un niveau de détail différent, tous deux reposent sur un système comptable intégré, et répondent aux mêmes contraintes de cohérence et d'équilibrage. La différence de fond entre ces deux systèmes tient à leur mode d'élaboration : les comptes annuels sont construits par la mobilisation de sources statistiques à champ exhaustif ou rendu exhaustif, qui sont successivement mises en cohérence conceptuelle, puis confrontées et arbitrées dans un cadre comptable unique, année par année. Les comptes trimestriels sont fondés sur une méthode économétrique et comptable qui "trimestrialise" le compte annuel et permet les prévisions ; l'information est élaborée, série par série, en utilisant des indicateurs, éventuellement partiels, dont la qualité fondamentale est la bonne représentativité à l'égard de la série annuelle correspondante.

Malgré ces différences sensibles dans leur méthodologie, les comptes annuels et trimestriels sont totalement articulés et cohérents. Pour leur première version d'un compte (en mars n+1), dite provisoire, les comptables annuels bénéficient de la réflexion conjoncturelle qu'ont déjà menée les comptables trimestriels mais apportent une information plus riche et plus détaillée. La confrontation des méthodes, des sources, des résultats et des diagnostics permet une convergence sur l'essentiel des éléments de compte.

Au fur et à mesure que l'information statistique et comptable sur l'année se développe, les comptables annuels revoient leurs premières estimations ; chaque année est ainsi révisée trois fois. Les comptes trimestriels se calent alors sur ces nouvelles évaluations et revoient leurs relations économétriques. Cette façon de faire assure que les comptes trimestriels ne sont pas entachés de biais auto-entretenus.

Dans la pratique, les comptables annuels élaborent une première version de leurs comptes annuels de façon presque entièrement indépendante des comptables trimestriels. C'est pourquoi les méthodes de base de chacun des intervenants sont d'abord décrites dans les deux premières parties. Les différences de méthodes génèrent naturellement des écarts entre les évaluations, dont la concertation tend à faire disparaître l'essentiel. Ceci est décrit en troisième partie.

Les comptes annuels et les comptes trimestriels comprennent des tableaux entrées-sorties, des comptes de secteurs institutionnels et des tableaux des opérations financières en flux et en encours, dont l'articulation est plus ou moins élaborée selon qu'il s'agit des comptes annuels ou trimestriels. Par souci de simplification, on s'est centré par la suite sur la synthèse relative aux biens et services.

2 Comment le compte provisoire des comptables annuels est-il construit ?

Dans le système de comptabilité nationale annuelle, les comptes de l'année n sont établis au printemps de l'année n+1, dans une version dite provisoire à partir de statistiques souvent partielles sur les produits et d'informations conjoncturelles qualitatives. Ces comptes sont ensuite révisés trois années de suite et donnent lieu à des versions dites semi-définitives 1 puis 2 et enfin définitive. Les comptes pour les années n-3 (définitif), n-2 et n-1 (semi-définitifs) et n (provisoire), sont élaborés lors de la "campagne de compte" qui s'étend de juillet n à avril n+1 et sont publiés dans le rapport sur les comptes de la Nation en juin n+1.

Les révisions sont rendues nécessaires par l'enrichissement au fil du temps de l'information statistique et économique ; en outre, la plus grande disponibilité des données de base permet un plus grand raffinement des méthodes utilisées.
Seule la version provisoire de l'année n donne lieu à une confrontation avec les résultats des comptes trimestriels, ces derniers étant exactement calés pour les années antérieures sur les comptes nationaux. Ce sont donc essentiellement les méthodes utilisées au compte provisoire que l'on va développer.

La méthode générale d'élaboration des comptes annuels français, consiste à évaluer le produit intérieur brut (PIB) selon trois approches [1]. La première est l'approche dite production dans laquelle on établit les comptes de production en 90 branches d'activité, c'est-à-dire au niveau 90 de la nomenclature d'activités et de produits (NAP). Dans la seconde, dite par la dépense, on élabore les

équilibres Ressources-Emplois (ERE) des produits au niveau 600 de la NAP. Dans la troisième, dite par le revenu, on établit les comptes d'exploitation des branches d'activité (CEB) au niveau 40 de la NAP, à partir des comptes de secteurs institutionnels.

Ces trois optiques ne sont pas élaborées indépendamment les unes des autres, mais au contraire confrontées en permanence. Un tel principe nécessite des redressements et arbitrages intermédiaires nombreux, qui permettent d'aboutir à une seule estimation du PIB. L'élaboration des comptes est menée chaque année dans le cadre du Tableau Entrées-Sorties (TES) et de l'articulation entre les comptes d'exploitation par secteurs institutionnels et par branche. Le TES est évalué en 90 branches et produits, dans trois systèmes de prix (aux prix courants, aux prix de l'année précédente et aux prix de 1980), dans deux systèmes de taxation (hors toutes taxes et y compris TVA non déductible) (*graphique 1).*

Dans l'élaboration du compte provisoire, faute d'informations suffisantes, on n'utilise pas l'approche par les revenus. Il n'y a ni compte d'exploitation par branche, ni décomposition par sous-secteurs d'activité des opérations des sociétés et quasi-sociétés non financières et des entreprises individuelles.

Dans l'optique de la production , le PIB aux prix du marché est la somme des valeurs ajoutées des différents agents économiques, qu'ils soient étudiés du point de vue des branches ou des secteurs institutionnels, et des impôts frappant les produits non déjà inclus dans les valeurs ajoutées des producteurs. La synthèse de cette approche est réalisée dans le cadre du TES par la mise en place des comptes de production par branche, qui donnent le montant de la production effective des branches, leur consommation intermédiares et leur valeur ajoutée.

On a :

PIB = $\sum$ Valeurs ajoutées des branches
+ TVA grevant les produits
+ Impôts à l'importation (nets des subventions à l'importation)

Les procédures d'évaluation de ces postes dépendent de la nature des sources disponibles. On peut distinguer trois situations différentes:

- il existe des informations comptables complètes et jugées suffisamment fiables en elles-mêmes : une analyse intégrée des informations permet une transposition de celles-ci à toutes les opérations non financières de la comptabilité nationale. On obtient simultanément les valeurs de la production et celles des consommations intermédiaires. Au moment du compte provisoire, cette méthode concerne les secteurs institutionnels des Administrations Publiques, des Institutions de crédit et des Entreprises d'assurance. Ces informations sont dites exogènes;
- la production est évaluée à partir de la connaissance des produits : cette procédure concerne surtout la production agricole;
- pour les autres secteurs institutionnels, il n'existe pas au moment du compte provisoire de sources comptables suffisamment développées pour permettre une évaluation intégrée des comptes par secteurs. La production des branches est évaluée à ce stade à partir de sources conjoncturelles, enquêtes de branche trimestrielles ou mensuelles, indices de la production industrielle...

L'élaboration du tableau des emplois intermédiaires se fait en deux temps :

- On met d'abord en place les informations partielles disponibles, lorsque le montant total de la consommation intermédiaire par branche est connu, ou lorsque le montant de certaines cases est connu.
- Les autres cases du tableau sont ensuite obtenues par une projection du TES au prix de l'année précédente, sous l'hypothèse de la fixité d'une année sur l'autre des coefficients techniques.

Les évaluations des comptes de production par branche sont confrontées à celles obtenues dans l'optique de la demande.

Dans l'optique de la demande , le PIB aux prix du marché est mesuré par la valeur des biens et services affectés à des emplois finals, nette de la valeur des importations. En pratique, on évalue pour chacun des produits au niveau 600 de la NAP, l'offre et la demande dans le cadre des équilibres ressources-emplois qui viennent alimenter le TES (*cf graphique 2*).

On a :

PIB = Consommation finale
+ Formation brute de capital fixe (FBCF)
+ Variations des stocks
+ Exportations de biens et services (nettes des importations) .

La production distribuée du produit est calculée à partir de la production effective de la branche. Après prise en compte des données relatives au commerce extérieur fournies par la Direction générale des Douanes, on évalue ensuite la demande intérieure, répartie entre les ménages (consommation finale et variations de stocks dans les commerces), la demande intermédiaire des entreprises et des administrations(consommation intermédiaire et variations de stocks utilisateurs et commerce) et la demande finale de ces mêmes agents (Formation Brute de Capital Fixe et variations de stocks commerce). Ces évaluations sont menées à partir des sources conjoncturelles disponibles.
A ce niveau de détail, il y a bon nombre de cas où une seule de ces catégories d'emplois existe. Ces équilibres sont ensuite agrégés au niveau 90, celui du TES, ce qui permet alors une meilleure estimation des variations de stocks.

A l'issue de ces deux étapes de travail, la mise en cohérence des évaluations dans le cadre du TES nécessite des arbitrages itératifs. La projection du Tableau des échanges intermédiaires, aux prix de l'année précédente, dans l'approche production donne une consommation intermédiaire potentiellement différente de celle obtenue directement dans l'approche par la demande.
En outre, on dispose d'une estimation de la FBCF par produits dans les équilibres ressources-emplois qui doit être confrontée à celle obtenue dans les comptes des secteurs institutionnels. Les comptes annuels sont établis en valeur, aux prix de l'année précédente, et aux prix de l'année 1980 par chaînage des indices de prix de l'année précédente.

Après ces arbitrages, on obtient une première estimation complète du PIB qui va être rapprochée de celle établie selon des méthodes radicalement différentes, mais bien souvent à partir de sources très proches, par les comptables trimestriels.

3 Méthodologie des comptes trimestriels

Les comptes du trimestre t, décrivant les opérations sur biens et services aux prix de l'année 1980, sont publiés au milieu du trimestre t+1 dans une version "allégée", les "Premiers Résultats". Les "Résultats Détaillés" au trimestre t+2 donnent un détail d'informations aux prix courants et aux prix de l'année 1980, pour les opérations sur biens et services ainsi que des éléments de comptes des sociétés et quasi-sociétés et des ménages. Un premier panorama de l'année n est ainsi publié en février n+1, puis détaillé en avril n+1. C'est pour cette dernière publication que la concertation entre comptables annuels et trimestriels est réalisée.

La méthode de base utilisée dans les comptes trimestriels consiste à relier économétriquement une information statistique infra-annuelle ou conjoncturelle à une grandeur comptable [2]. A chaque série de comptes est associée une série de périodicité trimestrielle disponible dans les délais d'élaboration des comptes et dont l'évolution est similaire à celle du poste comptable: cette série est appelée indicateur du compte. Le plus souvent, l'indicateur trimestriel utilisé diffère de l'évaluation de la donnée comptable pour des raisons de définition et de champ couvert.

Les comptes sont d'abord réalisés en volume, c'est-à-dire dans le cas des comptes trimestriels **aux prix de l'année 1980**, à partir d'indicateurs en volume. Ils sont ensuite évalués en valeur. Ce sont de plus des **comptes corrigés des variations saisonnières** (CVS), construits directement à partir d'indicateurs eux-mêmes désaisonnalisés. Cette façon de corriger la saisonnalité des comptes a été préférée à celle qui aurait consisté à construire d'abord des comptes bruts puis à les désaisonnaliser, parce que la saisonnalité de l'indicateur ne coïncide pas nécessairement avec celle du phénomène économique qu'on cherche à mesurer. Les comptes trimestriels n'existent que dans une version CVS ; ils ne sont pas calculés dans leur forme brute et ne sont pas non plus corrigés des jours ouvrables (CJO).

Plus précisément, l'élaboration des comptes trimestriels se fait en trois étapes. Les deux premières portent sur les années révolues. Ce sont l'étalonnage et le calage. La dernière consiste à évaluer les comptes pour la période séparant le compte calé du temps présent.

L'étalonnage est la transformation des indicateurs trimestriels en éléments de comptes trimestriels. Il consiste en un ajustement économétrique qui permet d'élaborer une donnée de comptabilité trimestrielle. On suppose qu'il existe une relation économétrique

simple (et non dynamique) qui lie, sur le passé, l'évolution annuelle du compte et l'indicateur annualisé retenu. Estimée sur des données annuelles, cette relation est supposée rester valable sur des données trimestrielles. Cette hypothèse conduit à effectuer la désaisonnalisation des séries non sur les comptes trimestriels étalonnés mais sur l'indicateur lui-même.

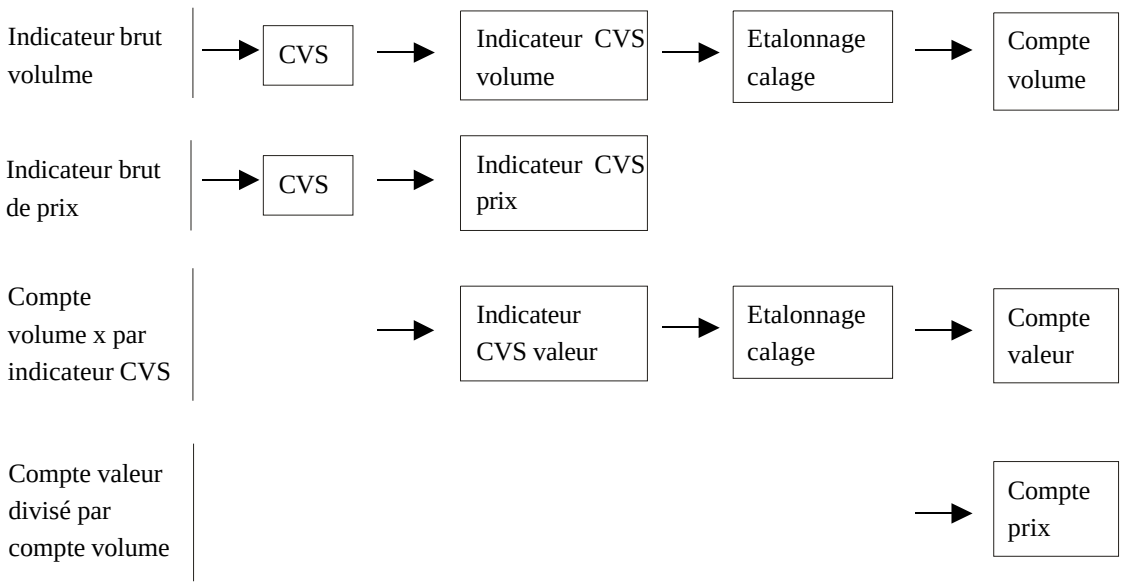
Le calage est la mise en conformité, sur le passé, des éléments des comptes trimestriels ainsi obtenus aux comptes annuels. Il n'est réalisé systématiquement que pour les versions semi-définitives et définitives des comptes annuels. L'écart constaté est réparti chaque année proportionnellement sur chaque trimestre de l'année, par ce qu'on appelle la procédure de lissage.

Lorsqu'on passe à l'évaluation de l'année en cours, on utilise la relation économétrique établie pour l'élaboration des comptes trimestriels étalonnés pour obtenir le compte du trimestre courant. En outre, pour préserver l'évolution entre le compte du dernier trimestre de l'année qui a pu être calée et ceux des trimestres suivants et ainsi éviter une "marche d'escalier", le dernier écart est conservé sur les comptes trimestriels suivants. **La procédure globale** consiste en l'articulation de ces opérations pour obtenir des comptes en volume, en valeur et en prix ; elle s'articule de la façon décrite dans le shéma 1.

Les comptes trimestriels sont sujets à des révisions au fur et à mesure de leurs publications successives. D'abord, parce que les comptables annuels révisent eux-aussi leurs estimations; à cette occasion, les relations économétriques sont revues, tout en préservant les profils infra-annuels. Ensuite, parce que la donnée comptable trimestrielle peut être modifiée, soit par rectification d'une première estimation de l'indicateur, soit par une mise à jour des corrections des variations saisonnières. Enfin, parce que la méthodologie du calcul dans les comptes trimestriels peut être modifiée à cause de la disparition d'une source ou l'adoption d'un nouvel indicateur considéré comme meilleur.

Les comptes trimestriels sont organisés d'abord selon des domaines de base (production, consommation des ménages, FBCF et commerce extérieur) et des domaines de synthèse du TES, qui permettent d'évaluer les séries qui ne sont pas calculées dans ces domaines de base. Pour ces derniers, les sources utilisées sont très proches de celles des comptes annuels; les estimations se font en règle générale au moins au niveau 40 de la NAP. Dans le TES, pour une branche donnée, les consommations intermédiaires en produits sont pour l'essentiel, calculées au niveau 16 de la NAP. Elles sont étalonnées sur la production de la branche correspondante en introduisant parfois un trend temporel. Seules quelques consommations

Schéma 1: La procédure globale



intermédiaires sont traitées différemment (agriculture par exemple, où on utilise des indicateurs directs). Il existe aussi un TEE (tableau économique d'ensemble), assez fruste.

La principale exception, due aux lacunes de l'information statistique et conjoncturelle dans ce domaine, apparaît dans les services.
Dans les services marchands d'abord, l'équilibre se construit à partir des emplois. La consommation intermédiaire de l'ensemble des branches en services marchands est évaluée directement, à partir d'indicateurs ad hoc. Celle-ci sert alors d'indicateur pour estimer la consommation intermédiaire en produit "services marchands" pour chaque branche utilisatrice. On peut alors déterminer la somme des emplois de ce produit et par conséquent la production.
Pour les consommations intermédiaires des branches non marchandes, on ne dispose pas d'indicateurs infra-annuels. Elles sont donc estimées annuellement, puis lissées trimestriellement.

Une fois ces traitements effectués, le TES est équilibré en calculant les variations de stocks comme solde du total des ressources et des emplois totaux hors stocks.

4 Quand vient le moment de la concertation

La concertation entre les comptables annuels et trimestriels est un exercice éminemment pratique et concret. Il ne s'agit donc pas dans cette partie de théoriser cet exercice mais au contraire de retracer comment la qualité de l'information macro-économique se trouve enrichie par la concertation.

Le rapprochement systématique entre les deux types de comptes est récent et pourrait d'ailleurs être étendu au compte annuel semi-définitif 1, une fois qu'une capitalisation de l'expérience acquise en matière de concertation aura été faite par les deux équipes. L'objectif actuel n'est pas de caler exactement et en détail un compte sur l'autre - ce qui ne poserait pas de problèmes techniques une fois les principaux écarts réduits -, mais d'obtenir un consensus sur le diagnostic économique que l'on peut tirer de chacun d'eux *(tableau 1)*. Ainsi, on vise à obtenir des taux de croissance globaux égaux, de faibles différences sur les composantes du PIB étant tolérées. La priorité lors du compte provisoire est donnée à la qualité de l'analyse macro-économique portée sur l'année qui vient de s'écouler.

Comme on l'a déjà vu (cf. partie 1), les comptables annuels et trimestriels élaborent quasi-indépendamment leur première version pour l'année écoulée n au printemps n+1. La confrontation de ces premiers résultats complets fait apparaître des écarts sur le PIB et ses contreparties, aussi bien dans l'optique production (les valeurs ajoutées des branches) que dans l'optique de la dépense (emplois finals nets des importations). Toutes choses égales par ailleurs, on est alors confronté, lors du rapprochement entre les deux comptes, à une difficulté mécanique inhérente à la nature matricielle du TES : l'intervention sur un poste particulier de celui-ci a des conséquences sur l'ensemble du tableau, qui plus est divergentes parfois entre comptes trimestriels et annuels. C'est pourquoi, là encore, les arbitrages sont conduits de manière itérative.

Les écarts sont d'abord générés par les différences de méthodes qui sont totalement "orthogonales". Là où les comptables annuels essaient de serrer au plus près, avec un grand niveau de détail, les informations économiques et statistiques, les comptables trimestriels se placent dans une perspective temporelle et une logique économétrique. Une fois ces méthodes mises en oeuvre, on constate des différences dont certaines sont de nature structurelle (comme le fait de travailler à un niveau de détail différent dans le TES) et donc délicates à réduire, tandis que les autres, qui portent sur des points particuliers (évolution des marges commerciales dans les industries agro-alimentaires par exemple) se résolvent par une analyse économique approfondie. Ces deux types d'écarts sont implicitement résolus par une procédure d'arbitrage en deux phases [3].

La première fait intervenir les responsables de domaines pour chaque type de comptes (par exemple, les responsables annuels et trimestriels de l'évaluation de la consommation des ménages). En premier lieu, afin d'anticiper et de minimiser les éventuelles divergences sur le diagnostic économique, des réunions d'information et de concertation se tiennent de façon précoce. Dès que les comptables trimestriels sont en mesure de porter un diagnostic sur l'année écoulée, un peu avant la fin de l'année n, ils informent leurs homologues annuels. Puis, les "négociations"

sont relancées, au fur et à mesure que ces derniers disposent à leur tour d'informations s'inscrivant dans leur cadre méthodologique. La concertation proprement dite se tient en mars n+1.

En restant sur l'exemple de la consommation des ménages, on constatait en mars 1994 un écart spontané de 0,5 point entre les deux estimations en volume *(tableau 2)*. L'étude approfondie des sources de base et de leur interprétation, ainsi que des méthodes fines employées, à un niveau plus détaillé que celui qui apparaît dans le tableau, a permis de ramener cet écart à 0,1 point (soit + 0,8 pour les comptes annuels et + 0,7 pour les comptes trimestriels). L'écart n'est pas totalement réduit, dans la mesure où sa réduction entraînait des problèmes d'équilibrage de certains équilibres ressources-emplois (sur le tabac par exemple). On butait dans ces cas sur le problème déjà souligné de l'interconnexion de toutes les évaluations.
Un autre exemple est celui de la branche automobile et matériel de transports, où les méthodes de construction de l'équilibre sont complètement différentes. En effet, la production est un solde pour les comptes annuels (on remonte l'équilibre), alors qu'elle est estimée directement à partir des indices de production industrielle (IPI) pour les comptes trimestriels.

La discussion est parfois compliquée par le fait que les comptables trimestriels élaborent des comptes étalonnés sur des indicateurs corrigés des variations saisonnières. Ce n'est bien sûr pas le cas des comptables annuels. Une même source statistique conduit donc parfois à des interprétations économiques ou à des estimations d'éléments de compte légèrement différentes. L'élaboration des comptes trimestriels bruts, étudiée actuellement par l'équipe qui en a la charge, facilitera naturellement les arbitrages avec les comptes annuels.
Enfin, une difficulté supplémentaire provient de l'organisation des équipes, dans lesquelles les domaines couverts ne sont pas exactement les mêmes, ce qui complique le dialogue. A titre d'illustration, on a un expert de la production trimestrielle et un autre pour la FBCF trimestrielle, alors que les comptables nationaux sont responsables de l'ensemble de l'équilibre pour des produits donnés. Il s'agit en quelque sorte de concilier dans ces cas une vision du TES en colonnes pour les premiers, et en ligne pour les seconds.

Cette nécessaire réconciliation est faite dans **la deuxième phase de l'arbitrage** , qui est une phase de synthèse et concerne à ce titre les responsables de la synthèse, et donc essentiellement du TES dans le cas du PIB et de ses composantes. Les discussions se fondent alors sur une mise en perspective et en relation de tous les éléments de comptes abordés auparavant de façon plus transversale. On rencontre dans cette phase les mêmes problèmes que ceux rencontrés dans la première, mais de manière exacerbée par la nécessaire cohérence de l'ensemble. On se heurte en outre à des difficultés structurelles récurrentes.

D'abord, **le problème d'agrégation du TES** . Pour les comptes nationaux annuels comme trimestriels, on a noté (cf. parties 1 et 2) que la synthèse s'articule autour d'une projection du TES en volume. Du côté des comptes annuels, elle s'appuie pour partie sur l'hypothèse de stabilité des coefficients techniques d'une année sur l'autre. Cette hypothèse est cependant largement tempérée par le fait que pour un certain nombre de produits (ou de branche), on connaît directement les consommations intermédiaires détaillées à partir des sources spécifiques. Ce sont les cases fixées, qui représentent près de 60 % du volume des consommations intermédiaires du TES annuel pour un compte semi-définitif (et plus encore sur un compte définitif).

Du côté des comptes trimestriels, l'hypothèse est légèrement différente, parce que le TES est réalisé à un niveau beaucoup plus agrégé (NAP 15) que dans les comptes annuels : les coefficients techniques sont calculés en appliquant la tendance observée auparavant. Des cases fixées sont aussi évaluées, qui s'appuient sur des relations économétriques directes.

Pour les comptes annuels, la projection s'effectue au niveau 90 de la NAP alors que pour les comptes trimestriels, elle s'effectue au niveau 15. Lorsqu'on aborde la deuxième partie de l'arbitrage, on doit donc agréger le TES des comptes annuels du niveau 90 au niveau 15 pour le comparer à celui des comptables trimestriels. Ce faisant, on introduit des distorsions par rapport au cas où le TES des comptes annuels aurait été projeté directement au niveau 15, du fait des effets de structure.

En effet, si on calcule le TES au niveau 90 (ce qu'on appelle par la suite le niveau des branches élémentaires), et si on l'agrège ensuite, la

consommation intermédiaire d'une branche du niveau 15 est la somme des consommations intermédiaires des branches élémentaires qui la composent pondérées par leurs taux de croissance respectifs. Si on calcule directement le TES au niveau 15, cela revient à faire la somme des consommations intermédiaires des branches élémentaires de la branche agrégée considérée, en leur appliquant uniformément le taux de croissance de cette branche agrégée.

De plus, le fait de fixer certaines consommations intermédiaires dans le tableau des échanges intermédiaires (les cases fixées), introduit une distorsion supplémentaire. En effet, ces tableaux ne sont pas construits dans le même détail dans les comptes annuels et les comptes trimestriels, et les consommations intermédiaires fixées ne se recouvrent alors pas nécessairement. Par exemple, les comptes annuels peuvent ne fixer en NAP 90 qu'une partie d'un niveau de la NAP 15 (en branche ou en produit), là où les comptes trimestriels fixeraient directement ce niveau.

Pour illustrer cet effet d'agrégation, variable en fonction de la conjoncture propre à chaque branche, on a comparé les résultats en volume de la projection du TES annuel de 1991, hors services non marchands, au niveau 90 agrégé au niveau 15, avec ceux qui auraient été obtenus par projection directe de ce tableau au niveau 15. On ne fait apparaître aucun écart sensible sur le total des consommations intermédiaires entre la première et la deuxième méthode, mais les écarts entre les branches sont plus dispersés (tableau 3). Ainsi, pour les consommations intermédiaires de la branche agriculture, l'écart est de 1,2 % et il est de -1,2 % pour la branche location immobilière.

Un autre problème structurel provient de ce que **les systèmes de prix sont différents** : les comptables annuels privilégient les prix de l'année précédente tandis que les comptables trimestriels s'appuient sur les prix de l'année 1980. Pour permettre les comparaisons, les comptables annuels chaînent leurs indices de prix. Non seulement le dialogue est délicat du fait de "cultures" différentes, non seulement cela pose le problème du choix des indices de prix, mais en outre ce dialogue est compliqué par les effets de modification de structure induits par le chaînage. Pour résoudre ce problème, il faut alors raisonner au niveau de détail le plus fin possible. De plus, les distorsions deviennent de plus en plus importantes au fur et à mesure que l'on s'éloigne de l'année de base.

Enfin, la réduction des écarts est compliquée par le fait que **les équilibres comptables n'ont pas la même fonction** selon qu'il s'agit des comptes annuels ou des comptes trimestriels. Pour les premiers (cf. partie 1), ils sont au coeur de l'évaluation de la cohérence d'ensemble des diverses estimations : l'examen et l'interprétation des soldes comptables sont à la base des procédures d'arbitrage. Cette importance apparaît a contrario lorsque l'absence d'arbitrage laisse subsister un ajustement, comme c'est le cas entre comptes financiers et comptes non financiers. Dans la confection des comptes trimestriels, la réalisation d'équilibres comptables sert parfois à estimer les séries pour lesquelles aucun indicateur n'est utilisable. Les soldes comptables sont dans un premier temps des résultats produits automatiquement par le système. L'examen de leur niveau et de leur évolution porte sur la vraisemblance de ce résultat et conduit éventuellement à une remise en cause des indicateurs utilisés en amont.

C'est le cas par exemple des variations de stocks (cf. partie 2), calculées comme solde du total des ressources et des emplois totaux hors stocks, dans les comptes trimestriels et appréhendées directement dans les comptes nationaux. Il s'agit typiquement d'un poste sur lequel la convergence est difficile et qui a un impact sur l'ensemble du système. Le problème est le même pour la production dans les services, calculée par solde pour les comptes trimestriels.

Malgré ces difficultés, l'expérience montre qu'on peut arriver à converger "raisonnablement" sur les principaux agrégats (*tableaux 4 et 5)*. Cela nécessite une étroite collaboration et un véritable dialogue entre toutes les équipes, mais la crédibilité des informations économiques mises à la disposition des utilisateurs se trouve ainsi renforcée.

Le rapprochement entre comptes annuels et trimestriels ne s'arrête pas là. On a vu que pour les années antérieures à l'année n, non seulement les comptes trimestriels sont calés sur les comptes annuels, mais, de plus, les relations économétriques sont revues à chaque révision, afin de mieux prendre en compte les trois nouveaux points annuels. Cette

façon de faire permet de s'assurer qu'il n'existe pas de biais systématique dans les comptes trimestriels.

5 Conclusion

Malgré des méthodes radicalement différentes, les comptes annuels et les comptes trimestriels français donnent une analyse de la situation économique du pays très convergente, aussi bien en termes qualitatifs que quantitatifs. Cela résulte d'un dialogue précoce et approfondi entre les diverses équipes responsables de chaque type de comptes.

Il faut cependant avoir à l'esprit que l'efficacité de l'organisation des comptes français a un coût matériel (notamment informatique) et humain. Par exemple, l'équipe des comptes trimestriels comporte 14 personnes et on peut estimer qu'une centaine de personnes travaillent sur les comptes annuels.
La concertation, facile en régime de croisière d'un cycle de croissance, est plus délicate au moment des retournements de conjoncture. Elle induit néanmoins une réflexion sur les méthodes employées par l'une ou l'autre équipe, ainsi que leur pertinence, dans un souci de lisibilité et d'interprétation des comptes. D'ores et déjà, des voies de progrès de la méthodologie des comptes trimestriels sont apparues : un niveau de travail plus fin (par exemple en NAP 40 pour les branches et les produits) et un perfectionnement des méthodes économétriques amélioreraient la qualité des premières estimations, de même que la réalisation de comptes bruts faciliteraient les arbitrages.

Références

[1] "Le Produit National Brut", ouvrage collectif, collection INSEE Méthodes, n°34-35-36, 1993.

[2] "Les comptes nationaux trimestriels", G. Dureau, collection INSEE Méthodes, n° 13,1991.

[3] "Notes internes" de l'INSEE, 1993 et 1994.

Graphique 1

Le tableau "Entrées-Sorties"
Outil de synthèse des opérations sur biens et services

Produc.	Import.	Marges Comm.	Droits de douane	TVA grevant les produits		Branches	Consom. finale	FBCF	Variation De stocks	Export.
Tableau des ressources					produits	Tableau des entrées intermédiaires	Tableau des emplois finals			

Compte de production par branche

Transferts de vente résiduelles

Compte d'exploitation par branche

PIB

Graphique 2

L'équilibre ressources-emplois (ERE) d'un produit

Production		Consommations intermédiaires
+		+
Importations		Consommation finale
+		+
Droits de douane (nets des subventions à l'import.)	=	FBCF
+		+
Marges commerciales		Variations de stocks
+		+
TVA grevant les produits		Exportations

Tableau 1

Evolution du PIB (en % aux prix de l'année 1980)

Année	Compte trimestriel (1)	Compte provisoire
1987	2,3	2,4
1988	3,8	3,5
1989	3,6	3,8
1990	2,8	2,8
1991	1,2	1,2
1992	1,4	1,2
1993	-1,0	-1,0

1) Evolutions des notes de conjoncture de juillet.

Tableau 2

Evolution de la consommation des ménages en 1993 avant et après concertation

en %

	avant concertation		après concertation	
	Comptes annuels	comptes trimestriels	Comptes annuels	Comptes trimestriels
Alimentation	0,6	0,8	0,8	0,7
Energie	0,5	-0,3	0,7	0,6
Produits manufacturés	-1,3	-0,8	-1,5	-1,3
Biens durables	-6,4	-6,3	-6,6	-6,6
Textiles et cuirs	-2,2	-1,6	-2,3	-1,9
Autres produits manufacturés	2,3	3,1	2,0	2,3
Services	1,4	2,2	2,0	2,2
Services marchands	1,3	2,2	2,0	2,3
Services non-marchands	2,9	2,5	1,8	1,8
Consommation totale (volume)	**0,3**	**0,8**	**0,6**	**0,7**
Prix de la consommation totale	**2,2**	**2,2**	**2,1**	**2,1**

Tableau 3

Consommations intermédiaires par branche obtenues par projection du TES 1991 SD2.

(écart en % entre la projection faite directement au niveau NAP15 et la projection faite au niveau 90 puis agrégée au niveau 15)

Agriculture	1,2
IAA	-0,2
Energie	0,8
Biens intermédiaires	-0,2
Biens d'équipement professionnels	=
Biens d'équipement ménagers	=
Automobile	-0,2
Biens de consommation courante	-0,5
Bâtiment, génie civil et agricole	0,3
Transports et Télécommunications	0,5
Services marchands	=
Location immobilière	-1,2
Assurances	-0,3
Organismes financiers	=
Total	=

Tableau 4

Evolution des ressources et emplois de biens et services en 1993 (en % aux prix de 1980)

	Comptes trimestriels (1)	Comptes annuels
PIB	-1,0	-1,0
Importations	-3,1	-3,4
Consommation finale des ménages	0,7	0,6
Consommation finale des administrations	0,5	0,9
FBCF totale	-5,1	-4,4
Variations des stocks (milliards de francs de 1980)	-41,5	-49,5
Exportations	0,1	0

1) Evolutions de la note de conjoncture de juillet 1994

Tableau 5

Evolution de la production par branches en 1993

	Comptes trimestriels (1)	Comptes annuels
Produits agro-alimentaires	-1,7	0,3
Energie	1,4	1,4
Produits manufacturés	-5,1	-5,2
Bâtiment génie civil et agricole	-4,1	-3,9
Commerce	-0,2	-1,1
Ensemble des services marchands	10,3	10,6

1) Evaluations de la note de conjoncture de juillet 1994.

Cyclical patterns of the Spanish Quarterly National Accounts: a VAR analysis

Ana M. ABAD GARCIA
Enrique Martin QUILIS
Istituto nacional de estadistica

1 Introduction

The Spanish economy has suffered deep changes in its level of activity and employment with adverse consequences on welfare and macroeconomic performance. This phenomenon is treated as a result of cyclical forces. Can we identify some basic sources of aggregate fluctuations?

In this paper we examine some evidence provided by the Spanish Quarterly National Accounts using multiple time series techniques. The cyclical signal is extracted applying annual rates of growth of the original data. This simple filter offers a good measure of the cyclical component of a time series, specially if it is free form irregular component (Spanish Quarterly National Accounts are free form seasonality and irregularity), (Melis, 1991; Cristóbal and Quillis, 1994; INE, 1192, 1993).

The statistical model employed is a Vector of Autorregressions (VAR). This type of model allows an easy characterisation of the data without unreliable a priori restrictions (Sims, 1980, 1981).

The list of variables is:

cpn: private national consumption
cpu: public consumption
fbe: fixed investment in equipment
fcs: fixed investment in buildings
xbs: export of goods and services
mbs: import of goods and services
pib: gross domestic product

All the data are expressed at 1986 constant prices. The sample covered from 1970. I to 1194.I (93 observations). The variables are expressed as annual rates of growth.

Section two deals with some descriptive measures of the data, including cointegration analysis. The model is presented in the third section and the structural analysis (response - impulse and variance decomposition) is presented in the fourth section. The paper finishes with a set of conclusions and a graphical appendix.

2 Descriptive and cointegration analysis

It is advisable to examine the sample properties of the data before identifying and estimating a VAR model. Special attention is devoted to the issue of unit roots and cointegrating relationships.

Attending to the variance and contemporaneous correlation with the gross domestic product (see table 1 and graph 1), we detect a group formed by the investment in equipment, in building and the imports, which are closely related to GDP and are remarkably more volatile than it. Private consumption is next to this group in terms of correlation (the highest: 0.9), but is slightly more volatile than aggregate output.
Public consumption is the least volatile serie and its association with GDP is not quite high. Exports are very volatile and almost unrelated to GDP.

Table 1: Volatility and correlation with PIB

Series	μ	σ	σ/ σ PIB	ρ., PIB
PIB	2.92	2.45	1.00	1.00
CPN	2.83	2.88	1.17	0.90
CPU	5.12	1.96	0.80	0.58
FBE	3.09	10.08	4.10	0.80
FCS	2.32	6.79	2.77	0.79
XBS	6.77	4.78	1.95	0.11
MBS	6.95	8.53	3.47	0.72

μ = sample mean

σ = sample standard deviation

ρ = contemporaneous correlation with PIB

Recursive correlations starting in 1975.l are depicted in grap 2. They show:

- public consumption exhibits increased procyclical behaviour;
- exports are less procyclical;
- the rest of the variables intensify their procyclical behaviour along the sample.

The series are non-stationary since all of them have a unit root. Their means and variances evolve through time (see graph 1) and their simple and partial autocorrelation functions exhibit the usual characteristics of non-stationarity (a high value of pact at lag 1 and a slow decay of act). The augmented Dickey-Fuller (ADF) test does not reject the hypothesis of a unit root. The regression used is:

$$\nabla x_t = \mu + \beta x_{t-1} + \sum_{j=1..4} \gamma_j \nabla x_{t-j} + \varepsilon_t \qquad (1)$$

where: $\nabla = 1 - B$ is the difference operator, B is the backward operator $B^j x_t = x_{t-j}$; μ, β and γ_j are parameters estimated using ordinary least squares (OLS) and ε_t is a iid perturbation with zero mean and constant variance.

Since all the series are I(1), is there a cointergration relationship among them? We use the Engle-Granger two-step method (Engle and Granger, 1987) to examine this issue. The regressions used are:

$$x_{it} = c_i + \beta_i' x_{(i)t} + \varepsilon_{it} \qquad \text{i=1..7} \qquad (2)$$

Table 2: Detection of unit roots

Series	ρ(1)	ρ(12)	ADF
PIB	0.96	0.00	-1.88
CPN	0.94	0.15	-1.72
CPU	0.87	-0.19	-0.73
FBE	0.93	-0.15	-1.58
FCS	0.95	-0.03	-2.02
XBS	0.86	-0.33	-2.91
MBS	0.92	0.08	-2.25

Critical value (1% level of significance) is -3.5 (MacKinnon, 1990)

where: c_i is a constant; β_i: 6x1 is a vector of parameters (to be estimated using OLS); $x_{(i)t}$ are all the variables excluding x_i and ε_{it} is a perturbation term.

Using the EG technique, we conclude that there are no cointegration relationship in the data. Another way (non inferential) to detect such relationships is by means of the analysis of the autovalues of the variance-covariance matrix of the residuals generated by VAR models of growing order (graph 3).

The results are:

- there are three autovalues close to zero, denoting a rank of cointegration of three;
- there is a dominant non-stationary autovalue (possibly) associated with the common cyclical pattern of the series;
- there are two intermediate autovalues which could be related to the idiosincrasic behaviour of some series (the first) and a deviation correction mechanism from main patterns (the second) (Luenberger, 1979).

We decided to model the cyclical signal of the series $(z_t = \nabla_4 x_t)$ without additional differences (that is $y_t = \nabla\nabla_4 x_t$) because of the discrepancies between EG

Table 3: Engle-Granger Test of Cointegration

PIB	CPN	CPU	FBE	FCS	XBS	MBS
-2.79	-3.36	-3.49	-2.31	-2.16	-3.30	-3.32

Critical value (1% level of significance) = -5.54

tests and autovalues analysis. Another reason is the inadequacy of differencing to achieve stationarity in multiple time series modeling (Box and Tiao, 1977, 1981, 1982; Tiao and Tsay, 1989; Tiao and al., 1993).

3 A VAR model

We consider the class of VAR models represented by:

$$\underset{m\times m}{\phi_p}(B)\,\underset{m\times 1}{z_t} = \underset{m\times 1}{c} + \underset{m\times 1}{a_t} \qquad (3)$$

where: $z_t = [PIB\,CPN\,CPU\,FBE\,FCS\,XBS\,MBS]_t'$ is the vector of variables to be modeled (m=7); $\phi_p(B) = I - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is a matrix polynomial of p order in the backward operator B; $\phi_i \;\; i = 1..p$ are m x m matrices of parameters; all the roots of polynomial determinant $|\phi_p(B)|$ are on or outside the unit circle; c is a vector of constants and a_t is a multivariate white noise shock with a matrix of variances-covariances $\sum$.

This kind of models are very popular in applied macroeconomic analysis because they distinguish clearly between data encapsulation, via reduced forms like (3), and structural analysis (through a reparameterisation of (3) via Cholesky decomposition of estimated $\sum$).

To specify the order p of the model we have the Akaike information criterion (Akaike, 1982), the likelihood ratio test (Box and Tiao, 1981, 1982; Liu, 1986), combined with the patterns provided by cross correlations matrices and stepwise autoregressions.

Although the message is not completely clear we have chosen p=3 as the appropriate order of the system. In graph 4 we have drawn the residual variances of each variable from stepwise autoregressions. They show a stabilisation of the decreases since p=3.

4 Structural analysis

In this section we expose the results of the structural analysis, performed using the response-impulse functions and the variance decomposition of the prediction errors. So, we get information about the mechanism of propagation of impulses through the system (response-impulse functions) and about the intensity of dynamic interactions among the variables (variance decomposition).

In order to perform this analysis we have ortogonalised the innovations via Cholesky decomposition of the matrix of variances and covariances of the residuals. This procedure is equivalent to identify a structural model with a recursive structure in its contemporaneous interactions (Lütkepohl, 1991; Sims, 1980). The correlation matrix of the innovation is:

P =

PIB	CPN	CPU	FBE	FCS	XBS	MBS
1.00	0.36	0.37	0.49	0.39	0.14	0.44
	1.00	0.44	0.32	0.35	-0.30	0.59
		1.00	0.31	0.11	-.032	0.20
			1.00	0.31	-0.01	0.37
				1.00	-0.08	0.47
					1.00	-0.08
						1.00

Contemporaneous interactions among variables are significant in most cases. It is noticeable the strong correlation between aggregate output (PIB) and investment, and the negative association of exports with the rest of variables (except output).

The scheme of identification used in the Cholesky decomposition is:
Variance decomposition of the prediction error is shown in table 4 and summarised visually in graph 5.
In the short run all the variables are explained by themselves: there is a clear predominance of proper shocks in short run dynamics. In the long run (40 quarters) these elements are less relevant although continue to keep a significant role.

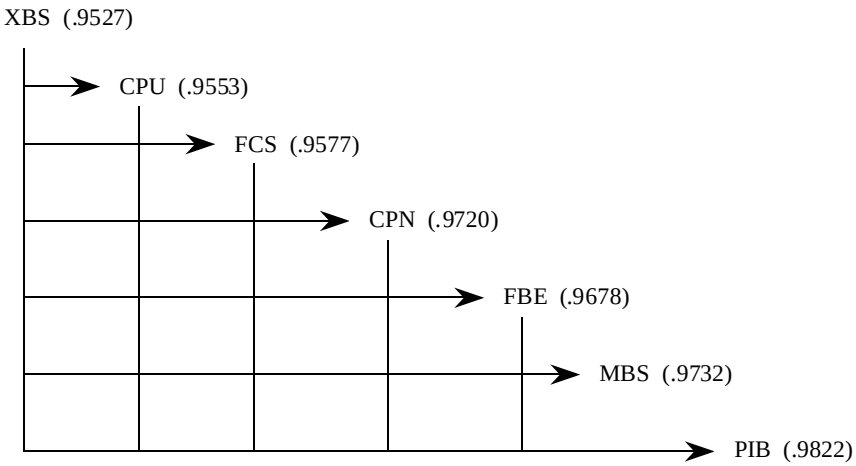
Table 4 contains the decomposition. The i, j element is the percentage of the variance of the prediction error of the variable i explained by an innovation in the variable j, in short term (up) and long term (down).
The main results from impulse-response analysis are presented in the table 5 and in graphs 6 to 12. They are:

- private consumption (CPN) is the main source of cyclical variability because of its strong effects on investment (specially in equipment), imports and gross domestic product,

- investment in equipment (FBE) and imports (MBS) are receivers if impulses because of its intense

CONTEMPORANEOUS INTERACTIONS

Adjusted R^2 in brackets



reactions to impulses in the rest of the system and the weak response of most of the variables to their innovations,

- investment in building (FCS) plays a very important role since it is a receiver (like FBE) and an emitter at the same time.

All the variables respond to its innovations (except public consumption and exports) and it reacts in shocks in the rest of the system (except public consumption and exports, again),

- public consumption (CPU) and exports (XBS) are the most idiosyncratic variables. The fist one is

Table 4: Decomposition of the variance of prediction errors, short (1 Q) and long (40 Q)

	PIB	CPN	CPU	FBE	FCS	XBS	MBS
PIB	77.22	2.60	2.23	6.50	6.38	1.55	3.52
	8.97	35.69	7.95	8.23	17.64	18.94	2.57
CPN	0.00	73.59	13.43	0.00	7.01	5.97	0.00
	5.25	40.56	8.92	3.57	13.54	26.31	1.86
CPU	0.00	0.00	100.00	0.00	0.00	0.00	0.00
	4.68	19.13	22.52	12.84	10.33	19.78	10.71
FBE	0.00	0.66	0.78	85.27	10.67	2.61	0.00
	6.87	10.06	4.07	27.86	36.01	9.56	5.57
FCS	0.00	0.00	0.01	0.00	96.15	3.85	0.00
	2.43	11.09	5.49	8.47	64.38	5.26	2.87
XBS	0.00	0.00	0.00	0.00	0.00	100.00	0.00
	12.60	4.57	13.49	19.34	7.48	38.05	4.47
MBS	0.00	4.88	5.02	8.00	15.16	20.84	46.11
	12.30	19.70	7.53	7.14	27.55	9.55	16.23

almost neutral in the sense that it is not affected by others and does not affect other variables. The second one is dominated by its own impulses and its shocks affects the rest of variables in a negative fashion, except gross domestic product. This is a strange result which deserves further analysis.

Table 5 illustrates the impulse-response analysis. The i, j element represents the maximum response (in absolute value) of i to an impulse in j. a '+'/'-' indicates the sign of the response: positive (movements in the same direction) or negative (movements in opposite directions). The number of signs denotes the intensity of the response: '+'/'-' weak (between 0 and 0.5), '++'/'--' medium (between 0.5 and 1) and '+++'/'---' strong (greater than 1). The quarter in which the maximum is located determines the delay of the response: short run (cp, between 1 and 4 quarters), medium run (mp, between 5 and 12 quarters) and long run (lp, more than 13 quarters).

5 Conclusions

The main results of the analysis are:

- Private consumption is a strongly procyclical serie. It is a very important source of aggregate fluctuations and its shocks play an important role in the evolution of investment, imports and gross domestic product.

- Public consumption is a weakly procyclical serie with low volatility and does not play a significant role in the explanation of the cyclical behaviour of the other variables.

- Investment in equipment and building are strongly and increasingly procyclical. Both are very volatile. The fist one receives impulses from consumption, aggregate output, imports and building. The second one receives impulses and send shocks to these variables. So, it plays an intermediate role in the explanation of the co-movements of the series.

- Exports are very idiosyncratic since do not share the common cycle of the rest of variables and affects them in a strange way.

- Imports are highly, very volatile and very sensitive to shocks in the whole system. They are a good summarising variable of the state of the economy.

Table 5: Impulse-Response Analysis

	PIB	CPN	CPU	FBE	FCS	XBS	MBS
PIB	+	++	+	+	++	+	+
	cp	mp	cp	mp	mp	mp	mp
CPN	+	++	+	+	++	-	+
	mp	cp	cp	mp	mp	cp	mp
CPU	+	+	++	+	+	-	0
	lp	lp	cp	mp	mp	mp	
FBE	++	+++	+++	+++	+++	-	+++
	mp	mp	cp	cp	mp	cp	mp
FCS	++	++	+	+++	+++	-	+
	cp	mp	cp	cp	cp	mp	cp
XBS	+++	++	--	--	-	+++	0
	mp	mp	mp	mp	lp	cp	
MBS	++	+++	+	+++	+++	-	+++
	mp	cp	cp	mp	mp	lp	cp

References

Akaike, H. (1982) "Likehood of a model and information criteria", Journal Econometrics, vol.20, n°1.

Box, G.E.P. y Tiao, G.C. (1977) "A canonical analysis of multiple time series", Biometrika, n°64, 2, pp. 355-365.

Box, G.E.P. y Tiao, G.C. (1981) "Multiple time series with applications", Journal of the American Statistical Association, vol. 76, pp. 802-816.

Box, G.E.P. y Tiao, G.C. (1982) "Introduction to applied multiple time series", Technical Report 582, University of Wisconsin, Madison, U.S.A.

Cristóbal, A. y Quilis, E.M. (1994) "Tasas de variación, filtros y análisis de la coyuntura", Boletín Trimestral de Coyuntura n°52, Instituto Nacional de Estadística.

Engle, R.F. y Granger, C.W.J. (1987) "Co-integration and error correction: representation, estimations and testing", Econometrica, n°55, pp. 251-276.

INE (1992) "Nota metodológica sobre la Contabilidad Nacional Trimestral", Boletín Trimestral de Coyuntura n°44, Instituto Nacional de Estadística.

INE (1993) "Contabilidad Nacional Trimestral de España. Metodología y serie trimestral 1970-1992", Instituto Nacional de Estadística.

Liu, L.M. (1986) "Multivariate time series analysis using vector ARMA models", SCA Working Paper.

Luenberger, D.G. (1979) "Introduction to dynamic systems", John Wiley, New York, U.S.A.

Lütkepohl, H. (1991) "Introduction to multiple time series analysis", Springer Verlag, Berlín, Alemania.

MacKinnon, J. (1990) "Critical values for cointegration tests", Working Paper, University of California, San Diego.

Melis, F. (1991) "La estimación del ritmo de variación en series económicas", Estadística Española, vol. 33, n°126.

Sims, Ch.A. (1980) "Macroeconomics and reality", Econometrica, vol.48, n°1, pp. 1-48.

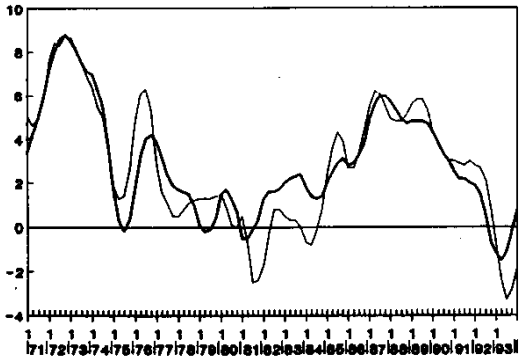
Sims, Ch.A. (1981) "An autoregressive index model for the U.S., 1948-75", en Kmenta, J. y Ramsey, J.B. (Eds) "Large scale macroeconometric models", North-Holland, Amsterdam, The Netherlands.

Tiao, G.C. y Tsay, R.S. (1989) "Model specification in multivariate time series", Journal of the Royal Statistical Society, series B, vol.51, n°2, pp.157-213.

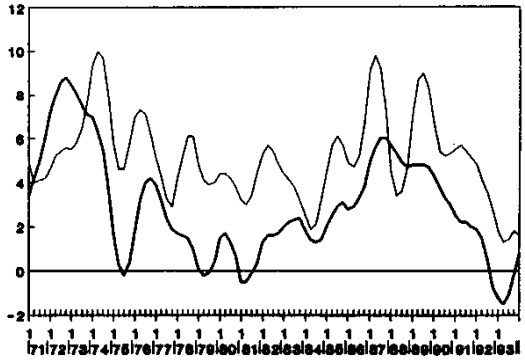
Tiao, G.C. y Tsay, R.S. y Wang, T. (1993) "Usefulness of linear transformation in multivariate time series analysis", Empirical Economics, pp. 567-595.

Graph 1: Annual rates of growth

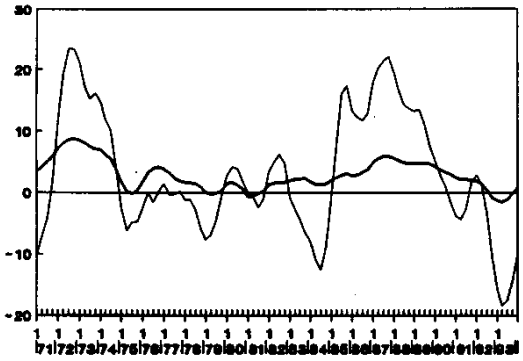
PIB & CPN



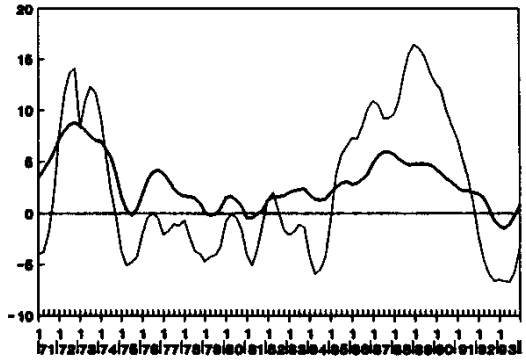
PIB & CPU



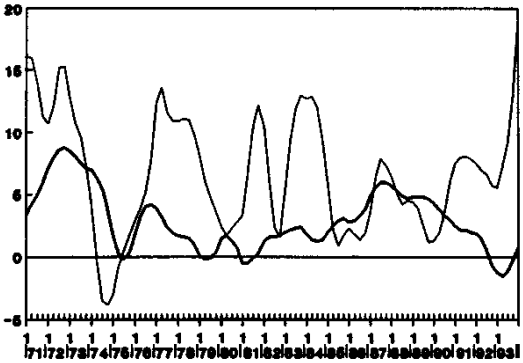
PIB & FBE



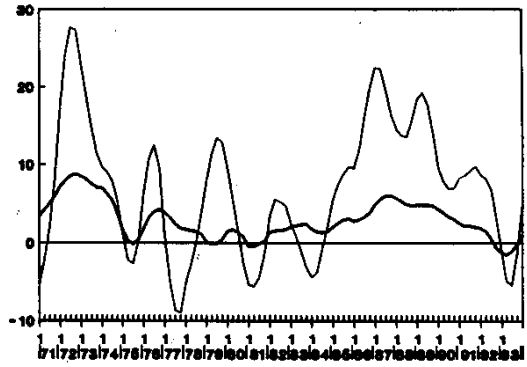
PIB & FCS



PIB & XBS

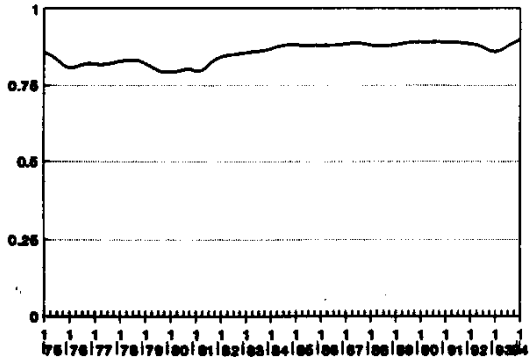


PIB & MBS

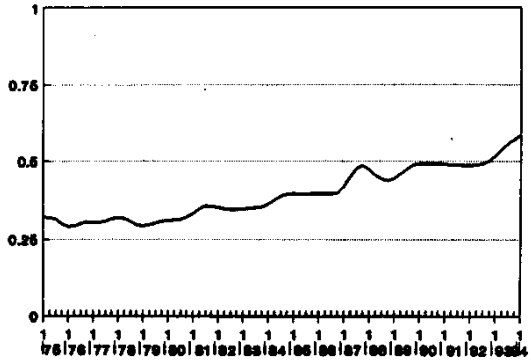


Graph 2: Recursive contemporaneous correlations with PIB

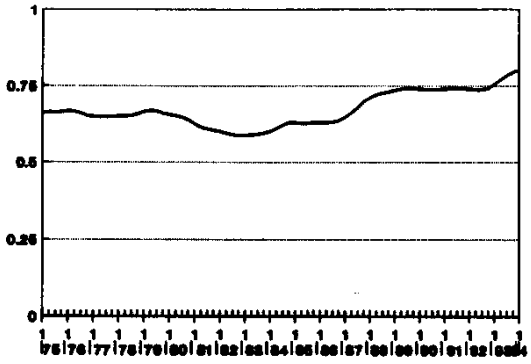
CPN



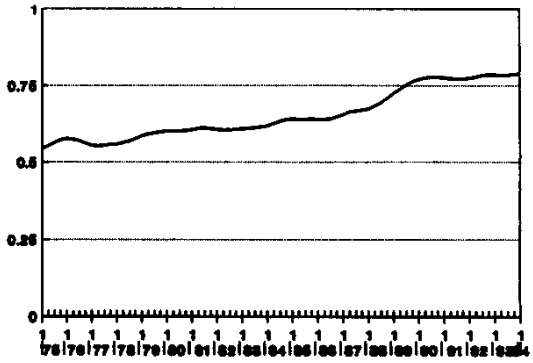
CPU



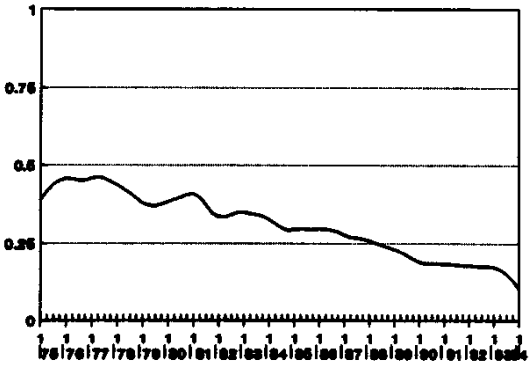
FBE



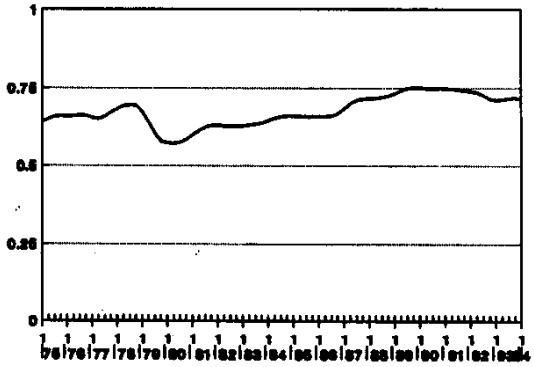
FCS



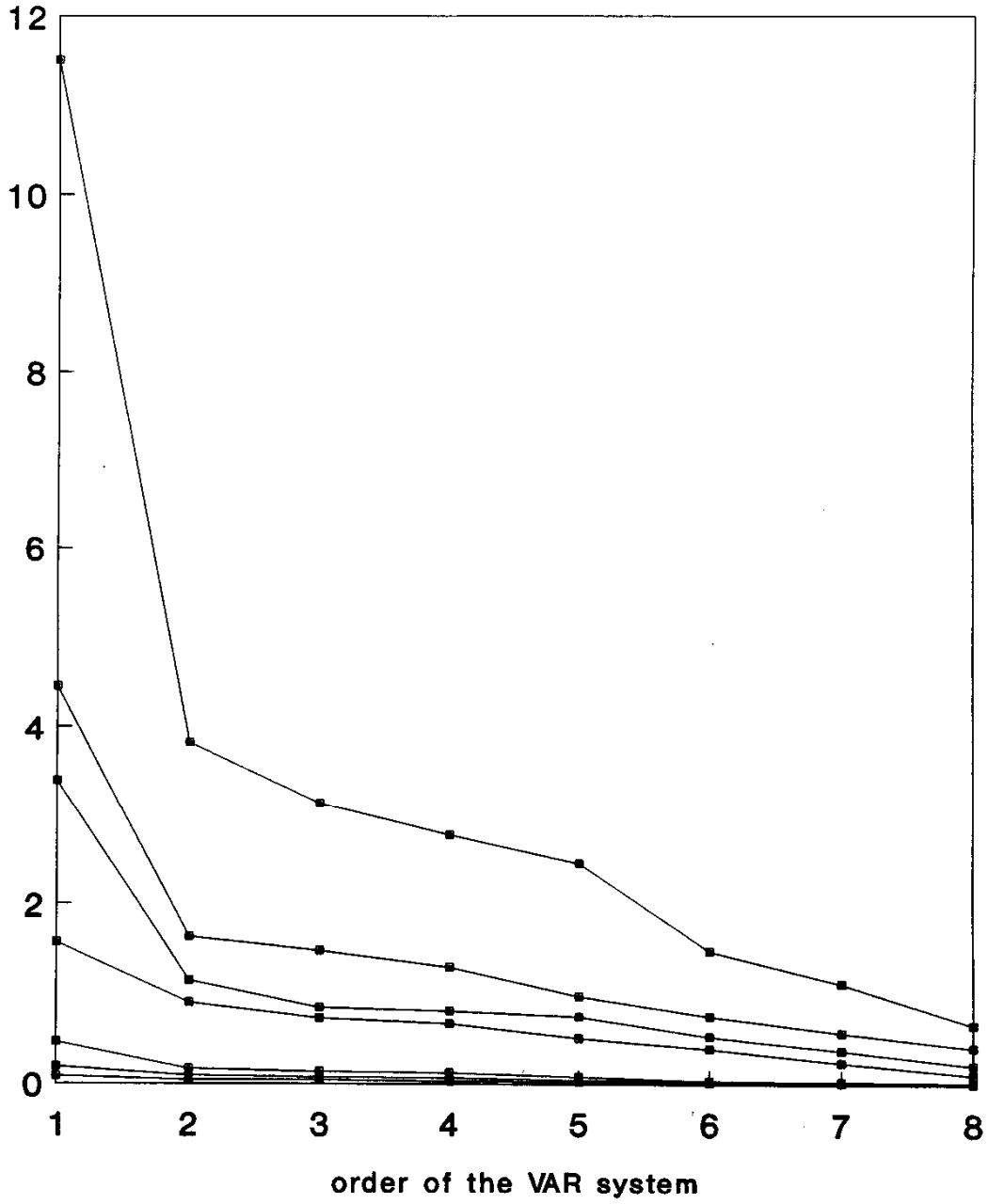
XBS



MBS

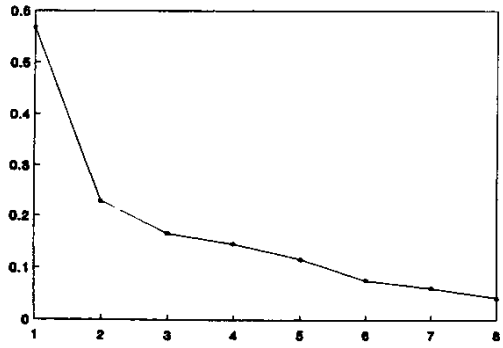


Graph 3: Autovalues of the residuals covariances matrix

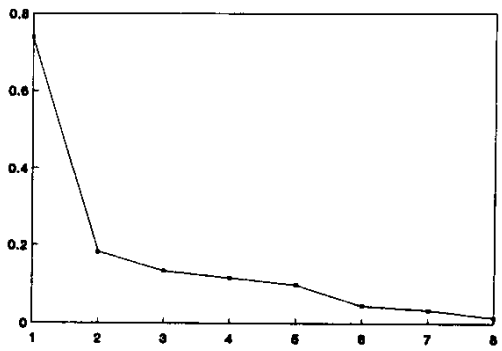


Graph 4: Residual variances of stepwise estimated VAR

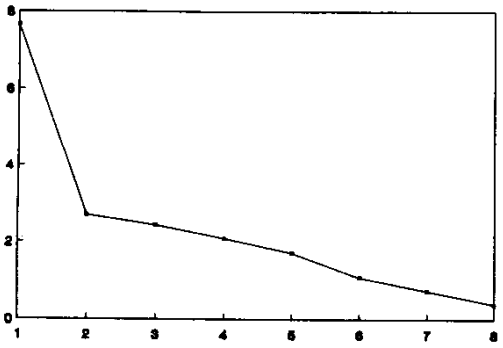
CPN



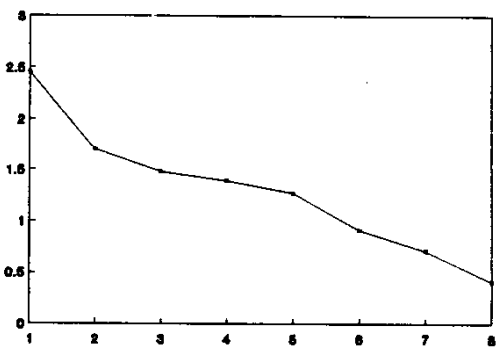
CPU



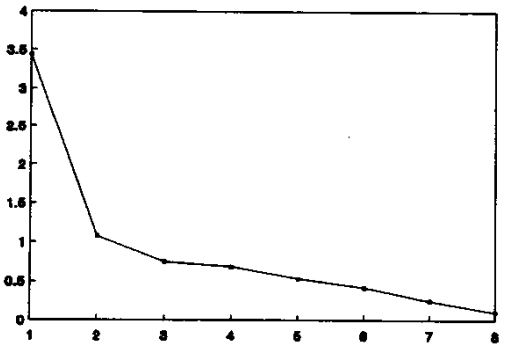
FBE



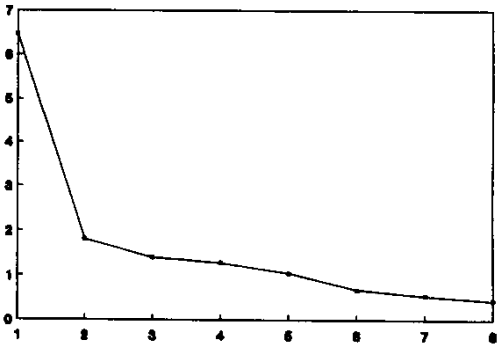
FCS



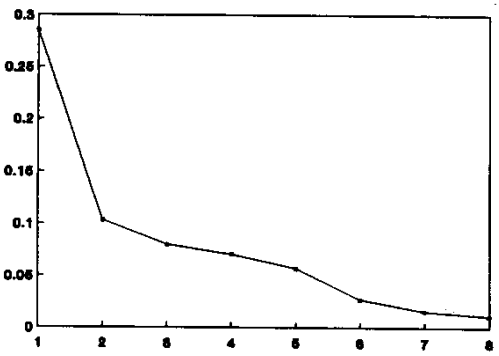
XBS



MBS

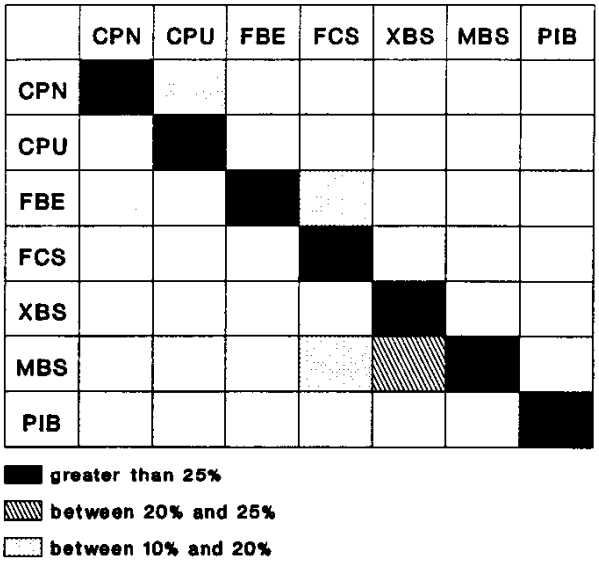


PIB

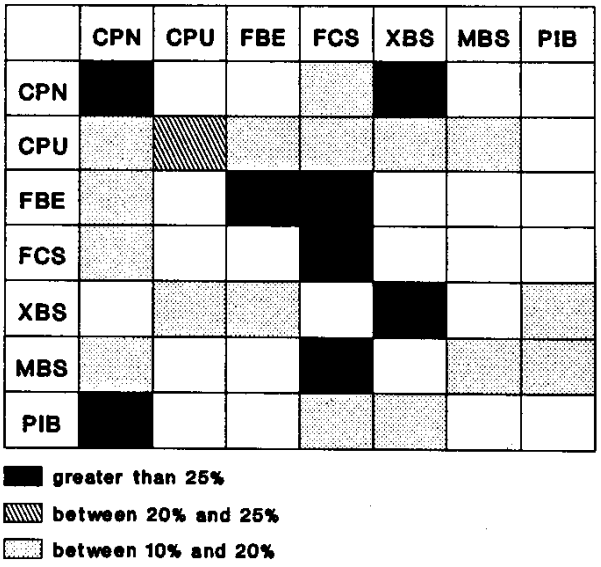


Graph 5: Variance decomposition of forecasting error

Short run (1 quarter)

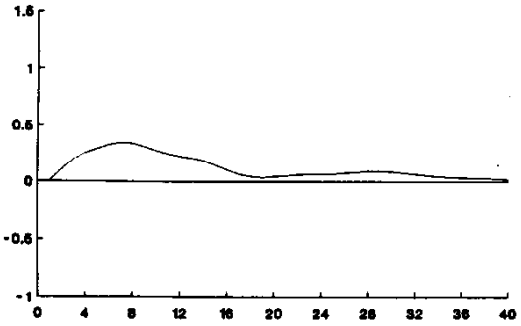


Long run (40 quarters)

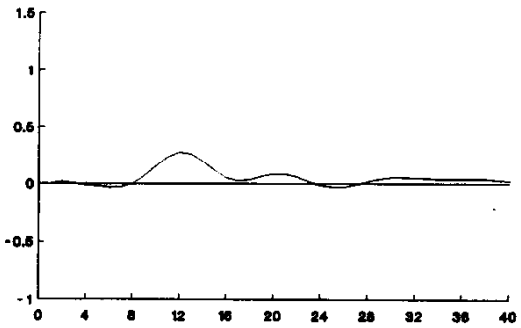


Graph 6: Response to a unit impulse in PIB

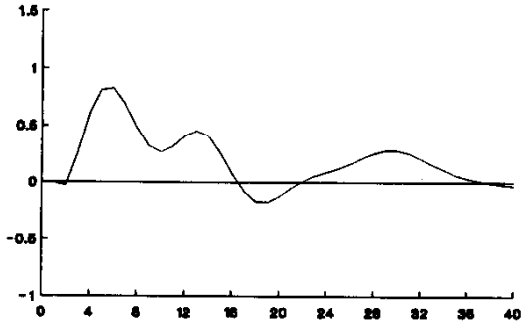
CPN



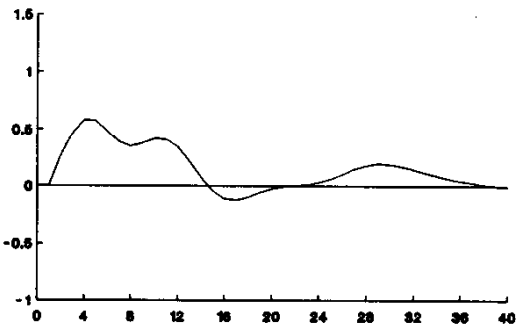
CPU



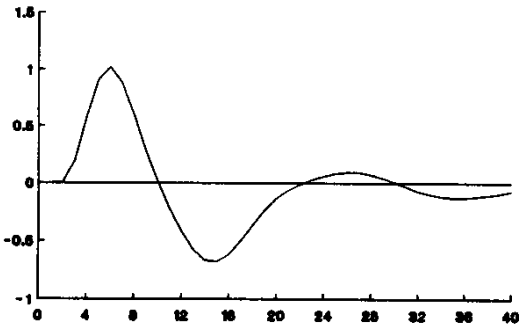
FBE



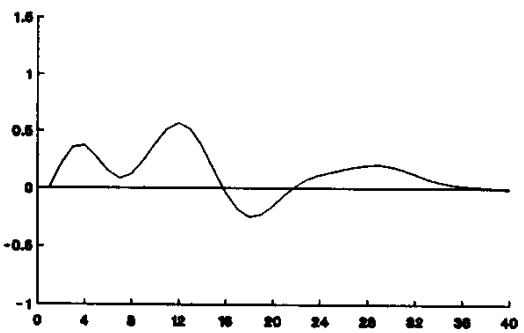
FCS



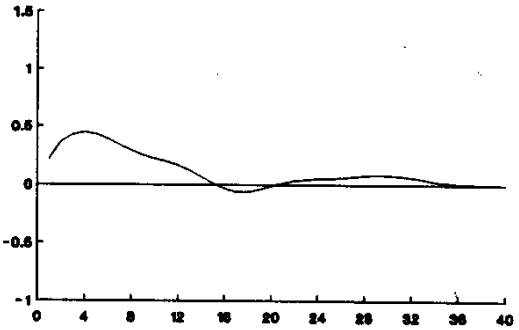
XBS



MBS

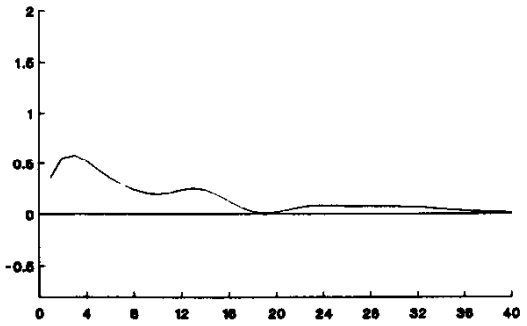


PIB

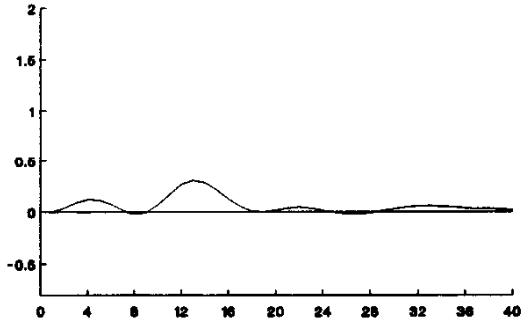


Graph 7: Response to a unit impulse in CPN

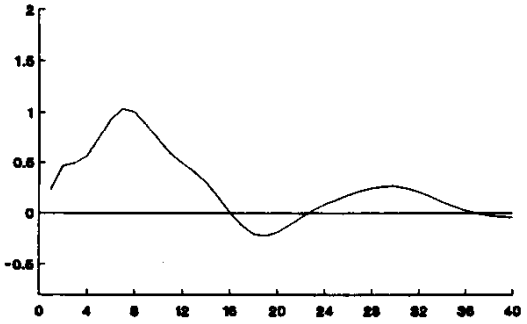
CPN



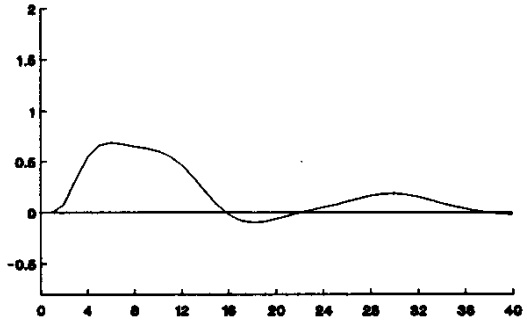
CPU



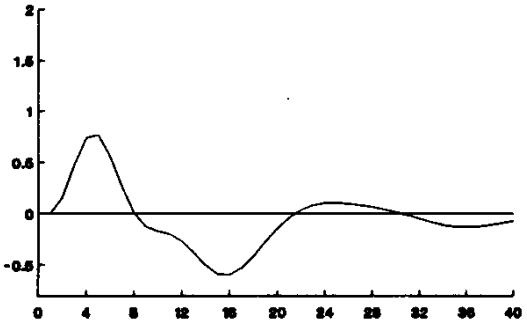
FBE



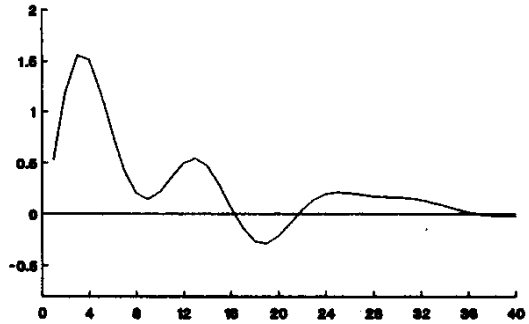
FCS



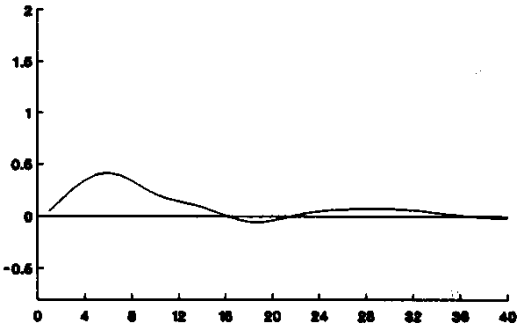
XBS



MBS

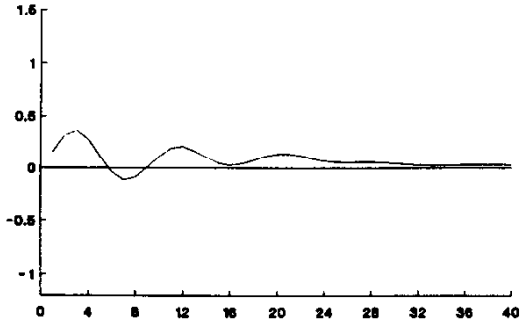


PIB

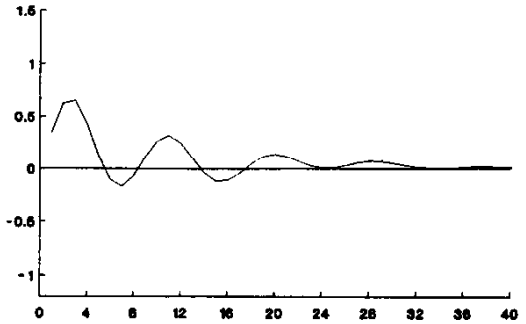


Graph 8: Response to a unit impulse in CPU

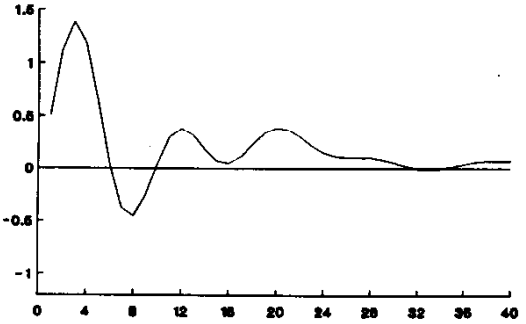
CPN



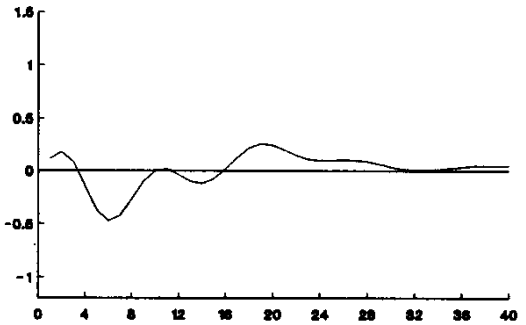
CPU



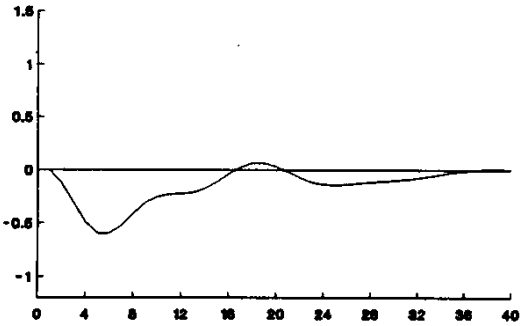
FBE



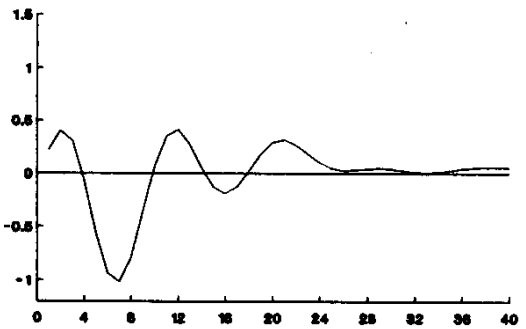
FCS



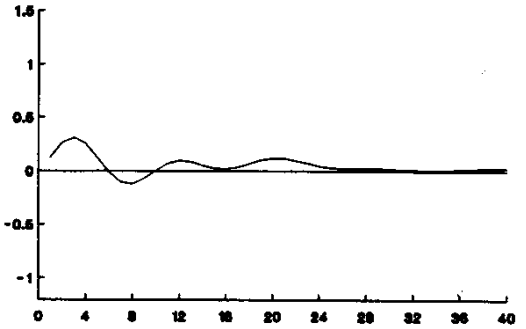
XBS



MBS

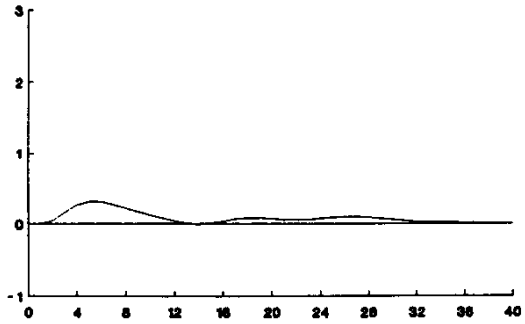


PIB

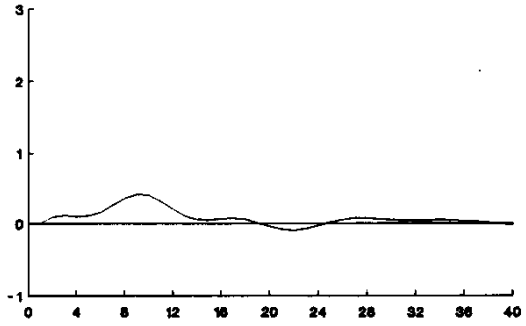


Graph 9: Response to a unit impulse in FBE

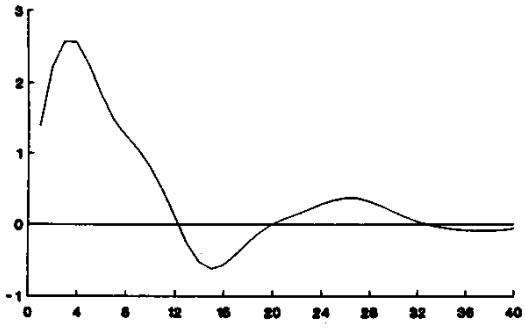
CPN



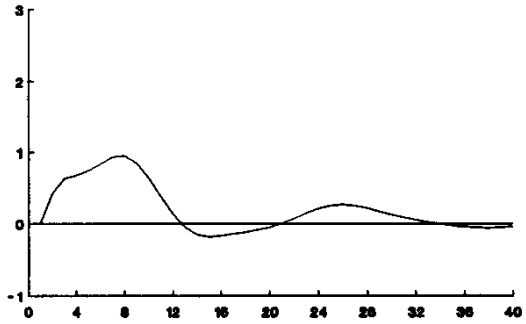
CPU



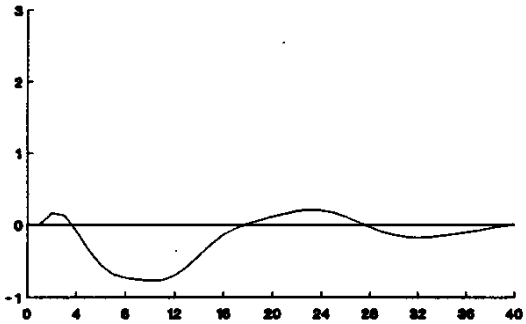
FBE



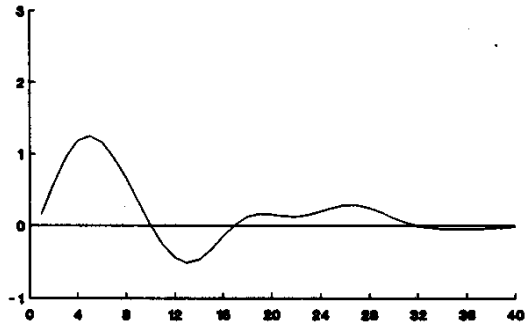
FCS



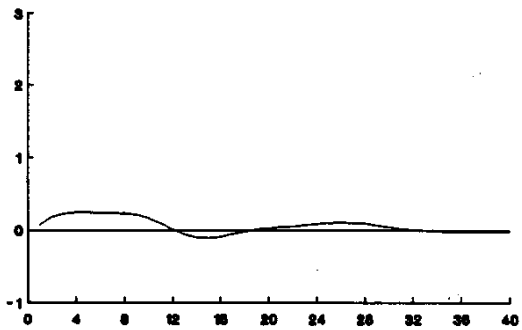
XBS



MBS

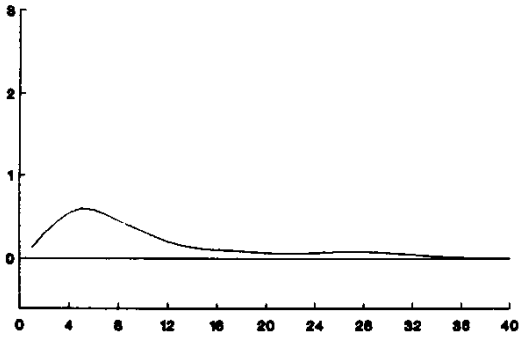


PIB

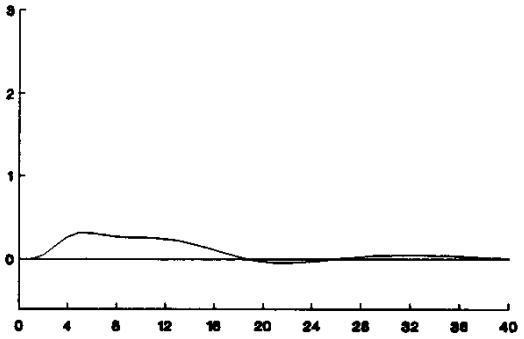


Graph 10: Response to a unit impulse in FCS

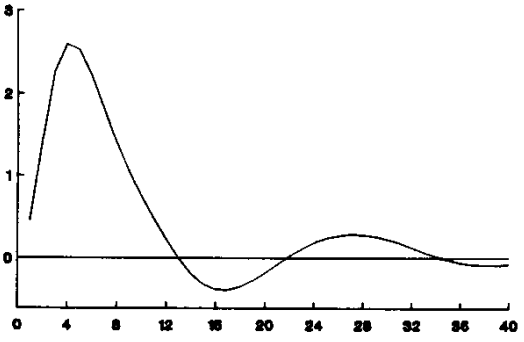
CPN



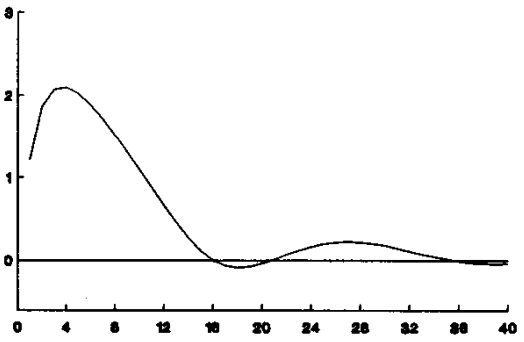
CPU



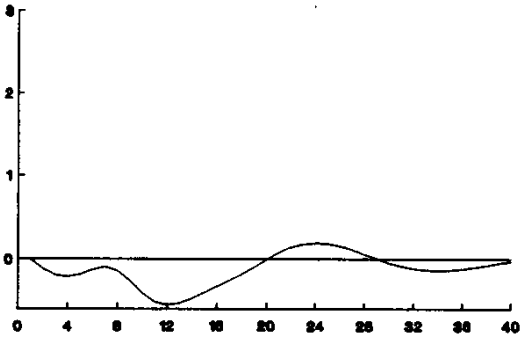
FBE



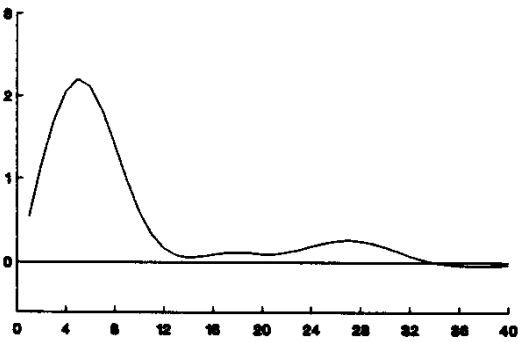
FCS



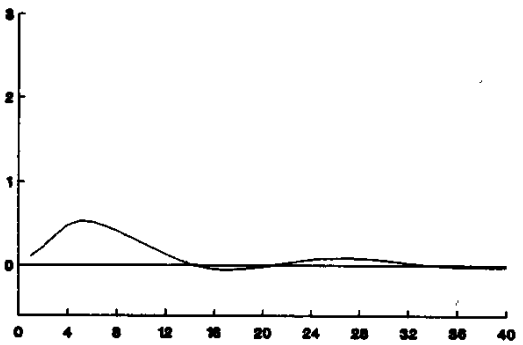
XBS



MBS

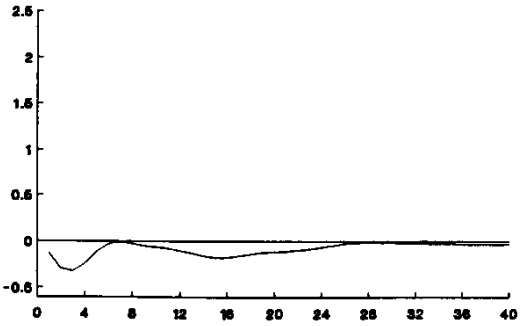


PIB

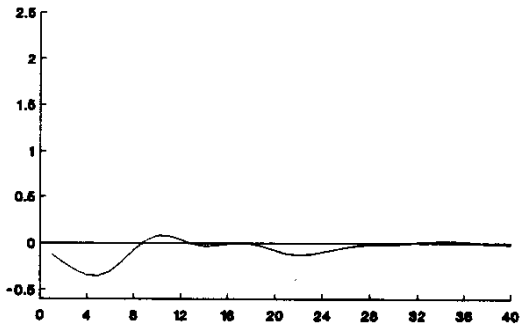


Graph 11: Response to a unit impulse in XBS

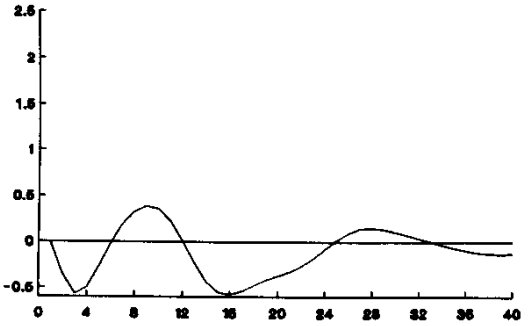
CPN



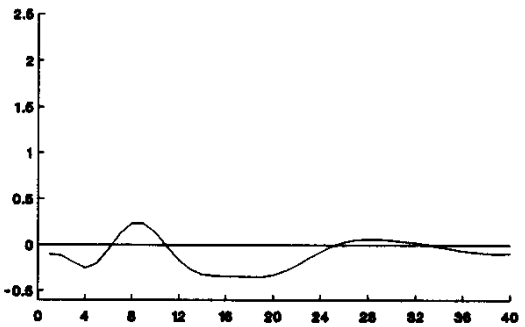
CPU



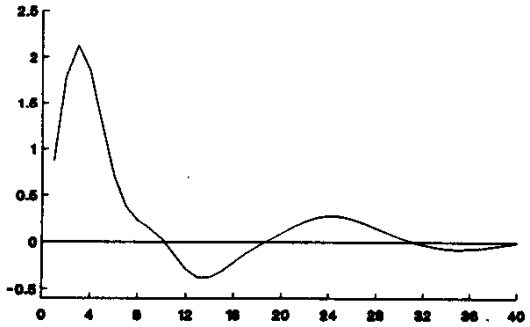
FBE



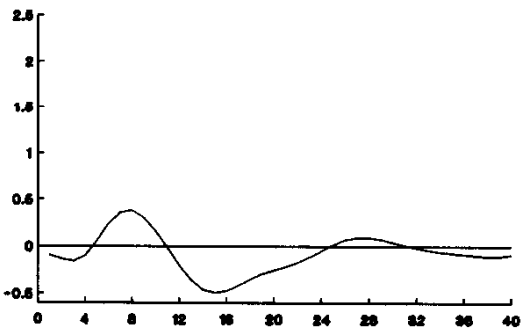
FCS



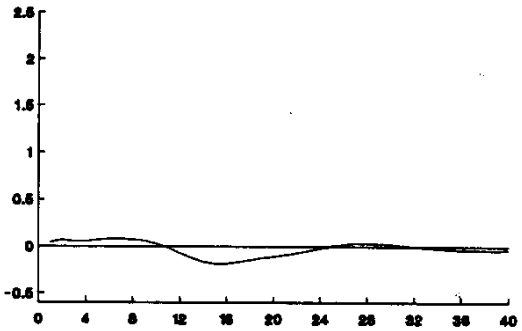
XBS



MBS

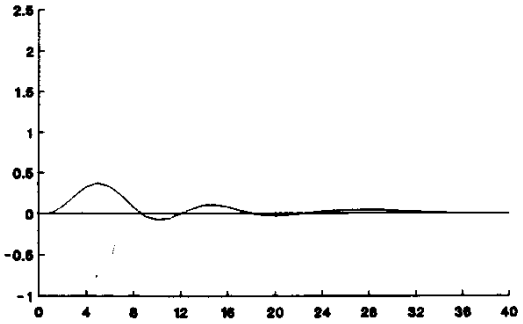


PIB

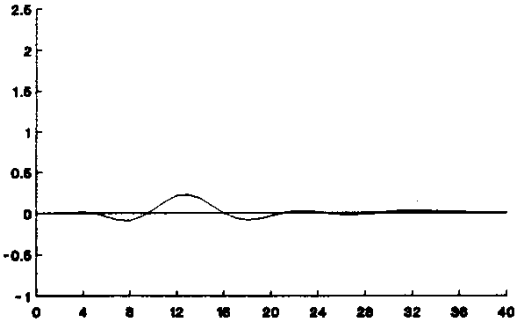


Graph 12: Response to a unit impulse in MBS

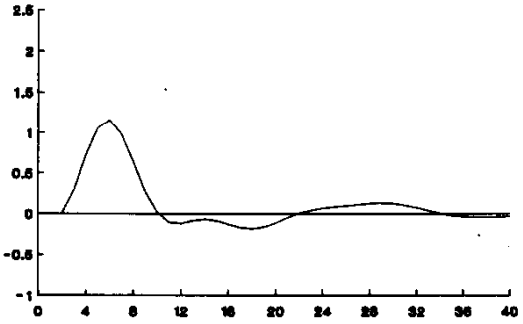
CPN



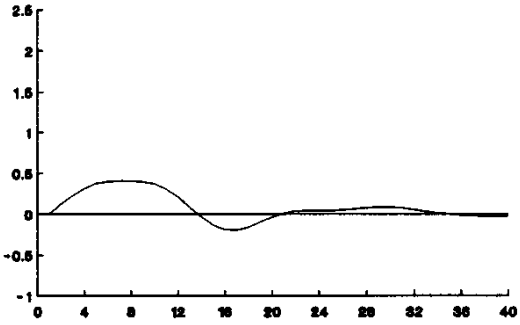
CPU



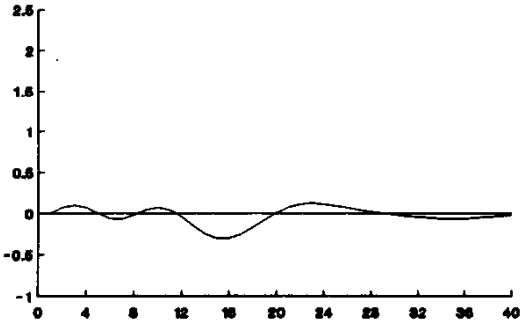
FBE



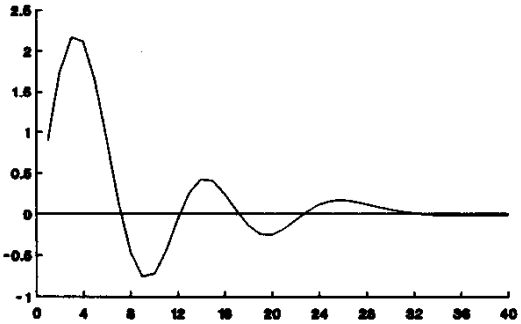
FCS



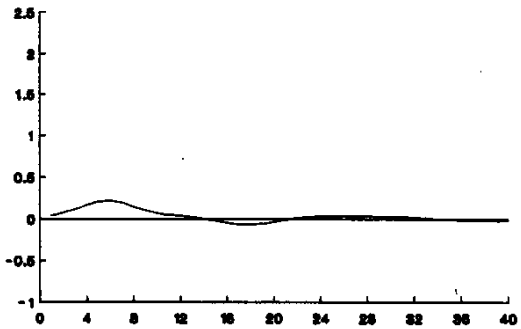
XBS



MBS



PIB



SECTION 2 -

NEW DEVELOPMENTS IN THE ECONOMETRICS-BASED APPROACH

Temporal disaggregation of a system of time series when the aggregate is known: Optimal VS. adjustment methods

Tommaso DI FONZO
Dipartimento di Scienze Statistiche
Università di Padova

1 Introduction

Most of the data obtained by statistical agencies have to be adjusted, corrected or somehow processed by statisticians in order to arrive at useful, consistent and publishable values.

A problem often faced by government agencies that collect and publish quarterly time series is that of obtaining subannual data that simultaneously comply with the relevant annual figures and satisfy accounting constraints. From a formal point of view, one is interested in estimating a number of individual variables, which are aggregated over units and over time.

We deal with a number of statistical procedures developed to solve this problem and available in ECOTRIM (Barcellan, 1994), a program for temporal disaggregating one or more time series. Much emphasis is placed on technical characteristics and on operational issues, in order to highlight distinctive features of the various procedures at disposal.

The subject is a classical missing data problem solved in an indirect estimation framework: given M high-frequency (say, quarterly) time series (z), in this paper we consider two distinct situations:

1. M Preliminary quarterly time series, p_j, $j = 1, \mathrm{K}, M$, are available, where $\sum_{j=1}^{M} p_j \neq z$ and/or p_j doesn't comply with y_{0j};
2. a set of quarterly related indicators is used to obtain indirect estimates of the M unknown time series.

It should be noted that the distinction is not necessarily as strict as it seems, in that preliminary quarterly series could have been individually obtained by using related indicators.

Benchmarking methods that generalize the work by Denton (1971) can be used in both situations (Cholette, 1988): the preliminary estimates are adjusted so that they comply with more reliable and separately obtained aggregated measurements [1]. Due to the nature of the paper, essentially developed as a technical support to ECOTRIM, we don't deal with adjustment issues according to some data reliability indicator (see Stone *et al.*, 1942, for the seminal idea). There is no doubt, however, that recent contributions about linear least squares adjustment of economic data subject to accounting constraints (see, among others, Weale, 1992, Solomou and Weale, 1993) could reveal very useful for the problem in hand.

1 Indeed, the adjustment procedures supported by ECOTRIM refer to a less general concept of benchmark ing than the one of Cholette (1987, p.14): "Benchmarking is the process of optimally combining the orig inal sub-annual series with the annual benchmarks and with the sub-annual benchmarks, in order to obtain a more reliable sub-annual series and a more reliable annual series".

As far as point 2. is concerned, assuming that each basic series satisfies a multiple regression relationship with a number of known related indicators, an optimal (in least squares sense) solution, consistent with the aggregation constraints, can be obtained (Di Fonzo, 1990). We develop this approach discussing some specific error covariance pattern.

In order to appreciate how the different procedures operate, the paper ends with a stylized empirical application to a real-data problem coming from Italian Quarterly National Accounts.

2 Terms of the problem

We deal with an indirect estimation problem involving a system of variables rather than a single one [2]. More precisely, we wish to estimate M unknown $(n\times 1)$ vectors of high-frequency data, each pertaining to M basic (i.e., disaggregate) time series which have to satisfy both contemporaneous and temporal aggregation constraints.

In order to solve this problem, we consider a number of procedures that can be characterized as either *pure adjustment* or *optimal* (in least squares sense). The information basis common to both cases is given by the following $M+1$aggregated vectors [3]:

- z, $(n\times 1)$ vector of contemporaneous aggregated data;
- y_{oj}, $j=1,\mathrm{K},M$, $(N\times 1)$ vectors of temporally aggregated data.

Denoting by y_j, $j=1,\mathrm{K},M$, the $(n\times 1)$ vectors to be estimated, the following accounting constraints hold:

$$\sum_{j=1}^{M} y_j = z, \qquad (1)$$

$$Dy_j = y_{0j},\ j=1,\mathrm{K},M, \qquad (2)$$

where **D** is the $(N\times n)$ aggregation matrix converting high-frequency in low frequency data. Each element of y_{0j} can be viewed as a non overlapping linear combination of y_j, with coefficients given by the $(m\times 1)$ vector c, m being the aggregation order. Thus, in general matrix **D** is equal to

$$D=[c'\otimes I_N \mathrm{M}\ \ 0], \qquad (3)$$

where **0** is a null $(N\times h)$ matrix[4], h=n-mN.

A graphical overview of such a data configuration [5] can be found in table 1.

Constraint (1) can be re-written in the following way:

$$(1_M\otimes \mathrm{I}_n)\,y = z, \qquad (4)$$

where 1_M is an $(M\times 1)$ vector of ones and

$$y=\begin{bmatrix} y_1 \\ \mathrm{M} \\ y_j \\ \mathrm{M} \\ y_M \end{bmatrix}$$

As far as temporal aggregation constraints (2) are concerned, in a more compact form we have

$$(\mathrm{I}_M\otimes D)\,y = y_0, \qquad (5)$$

where

$$y_0=\begin{bmatrix} y_{01} \\ \mathrm{M} \\ y_{0j} \\ \mathrm{M} \\ y_{0M} \end{bmatrix}.$$

2 See Di Fonzo (1987) for a review of univariate procedures.

3 The notation is as close as possible to that of Barcellan (1994).

4 The presence of the null matrix in **D** permits to deal with extrapolation. Obviously, when n=mN, $D = c'\otimes \mathrm{I}_N$

5 The available data are represented in italic font, whereas for the unknown $y_{j,t}$'s the normal font is used.

Note that the contemporaneous aggregation of temporally aggregated series implies

$$\sum_{j=1}^{M} y_{0j,T} = \sum_{k=1}^{M} z_{m(T-1)+k} = z_{0,T}, \quad T = 1, \mathsf{K}, N,$$

or, in matrix form,

$$\left(1_M' \otimes \mathrm{I}_n\right) y_0 = D^* z, \qquad (6)$$

where D^* is the $(N \times n)$ aggregation matrix (3) where $c = 1_m$. Let **H** be the following $[(n+NM) \times nM]$ aggregation matrix,

$$H = \begin{bmatrix} 1_M' \otimes \mathrm{I}_n \\ \mathrm{I}_M \otimes D \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix},$$

and y_a the $[(n+NM) \times 1]$ vector

$$y_a = \begin{bmatrix} z \\ y_0 \end{bmatrix}.$$

The complete set of constraints between the unknown values and the available aggregated information can be expressed in matrix form as

$$Hy = y_a \qquad (7)$$

that is:

$$H_1 y = z, \qquad H_2 y = y_0.$$

It has to be noted that, given the relationship (6) between the $M+1$ aggregated vectors, matrix **H** has rank $r = n + N(M-1)$, N aggregated observations being redundant.

3 Multivariate adjustment procedures

Suppose we have M preliminary series, p_j, $j = 1, \mathsf{K}, M$, that need to be adjusted in order to satisfy the accounting constraints. This has to be accomplished by distributing the discrepancies

$$z - H_1 p, \qquad y_0 - H_2 p,$$

Tab. 1: The elements of a multivariate indirect estimation problem

Year T	Quarter k	$y_{1,t}$	$y_{01,T}$	L	$y_{j,t}$	$y_{0j,T}$	L	$y_{M,t}$	$y_{0M,T}$	z_t	$Z_{0,T}$
	1	$y_{1,1}$		L	y_{j1}		L	$y_{M,1}$		z_1	
	2	$y_{1,2}$		L	$y_{j,2}$		L	$y_{M,2}$		z_2	
	3	$y_{1,3}$		L	$y_{j,3}$		L	$y_{M,3}$		z_3	
	4	$y_{1,4}$		L	$y_{j,4}$		L	$y_{M,4}$		z_4	
1			$y_{01,1}$	L		y_{0j1}	L		$y_{0M,1}$		$z_{0,1}$
	L	L			L			L		L	
	1	$y_{1,4T-3}$		L	$y_{j,4T-3}$		L	$y_{M,4T-3}$		z_{4T-3}	
	2	$y_{1,4T-2}$		L	$y_{j,4T-2}$		L	$y_{M,4T-2}$		z_{4T-2}	
	3	$y_{1,4T-1}$		L	$y_{j,4T-1}$		L	$y_{M,4T-1}$		z_{4T-1}	
	4	$y_{1,4T}$		L	$y_{j,4T}$		L	$y_{M,4T}$		z_{4T}	
T			$y_{01,T}$	L		$y_{0j,T}$	L		$y_{0M,T}$		$z_{0,T}$
	L	L			L			L		L	
	1	$y_{1,4N-3}$		L	$y_{j,4N-3}$		L	$y_{M,4N-3}$			
	2	$y_{1,4N-2}$		L	$y_{j,4N-2}$		L	$y_{M,4N-2}$		z_{4N-3}	
	3	$y_{1,4N-1}$		L	$y_{j,4N-1}$		L	$y_{M,4N-1}$		z_{4N-2}	
	4	$y_{1,4N}$		L	$y_{j,4N}$		L	$y_{M,4N}$		z_{4N-1}	
N			$y_{01,N}$	L		$y_{0j,N}$	L		$y_{0M,N}$	z_{4N}	$z_{0,N}$
	1	$y_{1,n-1}$		L	$y_{j,n-1}$		L	$y_{M,n-1}$		z_{n-1}	
	2	$y_{1,n}$		L	$y_{j,n}$		L	$y_{M,n}$		z_n	

where

$$P = \begin{bmatrix} P_1 \\ \mathsf{M} \\ P_j \\ \mathsf{M} \\ P_M \end{bmatrix},$$

according to some reasonable criterion. In this section we deal with two multivariate adjustment procedures:

- proportional adjustment;
- Denton's multivariate adjustment.

The former is a very simple and widely used technique, while the latter generalizes the univariate procedure worked out by Denton (1971) by taking into account some technical devices about (i) the treatment of starting values (Cholette, 1984, 1988) and (ii) the nature of the accounting constraints.

As we shall see, proportional adjustment is less generally applicable than Denton's one, because it assumes that only the contemporaneous aggregation constraints must be fulfilled (that is, $Dp_j = y_{0j}$, $j = 1, \mathsf{K}, M$).

3.1 The proportional adjustment

A simple and fairly reasonable way to eliminate the discrepancy between a contemporaneously aggregated value and the corresponding sum of disaggregated preliminary estimates consists in distributing such discrepancy according to the weight of each single temporally aggregated series with respect to the contemporaneously aggregated one. Thus, let

$$w_{j,T} = \frac{y_{0,T}}{z_{0,T}}, \quad T = 1, \mathsf{K}, N$$

be a set of cefficients such that $\sum_{j=1}^{M} wj, T = 1$. If $Dp_j = y_{0j}$, adjusted estimates are given by [6]:

$$\hat{y}_{j,m(T-1)+k} = p_{j,m(T-1)+k} + w_{j,T}\left[z_{m(T-1)+k} - \sum_{r=1}^{M} p_{r,m(T-1)+k}\right],$$

that is

$$\hat{y}_{j,m(T-1)+k} = p_{j,m(T-1)+k} + w_{j,T}\, d_{m(T-1)+k},$$

$$\begin{cases} k = 1, \mathsf{K}, m \\ T = 1, \mathsf{K}, N. \\ j = 1, \mathsf{K}, M \end{cases} \qquad (8)$$

It's immediate to recognize that

$$\sum_{j=1}^{M} \hat{y}_{j,m(T-1)+k} = \sum_{j=1}^{M} p_{j,m(T-1)+k} + d_{m(T-1)+k} \sum_{j=1}^{M} w_{j,T} = z_{m(T-1)+k}$$

and

$$\sum_{j=1}^{M} \hat{y}_{j,m(T-1)+k} = \sum_{j=1}^{M} p_{j,m(T-1)+k} + w_{j,T}\left(\sum_{k=1}^{m} z_{m(T-1)+k} - \sum_{k=1}^{m}\sum_{j=1}^{M} p_{j,m(T-1)+k}\right),$$

that is

$$\sum_{j=1}^{M} \hat{y}_{j,m(T-1)+k} = y_{0j,T} + w_{j,T}\left(z_{0,T} - z_{0,T}\right) = y_{0j,T},$$

The adjusted estimates can be expressed in matrix form as follows:

$$\hat{y} = p + \left\{\begin{bmatrix} diag(y_{01}) \\ diag(y_{02}) \\ \mathsf{M} \\ diag(y_{0M}) \end{bmatrix}\left[diag(z_0)\right]^{-1} \otimes \mathrm{I}_m\right\}\left(z - \sum_{j=1}^{M} p_j\right)$$

If n>mN, disaggregated figures that comply only with the contemporaneous constraints are to be estimated. In this case the discrepancy can be modulated according to the relative weight of each preliminary series, that is

$$w_{j,r} = \frac{Pj, r}{\sum_{q=1}^{M} P_{q,r}}, \qquad j = 1, \mathsf{K}, M,$$

$$r = mN + 1, \mathsf{K}, n,$$

6 We assume n=mN

with $\sum_{j=1}^{M} w_{j,r} = 1, \quad r = mN+1, \mathrm{K}, n$ and

$$\hat{y}_{j,r} = p_{j,r} + w_{j,r}\left[z_r - \sum_{q=1}^{M} p_{q,r} \right], \qquad j = 1, \mathrm{K}, M, \qquad r = mN+1, \mathrm{K}, n.$$

3.2. The Denton's multivariate adjustment

Denton's multivariate procedure is a straightforward extension of the univariate counterpart that, as it's well known, is based on a *movement preservation principle* (Cholette, 1984). In general, like any other least squares technique, the adjusted estimates are found by solving the following quadratic loss function minimization:

$$\min_{y} (y-p)' M (y-p) \tag{9}$$

subject to the contemporaneous and temporal aggregation constraints (7). These constraints being linearly dependent, particular attention has to be given to the rank of the matrices involved in the procedure. The solution to problem (9) is actually given by

$$\hat{y} = p + M^{-1}H'\left(HM^{-1}H'\right)^{-}(y_a - Hp), \tag{10}$$

where $\left(HM^{-1}H'\right)^{-}$ denotes the Moore-Penrose generalized inverse of $\left(HM^{-1}H'\right)$. A solution, equivalent to (10) and which does not involve singular matrices to be inverted, can be expressed in terms of $n + m(N-1)$ "free" observations (see the Appendix).

In practicce the choice of **M** is restricted to the matrices

- $\mathrm{I}_M \otimes \Delta_{1,n}\Delta_{1,n}$
  Denton additive first difference (AFD)
- $\mathrm{I}_M \otimes \Delta_{1,n}\Delta_{1,n}\Delta_{1,n}\Delta_{1,n}$
  Denton additive second difference (ASD)
- $P^{-1}\left(\mathrm{I}_M \otimes \Delta_{1,n}\Delta_{1,n}\right)P^{-1}$
  Denton proportional first difference (PFD)
- $P^{-1}\left(\mathrm{I}_M \otimes \Delta_{1,n}\Delta_{1,n}\Delta_{1,n}\Delta_{1,n}\right)P^{-1}$
  Denton proportional first difference (PSD)
- $(S \otimes \mathrm{I}n)^{-1}$
  Rossi (1982) multivariate adjustment

where $\Delta_{1,n}$ is the following $(n \times n)$ matrix[7] performingfirst differences:

$$\Delta_{1,n} = \begin{bmatrix} 1 & 0 & 0 & \mathrm{L} & 0 & 0 \\ -1 & 1 & 0 & \mathrm{L} & 0 & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{M} \\ 0 & 0 & 0 & \mathrm{L} & -1 & 1 \end{bmatrix}$$

S is a positive definite $(M \times M)$ matrix and

$$P = \begin{bmatrix} diag(p_1) & 0 & \mathrm{L} & 0 & \mathrm{L} & 0 \\ 0 & diag(p_2) & \mathrm{L} & 0 & \mathrm{L} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \mathrm{L} & diag(p_j) & \mathrm{L} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \mathrm{L} & 0 & \mathrm{L} & diag(p_M) \end{bmatrix}$$

Both additive and proportional variants of Denton's procedure operate according a movement preservation principle focusing on:

- the simple period-to-period change in the additive case;
- the period-to-period percentage change in the proportional case.

Finally, it should be noted that, in order to perform correctly, Rossi's procedure needs that the preliminary estimates comply with the temporal aggregation constraints. In this case (Di Fonzo, 1987, 1990), the adjusted estimates are given by

7 For notational convenience, we present the original difference matrix used by Denton (1971). As Cho lette (1984, 1987) pointed out, the computational reasons behind this specification, which involves $p_{0,j} - y_{0,j} = 0$ and $p_{0,j} - y_{0,j} = p_{-1,j} - y_{-1,j} = 0$ for the FD and SD variants, respectively, have become obsolete. ECOTRIM is currently being upgraded to correctly handle starting values.

$$\hat{y} = p + s^{-1}\left(S1_M \otimes \mathrm{I}_n\right)\left(z - \sum_{j=1}^{M} p_j\right)$$

with $s = 1_M' S 1_M$.

4 The BLUE approach

If we have at disposal a set of high-frequency indicators related to the unknown disaggregated series, we may set up M regression models

$$y_j = X_j\beta_j + u_j, \qquad (12)$$

with

$$\mathrm{E}(u_j) = 0, \qquad \mathrm{E}(u_i u_j') = V_{ij}, \qquad i, j = 1, \ldots, M,$$

where X_j, $j = 1, \ldots, M$ are $(n \times p_j)$ matrices of related series. Models (12) can be grouped as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & 0 & \vdots \\ 0 & 0 & \ldots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}$$

or

$$y = X\beta + u, \qquad (13)$$

where $\mathrm{E}(u) = 0$ and $\mathrm{E}(uu') = V = \{V_{ij}\}$. Pre-multiplying (13) by the aggregation matrix **H** we obtain the observed, aggregated regression model:

$$y_a = X_a\beta + u_a,$$

where $X_a = HX$. The aggregated disturbance vector, $u_a = Hu$, has zero mean and singular covariance matrix $\mathrm{E}(u_a u_a') = V_a = HVH'$.

The high-frequency estimates can be obtained as a solution of a linear prediction problem in a generalized regression model with singular covariance matrix (Di Fonzo, 1990):

$$\hat{y} = X\hat{\beta} + L\left(y_a - X_a\hat{\beta}\right), \qquad (14)$$

$$\hat{\beta} = \left(X_a' V_a^{-} X_a\right)^{-1} X_a' V_a^{-} y_a, \qquad (15)$$

with $L = VH'V_a^{-}$. As it's clear, the estimator (14) is a natural extension of the optimal univariate counterpart worked out by Chow and Lin (1971).

The estimation errors covariance matrix,

$$\mathrm{E}\left[(\hat{y} - y)(\hat{y} - y)'\right] = (\mathrm{I}_n - LH)V + (X - LX_a)\left(X_a' V_a^{-} X_a\right)^{-1}(X - LX_a)'$$

Depends on two components: the former is only related to **H** and **V**, the latter is a systematic one and rises with $(X - LX_a)$ (Bournay and Laroque, 1979).

In most practical problems matrices V_{ij} are unknown and must, therefore, be estimated according to proper assumptions on the u_j's. Two cases seem to be interesting from both a theoretical and a computational point of view.

- **Multivariate white noise**

We assume

$$\mathrm{E}(u_i u_j') = \sigma_{ij}\mathrm{I}_n, \qquad i, j = 1, \ldots, M$$
$$\mathrm{E}(uu') = \Sigma \otimes \mathrm{I}_n, \qquad \Sigma = \{\sigma_{ij}\}.$$

The elements of Σ can be estimated using the OLS residuals of the temporally aggregated regressions $y_{0j} = X_{0j}\beta_j + u_{0j}$. Furthermore, in this case the inversion of V_a can be notably semplified [8].

- **Multivariate random walk**

A straightforward generalization of the univariate approach of Fernández (1981) is the following:

$$u_t = u_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n,$$

$$u_0 = 0, \qquad \mathrm{E}(\varepsilon_t) = 0,$$

$$\mathrm{E}(\varepsilon_l \varepsilon_s') = \begin{cases} 0 & \text{if } l \neq s \\ \Sigma & \text{if } l = s \end{cases} \qquad r, s = 1, \ldots, n,$$

where u_t and ε_t are $(M \times 1)$ contemporaneous disturbances vector. These assumptions imply

8 By suitable partition of V_a only an $\left[(M-1)\times(M-1)\right]$ matrix needs to be inverted (see Di Fonzo, 1990, p.180, for details).

$E(u_t) = 0$ and $E(u_l u_s') = \Sigma \min(l,s)$, that is $E(uu') = \Sigma \otimes (\Delta'_{l,n} \Delta_{l,n})$, where Σ can be estimated as in the multivariate white noise approach.

4.1 Extrapolation

When $n > MN$ we need to estimate data for which the relevant temporally aggregated values are not available. We distinguish whether the contemporaneously aggregated information is or is not available. In the former case we talk about *constrained extrapolation* while in the latter we have a *pure extrapolation* problem. In both cases we look for the BLU estimator of:

$$y_{j,mN+r} = x'_{j,mN+r}\beta_j + u_{j,mN+r}, \quad j = 1,\mathrm{K}, M, \quad r = 1,\mathrm{K}, h,$$

where $x_{j,mN+r}$ is a $(p_j \times 1)$ vector of related indicators for the j-th series at time mN+r and $u_{j,mN+r}$ is a zero mean unobservable random error. Following Di Fonzo (1990), denote by

$$y_e = \begin{bmatrix} y_{el} \\ \mathrm{M} \\ y_{ej} \\ \mathrm{M} \\ y_{eM} \end{bmatrix}$$

the $(hM \times 1)$ vector of values to be estimated and, with obvious notation, by X_e and u_e the relevant matrix of related series and vector of stochastic disturbances, respectively. Now, let us consider the enlarged model

$$y^* = X^*\beta + u^*,$$

where

$$y^* = \begin{bmatrix} y \\ y_e \end{bmatrix}, \quad X^* = \begin{bmatrix} X \\ X_e \end{bmatrix}, \quad u^* = \begin{bmatrix} u \\ u_e \end{bmatrix},$$

$E(u^*) = 0$ and

$$E[u^*(u^*)'] = V^* = \begin{bmatrix} V & \Omega' \\ \Omega & V_e \end{bmatrix},$$

with $\Omega = E(u_e u')$ and $V_e = E(u_e u_e')$.

If a contemporaneously aggregated series is available, z_e such that $H_e y_e = z_e$, the complete set of estimates must be re-calculated[9] as follows:

$$\hat{y}^* = X^*\hat{\beta}^* + L^*(y_a^* - X_a^*\hat{\beta}^*), \qquad (16)$$

$$\hat{\beta}^* = \left[(X_a^*)'(V_a^*)^- X_a^*\right]^{-1}(X_a^*)'(V_a^*)^- y_a^*, \qquad (17)$$

where $L^* = V^*(H^*)'(V_a^*)^-$, $H^* = \begin{bmatrix} H & 0 \\ 0 & H_e \end{bmatrix}$, $y_a^* = H^* y^*$, $X_a^* = H^* X^*$ and $V_a^* = H^* V^* (H^*)'$.

In the pure extrapolation case the temporally constrained data need not to be re-estimated, while the extrapolations are given by

$$\hat{y}_e = X_e\hat{\beta} + \Omega H' V_a^- (y_a - X_a\hat{\beta}),$$

with $\hat{\beta}$ as in (15).

5 A stylized empirical application

In this section we present a typical session of multivariate indirect estimation. Three annual Italian employment series, each pertaining to the transport and communication branch of economic activity, have been disaggregated so that the quarterly estimated series comply with the known benchmark **z**.
The series in hand, observed in the period 1970-1990, are:

Y_{01}: labour units, transports;
Y_{02}: labour units, activities related to transport;
Y_{03}: labour units, communication.

We used three related indicators supplied (and used) by Istat[10]: the first one (X_1) comes from the Italian Quarterly Labour Force Survey and has been seasonally adjusted, while the remaining ones are a deterministic trend (X_2) and a dummy variable (X_3) equal to 1 from 1987.1 to 1991.1, respectively.

Each preliminary has been estimated according to the AR(1) version of the optimal univariate approach [11]

9 In the multivariate white noise case expression (16) can be notably simplified (Di Fonzo, 1990, p. 181).

worked out by Chow and Lin (1971). The GLS regression results [12] are in table 2, while in table 3 the OLS regression results are showed [13]. In both cases Y_{02} gave the worst fitting ($\overline{R}^2 < 0.8$) even if, on the whole, the results seem acceptable from a pure statistical point of view.

The multivariate regression estimates are reported in table 4 and 5 respectively. Both in terms of significativity of the estimated parameters and of global fitting, the results appear generally rather satisfactory.

Tab. 2: Estimated parameters of the auxiliary annual univariate regressions. AR(1)

Series	Constant	X_1	X_2	X_3	$\hat{\rho}$	$\overline{R}^2$	*F*
Y_{01}	450.2315 (5.9400)	0.1037 (1.1837)	2.4453 (11.4299)		0.8316	0.9247	123.86
Y_{02}	-24.5809 (-1.1277)	0.1692 (7.1464)		-12.6245 (-4.3378)	0.7524	0.7185	26.52
Y_{03}	144.1347 (5.2019)	0.0513 (1.6061)	1.4635 (20.2552)		0.7524	0.9774	432.97

Tab. 3: Estimated parameters of the auxiliary annual univariate regressions. OLS

Series	Constant	X_1	X_2	X_3	$\overline{R}^2$	*F*
Y_{01}	463.3512 (6.5410)	0.0869 (1.0693)	2.4612 (14.6587)		0.9599	240.41
Y_{02}	-30.9303 (-1.6905)	0.1759 (8.8592)		-12.8261 (-5.0653)	0.7946	39.69
Y_{03}	152.5776 (6.6925)	0.0421 (1.6105)	1.4753 (27.3027)		0.9879	817.70

Tab. 4: Estimated parameters of the multivariate white noise regression

Series	Constant	X_1	X_2	X_3
Y_{01}	425.4570 (5.2928)	0.1300 (1.4108)	2.4117 (12.5843)	
Y_{02}	-33.5542 (-1.6206)	0.1789 (7.9824)		-13.2682 (-5.3142)
Y_{03}	147.7660 (5.8801)	0.0484 (1.6965)	1.4498 (26.4409)	

$\overline{R}^2 = 0.9984$

$F = 11422.17$

10 No specification strategy was adopted. Besides the error term structure, the annual auxiliary regre ssion are as close as possible to those used by Istat. As it is clear, this is a classical 'poor indicators' sit uation.

11 See Istat (1985) for details.

12 t-ratios are in parentheses.

13 The OLS residuals are used to estimate .

Tab. 5: Estimated parameters of the multivariate random walk regression

Series	Constant	X_1	X_2	X_3
Y_{01}	-6.6290 (-0.1008)	0.6886 (8.5019)	1.2496 (1.4415)	
Y_{02}	-43.0699 (-1.3820)	0.1984 (5.2268)		-12.7544 (-4.4399)
Y_{03}	177.0377 (5.4066)	0.0027 (0.0678)	1.5020 (6.2557)	

$\overline{R}^2 = 0.9358$
$F = 268.98$

The discrepancies to be distributed by the adjustment methods are represented in fig. 1 as percentage of the contemporaneous benchmark. In fig. 2 the series estimated by the multivariate random walk procedure are represented along with their confidence intervals. Unlike the other two, series $\hat{Y}_2$ does not exhibit a marked upward trend and is characterized by relatively wider confidence intervals. This fact confirms the perplexities raised by the annual auxiliary regression results, and advises to be cautious when using this variable.

A simple, graphical comparison of the performances of the two approaches described so far is conducted in terms of quarterly growth rates (figg. 3, 4 and 5).
Two adjustment (proportional and Denton AFD) and two optimal multivariate (white noise and random walk) procedures have been considered. We consider also the preliminary series, whose dynamics should be in some sense 'closer' to that of the related series. The results are not univocal and can be summarized as follows:

The final estimates of Y_1 give raise to generally more irregular growth rates than those generated by the preliminary series. However, Denton's AFD procedure behaves well as compared to the remaining ones, in that the growth rates are characterized by less marked peaks.

As for series 2, both multivariate procedures work well as compared to the preliminarly estimated growth rates, while the adjustment procedures, notably Denton's, show unexpected marked peaks.

Multivariate random walk estimates behave fairly well for series 3, in the sense that they are close enough to the preliminary counterparts, while very irregular dynamics originate from the remaining procedures.

Fig. 1: Discrepancy as percentage of the contemporaneous benchmark

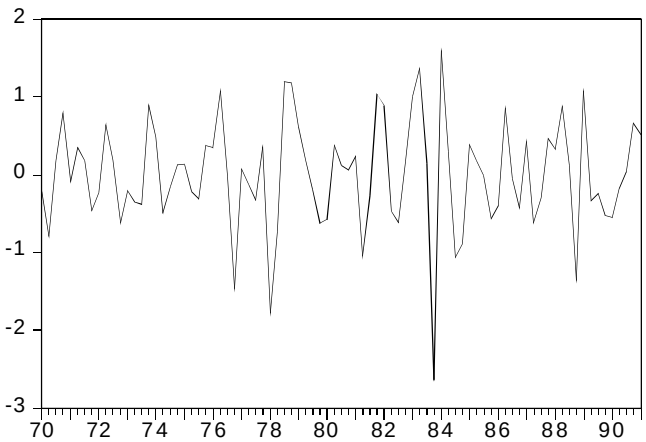


Fig. 2: Estimated series and confidence intervals: multivariate random walk

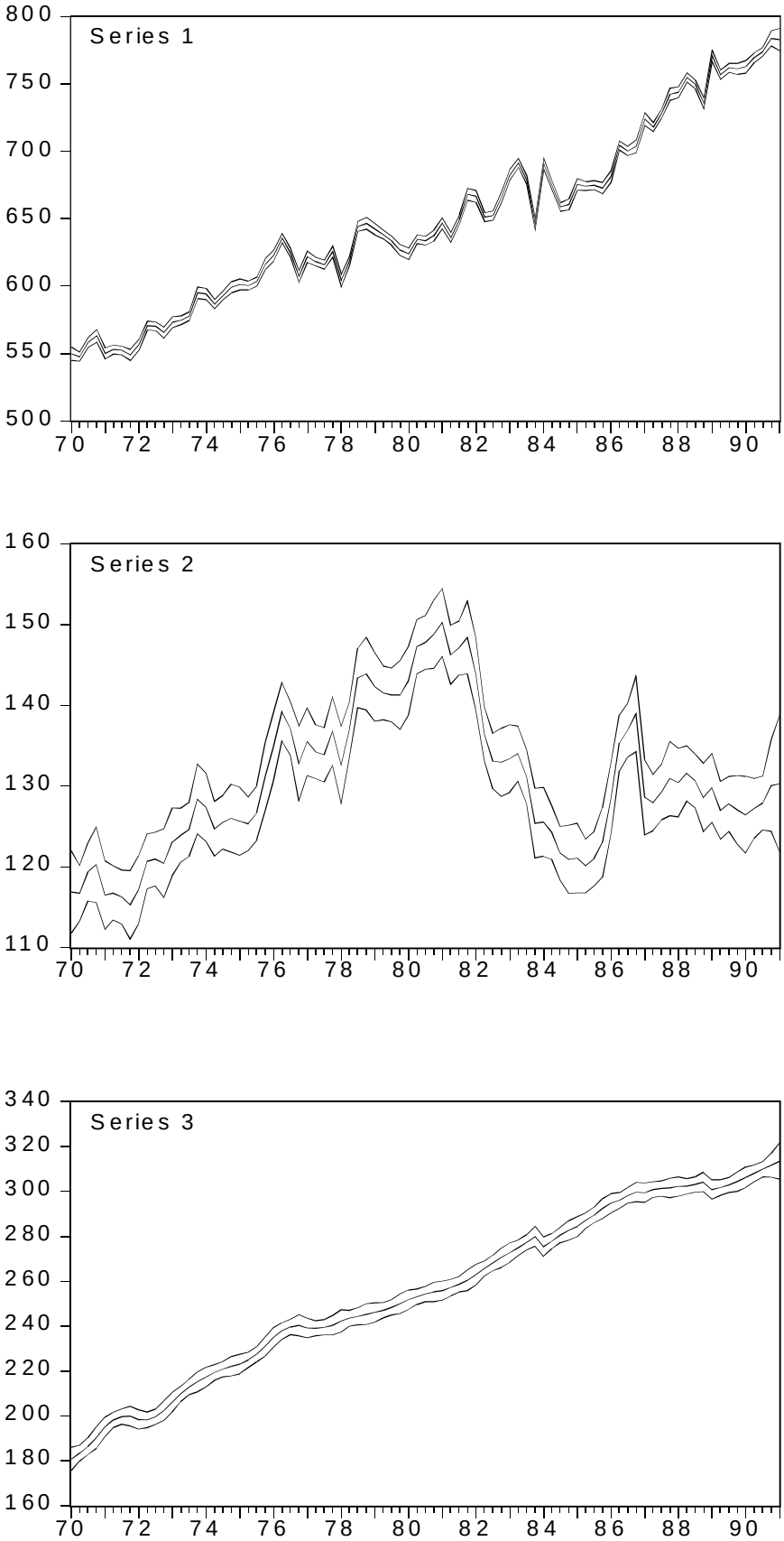
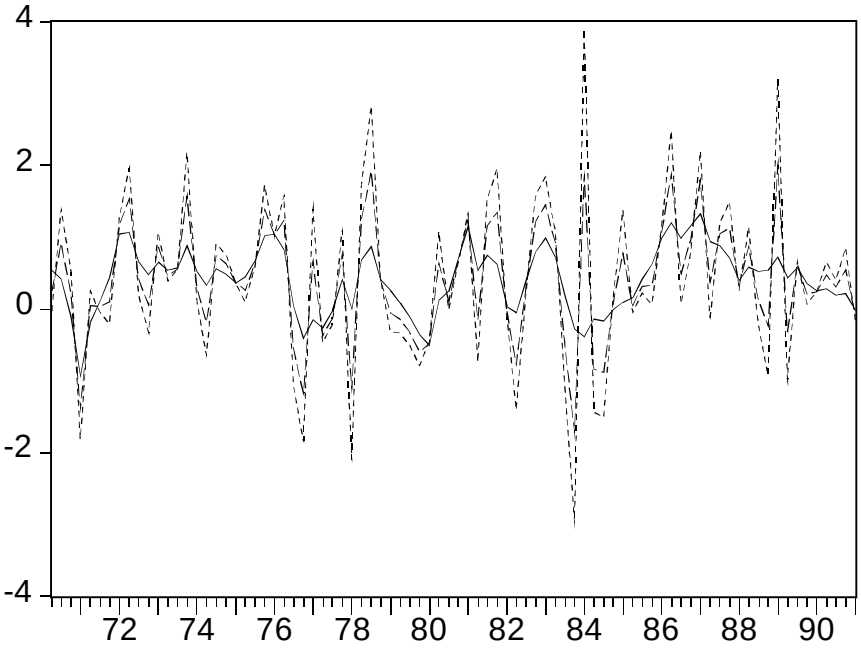
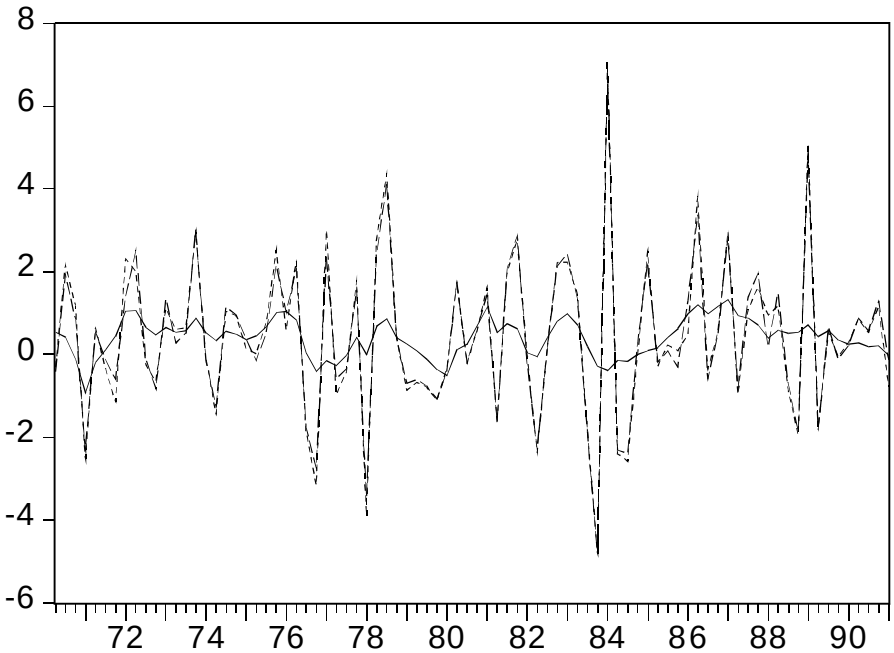


Fig. 3: Quarterly growth rates: series 1

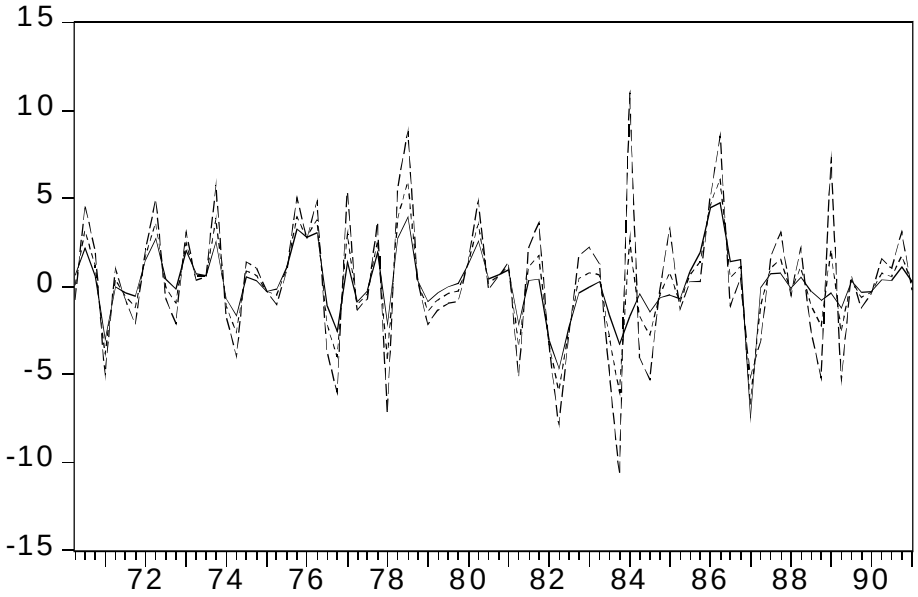


——— Univariate AR(1) — — — Proportional ——- Denton AFD

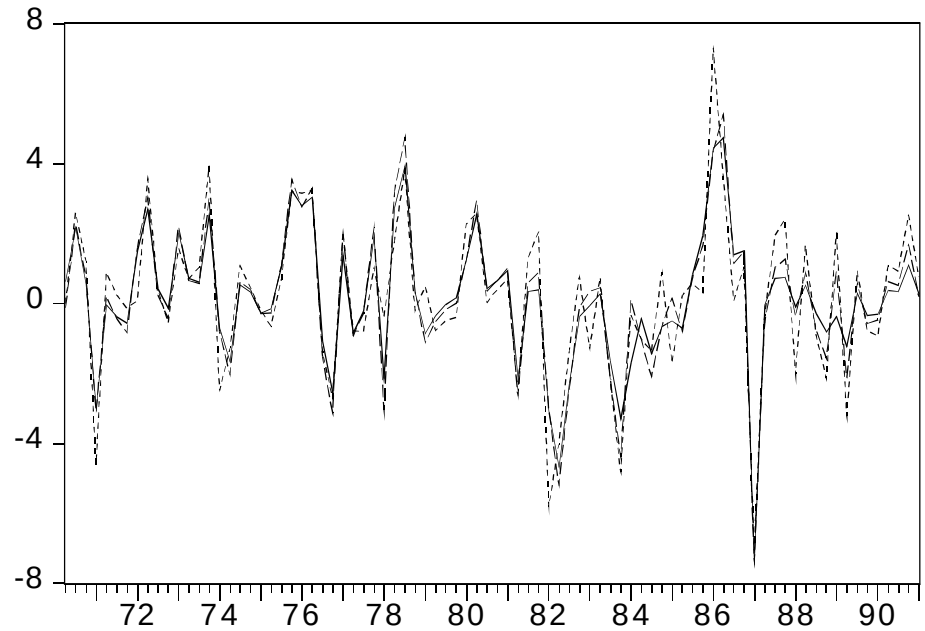


——— Univariate AR(1) — — — Multiv. white noise ——- Multiv. random walk

Fig. 4: Quarterly growth rates: series 2

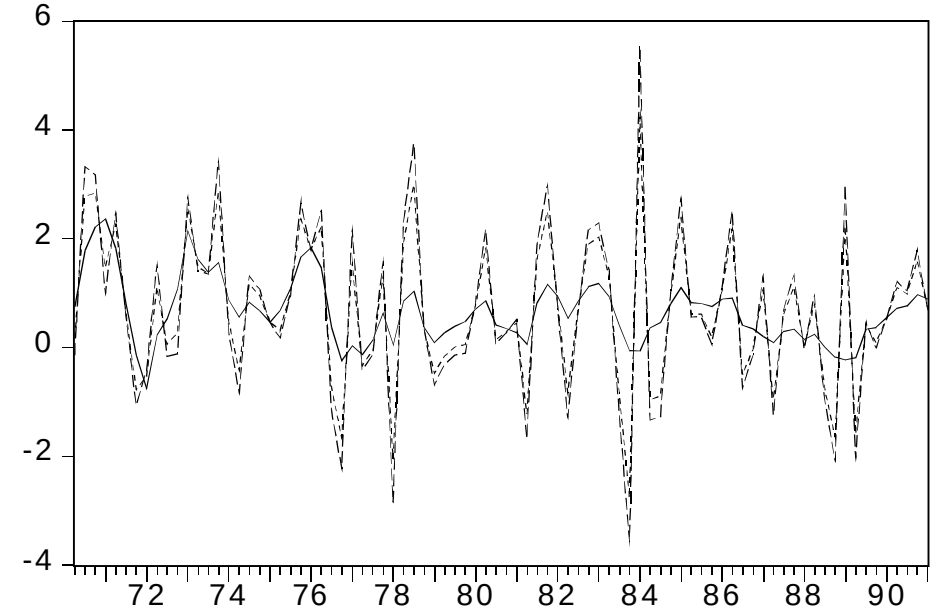


——— Univariate AR(1) — — — Proportional ——- Denton AFD

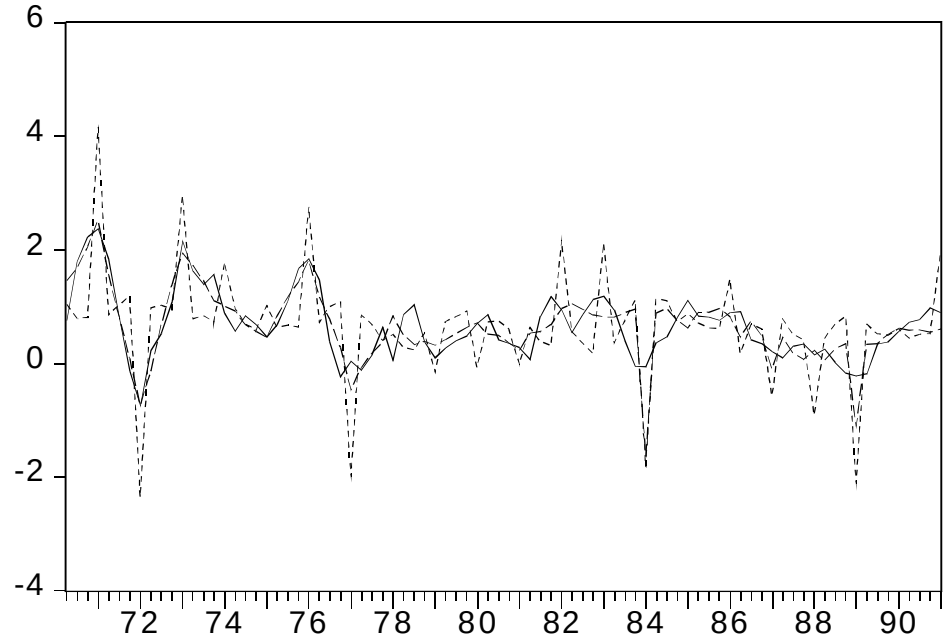


——— Univariate AR(1) — — — Multiv. white noise ——- Multiv. random walk

Fig. 5: Quarterly growth rates: series 3



——— Univariate AR(1) — — — Proportional ——- Denton AFD



——— Univariate AR(1) — — — Multiv. white noise ——- Multiv. random walk

Appendix

Following Di Fonzo (1990), matrix **H** can be partitioned as

$$H = \begin{bmatrix} 1_M' \otimes \mathrm{I}_n & & \\ \mathrm{I}_{M-1} \otimes D & \mathbb{M} & 0 \\ 0 & \mathbb{M} & D \end{bmatrix} = \begin{bmatrix} H_w \\ \mathbb{L} \\ H_M \end{bmatrix}.$$

Denoting **W** the $(n \times r)$ matrix, $r = n + N(M-1)$,

$$W = \left[D\mathbb{M} - \left(1_{M-1}' \otimes \mathrm{I}_n \right) \right]$$

and **R** the $\left[(r+n) \times r \right]$ matrix

$$R = \begin{bmatrix} \mathrm{I}_r \\ W \end{bmatrix},$$

we have

$$H_M = ZH_w, \qquad H = RH_w,$$

the singular matrix $HM^{-1}H'$ can thus be written as $RH_w M^{-1} H_w' R' = RM_w^{-1} R'$, where $M_w = H_w M^{-1} H_w'$ is a full rank $(r \times r)$ matrix. Furthermore, it can be readily checked that

$$\left(HM^{-1}H' \right)^{-} = R(R'R)^{-1} M_w^{-1} (R'R)^{-1} R'.$$

Given that $R'H = R'RH_w$, the adjusted estimates (10) can thus be expressed as

$$\hat{y} = p + M^{-1} H_w' M_w^{-1} (y_w - H_w p),$$

where $y_w = H_w y$ is the $(r \times 1)$ vector containing r "free" aggregated observations.

References

Barcellan, R. (1994), ECOTRIM: a program for temporal disaggregation of time series, paper presented at the INSEE-EUROSTAT workshop on Quarterly National Accounts, Paris, December 5-6, 1994.

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

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

Cholette, P.A. (1987), "Concepts, definitions and principles of benchmarking and interpolation of time series", Working Paper, Time Series Research and Analysis Division, Statistics Canada.

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

Chow, G.C. and Lin, A. (1971), "Best linear unbiased interpolation, distribuction and extrapolation of time series by related series", The Review of Economics and Statistics, 53, 4, pp. 372-375.

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, 333, pp. 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, pp. 178-182

Fernández, R.B. (1981), "A methodological note on the estimation of time series", The Review of Economics and Statistics, 63, 3, pp. 471-478.

Istat (1985), "I conti economici trimestrali. Anni 1970-1984", Supplemento al Bollettino Mensile di Statistica, 12.

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, pp. 695-696.

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, pp. 89-105.

Stone, J.R.N., Champernowne, D.A. and Meade, J.E. (1942), "The precision of national income accounting estimates", The Review of Economics Studies, 9, pp. 111-125.

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

Ecotrim: A program for temporal disaggregation of time series

Roberto BARCELLAN
Università di Padova

Applied researchers often face the problem of deriving disaggregated time series data when only information in temporally aggregated form is available. ECOTRIM is a program which supplies a set of mathematical and statistical techniques to carry out temporal disaggregation. It performs an indirect approach to the disaggregation problem: the disaggregated series are estimated by using the available aggregated series and, in case, a set of known related indicators. This paper offers a methodological support to the use of the package.

1 Introduction

Many economic and social variables are in principle generated by continuous processes, but are observed into discrete temporal units. Even one assumes (as we shall do in the rest of the paper) that the original (or basic) data generating process is discrete one, in time series analysis we often deal with temporally aggregated data. Thus we refer to a set of observations which are recorded in a lower-frequency time unit than that of the relevant original process.

The aggregated data are generally available in the form of (i) sum over a period of time of, or (ii) systematic sampling on m (m>1) consecutive original observations. In the former case we have a distribution problem which arises with reference to flow variables, while in the latter we treat an interpolation problem which deals with stocks variables.

We will refer to the process underlying the basic (high frequency) series as original (or disaggregated) process as opposed to the derived (or aggregated) process. The analyst, therefore, faces the problem of deriving the disaggregated data in reasonable way from the available information given by the low-frequency time series at disposal and, sometimes, by one or more related series.

Several approaches to the disaggregation problem have been proposed in literature. We can group them in the following way:

- *Mathematical approach:* the disaggregated values are estimated according to a mathematical criterion. We usually consider a minimum quadratic loss function problem whose solution gives raise to the estimated series (Boot, Feibes and Lisman, 1967, Cohen, Müller and Padberg, 1971);
- *Adjustment approach:* these methods forces a preliminary series of estimates which does not fulfill the aggregation constraints to fulfill them by modulating the discrepancies (Denton, 1971, Ginsburgh, 1973, Cholette, 1988);
- *optimal statistical approach:* the estimates are BLU with respect to a regression model which involves the unknown series and the related series (Chow and Lin, 1971, Bournay and Laroque 1979, Barbone,

Bodo and Visco, 1981, Fernandez, 1981, Litterman, 1983, Di Fonzo 1987);

- *ARIMA model based approach:* the relationships between the aggregated and the disaggregated ARIMA model which describe respectively the original and the derived process are used in order to estimate the derived series. We can use the relationships between the autocovariance structures or with the preliminary series to solve the disaggregation problem (Stram and Wei, 1986b, Al-0sh, 1989, Wei and Stram, 1990, Guerrero, 1990, Barcellan and Di Fonzo, 1994).

The above considered techniques refer only to the univariate disaggregation problem while in practice we often deal with a multivariate problem, that is we want to estimate M disaggregated series which fulfill an accounting constraint. In this case the series must fulfil both temporal and contemporaneous aggregation constraints. In this case too, adjustment methods or a BLUE approach can be used.

1.1 Ecotrim

Economic data estimation issues like those discussed until now are strongly concerned by most statistical agencies, which could find useful to have at their disposal a computational support which permits them to use the disaggregation techniques proposed in literature.

ECOTRIM is a program system which supplies a set of mathematical and statistical techniques to carry out temporal disaggregation. It has been created by R. Barcellan and T. Di Fonzo on behalf of the European Commission, Statistical 0ffice [1]. This paper represents a methodological support to the use of the package.

The present version of ECOTRIM is written in Fortran 90. It offers a menu driven approach in which the user is asked to choice the disaggregation procedure he wants to run. It can be used in an interactive mode or in batch mode according to the researcher's requirements. A typical ECOTRIM session proposes to the user a beginning choice between

- interactive mode;
- batch mode.

The former option allows the user to run an interactive session. This mode requires an active role to the researcher, that is, by using a menu driven path, he has to specify to the program all the information needed, like aggregated data files, in case, related series files, the disaggregation procedure he wants to use, contemporaneous aggregated series, if required, and so on.

The latter option corresponds to the choice of a batch session. In this case the user has only to write, by using an editor, a batch file, according to a set of rules, and then to run it from the apposite menu. The batch mode approach to the program is particularly suitable to treat large number of series to be disaggregated. In fact the batch command file can contains more than one set of instructions to disaggregate aggregated time series. So the program can be used in an automatic way to disaggregate several series according to the procedures chosen by the user and specified in the batch command file.

In order to give an idea of the opportunities that ECOTRIM offers and to introduce the methodologicaI aspects which will be treated in this paper, we briefly consider now a typical interactive ECOTRIM session showing the main choices the user is called to express (for more details see the manual).

Besides the option which allows to load the series required by the analysis (like series to be disaggregated, related series, other series to be used to carry out graphical comparison between original, if available, and estimated series), the user must chose the procedure he wants to use, according to the problem he wants to solve and to the set of information he disposes.

First of all he has to specify if he is dealing with univariate or multivariate case choosing between

- univariate methods;
- multivariate methods.

1 Directorate B, Economic Statistics and Economic and Monetary Convergence.

The univariate methods option allows to treat the univariate disaggregation problem, that is to disaggregate a single series in order to fulfill temporal aggregation constraints.

According to the information available we can choose

- disaggregation without related series;
- disaggregation by related series.

In the former case we suppose to deal with problems where the only available information corresponds to the aggregated series and, sometimes; the ARIMA model of its generating process. The procedures offered by ECOTRIM are, therefore, mathematical procedures or ARIMA model based procedure:

- Boot, Feibes and Lisman's procedure;
- Denton's procedure;
- Wei and Stram's procedure;
- Al-Osh's procedure.

While the first two methods are mathematical ones, the last two are ARIMA model based techniques and require the knowledge of the aggregated ARIMA model. The methodological concepts underlying these methods are illustrated in this paper in a slight different manner, that is according to the theoretical approach used (either mathematical adjustment or ARIMA model based).

The disaggregation by related series uses the idea that a logically correlated high-frequency series is available (note that we do not pose the problem to verify this hypothesis). In this case ECOTRIM offers several techniques: some of them correspond to the mathematical adjustment approach, some of them to the optimal, in the least squares sense, approach and some of them to the ARIMA model based approach.

Thus we may choose among .

- Denton's procedure;
- Ginsburgh's procedure;
- AR(1) procedure;
- Fernandez's procedure;
- Litterman's procedure;
- Guerrero's procedure.

While Denton and Ginsburgh's are adjustment methods, AR(1), Fernandez and Litterman's are optimal procedures and Guerrero's is an ARIMA model based techniques. All the techniques considered until now produce an estimated disaggregated series and some of them offer other information like confidence intervals (not available for mathematical/adjustment approach), diagnostics or the disaggregated estimated ARIMA model (obviously, only for ARIMA model based approach). ECOTRIM interactive mode also allows to plot the available series aggregated series, related series, other series and estimated series) in order to see their graphical representation and to make comparison between them. It is possible to save the obtained results, too. As to the multivariate methods, ECOTRIM permits to estimate disaggregated series and to fulfil temporal and contemporaneous aggregation constraints. Both pure adjustment and optimal, in the least squares sense. techniques are offered:

- white noise procedure;
- random walk procedure;
- Rossi's procedure;
- Denton's procedure;
- weighted adjustment procedure.

The first two options correspond to the optimal approach, while the remaining ones to the adjustment approach, It should be noted that in the last three cases preliminary estimated series are needed and, as to Rossi and weighted adjustment procedures, these series must fulfill the temporal aggregation constraints.

ECOTRIM batch mode essentially uses the same idea stated above and properly indicated in the batch command file (for more technical details see the manual). The present version of ECOTRIM is a preliminary version to be developed, ECOTRIM is also available in a more advanced version written in GAUSS which requires GAUSS 3.1.5 version. We are going to implement several improvements like data manipulation, a Fortran Sun version running under UNIX, a more detailed diagnostic analysis, state space and Kalman filter based methods.

All the univariate techniques available in the program are described from a theoretical point of view, pointing out the approaches and the solutions proposed by the authors. It should be stressed that ECOTRIM offers the opportunity to obtain

extrapolations, that is to estimate future values, when the related aggregated value is not yet available.
The issue is addressed for the methods which present this option. As far as the multivariate procedures are concerned, we refer to Di Fonzo (1994). Section 2 is devoted to establish the notation. The mathematical and adjustment procedures available in ECOTRIM are analyzed in section 3, while in section 4 the ARIMA model based techniques are described. Optimal procedures and the multivariate disaggregation problem are considered in section 5 and 6, respectively. Finally, section 7 contains some concluding remarks.

2 Notation

Let $y_t, t = 1, \mathrm{K}, n$ an equally spaced time series that admits the ARIMA (p, d, q) representation

$$\phi_p(B)(1-B)^d y_t = \theta_q(B) a_t, \qquad (1)$$

where B is the backshift operator, such that, $By_t = y_{t-1}$, $\phi_p(B) = (1-\phi_1 B - \mathrm{K} - \phi_p B^p)$, $\theta_q(B) = 1-\theta_1 B - \mathrm{K}$ $\mathrm{K}\,\theta_q B^q$) and $\{a_t\}$ is a white noise process with variance σ_a^2. Further assume that $\phi_p(B)$ and $\theta_q(B)$ have roots that lie outside the unit circle and don't share common roots.

Let $y_t, t = m, \mathrm{K}, n$, the series obtained by a linear combination of m consecutive values of the basic series y_t:

$$\mathrm{Y}_t = \sum_{j=0}^{m-1} c_{m-j} x_{t-j}, \qquad t = m, \mathrm{K}, n, \qquad (2)$$

Where c_j is the $j-th$ coefficient of the linear combination considered, $j = 1, \mathrm{K}, m$.
Let y_T, $\mathrm{T} = 1, \mathrm{K}, N$, be the series obtained by systematic sampling of Y_t every m time periods, $N = \left[\frac{n}{m}\right]$,

$$y_{0,\mathrm{T}} = \mathrm{Y}_{m\mathrm{T}}, \qquad \mathrm{T} = 1, \mathrm{K}, N.$$

It is easy to recognise that each value of the basic series y_t concurs to the determination of only one of the values of the series $y_{0,\mathrm{T}}$. Whereas each y_t can concur to form several values of Y_t. We will refer to the transformation that gives rise to Y_t and $y_{0,\mathrm{T}}$ as overlapping and non-overlapping linear combination, respectively. In this paper we consider only non-overlapping linear combinations.

Let $y_{0,T}$, T=1,...,N admit the ARIMA (P,D,Q) representation:

$$\psi_p(\mathrm{B})(1-\mathrm{B})^D y_{0,T} = \eta_Q(\mathrm{B}) e_T, \qquad (3)$$

where B is the bachshift operator such that $\mathrm{B}y_{0,T} = y_{0,T-1}$ and $\mathrm{B} = B^m$, $\psi_p(\mathrm{B}) = 1-\psi_1\mathrm{B}-\mathrm{K}$ $\mathrm{K}\,\psi_P\mathrm{B}^{\mathrm{P}})$, $\eta_Q(\mathrm{B}) = (1-\eta_1\mathrm{B}-\mathrm{K}-\eta_Q\mathrm{B}^Q)$ and $\{e_T\}$ is a white noise process with variance σ_e^2.

The relation between y_t and $y_{0,T}$ can be expressed as follows:

$$\begin{aligned} y_{0,T} &= \sum_{j=0}^{m-1} c_{m-j}\, y_{mT-j} \\ &= \left(\sum_{j=0}^{m-1} c_{m-j} B^j\right) y_{mT} \qquad (4) \\ &= \left(c_1 B^{m-1} + c_2 B^{m-2} + \mathrm{K} + c_m\right) y_{mT} = c'\, y_{mT}, \\ & \qquad T = 1, \mathrm{K}, N, \end{aligned}$$

where m is the aggregation order, $c = [c_1 c_2 \mathrm{K}\ c_m]'$ is the $(m \times 1)$ vector of the aggregation coefficients and $y_{mT} = [\,y_{m(T-1)+1}\ \mathrm{K}\ y_{mT}\,]'$. The vector **c** generally assumes one of the following forms:

$$\begin{aligned} \mathrm{c} &= \quad\ (1\ \ 1\ \ \mathrm{K}\ \ 1)', \\ \mathrm{c} &= 1/\mathrm{m}\ (1\ \ 1\ \ \mathrm{K}\ \ 1)', \\ \mathrm{c} &= \quad\ (1\ \ 0\ \ \mathrm{K}\ \ 0)', \\ \mathrm{c} &= \quad\ (0\ \ 0\ \ \mathrm{K}\ \ 1)'. \end{aligned} \qquad (5)$$

The first configuration, typical for flow variables as consumption, production and so on, is referred to as linear temporal aggregation in strict sense. The second configuration, suitable for index variables (e.g., prices) gives rise to temporal averages of the basic series. The last two configurations, finally, correspond to the systematic sampling of the first and the last, respectively, of m consecutive values of a stock variable.

In practice we may find other kinds of coefficient vectors, let's think, for example, to the weights used to

calculate an annual functions as linear combination of monthly values.

We can represent all the above relationships in a more compact form. Let $y = [y_1 y_2 \mathrm{K}\ y_n]'$ the $(n \times 1)$ vector of basic values, $y_0 = [y_{0,1} \mathrm{K}\ y_{0,N}]'$ the corresponding $(N \times 1)$ aggregated vector and $D = c' \otimes \mathrm{I}_N$, the $(N \times mN = n)$ aggregation matrix (where $\otimes$ denotes the Kronecker product). Then

$$y_0 = Dy. \tag{6}$$

We are usually interested, besides the distribution or interpolation problem, in forecasting future values of the unknown derived series (extrapolation). In this case no corresponding aggregated value is available, so we have to forecast future disaggregated values by using all past information.

Let y_F the $((n+r) \times 1)$ vector unknown data, where r is the forecast horizon. Then the temporal aggregation constraint becomes

$$y_0 = [D | 0_r]\, y_F = D_F\, y_F.$$

If the basic and the aggregated process are described by non-stationary ARIMA models, respectively, we pose $w_t = (1-B)^d\, y_t$, $t = d, \mathrm{K}, n$ and $u_T = (1-\mathrm{B})^D\, y_{0,T}$, $T = D, \mathrm{K}, N$, note that D=d due to the relationships between aggregated and disaggregated ARIMA models (see Barcellan and Di Fonzo, 1993). Thus the following relationships hold:

$$w = \Delta_n^d\, y, \qquad u = \Delta_N^d\, y_0 \tag{8}$$

with $\Delta_l^d [(l-d) \times l]$ matrix of differentiation of the form

$$\Delta_l^d = \begin{bmatrix} \delta_1 & \delta_2 & \mathrm{K} & \delta_d & 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 \\ 0 & \delta_1 & \delta_2 & \mathrm{K} & \delta_d & 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 & \delta_1 & \delta_2 & \mathrm{K} & \delta_d \end{bmatrix} \tag{9}$$

where δ_i are the coefficients of L in $(L-1)^d$, with L equal to B or B according to the kind of difference working in practice.

We often face with problems where, besides the available low-frequency series, we dispose of other information such as related series, that is high-frequency series logically correlated with the unknown basic one.

Let $x_{j,t}$, $t = 1, \mathrm{K}, n$, a related high-frequency series, x_j, $j = 1, \mathrm{K}, p$, the corresponding $n \times 1$ vector, and X the associated $(n \times p)$ matrix of all the available related series. Using the same notation as for the original series, we denote with x_{0j} the aggregated $(N \times 1)$ vector corresponding to the related series x_j and with X_0 the aggregated $(N \times p)$ matrix of all related series.

We suppose that the unknown series and the available high-frequency series are related by the linear regression model

$$y = \mathrm{X}\beta + u, \tag{10}$$

with $u\,(n \times 1)$ zero mean vector of stochastic disturbances.

3 Mathematical and adjustment procedures

ECOTRIM offers a set of univariate mathematical and adjustment procedures to get the disaggregated series from an aggregated one. These methods usually correspond to a quadratic loss function minimisation problem with reference to the disaggregated series values. While they do not offer, like other approaches, related information as diagnostics or confidence intervals, due to their mathematical nature, they represent a simple and quick way of obtaining series which can be used as preliminary with other techniques (for example, in multivariate adjustment methods). The obtained series often supply good estimates for the problem we face with.

3.1 Boot, Feibes and Lisman's procedure

Boot, Feibes and Lisman (1967, hereafter BFL) proposed a mathematical procedure to solve the univariate problem of temporal disaggregation. The main idea is that a plausible criterion is to minimise the sum of squared differences between successive disaggregated values (first differences model, FD) or to minimise the sum of squared second differences (second differences model, SD), subject to the aggregation constraints. The subperiod values are determined on the

basis of the values of the previous subperiods, that is, in an autoregressive manner.

BFL developed this idea for the particular case of temporal aggregation in strict sense of quarterly values. A generalised approach, for the strict sense case, has been proposed by Cohen, Müller and Padberg (1971). More generally, here we consider linear combinations of m disaggregated values. From a mathematical point of view the disaggregation problem corresponds to the following

$$\min_{y_j} \sum_{j=d+1}^{n=mN} \left[(1-B)^d \, y_j \right]^2,$$

subject to the aggregation constraints (4) or, in matrix form,

$$\min_{y} \; y' \left(\Delta_n^d \right)' \Delta_n^d y,$$

subject to

$$y_{0,T} = D' y,$$

where $(1-B)^d \, y_j$ is a general expression for the dth differences version. In the special case of the first second differences versions to be treated here d assumes the values 1 and 2 respectively.

The problem is solved by considering the Lagrangean expression

$$L^d \left(y, y_0, \lambda \right) = y' \left(\Delta_n^d \right)' \Delta_n^d y - \lambda \left(Dy - y_0 \right).$$

Upon differentiating with respect to $y's$ and $\lambda's$ and equating the resulting expression to zero we obtain a system of $n+N$ linear equations of the general form

$$\begin{bmatrix} 2\left(\Delta_n^d \right)' \Delta_n^d & -D \\ D & 0 \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ y_0 \end{bmatrix}. \qquad (11)$$

It is possible to show that the system matrix has full rank so the system (11) can be easily solved.

3.2 Ginsburgh's procedure

Ginsburgh (1973) proposed an adjustment procedure to solve the disaggregation problem by using related series.

He supposed that a linear regression model, which involves the aggregated values y_0, T and the aggregated related values $x_{0,\mathrm{T}}, \mathrm{T} = 1, \mathrm{K}, N$, holds:

$$y_{0,T} = \beta_0 + \beta_1 x_{0,T} + u_{0,T}, \qquad T = 1, \mathrm{K}, N. \qquad (12)$$

Then he supposed that the unknown $y_t's$ can be estimated by

$$\tilde{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t, \qquad t = 1, \mathrm{K}, n.$$

Obviously these estimates do not generally fulfill the aggregation constraints. Thus the $\tilde{y}'s$ have to be adjusted in order to fulfill them [2].

In practicethe required steps are:

1 Interpolation by BFL method of both the aggregated series and the related aggregated indicators;

2 Computation of the disaggregated final values as follows

$$\hat{y}_{m(t-1)+j} = y^{BFL}_{m(t-1)+j} + \hat{\beta}_1 \left(x_{m(t-1)+j} - x^{BFL}_{m(t-1)+j} \right)$$
$$j = 1, \mathrm{K}, m. \qquad (13)$$

where $\hat{\beta}_1$ represents the OLS estimate of β_1 in the aggregated regression equation (12), y^{BFL} and x^{BFL} represent the preliminary estimates obtained by BFL.

It can be easily shown that the estimates $\hat{y}'s$ are obtained by using an adjustment method. The preliminary estimates $\tilde{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t, \; t = 1, \mathrm{K}, n,$ with $\hat{\beta}_0$ and $\hat{\beta}_1$ OLS parameter estimates in (12), are

2 Such a kind of method was first proposed by Vangrevelinghe (1966). The main difference between the two methods consists of using a different mathematical procedure for the first step (BFL for Ginsburgh's method and Lisman and Sandee, 1964, for Vangrevelinghe's method).

adjusted by modulating the discrepancies between the aggregated values $y_{0,T}$'s and the corresponding estimated aggregated values $\hat{\beta}_0 + \hat{\beta}_1 x_{0,T}$ according to the BFL procedure.

3.3 Denton's procedure

Denton (1971) developed a pure adjustment method. In fact, he did not consider the issue of estimating the preliminary series, but he dealt with the problem of distributing the discrepancies in order to fulfill the aggregation constraints.

The problem is to adjust a vector **p** of preliminary data to obtain estimates which minimise, in some sense, the preliminary series distortion and jointly are coherent with the temporally aggregated data.

More generally we specify a penalty function $f(y, p)$ and express the problem as that of choosing y in order to minimise this penalty function. We wish to solve the problem of minimising a quadratic loss function in the differences between the preliminary and the adjusted series:

$$\min_y (y-p)'M(y-p), \qquad (14)$$

subject to the aggregation constraints (14), where **M** is a symmetric $(n \times n)$ nonsingular matrix to be specified. We set up a Lagrangean expression and write

$$L = (y-p)'M(y-p) - 2(\lambda)'(y_0 - Dy).$$

Differentiating with respect to y's and λ's and equating the resulting expression to zero we obtain a system of $n+N$ linear equations of the general form

$$\begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} M & D' \\ D & 0 \end{bmatrix}^{-1} \begin{bmatrix} Mp \\ y_0 - Dp \end{bmatrix},$$

whose solution is

$$y = p + \mathrm{K}(y_0 - Dp),$$
$$\mathrm{K} = M^{-1}D'\left(DM^{-1}D'\right)^{-1}.$$

Therefore, the preliminary estimates are adjusted distributing the aggregated discrepancies according to the matrix of weights K.

If we suppose that M is the identity matrix, we simply distribute the discrepancies in equal amounts among the m subperiod values. A more interesting possibility is to employ a penalty function based on the *d*th differences of the preliminary and adjusted series. Let $\Delta_{d,n} = (\Delta_{1,n})^d$, with

$$\Delta_{1,n} = \begin{bmatrix} k \\ \Delta_1^d \end{bmatrix},$$

where **k** is the $(1 \times n)$ vector equal to $[1 0 \mathrm{K} 0]$; thus the penalty function is

$$f(y, p) = (y-p)'\Delta'_{d,n}\Delta_{d,n}(y-p),$$

with $M = \Delta'_{d,n}\Delta_{d,n}$. Thus for the *additive first differences* and for *additive*, *second difference* s cases, which are of interest to us, the matrix M assumes the form $M = \Delta'_{1,n}\Delta_{1,n}$ and $M = \Delta'_{1,n}\Delta'_{1,n}\Delta_{1,n}\Delta_{1,n}$, respectively.

The calculation of the estimated values, as pointed out by Denton, can be carried out without difficulty since the form of the matrices involved makes it easy to deal with (see Denton, 1971, Di Fonzo, 1987). Note that we have to eliminate from the penalty function all values outside the adjustment range. Thus we have set $y_t = p_t$ for $t = 0, -1, \mathrm{K}, 1-d$[3].

Finally, we note that the main problem of this approach are initial values. Cholette (1988) gives a solution to such a problem.

4 Arima model based approach

The main idea of the ARIMA model based approach is that we can use, to estimate the high-frequency series, the information supplied by the aggregated ARIMA model associated to the low-frequency available series

3 That is we assume that $(y_0 - p_0) = (y_{-1} - p_{-1}) = \mathrm{K} = 0$

and by the theoretical relationships which link aggregated and disaggregated ARIMA models.
As a result of the ARIMA model based approach some further information can be obtained using these methods. In particular, all the techniques considered provide for confidence intervals and some of them supply an estimate of the disaggregate ARIMA model besides the opportunity to predict future values. Let's note that, with reference to the first two procedure we are going to analyse, the only information available corresponds to the aggregated series and to its ARIMA model, so we deal with a problem with strong information lack.

Despite of this, in practice these methods require more computational resources and a more active role of the researcher.

4.1 Wei and Stram's procedure

Since the ARIMA model and its autocovariance structure are closely related, Stram and Wei (1986b) and Wei and Stram (1990) considered the possibility of estimating the autocovariance structure for the unknown series from the available autocovariance of the aggregated model.

The relationship between the disaggregated and the aggregated structure of covariance is the basis of the approach. We refer to the stationary series $w_t = (1-B)^d y_t$ and $u_T = (1-B)^d y_{0,T}$ obtained by differentiating the basic and the aggregated series respectively. The following result (Barcellan e Di Fonzo, 1993) states the relationship between the covariances of $\{w_t\}$ and $\{u_T\}$:

$$\gamma_u(k) = \left(1 + B + \mathrm{K}\, B^{m-1}\right)^{2d} \sum_{i=0}^{m-1} c_{m-i} \sum_{j=0}^{m-1} c_{m-j}\gamma_w\left[mk + (m-1)d + i - j\right] \quad (15)$$

Thus each autocovariance $\gamma_u(k)$ can be expressed as a linear combination of $\gamma_w(j)$'s with j varying from $km-(d+1)(m-1)$ to $km+(d+1)(m-1)$. We can use a matrix form to express (15), that is

$$\gamma_u^* = \mathrm{A}^*\gamma_w^*,$$

where $\gamma_u^* = [\gamma_u(0)\gamma_u(1)\mathrm{K}]'$,
$\gamma_w^* = [\gamma_w(-(d+1)(m-1))\mathrm{K}\,\gamma_w(0)\mathrm{K}]'$ and

$$\mathrm{A}^* = \begin{bmatrix} v' & 0' & 0' & \mathrm{K} & 0' & \mathrm{K} \\ 0_m' & v' & 0' & \mathrm{K} & 0' & \mathrm{K} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & 0 & \mathrm{M} & 0 \\ 0' & 0' & 0' & \mathrm{K} & v' & \mathrm{K} \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & 0 & \mathrm{M} & 0 \end{bmatrix},$$

with v $\left[(2(d+1)(m-1)+1)\times 1\right]$ vector of the coefficients of B^i, $i = 0, \mathrm{K}, 2(d+1)(m-1)$, in the polynomial

$$\left(1 + B + \mathrm{K} + B^{m-1}\right)^{2d} \sum_{i=0}^{m-1}\sum_{j=0}^{m-1} c_{m-i}\, c_{m-j}\, B^{(m-1)-i+j}.$$

Let's suppose to be interested only in the first h autocovariances of $\{u_T\}$. Then we consider only the "adjusted" upper left corner matrix of the matrix A^*, let's indicate it with A, $[(h+1)\times(l+1)]$, with $l = hm + (m-1)(d+1)$,), (Stram and Wei, 1986a, Barcellan and Di Fonzo, 1993). Let $\gamma_u = [\gamma_u(0)\mathrm{K}\,\gamma_u(h)]'$ and $\gamma_w = [\gamma_w(0)\mathrm{K}\,\gamma_w(l)]'$ then

$$\gamma_u = \mathrm{A}\gamma_w. \quad (16)$$

Then the disaggregation problem is dealt with according two components:

- the identification and estimation of the basic ARIMA model;
- the estimation of the disaggregated series values.

In order to obtain the estimates for y's, Stram and Wei (1986b) and Wei and Stram (1990) developed a generalised least squares procedure, for the strict sense aggregation case, which minimises the quadratic form $w'V_\omega^{-1}w$ subject to the temporal aggregation constraint $y_0 = Dy$.

Barcellan and Di Fonzo (1994) extended these results to general linear combinations obtaining.

$$\begin{aligned} \hat{w} &= V_w\left(C^d\right)'V_u^{-1}u \\ &= V_w\left(C^d\right)'V_u^{-1}\Delta_N^d\, y \end{aligned}.$$

$$\hat{x} = \begin{bmatrix} \Delta_n^d \\ 0 \mathrm{M} \mathrm{I}_d \otimes c' \end{bmatrix}^{-1} \begin{bmatrix} V_w \left(C^d \right)' V_u^{-1} \Delta_N^d \\ 0 \quad \mathrm{M} \quad \mathrm{I}_d \end{bmatrix} y_0. \qquad (17)$$

and

$$= \begin{bmatrix} g_0 & g_1 & \mathrm{K} & g_{m-1} & g_m & \mathrm{K} & g_{(m-1)(d+1)} & 0 & \mathrm{K} & 0 & \mathrm{K} & 0 \\ 0 & 0 & \mathrm{K} & 0 & g_0 & \mathrm{K} & g_{(m-1)d-1} & g_{(m-1)d} & \mathrm{K} & g_{(m-1)(d+1)} & \mathrm{K} & 0 \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 0 & 0 & \mathrm{K} & 0 & 0 & \mathrm{K} & 0 & 0 & \mathrm{K} & 0 & \mathrm{K} & g_{(m-1)(d+1)} \end{bmatrix}$$

is a $\left[(N-d) \times (n-d) \right]$ matrix where g_j represents the coefficient of B, $j = 0, \mathrm{K}, (m-1)(d+1)$, in the polynomial $\left(1 + B + \mathrm{K} + B^{m-1}\right)^d \left(\sum_{i=0}^{m-1} c_{m-i} B^i \right)$

In practice the matrices V_w and V_u are usually unknown and have to be estimated using the available data. While V_u can be easily calculated by using the aggregated model parameter estimates, the estimation of V_w is not so immediate. Wei and Stram (1990) founded their disaggregation procedure over the relationship (16). This is not generally a one to one relationship but, under suitable assumptions, it permits to estimate the disaggregated ARIMA model and then the associated autocovariance matrix (see Wei and Stram, 1990).

The extrapolation problem can be treated noting that the aggregation constraints are fulfilled if we use the relationship (7). As we shall see for the optimal approach (see par. 5.7) the extrapolation problem can be solved by using the augmented aggregation matrix $[0_r \,|\, D]$ instead of D.

4.2 Al-Osh's procedure

Al-Osh (1989) considered a dynamic linear model approach for disaggregating time series. He used an appropriate state space representation of the ARIMA model which describes the unknown series and which takes care of the temporal aggregation constraint.

Let $\phi_p(B)(1-B)^d = \Xi(B) = \left(1 - \xi_1 \mathrm{B} - \mathrm{K} - \xi_{p+d} B^{p+d}\right)$, $r = \max(p+d, q+1)$, $h = m + r - 1$. Thus

$$y_T = z' \alpha_{mT}, \qquad T = 1,2, \mathrm{K}, N \quad (18)$$

with $\mathbf{z}$ a $(h \times 1)$ vector whose elements are

$$z_i \begin{cases} c_m & for\ i = 1 \\ 0 & for\ i = 2, \mathrm{K}, r \\ c_{m-j} & for\ i = r+1, \mathrm{K}, h \quad and\ j = 1, \mathrm{K}, m-1 \end{cases}$$

and α_{mT} a $(h \times 1)$ vector of h state variables with the following form:

$$\alpha_{mT}(i) = \begin{cases} y_{mT} & for\ i = 1 \\ \sum_{j=1}^{r} \xi_j y_{mT+i-1+j} - \sum_{j=i-1}^{r-1} \vartheta_j a_{mT+i-1-j} & for\ i = 2, \mathrm{K}, r \\ y_{mT+r-i} & for\ i = r+1, \mathrm{K}, h \end{cases}$$

Note that some ξ_i or θ_j are zeros unless $p + d = q + 1$. The vector α_{mT} is dynamic and changes according to the *state* equation

$$\alpha_{mT} = F \alpha_{m(T-1)} + G \eta_{mT}, \quad T = 1,2, \mathrm{K}, N, \qquad (19)$$

with $\eta_{mT} = \left[a_{mT} a_{mT-1} \mathrm{K} a_{m(t-1)+1} \right]'$ a $(m \times 1)$ gaussian vector with zero and covariance matrix **Q; F** and **G** are a $(h \times h)$ and a $(h \times m)$ matrices derived (Al-Osh, 1989, p. 88) from the state space representation of the disaggregated ARIMA model and from the temporal aggregation constraints.

The estimated state vectors and the corresponding covariance matrices are obtained by using the Kalman filter with reference to the state space representation (18) and (19).

In order to estimate the values of the maximum likelihood estimates of the parameters of the identified model by using the available information and the recursive Kalman filter equations (Al-Osh, 1989, p. 90).

4.3 Guerrero's procedure

Let y_t, $t = 1, \mathrm{K}, N$ the unknown high-frequency series which we suppose admits the ARIMA representation (1). If we consider the minimum square error (MMSE) linear estimator of y_t, based on information up to time 0<t (Box and Jenkins, 1976), given by the conditional expectation

$$\mathrm{E}\left(y_t \,|\, y_0, y_1, \mathrm{K} \right) = \mathrm{E}_0 \left(y_t \right),$$

The forecast error, in terms of the pure MA representation, is

$$y_t - \mathrm{E}_0(y_t) = \sum_{j=0}^{t-1} \tau_j a_{t-j}, \qquad (\tau_0 \equiv 1),$$

with $\tau's$ obtained equating coefficients of the powers of B in the equation $\tau(B)\varnothing(B)(1-B)^d = \theta(B)$. In matrix form

$$y - \mathrm{E}_0(y) = \mathrm{T}a,$$

where **T** is a low triangular matrix formed by the sequences of weights $1, \tau_1, \mathrm{K}, \tau_{n-1}$ in the first column, the weights $0,1, \mathrm{K}, \tau_{n-2}$ in the second column, and so on. The linear MMSE estimator of **y** is

$$\hat{y} = \mathrm{E}_0(y) + \hat{A}\left[y_0 - D\mathrm{E}_0(y)\right],$$

with

$$\hat{A} = TT'D'\left(DTT'D'\right)^{-1}$$

and the estimatior error covariance matrix is

$$E_0\left[(\hat{y}-y)(\hat{y}-y)'\right] = \sigma_a^2\left(\mathrm{I} - \hat{A}D\right)TT'.$$

In practice, we assume that a preliminary estimate **p** exists and that $\{p_t\}$and $\{y_t\}$have essentially the same autocorrelation structure. So they admit basically the same ARIMA model but **p** does not fulfill the aggregation constraints.

Let's assume that $E(y_t|p) = p_t$ and $E_0(y)$ is independent of **p**. Then, taking the conditional expectation of (20), given **p**, we have

$$p - E_0(y) = E(y|p) - E\left[E_0(y)|p\right] = TE(a|p),$$

and subtracting this form (20) we obtain

$$y - p = Ta^*, \qquad (21)$$

with $a^* = a - E(a|p)$ a random vector such that

$$E(a^*|p) = 0, E\left[a^*(a^*)'|p\right] = \sigma^2 Q,$$

with **Q** a positive definite matrix.

The BLUE of **y**, given **p** and the aggregation constraints, is

$$\hat{y} = p + \hat{A}\left[y_0 - Dp\right]$$

with

$$\hat{A} = TQT'D'\left(DTQT'D'\right)^{-1} \qquad (22)$$

and

$$\mathrm{cov}\left[(\hat{y}-y)|p\right] = \sigma^2\left(\mathrm{I} - \hat{A}D\right)TQT' \qquad (23)$$

As we can see this estimator consists of two components: the former corresponds to the preliminary estimated series, the latter can be interpreted as a term which adjusts the first component by using a weighted combination, whose weights are derived from the ARIMA representation by $\hat{A}$, of the discrepancy between the aggregated and the preliminary aggregated series.

Obviously, we should be able to estimate σ^2 and TQT' in order to obtain the estimator. A simple case occurs when $Q = I$ since expressions (22) and (23) become easier and σ^2 can be estimated from

$$(n-k)\hat{\sigma}^2 = (a^*)'a^* = (\hat{y}-p)'(TT')^{-1}(\hat{y}-p),$$

where *k* is the number of estimated parameters in the ARIMA model. If this is not the case, Guerrero (1990) proposed the following EGLS procedure. Given (21)

$$e = Ka^* = K^{-1}T(y-p) = \Omega^{-1}(y-p),$$

with the non-singular matrix **K** such that **KQK'=I** according to the steps:

1. assume Q=I and calculate the BLU estimator. Construct the series $\hat{a} = T^{-1}(\hat{y}-p)$. If this series can be assumed as generated by a white noise, the assumption Q=I is valid and so are the estimator and the estimated error covarariance matrix. If it is not, we must build an ARIMA model for

$\{y_t - p_t\}$ and obtain its pure MA representation in order to get an estimate of the matrix Ω:

2. make use of the relationship

$$\Omega\Omega' = T(K'K)^{-1}T' = TQT'$$

To calculate $\hat{A}$ for the BLU estimator and then calculate $\hat{y}$.

Finally estimate σ^2 by means of

$$(n-k)\hat{\sigma}^2 = (e^*)'e^* = (\hat{y}-p)'(TQT')^{-1}(\hat{y}-p),$$

with *k* equal to the number of parameters of the ARIMA model used to estimate Ω.

5 The best linear unbiased estimator approach

This approach was developed by Chow and Lin (1971). The idea is to relate the regression model for the disaggregated series and the corresponding aggregated regression model. Obviously, this method requires related series.

Let's consider the disaggregated regression model (10), which relates the series to be estimated and the available related series, and the corresponding aggregated regression model obtained by pre-multiplying this equation by the aggregation matrix **D**

$$D_y = DX\beta + Du,$$

that is

$$y_0 = X_0\beta + u_0.$$

We consider the problem of predicting y in a general linear model (Goldberger, 1962). A linear unbiased estimator $\hat{y}$ of *y* satisfies; for some matrix **A**,

$$\hat{y} = Ay_0 = A(X_0\beta + u_0), \tag{24}$$

and

$$\begin{aligned} E(\hat{y}-y) &= E\left[A(X_0\beta + u_0) - (X\beta + u)\right] \\ &= A(X_0 - X)\beta = A(DX - X)\beta = 0 \end{aligned} \tag{25}$$

The conditions (24) and (25) imply

$$AX_0 - X = 0; \tag{26}$$

$$\hat{y} - y = Au_0 - u.$$

The covariance matrix of $(\hat{y}-y)$ is therefore

$$\begin{aligned} Cov(\hat{y}-y) &= E\left[(\hat{y}-y)'(\hat{y}-y)\right] \\ &= AV_0A' - AV_u - V_uA' + V \end{aligned} \tag{27}$$

where V_0 denotes $E[u_0u_0']$, **V** denotes $E(uu')$ and V_u denotes $E(u_0u')$.

To find the best linear unbiased estimator $\hat{y}$, Chow and Lin minimised the trace of (27) with respect to A subject to the matrix equation $AX_0 - X = 0$. Using a matrix **H** of Lagrange multipliers we form the Lagrangean expression

$$L = \frac{1}{2}tr\left[AV_0A' - AV_u - V_uA' + V\right] - tr\left[H'(AX_0 - X)\right]$$

and setting its partial derivatives with respect to A equal to 0 we obtain

$$AV_0 - V_u = HX_0 \tag{28}$$

Solving (28) for **A** gives $A = HX_0V_0^{-1} + V_uV_0^{-1}$, which, when subsituted in (26) gives the solution for **H**:

$$HX_0'V_0^{-1}X_0 + V_uV_0^{-1}X_0 - X = 0,$$

or

$$H = X\left(X_0'V_0^{-1}X_0\right)^{-1} - \left(V_uV_0^{-1}X_0\right)\left(X_0'V_0^{-1}X_0\right)^{-1}.$$

The solution for **A** is then

$$\begin{aligned} A = X\left(X_0'V_0^{-1}X_0\right)^{-1}X_0'V_0^{-1} + \left(V_uV_0^{-1}\right) \\ \left[\mathrm{I} - X_0\left(X_0'V_0^{-1}X_0\right)^{-1}X_0'V_0^{-1}\right] \end{aligned} \tag{29}$$

The resulting estimator is

$$\hat{y} = Ay_0 = X\hat{\beta} + \left(V_uV_0^{-1}\right)\hat{u}_0, \tag{30}$$

where $\{y_t - p_t\}$

$$\hat{\beta} = \left(X_0'V_0^{-1}X_0\right)^{-1}X_0'V_0^{-1}y_0 \tag{31}$$

Is the generalised least squares estimate of the regression coefficients for the aggregated model, and

$$\begin{aligned}\hat{u}_0 &= \left[I - X_0 \left(X_0' V_0^{-1} X_0 \right)^{-1} X_0' V_0^{-1} \right] y_0 \\ &= y_0 - X_0 \hat{\beta}\end{aligned}$$

represents the $(N \times 1)$ vector of residuals in the aggregated regression.

Since

$$\begin{aligned}V_u &= E(uu'_0) = E\left[(y - X\beta)(y_0 - X_0\beta) \right] \\ &= E\left[(y - X\beta)\left(D_y - DX\beta \right)' \right] \\ &= VD'\end{aligned} \quad ,$$

Then,

$$\begin{aligned}\hat{y} &= X\hat{\beta} + \left(VD'V_0^{-1} \right) \hat{u}_0 , \\ &= X\hat{\beta} + L\hat{u}_0\end{aligned} \qquad (32)$$

with $L = VD'V_0^{-1}$.

We note that the estimate (32) consits of two components:

1. The former $X\hat{\beta}$, applies the estimated regression coefficient to the disaggregated observations of the related variables in order to obtain a preliminary estimate of the dependent variable;

2. The latter, can be interpreted as a term which adjusts the first component by using a weighted combination, whose weights are given by the matrix L, of the estimated residuals of the aggregated regression model.

It is easy to show that the estimates produced by this method fulfill the aggregation constraints:

$$\begin{aligned}D\hat{y} &= DX\hat{\beta} + D\left(VD'V_0^{-1} \right) \hat{u}_0 \\ &= X_0\hat{\beta} + \hat{u}_0 = y_0\end{aligned}$$

Recalling that $V_0 = Cov(Du) = DVD'$.

The covariance matrix of the errors $(\hat{y} - y)$ is (Bournay and Laroque, 1979)

$$\begin{aligned}Cov(\hat{y} - y) &= \mathrm{E}\left[(\hat{y} - y)(\hat{y} - y)' \right] \\ &= (\mathrm{I} - LD)V + \left(X - LX_0 \right) \left(X_0 V_0^{-1} X_0 \right)^{-1} \left(X - LX_0 \right)'\end{aligned}$$

As we can it is formed by two components: the former which depends only on D and V, the latter increases with $(X - LX_0)$. This result can be used to evaluate reliability indicators (Di Fonzo, 1987, p. 44) and confidence intervals.

5.1 Estimation of the covariance matrix of residuals

Like any other generalised least squares estimator, we need to know the covariance matrix V to calculate the estimated series (32). In practice this matrix is unknown and has to be estimated. The usual practice is to assume some structure for the high-frequency disturbances.

The simplest case assumes that the disturbances are serially uncorrelated, each with variance σ^2.
In this case $V = \sigma^2 I$. A more attractive case is that of an autoregressive structure.

5.2 AR(1) model

If the disaggregated disturbances follow a first order autoregressive process

$$u_t = \rho u_{t-1} + v_t, \qquad t = 1, \mathrm{K}, n,$$

with $|\rho| < 1$ and v_t white noise of variance σ_v^2, the variance-covariance matrix is equal to

$$V = \frac{\sigma_v^2}{\left(1 - \rho^2 \right)} \begin{bmatrix} 1 & \rho & \rho^2 & \mathrm{K} & \rho^{n-1} \\ \rho & 1 & \rho & \mathrm{K} & \rho^{n-2} \\ \rho^2 & \rho & 1 & \mathrm{K} & \rho^{n-3} \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ \rho^{n-1} & \rho^{n-2} & \mathrm{K} & \mathrm{K} & 1 \end{bmatrix}.$$

In practice ρ is unknown and has to be estimated in a suitable way. If we suppose that the disaggregated disturbances are normally distributed, we can estimate ρ, β and σ_v^2 using likelihood function

$$\left(2\pi\sigma_v^2\right)^{-N/2}\left|V_0\right|^{-1/2}\exp\left\{-\frac{1}{2\sigma_v^2}\right.$$
$$\left.\left(y_0 - X_0\beta\right)'V_0^{-1}\left(y_0 - X_0\beta\right)\right\}. \qquad (33)$$

As noted by Bournay and Laroque (1979) this corresponds to a problem of maximisation with respect to ρ of the log-likelihood function.

Barbone, Bodo and Visco (1981) proposed to minimise the weighted sum of square residuals, that is

$$\hat{u}_0' V_0^{-1} \hat{u}_0 \qquad (34)$$

which corresponds to the argument of the exponential function in (33). As we can see this is an EGLS approach.

From a technical point of view a scanning procedure is adopted to estimate ρ. We assign a set of values between –1 and 1 to ρ and then we determine $V, \hat{\beta}, \sigma_v^2$ choosing the value which maximises the log-likelihood function (for more details see Di Fonzo, 1987).

5.3 Fernandez's procedure

Fernandez (1981) considered the usual regression model assuming that the disturbance u_t follows a random walk process

$$u_t = u_{t-1} + v_t, \qquad t = 1, \mathrm{K}, n,$$

with v_t white noise with variance σ^2. Thus

$$Cov\left(u_t u_l\right) = \sigma^2 \min\left(t, l\right), \qquad t, l = 1, \mathrm{K}, n.$$

In matrix form it can be showed that

$$V = \sigma^2\left(\Delta_{1,n}'\Delta_{1,n}\right)^{-1}. \qquad (35)$$

Subsituting the right-hand side of (35) in (31) and (32) yields

$$\hat{y} = X\hat{\beta} + \left(\Delta_{1,n}'\Delta_{1,n}\right)^{-1} D'\left[D\left(\Delta_{1,n}'\Delta_{1,n}\right)^{-1} D'\right]^{-1}\hat{u}_0$$

with

$$\hat{\beta} = \left(X_0'\left[D\left(\Delta_{1,n}'\Delta_{1,n}\right)^{-1} D'\right]^{-1} X_0\right)^{-1}$$
$$X_0'\left[D\left(\Delta_{1,n}'\Delta_{1,n}\right)^{-1} D'\right]^{-1} y_0$$

This procedure presents some advantages like the computationally easy procedure which allows to estimate β and σ^2 in a generalised regression model framework with known disturbances covariance matrix.

5.4 Litterman's procedure

Litterman (1983) supposed that the disturbances process is described by

$$u_t = u_{t-1} + e_t, \qquad t = 1, \mathrm{K}, n,$$
$$e_t = \mu e_{t-1} + v_t \qquad t = 1, \mathrm{K}, n,$$

with $|\mu| < 1$ and initial conditions $u_0 = e_0 = 0$. Letting **H** a $(n \times n)$ matrix

$$H = \begin{bmatrix} 1 & 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 \\ -\mu & 1 & 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 \\ 0 & -\mu & 1 & 0 & \mathrm{K} & \mathrm{K} & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 & -\mu & 1 \end{bmatrix},$$

The covariance matrix of **u** is equal to

$$V = \sigma^2\left(\Delta_{1,n}'H'H\Delta_{1,n}\right)^{-1}. \qquad (36)$$

Substituting the right-hand side of (36) in (31) and (32) yields

$$\hat{y} = X\hat{\beta} + \left(\Delta_{1,n}'H'H\Delta_{1,n}\right)^{-1} D'\left[D\left(\Delta_{1,n}'H'H\Delta_{1,n}\right)^{-1} D'\right]^{-1}\hat{u}_0$$

with

$$\hat{\beta} = \left(X_0'\left[D\left(\Delta_{1,n}'H'H\Delta_{1,n}\right)^{-1} D'\right]^{-1} X_0\right)^{-1}$$
$$X_0'\left[D\left(\Delta_{1,n}'H'H\Delta_{1,n}\right)^{-1} D'\right]^{-1} y_0$$

Some computational problems can arise when estimating μ. Di Fonzo (1987) proposed a scanning approach like the one followed for the AR(1) model [4].

5.5 Extrapolation

To solve the extrapolation problem we look for the best linear predictor, according to y_0, of

$$y_{n+j} = X'_{n+j}\beta + u_{n+j}, \qquad j = 1,2,\mathrm{K}$$

with $E(u_{n+j}) = 0$ and $E(u^2_{n+j}) = \sigma^2_{n+j} < +\infty$.

According to sec. 5 the solution is

$$\hat{y}'_{n+j} = x'_{n+j}\hat{\beta} + w'_{n+j}V_0^{-1}u_0,$$

with $w_{n+j}(n\times 1)$ vector of the covariances between u_{n+j} and u_0.

Di Fonzo (1987, p. 55) illustrates how to calculate forecasts according to the chosen estimation procedure. In practice we use the same approach of the estimation problem where we substitute the aggregation matrix **D** with the augmented matrix D_F in (7).

6 Multivariate disaggregation

ECOTRIM permits to estimate *M* high-frequency series which have to satisfy both contemporaneous and temporal aggregation constraints (for technical details see Di Fonzo, 1994). Two different kinds of multivariate procedures have been implemented:

- Adjustment methods;
- Optimal methods.

In the former case the options are the following:

- Denton's multivariate procedure;
- Weighted adjustment procedure;
- Rossi's procedure.

Denton's procedure represents a multivariate version of the univariate one and ECOTRIM offers the following choices for the matrix of weights:

- Denton additive first differences;
- Denton additive second differences;
- Denton proportional first differences;
- Denton proportional second differences.

Rossi's procedure can be viewed as a sub-case of Denton's. The user can supply preliminary series, that satisfy or not the temporal aggregation constraints, otherwise he can calculate them in a preliminary univariate session of ECOTRIM. Note that Rossi's

---

4 The Stram and Wei's procedure offers a general form for the optimal approach which handles with a l arge class of error assumptions that corresponds to the ARIMA models whose sub-cases are the procedures illust rated in this section, unless initial conditions. The optimal solution to the disaggregation problem is form ally equal to (32) with $\hat{\beta}$ and **L** given by

$$\hat{\beta} = \left[\left(\Delta^d_N X_0\right)' \left(C^d V_w \left(C^d\right)'\right)^{-1} \left(\Delta^d_N X_0\right) \right]^{-1} \left(\Delta^d_N X_0\right)' \left(C^d V_w \left(C^d\right)'\right)^{-1} \Delta^d_N y_0 \qquad (37)$$

and

$$L = \begin{bmatrix} & \Delta^d_N & \\ 0 & \mathrm{M} & \mathrm{I}_d \otimes c' \end{bmatrix}^{-1} \begin{bmatrix} V_w \left(C^d\right)' \left(C^d V_w C^d\right)^{-1} \Delta^d_N \\ 0 \quad \mathrm{M} \quad \mathrm{I}_d \end{bmatrix} \qquad (38)$$

We immediately recognise in $\hat{\beta}$ the generalised least squares estimator of β in the regression model

$$\Delta^d_N y_0 = \Delta^d_N X_0 \beta + \Delta^d_N u_0,$$

with error covariance matrix

$$\mathrm{E}\left[\Delta^d_N u_0 (u_0)' \left(\Delta^d_N\right)'\right] = \mathrm{E}\left[\left(C^d w w' \left(C^d\right)'\right)\right] = C^d V_w \left(C^d\right)'.$$

and weighted adjustment procedures require preliminary series that fulfill the temporal aggregation constraints while Denton's do not. In fact both methods force the estimated series only to satisfy the contemporaneous constraints and therefore, in order to fulfill the temporal constraints too, the preliminary series must already fulfill them. On the other hand, Denton's procedure force the estimated series to satisfy both the aggregation constraints.

The adjustment step is related to a high-frequency series, say z, with which the estimated series have to be coherent. Then the adjustment procedures modulate the discrepancies between this contemporaneous aggregated series and the sum of the available preliminary series. As far as the multivariate methods are concerned, ECOTRIM has two options:

- white noise:
- random walk,

that can be viewed as a straightforward extension of the univariate counterpart. Multivariate AR(1) models are being developed to enlarge this section of the program. Finally, in the multivariate case too, ECOTRIM offers the extrapolation option as regards the optimal methods and the weighted adjustment method.

7 Concluding remarks

In this paper I have presented the techniques offered by ECOTRIM to estimate the unknown originally series by using the available information. Both univariate and multivariate issues have been considered and the related solutions that ECOTRIM offers have been analyzed.

According to the disaggregation method chosen, ECOTRIM can offer further disaggregation information such as diagnostics.

In particular, if we use an ARIMA model based approach we obtain, besides the estimated disaggregated series, an estimated ARIMA model for the high-frequency process and an estimate of the covariance matrix of estimated values. Note that we do not give a significance level for the ARlMA parameters since, due to the non-linear relationships which link aggregated and disaggregated models, it is straightforward to estimate their covariances using the corresponding aggregated information.

The optimal approach supplies much more information about regression results, such as regression diagnostics, significance parameter levels, confidence intervals and some other useful statistics. Similar indicators are obtained for the multivariate optimal approach.

Mathematical and adjustment procedures neither allow to calculate diagnostic nor permit to get any extrapolation, except for the multivariate adjustment method as stated in sec. 6.1.1.

The choice of the technique to be used in practice is devolved to the analyst. He must consider also the computational time required, besides the information at disposal and the problem he is facing with.

If related series are available the optimal approach seems to offer many advantages: diagnostics, several approaches, extrapolation. If it is not, ARIMA model based procedures seem to offer the same opportunities (extrapolation, several approaches, some diagnostics) but they require a preliminary analysis in order to estimate the ARIMA aggregated model. This may represent a problem if the analysis must be developed in an automatic way.

References

Al-Osh, M. (1989) "A dynamic linear model approach for disaggregating time series data", Journal of Forecasting. 8, pp. 65-96. '

Barbone, L., Bodo. G. and Visco, L. (1981) , "Costi e profitti in senso stretto: un'analisi su serie trimestrali, 1970-1980", Bollettino delta Banca d'ltalia, 36, numero unico, pp. 465-510.

Barcellan. R. and Di Fonzo, T. (1993) , Temporal aggregation and time series: a survey, unpublished paper

Barcellan, R. and Di Fonzo, T. (1994) , Disaggregazione temporale e modelli ARIMA", Atti della XXXVII Riunione Scientifica della S.I.S., San Remo 6-8th April 1994, vol. II, pp. 355-362,

Bassie, V.L. (1958) , Economic Forecasting, New York, Mc Grow-Hill

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

Bournay, J. and Laroque, G. (1979) , "Reflexions sur la méthode d'elaboration des comptes trimestriels", Annales de I'INSEE, 36, pp. 3-30.

Box, G.E.P. and Jenkins, G.M. (1976) , Time series analysis, forecasting and control, San Francisco. Holden-Day.

Cholette, P.A. (1988), "Benchmarking and interpolation of time series"', Working Paper, Time Series Research and Analysis Division, Statistics Canada.

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

Cohen, G.C., Müller. W.M. and Padberg, M.W. (1971), "Autoregressive approaches to disaggregation of time series data", Applied Statistics, 20 (2), pp. 119-129.

Denton, F.T. (1971) , "Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimisation", Journal of the America Statistical Association, 66 (333), pp. 99-102.

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

Di Fonzo. T. (1990a) , "The estimation of M disaggregated time series when contemporaneous and temporal aggregates are known"', The Review of Economics and Statistics, 72 (1), pp. 178-182

Di Fonzo. T. (1910b) , "Stima indiretta di pill serie economiche legate da un vincolo contabile", Statistica Applicata, vol. 2, n. 2, pp. 149-161.

Di Fonzo, T. (1994) , '"Temporal disaggregation of a system of time series when the aggregate is known: optimal vs adjustment methods", Proceedings of National Quarterly Accounts Workshop, Paris 5-6th December 1994.

Fernandez. R.B. (1981) , "A methodological note on the estimation of time series", The Review of Economics and Statistics, 63 (3), pp. 471-478.

Ginsburgh, V.A. (1973) , "A further not on the derivation of quarterly figures consistent with annual data", Applied Statistics, 22 (3), pp. 368-374.

Goldberger, A. S. "Best linear unbiased prediction in the generalised linear regression model", Journal of the American Statistical Association 57, pp. 369-375.

Guerrero, V.M. (1990) , "Temporal disaggregation of time series: an ARIMA-based approach", International Statistical Review, 58 (1), pp. 29-46.

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

Litterman, R.B. (1983) , "A random walk, Markov model for the distribution of time series". Journal of Business and Economic Statistics, 1 (2), pp. 169-173.

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

Stram, D.O. and Wei, W.W.S. (1986a) , "Temporal aggregation in the ARIMA process", Journal of Time Series Analysis, 7 (4), pp. 279-292.

Stram, D.O. and Wei, W.W.S. (1986b) , "A methodological note on the disaggregation of time series totals", Journal of Time Series Analysis, 7 (4), pp. 293-302.

Vangrevelinghe, G. (1966) , "L'évolution a court terme de la consommation des menages: connaissance, analyse et prevision", études et Conjoncture, 9, pp. 54-102.

Wei, W.W.S. and Stram, D.O. (1990) , "Disaggregation of time series models", Journal of the Royal Statistical Society, B, 52 (3), pp. 453-467.

Estimating observations in ARIMA models with the Kalman filter

Víctor GÓMEZ

Two problems regarding missing observations are considered. The first concerns the estimation of missing observations in time series. The second concerns the disaggregation of time series totals, like, for example, the disaggregation of annual time series data to quarterly figures. Both problems can be solved by setting up the model in state space form and applying the Kalman filter.

KEY WORDS: Time Series, Missing Observations, Nonstationarity, ARIMA Models, Temporal Aggregation

1 Introduction

This paper considers two problems that are usually encountered when dealing with economic time series. The first concerns the interpolation of missing observations in a series, while the second has to do with the distribution of time series data subject to temporal aggregation.

As an example of the first problem, one can think of a stock variable, such as the money supply, which is published at monthly intervals but is only available on a quarterly basis. An example of the second problem is a flow variable, like investment, which is published at quarterly intervals but is only available on an annual basis. Note that a stock is the quantity of something at a particular point in time, while a flow is a quantity that accumulates over a given period of time.

For the problem of missing observations on a stock variable, the methodology of Gomez and Maravall (1994b) is used, which is implemented in a program called "Time Series Regression with ARIMA noise, Missing observations and Outliers" (or TRAMO, in short). This methodology is a generalization to non-stationary series of the skipping strategy of Jones (1980) for ARMA models. It uses the Kalman filter for prediction and likelihood evaluation, skipping the missing observations whenever they are encountered. In this way, estimation of model parameters is possible and then, at a later stage, the fixed point smoother can be used to interpolate unobserved values of the series. The methodology can be easily extended to the case of regression models with ARIMA noise and missing observations. Initial missing observations (i.e., missing observations among the first observations of the series lost by differencing) are considered fixed and are, therefore, treated as regression parameters.

To handle the problem of temporal aggregation of a flow variable, the same methodology of Gomez and Maravall (1994b) is applicable. Only a change in the state space representation is needed. To illustrate, a program in Fortran has been written and applied to some regression models with ARIMA noise. It is to be noted that this methodology does not require the construction and inversion of large covariance matrices. Also, the specification of the model for the noise is not constrained to an AR(1) process or a random walk, like in Chow and Lin (1971, 1976), Denton (1971) and Fernandez (1981).

2 State space representation and the Kalman filter

2.1. State space representation

Le the process $\{z_t\}$ follow the ARIMA (p, d, q) model

$$\phi(L)\alpha(L)z_t = \theta(L)a_t,$$

Where L is the lag operator, $L(z_t) = z_{t-1}$, all the roots of the polynomial in L $\alpha(L) = 1 + \alpha_1 L + \cdots + \alpha_d L^d$ are in the unit circumference, all the roots of $\phi(L) = 1 + \phi_1 L + \cdots + \phi_p L^p$ are outside the unit circle and those of $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$ are outside the unit circle or on the unit cirumference. It is assumed tha t $\phi(l), \alpha(l)$ and $\theta(L)$ have no common factors and that $\{a_t\}$ is a sequence of i.i.d. $N(0, \sigma^2)$ Variables. Let $r = \max\{p + d, q + 1\}$ and define $\phi^*(L) = \phi(L)\alpha(L)$, $\psi^*(L) = \theta(L)/\phi^*(L) = \sum_{i=0}^{\infty} \psi_i^* L^i$ and $\phi_i = 0$ if $i > p$. Then, one state space representation for $\{z_t\}$ which has minimum dimension is given by

$$x_t = F x_{t-1} + G a_t \qquad (2.1a)$$
$$z_t = H' x_t \qquad (2.1b)$$

where $t = 1, \ldots, N$, $x_t = (z_t, z_{t+1,t}, \ldots, z_{t+r-1,t})'$, $G = (1, \psi_1^*, \ldots, \psi_r^*)'$, $H' = (1, 0, \ldots, 0)$,

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\phi_r^* & -\phi_{r-1}^* & -\phi_{r-2}^* & \cdots & -\phi_1^* \end{bmatrix},$$

and $z_{t+i,t} = z_{t+i} - \psi_0^* a_{t+i} - \cdots - \psi_{i-1}^* a_{t+1}$, $i = 1, \ldots, r-1$.

The expression $z_{t+i,t}$ denotes the prediction of z_{t+i} based on $\{z_s : s \le t\}$. Thus, the state vector x_t contains z_t and its (r-1)-periods-ahead forecast function with respect to the semi-infinite sample $\{z_s : s \le t\}$. Lemma 3 of Gómez and Maravall (1994b) ensures that the state space representation (2.1) is correct.

The state space representation (2.1) is defined in terms of the original series. This will allow for the use of the Kalman filter for prediction and likelihood evaluation, and the fixed point smoother for interpolation, even in the case where there are missing observations.

The following example will be used throughout the paper to illustrate matters.

$$z_t - z_{t-4} = a_t + \theta a_{t-1}.$$

For this model, $G = (1, \theta, 0, 0)'$ and

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

2.2 Prediction error decomposition and the Kalman filter

Suppose first that there are no missing observations in the series $z = (z_1, \ldots, z_n)'$. Letting $u = (u_{d+1}, \ldots, u_N)'$, where $u_t = \alpha(L)z_t$ denotes the transformation that renders $\{z_t\}$ stationary, the likelihood of the ARIMA model is usually defined as the likelihood of the defferenced series, $L(u)$ (see Box and Jenkins 1976, chap. 7). In order to define the likelihood in terms of the original series, the following two assumptions will be made:

Assumption A:
the variables $\{z_1, \ldots, z_d\}$ are independent of the variables $\{u_t\}$.

Assumption B:
The variables $\{z_1, \ldots, z_d\}$ are jointly normally distributed.

The first assumption is a starndard one when forecasting with ARIMA models (see Brockwell and Davis, 1992, pp. 314-317).

It is shown in Gómez and Marvall (1994b), pg. 614, that, under assumptions A and B, the conditional density

$$L(z_{d+1}, \ldots, z_N | Z_d), \qquad (2.2)$$

Where $Z_d = \{z_1, \ldots, z_d\}$, is a well defined density and coincides with the Box-Jenkins likelihood $L(u)$. The conditional likelihood (2.2) admits the prediction error decomposition

$$L(z_{d+1},\K,z_N|Z_d)=\prod_{t=d+1}^{N}L(z_t|z_{t-1},\K,z_1),$$

which in turn suggests the application of the Kalman filter to the state space representation (2.1) to evaluate (2.2). Following Bell (1984), pg. 650, the variable z_t can be expressed as

$$z_t=A'_t\,z_I+\sum_{i=0}^{t-d-1}\xi_i u_{t-1},\qquad t>d,$$

where $z_I=(z_1,\K,z_d)'$, the $A_t=(A_{1t},\K,A_{dt})'$ can be calculated recursively from

$$\begin{aligned}&A_{ij}=\delta_{ij,}\qquad i,j=1,\K,d,\\&A_{it}=-\alpha_1A_{it-1}-\K-\alpha_dA_{it-d},\quad t>d,\quad i=1,\K,d,\end{aligned}\tag{2.3}$$

δ_{ij} Is the Kronecker delta and $\xi(L)=1/\alpha(L)=\sum_{i=0}^{\infty}\xi_iL^i$. From this, it is obatined that $x_{d+1}=A_Iz_i+\Xi U$, where $A_I=[A_{d+1},\K,A_{d+r}]'$, Ξ is the lower triangular $r\times r$ matrixwith rows the vectors $(\xi_{j-1},\xi_{j-2},\K,1,0,\K,0)$, $j=1,\K,r$, $U=(u_{d+1},u_{d+2,d+1},\K,u_{d+r,d+1})'$ and $u_{d+i,d+1}=E(u_{d+i}|u_s:s\le d+1)$. The Kalman filter can then be initialized with

$$\hat{x}_{d+1,d}=A_Iz_I\ \text{ and }\ \Sigma_{d+1,d}=\Xi E(UU')\Xi',\tag{2.4}$$

where $E(UU')$ can be computed from the stationary process $\{u_t\}$ as in Jones (1980), $\hat{x}_{d+1,d}=E(x_{d+1}|Z_d)$ and $\Sigma_{d+1,d}=Var(x_{d+1}-\hat{x}_{d+1,d})$.

For the example of Section 2.1, it is not difficult to check that $z_I=(z_1,z_2,z_3,z_4)'$and

$$A_I=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},$$

$$\Sigma_{5,4}=\Xi E(UU')\Xi'=I_4E(UU')I_4=$$
$$=\begin{bmatrix}1+\theta^2&\theta&0&0\\\theta&\theta^2&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}$$

2.3 Missing observations

Suppose now that there are missing observations in the series $z=(z_1,\K,z_N)'$ and let $z_0=(z_{t_1},\K,z_{t_M})'$, $1\le t_1,\K,t_M\le N$, be the observed series. Than the observation equation (2.1b) can be replaced by

$$z_t=H'x_t+\gamma_tW_t,\qquad t=1,\K,N\tag{2.5}$$

where $H'_t=(1,0,\K,0)$, $\gamma_t=0$ if z_t is observed, and $H'_t=(0,0,\K,0)$, $\gamma_t=1$ if z_t is missing (Brockwell and Davis 1992, p. 483). The variable W_t represents a sequence of i.i.d. $N(0,1)$ variables independent of z_t, $t=1,\K,d$ and a_t, $t=0,\pm1,\K$.

If there are no missing observations among the first d values of the series, the Kalman filter can be applied to the state space representation given by (2.1a) and (2.5), with starting conditions (2.4), like in Jones (1980) to evaluate the likelihood. At a later stage, after having estimated all unknown parameters in the model, the fixed point smoother can be used to interpolate the missing observations.

The general case, where there may be missing observations among the first d values of the series, can be handled as follows. Let $z_{I_0}=(z_{t_1},\K,z_{t_k})'$ be the vector of observations in $z_I=(z_1,\K,z_d)'$. If z_{Im} denotes the vector of missing observations in z_I, then the observed values in the series z can be expressed as

$$\begin{aligned}z_{t_i}&=A'_{t_i}\,z_I\\&=B'_{t_i}\,z_{I0}+C'_{t_i}\,z_{t_i},\quad i=1,\K,k,\end{aligned}$$

and

$$\begin{aligned}z_{t_i}&=A'_{t_i}\,z_I+\tilde{u}_{t_i}\\&=B'_{t_i}\,z_{I0}+C'_{t_i}\,z_{\mathrm{Im}}+\tilde{u}_{t_i},\quad i=k+1,\K,M,\end{aligned}$$

where $\tilde{u}_t=\sum_{j=0}^{t-d-1}\xi_ju_{t-j}$, $t>d$, the vectors A'_s have been defined in (2.3), and B'_s and C'_s are the subvectors of A'_s such that $A'_s\,z_I=B'_s\,z_{I0}+C'_s\,z_{\mathrm{Im}}$, $s>0$. Let z_{II} and $\tilde{u}$ denote the vectors $(z_{t_{k+1}},\K,z_{t_M})'$ and $(\tilde{u}_{t_{k+1}},\K,\tilde{u}_{t_M})'$, and let B and C denote the $(M-k)\times k$ and $(M-k)\times(d-k)$ matrices with rows B'_s and C'_s, $s=t_{k+1},\K,t_M$. Then, the preceding equations can be written as

$$z_0 = \begin{bmatrix} z_{I0} \\ z_{II} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ B & C \end{bmatrix} \begin{bmatrix} z_{I0} \\ z_{\mathrm{Im}} \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{u} \end{bmatrix}. \qquad (2.6)$$

Defining $y = z_{II} - Bz_{I0}$, (2.6) implies that

$$y = Cz_{\mathrm{Im}} + \tilde{u}. \qquad (2.7)$$

A natural way of extending the definition of the conditional likekihood (2.2) to the case of missing observations is to consider the likelihood of the observations z_{II} conditional on $Z_d = \{z_1, \mathsf{K}, z_d\}$ and to treat the vector of initial missing observations z_{Im} as additional parameters. This is the definition of Gómez and Maravall (1994b). It is equivalent to considering (2.7) as a regression model whose errors $\tilde{u}$ have a known covariance matrix $\sigma^2\Delta$. The likelihood associated with (2.7) is

$$\frac{1}{(2\pi\sigma^2)^{(M-k)/2}} |\Delta|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y - Cz_{\mathrm{Im}})'\Delta^{-1/2}(y - Cz_{\mathrm{Im}})\right) \qquad (2.9)$$

Where the unknown parameters are $\sigma^2, z_{\mathrm{Im}}$ and the coefficients $(\phi, \theta) = (\phi_1, \mathsf{K}, \phi_p, \theta_1, \mathsf{K}, \theta_p)$ of the ARIMA model.

The parameters σ^2 and z_{Im} can be concentrated out of the likelihood (2.9) using the Kalman filter. Specifically, let $\Delta = LL'$, whit L lower triangular, be the Cholesky decomposition of Δ and suppose that $z_{\mathrm{Im}} = 0$. The Kalman filter, applied as in Hones (1980) to (2.1a) and (2.5), with starting conditions (2.4), yields $L^{-1}y$ and $|L|$. Note that $\hat{x}_{d+1,d} = [B_{d+1}, \mathsf{K}, B_{d+r}]' z_{I0}$ in this case because of the assumption $z_{\mathrm{Im}} = 0$ (see Gómez and Marvall,(1994b). The same algorithm applied to the columns of the matrix C, with starting conditions $\hat{x}_{d+1,d} = 0$ and $\Sigma_{d+1,d}$ as given by (2.4), also permits the computation of $L^{-1}C$. Then, it is possible to move from the generalised least squares regression model (2.7) to the model

$$L^{-1}y = L^{-1}Cz_{\mathrm{Im}} + L^{-1}\tilde{u}, \qquad (2.10)$$

where $Var(L^{-1}\tilde{u}) = \sigma^2 I_{M-k}$. Therefore, (2.10) is an ordinary least squares regrerssion model. The maximum likelihood estimators $\hat{\sigma}^2$ and $\hat{z}_{\mathrm{Im}}$ of σ^2 and z_{Im} in (2.7) can now be efficiently and accurately obtained using the QR algorithm. Supposing C is of full column randk, this last algorithm premultiplies both $L^{-1}y$ and $L^{-1}C$ by an orthogonal matrix Q to obtain $v = QL^{-1}y$ and $(U', 0')' = QL^{-1}C$, where U is a nonsingular $(d-k)\times(d-k)$ upper triangular matrix. Then $z_{\mathrm{Im}} = U^{-1}v_1$ and $\hat{\sigma}^2 = v'_2 v_2/(M-k)$, where $v = (v'_1, v'_2)'$, v_1 has dimension d-k and v_2 has dimension M-d. Replacing σ^2 and z_{Im} in (2.9) by its estimators $\hat{\sigma}^2$ and $\hat{z}_{\mathrm{Im}}$, yields the concentrated likelihood. It is not difficult to verify that maximizing the concentrated likelihood with respect to the parameters (ϕ, θ) is equivalent to minimizing the non-linear sum of squares

$$S(\phi, \theta) = |L|^{1/(M-k)} v'_1 v_2 |L|^{1/(M-k)}.$$

Minimising $S(\phi, \theta)$, one can obtain the estimators $(\hat{\phi}, \hat{\theta})$; then, $\hat{\sigma}^2$ and $\hat{z}_{\mathrm{Im}}$ are estimated by $\hat{\sigma}^2$ and $\hat{z}_{\mathrm{Im}}$.

Once the parameters of the model have been estimated, one can use the Kalman filter for prediction and the fixed point smoother for interpolation of the missing observations z_t with $t_k < t \le N$. Note that the missing observations z_t with $1 < t \le t_k$ are estimated by $\hat{z}_{\mathrm{Im}}$. See Gómez and Maravall (1994b), pg. 617, for the details. See also Kohn and Ansley (1985).

Suppose that for the example of Section 2.1 the observed series is $z_o = (z_1, z_4, z_5, z_6, z_7, z_8, z_{10})'$. Then, $z_{Io} = (z_1, z_4)'$, $z_{\mathrm{Im}} = (z_2, z_3)'$, $z_{II} = (z_5, z_6, z_7, z_8, z_{10})'$ and

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

To obtain $L^{-1}y$, the Kalman filter should be initialised with $\hat{x}_{5,4} = (z_1, 0, 0, z_4)'$ and $\Sigma_{5,4}$ as given in Section 2.2.

2.4 Regression Models with Missing Observations and ARIMA errors

The methodology described in the preceding Sections can be extended easily to the case of regression models with missing observations and ARIMA errors. Consider the regression model

$$z_t = \omega'_t \beta v_t, \qquad (2.11)$$

where $\beta = (\beta_1, \ldots, \beta_h)'$ is a vector of parameters, ω_t' is a vector of h independent variables, z_t is the dependent variable, and ν_t is assumed to follow the ARIMA model

$$\phi(L)\alpha(L)\nu_t = \theta(L)a_t \qquad (2.12)$$

The state space reprensentation is given by (2.1a), where x_t is given by the state vector of Section 2.1 with z_t replaced by ν_t, and the observation equation

$$z_t = \omega_t'\beta + H'_t x_t + \gamma_t W_t, \qquad t = 1, \ldots, N,$$

where H'_t and γ_t are as in Section 2.3. Let $z_o = (z_{t_1}, \ldots, z_{t_M})'$, $1 \le t_1, \ldots, t_M \le N$, z_I, z_{Io}, z_{Im} and z_{II} be as in Section 2.3 and define the corresponding vectors $\nu_{Io}, \nu_{\mathrm{Im}}$ and ν_{II}. Let W_{Io}, W_{Im} and W_{II} be the matrices with rows the vectors ω_t' corresponding to the vectors $\nu_{Io}, \nu_{\mathrm{Im}}$ and ν_{II}. Then, with the notation of Section 2.3, one can write $\nu_{II} = B\nu_{Io} + C\nu_{\mathrm{Im}} + \tilde{u}$ and, replacing ν_{Io} by $z_{Io} - W_{Io}\beta$, ν_{Im} by $z_{\mathrm{Im}} - W_{\mathrm{Im}}\beta$ and ν_{II} by $z_{II} - W_{II}\beta$, the following regression model is obtained

$$z_{II} = Bz_{Io} + Cz_{\mathrm{Im}} + W_{II}\beta - BW_{Io}\beta - CW_{\mathrm{Im}}\beta + \tilde{u},$$

where the regression parameters are z_{Im} and β. Letting $y = z_{II} - Bz_{Io}$, this model can be rewritten as

$$\begin{aligned} y &= [C, W_{II} - BW_{Io} - CW_{\mathrm{Im}}]\,[z'_{\mathrm{Im}}, \beta']' + \tilde{u} \\ &= [C, W_{II} - A_{II}W_I]\,[z'_{\mathrm{Im}}, \beta']' + \tilde{u}, \end{aligned} \qquad (2.13)$$

where W_I is the $d \times h$ matrix formed by the rows ω_t', t=1,...,d, and A_{II} is the $(M-k)\times d$ matrix with rows the vectors A'_s, $s = t_{k+1}, \ldots, t_M$ defined in (2.3). The parameters σ^2, z_{Im} and β can be concentrated out of the likelihood. Specifically, the Kalman filter applied to the model $y = z_{II} - Bz_{Io} = \tilde{u}$, which coincides with (2.13) under the assumption that $z_{\mathrm{Im}} = 0$ and $\beta = 0$, yeilds $L^{-1}y$ and $|L|$, where L is as in Section 2.3. The starting conditions to use are $\hat{x}_{d+1,d} = [B_{d+1}, \ldots, B_{d+r}]'z_{Io}$ and $\Sigma_{d+1,d}$ as given by (2.4). The same algorithm applies to the columns of the matrix $[C, W_{II} - A_{II}W_I]$ with starting conditions $\hat{x}_{d+1,d} = 0$ and $\Sigma_{d+1,d}$ as given by (2.4), permits the caluclation of the product of L^{-1} by this matrix. Then the *QR* algorithm can be applied to the transformed model

$$L^{-1}y = L^{-1}[C, W_{II} - A_{II}W_I]\,[z'_{\mathrm{Im}}, \beta']' + L^{-1}\tilde{u},$$

and one can proceed as described after having obatined (2.10) in Section 2.3.

3 Temporal Aggregation

Consedr again the regression model (2.11), where the errors $\{\nu_t\}$ follow the ARIMA model (2.12). Suppose that for $t = t_1, \ldots, t_M$, the aggregate

$$\tilde{z}_t = \sum_{j=1}^{n(t)} z_{t-j+1} \qquad (3.1)$$

is observed, where $1 \le n(t) \le n$ and n is the maximum number of time periods over which the flow variable z_t is aggregated. Let $\tilde{r} = r + n - 1$, where r is as in Section 2.1, and define the $\tilde{r}$ dimensional state vector $\tilde{x}_t = (\nu_{t-n+1}, \ldots, \nu_{t-1}, x'_t)'$, where x_t is the state vector of Section 2.1 with z_t replaced by ν_t. Then, one state space representation is given by

$$\begin{aligned} \tilde{x}_t &= \tilde{F}\tilde{x}_{t-1} + \tilde{G}a_t \\ \tilde{z}_t &= \sum_{j=1}^{n(t)} \omega_{t-j+1}\beta + \tilde{H}'_t \tilde{x}_t + \gamma_t W_t, \end{aligned}$$

Where $\tilde{G} = (0, \ldots, 0, G')'$, G is as in Section 2.1, $\tilde{H}'_t = (0, \ldots, 0, 1, 1, 0, \ldots, 0)$, $\gamma_t = 0$ if $\tilde{z}_t$ is observed, $\tilde{H}'_t = 0, \gamma_t = 1$ if $\tilde{z}_t$ is missing, W_t is a sequence of i.i.d. $N(0,1)$ variables independent of ν_t, t=1,...,d and a_t, $t = 0, \pm 1, \ldots$,

$$\tilde{F} = \begin{bmatrix} 0 & & 0 & \cdots & 0 \\ 0 & 0 & & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\phi^*_{\tilde{r}} & -\phi^*_{\tilde{r}-1} & -\phi^*_{\tilde{r}-2} & \cdots & -\phi^*_1 \end{bmatrix},$$

And the ϕ_i^* are as in Section 2.1. Note that $\phi_i = 0$ if $i > p$, which implies $\phi_j^* = 0$ if $j > p + d$, and that there are $n(t)$ ones in $\tilde{H}_t \neq 0$. When $q = 0$, a more economical state space representation is obtained by redefining $\tilde{r}$ as $\max\{p + d, n\}$ and letting the state vector be $\tilde{x}_t = (\nu_{t-\tilde{r}+1}, \ldots, \nu_t)'$ and $\tilde{G} = (0, \ldots, 0, 1)'$.

As in Section 2.3, the likelihood is defined by conditioning on $Z_d = \{z_1, \mathsf{K}, z_d\}$. In order to get a regression model similar to (2.13), whose likelihood will be the likelihood of the aggregated series, suppose first that $\beta = 0$ and let $z = (z_1, \mathsf{K}, z_N)'$ be the disaggregated series. Then, $z = Az_I = [0, \tilde{u}']'$, where A is the $N \times d$ matrix with rows the vectors A'_t, $t = 1, \mathsf{K}, N$, defined in Section 2.3, $\tilde{u}$ is the $(N-d)\times 1$ vector with elements $\tilde{u}_t, t = d+1, \mathsf{K}, N$, also defined in Section 2.3, and $z_I = (z_1, \mathsf{K}, z_d)'$. The observed aggregated series $\tilde{z} = (\tilde{z}_{t_1}, \mathsf{K}, \tilde{z}_{t_M})'$ is

$$\tilde{z} = Jz = JAz_I + J[0, \tilde{u}']', \tag{3.2}$$

where the rows of the matrix J are formed with zeros and ones. For example, in the case of a quarterly series aggregated to yearly totals, if $N = 4M$, the J is the $M \times 4M$ matrix

$$J = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & \mathsf{K} & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & \mathsf{K} & 0 & 0 & 0 & 0 \\ \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{O} & \mathsf{M} & \mathsf{M} & \mathsf{M} & \mathsf{M} \\ 0 & 0 & 0 & 0 & 0 & \mathsf{K} & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Equation (3.1) should be used each time there is an observation $\tilde{z}_t$ with $t \le d$ to reduce the number of initial missing values z_1, K, z_d, which are considered fixed and are treated as parameters. That is, each time there is an observation $\tilde{z}_t$ with $t \le d$, equation (3.1) should be used to put one of the initial missing values as linear combination of $\tilde{z}_t$ and the other initial missing values. Let $\tilde{z}_I = (\tilde{z}_{t_1}, \mathsf{K}, \tilde{z}_{t_k})'$, where $t_k \le d$ and $t_{k+1} > d$. By selecting k appropriate variables in z_I, it is always possible to find a representation of the form $\tilde{z}_I = z_{I1} + Pz_{I2}$, where P is a $k \times (d-k)$ matrix and z_{I1} and z_{I2} are $k \times 1$and $(d-k)\times 1$subvectors of z_I which have no elements in common. Then one can write $z_{I1} = \tilde{z}_I - Pz_{I2}$. Let A_1 and A_2 be the submatrices of A such that $Az_I = A_1 z_{I1} + A_2 z_{I2}$ and define $\tilde{z}_{II} = (\tilde{z}_{t_{k+1}}, \mathsf{K}, \tilde{z}_{t_M})'$. Partition $J = [J'_I, J'_{II}]'$ conforming to $\tilde{z} = (\tilde{z}'_I, \tilde{z}'_{II})'$. Then,

$$\tilde{z} = Jz = \begin{bmatrix} \tilde{z}_I \\ \tilde{z}_{II} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ \tilde{B} & \tilde{C} \end{bmatrix} \begin{bmatrix} \tilde{z}_I \\ z_{I2} \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{\tilde{u}} \end{bmatrix},$$

where $\tilde{B} = J_{II}A_1$, $\tilde{C} = J_{II}A_2 - J_{II}A_1 P$ and $\tilde{\tilde{u}} = J_{II}[0, \tilde{u}']'$. Defining $\tilde{y} = \tilde{z}_{II} - \tilde{B}\tilde{z}_I$, the following regression model is obtained

$$\tilde{y} = \tilde{C}z_{I2} + \tilde{\tilde{u}}. \tag{3.3}$$

If $t_1 > d$, then $\tilde{y} = \tilde{z}, \tilde{C} = JA$ and $z_{I2} = z_I$ in (3.3). For the general case, where $\beta \neq 0$, an argument similar to that used in Section 2.4 to obtain (2.13) leads to

$$\tilde{y} = [\tilde{C}, \tilde{W}_{II} - \tilde{A}_{II}W_I][z'_{I2}, \beta']' + \tilde{\tilde{u}}, \tag{3.4}$$

where $\tilde{W}_{II} = J_{II}W$, $\tilde{A}_{II} = J_{II}A$, W is the matrix with rows the vectors ω'_t of (2.11), $t = 1, \mathsf{K}, N$, and W_I is, like in Section 2.4, the submatrix of W formed withits first d rows. Again, if $t_1 > d$, , then $\tilde{y} = \tilde{z}$, $\tilde{C} = JA$ and $z_{I2} = z_I$ in (3.4).

In the caseof missing observation sof a stock variable, the Kalman filter was always initialize at $t = s$, where $s = \max\{t_1, d+1\}$. The initial state vector is

$$\tilde{x}_s = \tilde{A}_I \nu_I + \tilde{\Xi}\tilde{U}, \tag{3.5}$$

where $\nu_I = (\nu_1, \mathsf{K}, \nu_d)'$ and the matrices $\tilde{A}_I, \tilde{\Xi}$ z and the vector $\tilde{U}$ can be constructed in a similar manner to that used in Section 2.2 to build up the matrices A_I, Ξ and the vector U. On very rare occasions, when some first elements of the intial state vector have a non -positive temporal subindex, Bell's backward representation (Bell, 1984) will be needed as well as the forward representation (2.3). Replacing $\nu_I$with $z_I - W_I\beta$ in (3.5), yields

$$\begin{aligned} \tilde{x}_s &= \tilde{A}_I z_I - \tilde{A}_I W_I \beta + \tilde{\Xi}\tilde{U} \\ &= \tilde{A}_{I1}\tilde{z}_I + (\tilde{A}_{I2} - \tilde{A}_{I1}P)z_{I2} - \tilde{A}_I W_I \beta + \tilde{\Xi}\tilde{U}, \end{aligned}$$

where $\tilde{A}_{I1}$ and $\tilde{A}_{I2}$ are the submatrices of $\tilde{A}_I$ formed with the columns of $\tilde{A}_I$ which correspond to the subvectors z_{I1} and z_{I2} of z_I. From this last expression , initial conditions for the Kalman filter can be obtained.

For likelihood evaluation, one can proceed like in Section 2.4. The parameters σ^2, z_{I2} and β can be concentrated out of the likelihood. The Kalman filter to be applied to the model $\tilde{y} = \tilde{\tilde{u}}$, which coincides with (3.4) under the assumption that $z_{I2} = 0$ and $\beta = 0$, should be initialised with $\hat{\tilde{x}}_{s,s-1} = \tilde{A}_{I1}\tilde{z}_I$ and $\tilde{\Sigma}_{s,s-1} = \tilde{\Xi}E(\tilde{U}\tilde{U}')\tilde{\Xi}'$. The starting conditions for the Kalman filter to be applied to the columns of

$[\tilde{C}, \tilde{W}_{II} - \tilde{A}_{II} W_I]$ are $\hat{\tilde{x}}_{s,s-1} = 0$ and $\tilde{\Sigma}_{s,s-1}$. After all unknown parameters in the model have been estimated, the Kalman filter can be used for prediction of the future observations and the fixed point smoother for interpolation of the missing observations. Note that the initial missing observations are estimated by linear regression.

Suppose that a quarterly series follows the model of the example of Secrion 2.1 and it is aggregated to yearly totals. Then, $n = 4$, $\tilde{r} = 7$, and if $x_t = (z_t, z_{t+1,t}, z_{t+2,t}, z_{t+3,t})'$ is the state vector corresponding to the disaggregated series, the state vector for the aggregated data is $\tilde{x}_t = (z_{t-3}, z_{t-2}, z_{t-1}, x'_t)'$. The Kalman filter should be initialized at $s = 5$ and

$$\tilde{x}_5 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & I_4 \end{bmatrix} \begin{bmatrix} 0 \\ v \end{bmatrix}, \qquad (3.6)$$

where $v = (u_5, u_6, u_7, u_8)'$. Given that $\tilde{z}_4 = \sum_{j=1}^{4} z_{5-j}$, one can put, for example, $z_1 = \tilde{z}_4 - \sum_{j=1}^{3} z_{5-j}$ to reduce the number of initial missing observations. Substituting back in (3.6), initial conditions for the Kalman filter can be obtained. The initial conditions for the Kalman filter to be applied to the model $\tilde{y} = \tilde{\tilde{u}}$, which is (3.3) under the assumption $z_{I2} = 0$, are $\hat{\tilde{x}}_{5,4} = (0,0,0,\tilde{z}_4,0,0,0)'$ and

$$\tilde{\Sigma}_{5,4} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{5,4} \end{bmatrix}, \qquad (3.7)$$

Where $\Sigma_{5,4}$ is the matrixgiven in Section 2.2. The Kalman filter to be applied to the columns of the matrix $\tilde{C}$ in the regression model (3.3), where $z_{I2} = (z_2, z_3, z_4)'$ are $\hat{\tilde{x}} = 0$ and (3.7).

4 Applications

The methodology of Gómez and Maravall (1994b) to handle the proble of missing observations on a stock variable, as described in the paper, has been implemented in a computer program, written in Fortran, called TRAMO ("Time Series Regression with ARIMA Noise, Missing Observations, and Outliers"). See Gómez and Maravall (1994a). The program performs estimation, forecasting and interpolation of regression models with missing observations and ARIMA errors, in the presence of possibly several types of outliers. The ARIMA model can be identified automatically (no restriction is imposed on the location of the missing observations in the series). The program fits the regression model (2.11), where the errors v_t follow the general ARIMA model (2.12). The polinomials $\delta(L)$, $\phi(L)$ and $\theta(L)$ in TRAMO are assumed to have the following multiplicative form

$$\delta(L) = (1-L)^d (1-L^s)^D$$
$$\phi(L) = (1+\phi_1 L+\ldots+\phi_p L^p)(1+\Phi_1 L^s+\ldots+\Phi_P L^{s\times P})$$
$$\theta(L) = (1+\theta_1 L+\ldots+\theta_q L^q)(1+\Theta_1 L+\ldots+\Theta_Q L^{s\times Q}),$$

where *s* denotes the number of observations per year.

The regression variables can be input by the user (such as economic variables thought to be related with z_t), or generated by the program. The variables that can be generated are trading day, easter effect and intervention variables of the type:

a) dummy variables;
b) any possible sequence of ones and zeros;
c) $1/(1-\delta L)$ of any sequence of ones and zeros, where $0 < \delta \leq 1$;
d) $1/(1-\delta_s L^s)$ of any sequence of ones and zeros, where $0 < \delta_s \leq 1$;
e) $1/(1-\delta L)(1-\delta_s L^s)$ of any sequence of ones and zeros.

Several algorithms are available for computing the likelihood or more precisely, the nonlinear sum of squares to be minimized. When the differenced series can be used, the algorithm of Morf, Sidhu and Kailath (1974) (with simplification simila ro that of Mélard, 1984) is employed.

For the nondifferenced series, it is possible to use the ordinary Kalman filter (default option), or its square root version (see Anderson and Moore, 1979). The

latter is adequate when numerical difficulties arise; however, it is markedly slower.

By default, the exact maximum likelihood method is employed, and the unconditional and conditional least squares methods are available as options. Nonlinear maximizaton of the likelihood function and computation of the parameter estimates standard errors is made using Marquardt's method and first numerical derivatives.

Estimation of regression parameters is made, as described in the paper, by using first the Cholesky decomposition of the inverse error covariance matrix to transform the regression equation (the Kalman filter provides an efficient algorithm to compute the variables in this transformed regression). Then, the resulting least squares problem is solved by orthogonal matrix factorisation using Householder transformation. This procedure yields an efficient and numerically stable method to compute GLS estimators of the regression parameters, which avoids matrx inversion.

For forescasting, the ordinary Kalman filter or the square root filter options are available. Interpolation of missin values is made by a simplified Fixed Point Smoother, and yields identical results to Kohn and Ansley (1986); for a more detailed discussion , see Gómez and Maravall (1993). When concentrating the regression parameters out of the likelihood, mean squared errors of the forecasts and interpolations are obtained following the approach of Kohn and Ansley (1985).

The program has a facility for detecting outliers and for removing their effect; the outliers can be entered by the user or they can be automatically detected by the program, using an approach similar to that of Chen and Liu (1993), with some important modifications incorporated. In brief, regression parameters are initialized by OLS, and then the ARMA model parameters are estimated with two regressions, as in Hannan and Rissanen (1982). Next, the Kalman filter provides the series residuals, and new regression parameter estimates are obtained. For each observation, *t*-tests are computed for four types of outliers, as in Chen and Liu (1993). Outliers are removed one by one and, each time, new model parameter estimates, are obtained. Once this first sequence has been completed, a multiple regression is performed and, if some outliers are eliminated, theprogram goes back to the first sequence, and iterates in this way until no more outliers are eliminated in the multiple regression. A notable feature of this algorithm is that all calculations are based on linear regression techniques, which reduces computational time. The four types of outliers considered are additive outlier, innovational outlier, level shift and transitory change.

The program contains also a facility for automatic identification of the ARIMA model. This is done in two steps. The first one yields the nonstationary polinomial $\delta(L)$ and detects whether a mean should be specified to the model (2.12). This is done by iterating on a sequence of AR and ARMA(1,1) models (with mean), which have a multiplicative structure when the data is seasonal. The procedure is based on results of Tiao and Tsay (1983, Theor. 3.2 and 4.1), and Tsay (1984, Corol. 2.1). Regular and seasonal differences are obtained, up to a maximum order of $\nabla^2\nabla_s$. The program also checks for possible complex unit roots at nonzero nonseasonal frequencies.

The second step identifies an ARMA model for the stationary series (corrected for outliers and regression-type effects) following the Hannan-Rissanen procedure, with some modifications. For the general multiplicative model

$$\phi_p(L)\Phi_P(L^s)x_t = \theta_q(L)\Theta_Q(L^s)a_t,$$

the search is made over the range $0 \le (p,q) \le 3$, $0 \le (P,Q) \le 2$. This is done sequentially (for fixed regular polynomials, the seasonal ones are obtained, and viceversa), and the final orders of the polinomials are chosen according to the BIC criterion, with some possible constraints aimed at increasing parsimony and favoring "balanced" models (similar AR and MA orders).

TRAMO has been designed so that it can be used with a companion program named SEATS ("Signal Extraction in ARIMA Time Series"), described in Maravall and Gómez (1994). SEATS is an ARIMA-model-based method for estimation of unobserved components (trend, cyclical, seasonal and irregular component), and in particular for seasonal adjustment. Since the method applies to linear time

series, TRAMO can be seen as a preadjustment program, that produces the linear series for SEATS. Since both programs can handle routine applications to a large number of series, they provide a fully model-based alternative to REGARIMA and X11ARIMA, that form the new Census X12 procedure (see Findley et al., 1992). Programs TRAMO and SEATS, and more detailed documentation on both programs, are available from the authors.

The first example illustrates the case of missing observations on a stock variable, where regression variables (intervention variables constructed by the program) are present. It is the series of ozone (O_3) levels of Bow and Tiao (1975). The model is given by the equations

$$z_t = \frac{\omega_0}{1-L} d_{1t} + \frac{\omega_2}{1-L^{12}} d_{2t} + \frac{\omega_3}{1-L^{12}} d_{3t} + \nu_t \qquad (4.1)$$
$$\nabla_{12} \nu_t = (1+\theta_1 L)(1+\theta_{12} L^{12}) a_t,$$

where

$$d_{1t} = \begin{cases} 1 & t = January\ 1960 \\ 0 & otherwise \end{cases}$$

$$d_{2t} = \begin{cases} 1 & months\ June-October, beginning\ in\ 1966 \\ 0 & otherwise \end{cases}$$

$$d_{3t} = \begin{cases} 1 & months\ November-May, beginning\ in\ 1966 \\ 0 & otherwise \end{cases}$$

8 observations were deleted from the 216 observations of the original data set. These were observations number 3, 21, 39, 43, 113, 142, 170 and 201. Noter that the first missing observation falls among the first 12 values.

The results of exact maixmum likelihood estimation are shown in Table 1.

Table 2 shows the estimates of the missing observations.

Table 1. Maximum Likelihood Estimates of Parameters for Ozone Model, (4.1)

Data Set	*Parameters**	
	θ_1	θ_{12}
Full Data	.241	-.765
	(0.068)	(0.061)
Missing Observations	.265	-.770
	(0.074)	(0.058)

* Figures in parentheses are standard errors.

To implement the approach proposed in this paper tohandle the problem of temporal aggregation of a flow variable, a computer program has been written in Fortran by the author. To illustrate, the first 196 observations of Series A of Box and Jenkins (1976) have been aggregated by groups of four and assigned to the last observation in each group. As in Box and Jenkins (1976), two models have been used for the disaggregated series: ARMA (1,1) with mean μ and IMA (1,1) without mean. The first one is a stationary mloedl, whereas the second one is not. The stationary model is a regression model, with the mean as regression parameter. The results of exact maximum likelihood estimation are shown in Table 3.

Table 4 shows the estimates of the last 8 missing observations.

Table 2. Estimates of Missing Observations and Associated Root Mean Squared Errors

Data Set	*Observation Number*							
	3	21	39	43	113	142	170	210
Missing Observation s	4.364	6.178	4.322	5.716	3.507	4.631	2.333	3.043
	(.752)	(.725)	(.716)	(.709)	(.701)	(.702)	(.707)	(.763)
Actual Values	3.600	8.700	2.500	4.400	3.500	4.800	1.700	3.400

Table 3. Maximum Likelihood Estimates of Parameters for Series A

Data Set	*Parameters**		
	μ	ϕ	θ
Full Data : ARMA (1,1)	17.065	-.905	-.566
	(0.098)	(0.047)	(.089)
Full Data: IMA (1,1)	-	-	-.695
	-	-	(.051)
Missing Observations: ARMA (1,1)	17.061	-.944	-.609
	(0.129)	(0.012)	(.052)
Missing Observations: IMA (1,1)	-	-	-.688
	-	-	(.030)

*Figures in parentheses are standard errors

Table 4. Estimates of the Last 8 Missing Observations and Associated Root Mean Squared Errors

Data Set	*Observation Number*							
	189	190	191	192	193	194	195	196
Missing Observation s	17.576	17.638	17.675	17.711	17.601	17.586	17.568	17.546
ARMA (1,1)	(.334)	(.204)	(.205)	(.200)	(.340)	(.206)	(.205)	(.217)
Missing Observations	17.591	17.641	17.675	17.692	17.575	17.575	17.575	17.575
IMA(1,1)	(.346)	(.213)	(.214)	(.208)	(.349)	(.215)	(.214)	(.223)
Actual Values	17.400	17.000	18.000	18.200	17.600	17.800	17.700	17.200

References

Anderson, B.D.A. and Moore, J.B., (1979) , Optimal Filtering, Prentice-Hall.

Ansley, C.F., (1979), "An algorithm for the exact likelihood of a mixed autoregressive-moving average process,"Biometrika, 66, 59-65.

Bell, W., (1984), "Signal Extraction for Nonstationary Series," The Annals of Statistics, 12, 646-664.

Brockwell, P., and Davis, R., (1992) , Time Series: Theory and Methods, Springer-Verlag, Berlin.

Box, G.E.P., and Jenkins, G.M., (1976) , Time Series Analysis, Forecasing and Control, Holden-Day, San Francisco.

Box, G.E.P., and Tiao, G.C., (1975) , "Intervention Analysis with Applications to Economic and Environmental Problems,", Journal of the American Statistical Association, 70, 70-79.

Chen, C. and Liu, L., (1993) , "Joint Estimation of Model Parameters and Outlier Effects in Time Series," Journal of the American Statistical Association, 88, 284-297.

Chow, G.C., and Lin., A. (1971) , "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series," Review of Economics and Statistics, 53, 372-375.

Chow, G.C., and Lin., A. (1976) , "Best Linear Unbiased Estimation of Missing Observation in an Economic Time Series," Journal of the American Statistical Association, 71, 719-721.

Denton, F.T. (1971) , "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," Journal of the American Stastical Association, 66, 99-102.

Fernández, R. (1981) , "A Methodological Note on the Estimation of Time Series," Review of Economics and Statistics, LXIII, 471-475.

Findley, D.F., Monsell, B., Otto, M., Bell, W. and Pugh, M. (1992), "Towards X-12 ARIMA," mimeo, Bureau of the Census.

Gómez, V., and Maravall, A., (1993) , "Initializing the Kalman Filter with Incompletely Specified Initial Conditions,"in Chen, G. R. (ed.), Approximate Kalman Filtering (Series on Approximation and Decomposition), World Scientific Publishing Co., Inc., London.

Gómes, V., and Maravall, A., (1994a) , "Program TRAMO (Time Series Regression with ARIMA noise, Missing observations and Outliers – Instructions for the user," EUI Working Paper ECO No. 94/31, Department of Economics, European University Institute.

Gómez, V., and Maravall, A., (1994b) , "Estimation, Prediction and Interpolation for Nonstationary Series With the Kalman Filter," Journal of the American Statistical Association, 89, 611-624.

Hannan, E. J. and Rissanen, J. , "Recursive Estimation of Mixed Autoregressive-Moving Average Order," Biometrika, 69, 81-94.

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

Jones, R., (1980) , Maximum Likelihood Fitting of ARMA Models to Time Series With MissingObservations," technometrics, 22, 389-395.

Kohn, R., and Ansley, C.F., (1985) , "Efficient Estimation and Prediction in Time Series Regression Models"Biometrika, 72, 694-697.

Kohn, R., and Ansley, C.F., (1986) , "Estimation, Prediction and Interpolation for ARIMA Models With Missing Data," Journal of the American Statistical Association, 81, 751-761.

Maravall, A., and Gómez, V., (1994) , "Program SEATS (Signal Extraction in ARIMA Time Series – Instructions for the user," EUI Working Paper ECO No. 94/28, Department of Economics, European University Institute.

Mélard, G., (1984), "A Fast Algorithm for the Exact Likehood of Autoregressive-Moving Average Models," Applied Statistics, 35, 104-114.

Morf, M., Sidhu, G.S., and Kailath, T., (1974) , "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete- Time Systems," IEEE Transactions on Automatic Control, AC-19, 315-323.

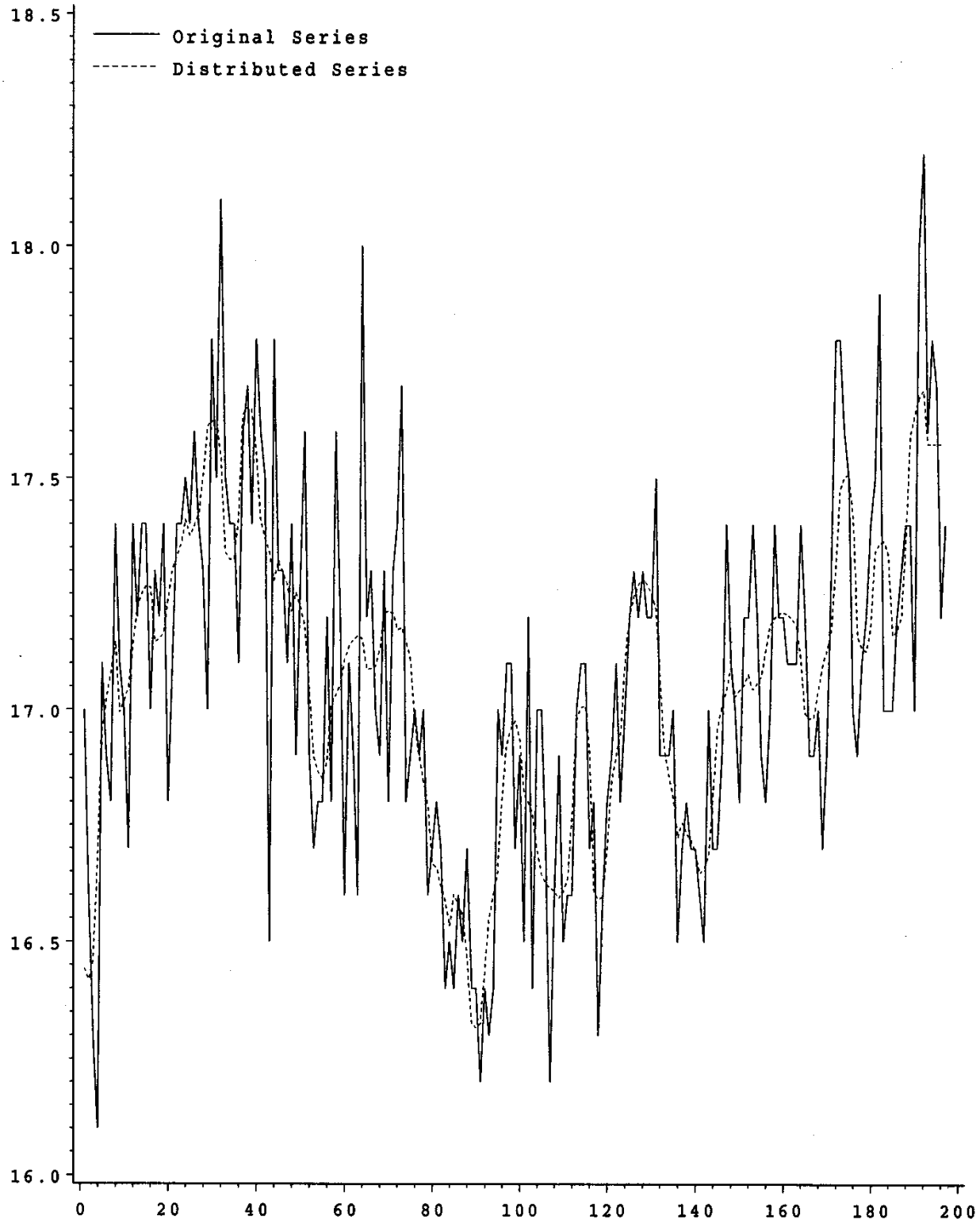
Pearlman, J.G., (1980) , "An Algorithm for the Exact Likelihood of a High-Order-Autoregressive-Moving Average Process," Biometrika, 67, 232-233.

Tiao, G. C. and Tsay, R. S., (1983) , "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models," The Annals of Statistics, 11, 856-871.

Tsay, R. S., (1984), "Regression Models With Times Series Errors," Journal of the American Statistical Association, 79, 118-124.

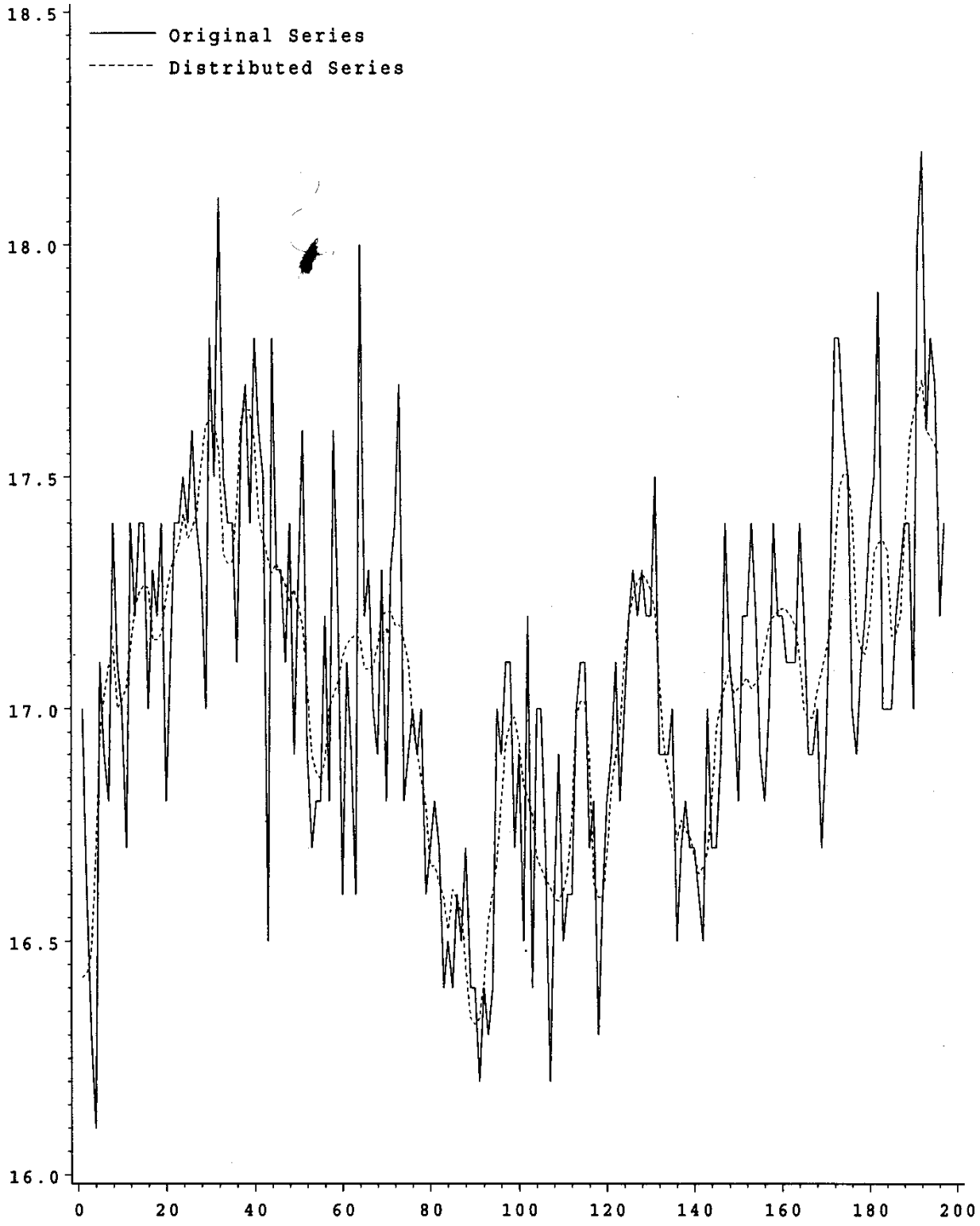
SERIES A: AGGREGATED BY GROUPS OF FOUR

MODEL: IMA(1,1) WITHOUT MEAN



SERIES A: AGGREGATED BY GROUPS OF FOUR

MODEL: ARMA(1,1) WITH MEAN



Temporal disaggregation of economic time series: Some econometric issues

Claudio LUPI; Istituto di Studi e Analisi Econometrica (ISAE)
Giuseppe PARIGI; Banca d'Italia, Servizio Studi

The paper critically reviews the most commonly used techniques to disaggreagate economic time series on a temporal basis. Different approaches are discussed in the attempt to assess their respective advantages and disadvantageous. Relatively new procedure are also described, such as the "missing data" approach, which sheds light on the general solutions offered by the Kalman filter estimator. The emphasis of the paper is on the econometric features of the different methods, and in particular on their characteristics and implications with respect to the integration/cointegration literature. This allows to highlight the inadequacy of some procedures to provide reliable results.

JEL Classifications Numbers: C22, C28
Key Words: Time Disaggregation, Missing Data, Cointegration

1 Introduction[1]

The knowledge of the short term evolution of most macroeconomic variables is crucial for the policy maker. In this context it is important to have high frequency information , quarterly, monthly or even daily in some cases. As it is not always possible to obtain direct measures at high frequency it is then necessary to apply statistical and mathematical procedures to temporally disaggregate the data. Several methods have been proposed in the literature, from purely mathematical to more sophisticated statistical procedures (a survey is in Di Fonzo, 1987; DiF henceforth).

A common feature of these methods is the neglect of proper econometric analysis, in the sense that economic interrelationships among variables are generally ignored. The aim of this paper is to re-examine the most influential and widely used procedures in the light of recent econometric developments, such as the unit root and cointegration

1 This paper was written while Claudio Lupi was working at ISTAT. The authors would like to thank Tom maso Di Fonzo, Enrico Giovannini, Stefano Pisani, Patrizia Ordine, Giovannini Savio for interesting obse rvations and fruitful discussions. However, none of them is responsible for any error remaining. We are grat eful also to Maria Rosaria Marino for her editing skill. The paper is the result of a close collaboration betwee n the two authors. However, sections 1-3 are to be attributed mainly to Giuseppe Parigi, sections 4-6 and the appendix to Claudio Lupi. The opinions expressed in the paper are of the authors only and not involve any respo nsibility on the part of the Bank of Italy, ISTAT and ISAE.

analysis. In this way, some well known methods for temporal disaggregation are critically reviewed.

The importance of the concept of cointegration, which is usually dependent on the particular theoretical model under examination, should not obscure the fact that the problem of temporal disaggregation has to be analysed within the framework of statistical accounts, without the influence of strong structural economic information.

The paper is organised as follows. A general description of the temporal disaggregation problem and the related methods is presented in the second section. The third one briefly sketches the problem of temporal disaggregation in the presence of a contemporaneous adding-up constraint. In the fourth section we focus on inference, with special emphasis on the econometric testing of hypotheses. The relationships among the different methods and the theory of integration and cointegration is described in the fifth section, while an empirical application is illustrated in the following. Some concluding remarks remarks are contained in the last section.

We try to maintain a uniform notation throughout the whole paper: y is the aggregated series of T observations; z, is the unknown series with frequency $m>1$; W is the $(m\mathrm{T}\times k)$ matrix of k indicators; ε and u two random disturbances with frequency m and 1 (e.g. yearly), respectively; $i=1,\mathrm{K},m\mathrm{T}$, and $t=1,\mathrm{K},\mathrm{T}$ are the temporal indices of the highest and lowest frequency respectively.

2 Indirect estimation methods

2.1 General remarks

The literature on temporal disaggregation is characterised by estimation methods which compute observations at high frequency by distributing or interpolating the low-frequency data.

More specifically, there is a problem of distribution when the aggregated series is computed by adding or averaging the high frequency data; this is the case of flow variables, such as the gross national product (GNP) and its components. The case of interpolation arises for stock variables, when the aggregated data coincide with one of the high frequency observations; the total of banking deposits or the initial capital stock of firm fall in this category.

In more formal terms, let z_i be the unknown series to be estimated for the periods $i=1,\mathrm{K},m\mathrm{T}$, and $y_t\,(t=1,\mathrm{K},\mathrm{T})$ the aggregated series; if $z_{m(t-1),t}=\left(z_{m(t-1)+1},\mathrm{K},z_{mt}\right)'$ and $c=(c_1,\mathrm{K},c_m)'$ are column vectors of length m (the elements of c are known constants), then:

$$y_t=c'\,z_{m(t-1),t}\qquad(t=1,\mathrm{K},\mathrm{T}) \tag{1}$$

where c is a $(m\times 1)$ vector which can be represented as:

interpolation

$$c=\begin{matrix}(1,0,\mathrm{K},0)' & \textit{initial period}\\ (0,0,\mathrm{K},1)' & \textit{final period}\end{matrix}$$

distribution

$$c=\begin{matrix}(1,1,\mathrm{K}1)' & \textit{sum}\\ \frac{1}{m}(1,1,\mathrm{K}1)' & \textit{mean}\end{matrix} \tag{2}$$

Indirect estimation methods may be classified according to the following categories: methods which do not use auxiliary information and methods based on indicators.

In the first case, the high frequency series is not obtained with a generally arbitrary mathematical-statistical procedure so that it can be applied when the information set is very limited (i.e. when no indicator is available).

The methods based on auxiliary information may be divided into two classes:

a) adjustment methods;
b) optimal methods.

Both exploit the idea that the dynamics of the series to be estimated is correlated with that of the indicators; in this case, z_i may be related to the indicator matrix $W(m\mathrm{T}\times k)$ through the statistical model:

$$z_i = W_i\beta + \varepsilon_i \qquad (i = 1, \mathrm{K}, m\mathrm{T}) \tag{3}$$

where W_i is the i-th row of W; β is the parameter $(k \times 1)$ vector and ε_i a random disturbance. For Instance, when we want to disaggregated the value added, W may be composed by the industrial production index (usually measured at high frequencies) and possibly by some trend and/or dummy variables for peculiar episodes. By aggregating (3) we have:

$$y = CW\beta + u \tag{4}$$

where C is the $(\mathrm{T} \times m\mathrm{T})$ aggregation matrix given by $\mathrm{I}_{\mathrm{T}} \otimes c'$ ($\otimes$ is the Kronecker product)[2]; $u = C\varepsilon$ is the aggregated error. Notice that the hypothesis of a static relationship between z and the indicators is imposed a priori. Dynamic elements may be contained in the error term, implicitly assuming the existence of common factors.

Disaggregation procedures may be generally interpreted as problems of constrained optimisation (see Stram and Wei, 1986b). If ε has mean 0 and covariance matrix Ω_ε (when ε_i follows a ARIMA (p,d,q) stochastic process, then Ω_ε is the covariance matrix of the d-th differences of ε, with dimension $(m\mathrm{T} - d) \times (m\mathrm{T} - d)$; see the Appendix), then:

$$\begin{aligned} &\min_{z,\beta} \ (z - W\beta)' \Omega_\varepsilon^{-1} (z - W\beta) \\ &s.t. \ \ y = C_z \end{aligned} \tag{5}$$

The minimand in (5) is the weighted sum of the squared residuals given by the unknown series z and its projection on the space generated by the indicators: $\hat{z} = W\hat{\beta}$, where $\hat{\beta}$ is any estimate of the parameter vector.

It can be shown that the solution for β in (5) corresponds to the Generalised Least Squares (GLS) estimate obtained from the aggregate model (4) and is given by:

$$\tilde{\beta} = \left[(CW)' \Omega_u^{-1} (CW) \right]^{-1} (CW)' \Omega_u^{-1} y \tag{6}$$

where Ω_u is the covariance matrix of u (if ε is integrated of order d, the same is true for u and Ω_u becomes the covariance matrix of the d-th differences of u). If Ω_u is known, $\tilde{\beta}$ is BLUE (best, linear, unbiased, GLS estimator).

The solution of (5) includes the simultaneous adjustment step, which consists of distributing the discrepancies at the aggregated level

$$\tilde{u} = Cz - C\tilde{z} = y - CW\tilde{\beta} \tag{7}$$

among the disaggregated observations so as to satisfy the constraint in (1). It is the smoothing matrix (L), which optimally distributes the aggregated errors among the high frequency observations. The final optimal solution for z is:

$$\tilde{z} = W\tilde{\beta} + L\tilde{u} \tag{8}$$

(8) is efficient only if L reflects the stochastic of $\varepsilon(u)$: this is the main difference between adjustment are distinct phases, while in the latter they are obtained simultaneously. In other words, the aggregation constraint in adjustment methods is obtained according to arbitrary hypotheses, not related to the procedure employed to compute the preliminary estimate. On the contrary, with optimal methods the properties of ε determine both the estimate of β and the form of the L matrix; moreover, β and z are estimated simultaneously.

Indirect estimation methods allow also to compute the values of z when aggregate is not yet known. This is the case of the extrapolation of z through the optimal

---

2 For quarterly flow variables:

$$c = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \mathrm{K} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \mathrm{K} & 0 & 0 & 0 & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & 1 & 1 & 1 & 1 \end{bmatrix}$$

solution (5) and the available information contained in the indicators. Extrapolated values may be considered as efficient forecasts if and only if the postulated model is based on a causal, theoretical, relationship among the variables. In the case of temporal disaggregation the link between the unknown series and the indicators should derive from statistical-accounting considerations and not from a particular economic theory. Moreover, the values extrapolated with indirect methods are generally liable to revisions because of the lack of aggregated information.

2.2 Methods which do not use indicators

These procedures do not employ auxiliary information to compute high frequency data. Generally the estimates must satisfy the following requirements:

a) aggregation constraint (1);

b) symmetry conditions: if $y_{t-1} < y_t < y_{t+1}$, then disaggregated observations must evolve opposite to the case when $y_{t-1} > y_t > y_{t+1}$;

c) zero variation condition: if $y_{t-1} = y_t = y_{t+1}$, then the *m*-th frequency observations must be equal to $(1/m)\,y_t$;

d) constant variation condition: if $y_t - y_{t-1} = y_{t+1} - y_t = a$, then the high frequency data must grow in the same proportion $(1/m^2)\,a$.

Most of these methods may be derived from that proposed by Boot et al. (1967), where the matrix W coincides with a constant value $(\bar{z})$. Following the procedure described in the preceding section, if:

$$\begin{aligned} z_i &= \bar{z} + \varepsilon_i \\ \varepsilon_i &= \varepsilon_{i-1} + y_i \end{aligned} \qquad (9)$$

the estimate $\tilde{z}$ is given by the solution to:

$$\begin{aligned} &\min_z \Delta z' \Omega_v^{-1} \Delta z \\ &s.t. \quad y = Cz \end{aligned} \qquad (10)$$

When v is a white noise process (mean zero and variance $\sigma_v^2 = 1$), then $\Omega_v = I$. In the general case, (9) coincides with the representation of time series integrated of order 1 (see Beveridge and Nelson, 1981).

The estimates derived from Boot et al. are not completely satisfactory for economic applications. First, when a variate grows, a constant preliminary estimate does not seem acceptable; moreover, the use of first differences does not allow to satisfy the condition of constant variation (point d above). The alternative proposal of Boot et al. is to employ second differences, which corresponds to the inclusion of a linear trend in (9). The trend however may be considered as an indicator so that a more efficient solution is given by optimal methods. Another flaw of the method just described is that the estimates of z may change dramatically according to the period and especially when a new aggregate value is available [3].

More recently, new disaggregation methods have been proposed, based on the relationships existing among the autocovariance matrices of the ARIMA processes of z and the temporal aggregate (see Stram and Wei, 1986a). In this context, the disaggregated series is obtained as a solution to a constraint minimisation problem similar to that described in the next section. What is needed are the autocovariance matrices of the d-th differences of y and z. While the matrices for $\Delta^d y$ (Δ is difference operator) may be obtained through the identification and estimation of an ARIMA model for y, this is not possible for z, which is not known. The problem is that there is not a one-to-one mapping from the aggregate ARIMA model to the disaggregate one. Only maximum orders may be obtained for the disaggregate model, when the orders (p,d,q) of the aggregate are known. The identification problem is tackled by Al-Osh (1989) by restricting the relevant class of models, and by selecting what they consider the simplest and most likely models for the disaggregate series. Wei and Stram (1990) solve elegantly this problem by proposing an approximate transformation to convert the autocovariances of $\Delta^d y$ into those of $\Delta^d z$. Barcellan and Di Fonzo (1994) applies the last two procedures to true economic time series and to simulated series, showing a slight, showing a slight superiority of the Wei and Stram method with respect to that of Al-Osh.

---

3 Alternative suggestions to overcome these problems have been proposed; see DiF for a survey.

2.3 Methods based on indicators: pure adjustment methods

The most famous and applied adjustment method is that by Denton (1971)[34], conceptually similar to that by Boot et al. According to this procedure, once a preliminary estimate of z has been obtained, the smoothing matrix can be computed by solving:

$$\min_{z}\left(z-\tilde{z}\right)' \mathrm{M}\left(z-\tilde{z}\right) \quad s.t. \quad y=Cz \tag{11}$$

M is a symmetric $(m\mathrm{T}\times m\mathrm{T})$ matrix, whose form depends on the hypotheses about the minimand or loss function. The smoothing matrix has the following form:

$$L=\mathrm{M}^{-1}C'\left(CM^{-1}C'\right)^{-1} \tag{12}$$

When the loss function is based on first differences (a case coincident with the problem (10) analysed by Boot et al. with a constant preliminary estimate of z), the matrix M (in this case, $\Omega_v^{-1}=\mathrm{M}$) is given by the product $\left(D'D\right)$, where D is[45]:

$$D=\begin{bmatrix} 1 & 0 & \mathrm{K} & \mathrm{K} & \mathrm{K} & 0 \\ -1 & 1 & 0 & \mathrm{K} & \mathrm{K} & 0 \\ 0 & -1 & 1 & 0 & \mathrm{K} & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0 & 0 & \mathrm{K} & \mathrm{K} & -1 & 1 \end{bmatrix} \tag{13}$$

(11) therefore becomes:

$$\min_{z}\left(z-\tilde{z}\right)'\left(D'D\right)\left(z-\tilde{z}\right) \quad s.t. \quad y=Cz \tag{14}$$

and the solution for L is:

$$L=\left(D'D\right)^{-1}C'\left[C\left(D'D\right)^{-1}C'\right]^{-1} \tag{15}$$

Denton's method is very simple to implement; in effects,

$$\left(D'D\right)^{-1}=\begin{bmatrix} 1 & 1 & 1 & \mathrm{K} & 1 \\ 1 & 2 & 2 & \mathrm{K} & 2 \\ 1 & 2 & 3 & \mathrm{K} & 3 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 1 & 2 & 3 & \mathrm{K} & m\mathrm{T} \end{bmatrix} \tag{16}$$

Denton's proposal may be interpreted as a two stage procedure, because only after the preliminary estimate is obtained the values are corrected so as to satisfy the aggregation constraint. Generally speaking, problem (5) may be seen as decomposed into two phases, the first one related to the estimate of β, without any consideration of Ω_ε; the second one related to the derivation of L. There is clearly an efficiency loss with respect to GLS methods, where β and L are estimated simultaneously (see DiF for a formal proof). Intuitively, the inefficiency of this method stems from the ignorance of the information contained in the covariance matrix Ω_ε, which is the base to compute the smoothing matrix. On the contrary, optimal methods solve this problem, but crucially depend on the hypotheses about the form of the matrix Ω_ε.

2.4 Methods based on indicators: the optimal methods

Optimal methods allow to compute the estimate of z subject to the aggregation constraint. The high frequency series is obtained through a statistical model similar to (5), where it is possible to apply the most common estimation techniques and, by computing the covariance matrix of the estimators, inference analysis. In this section we will describe three different models, which stem from some restrictive hypotheses imposed on the stochastic process for ε. In particular, we will consider ε generated as:

- a white noise, with zero mean and variance σ_ε^2 (section 2.4.1);
- an AR(1) process (section 2.4.2);
- a random walk (section 2.4.3).

The first case does not differ from the two stage methods; when the disturbance is distributed as a white noise, the optimal method coincides with the

4 Other adjustment methods are: Bassie (1958), Vangrevelinghe (1966) e Ginsburgh (1973).

5 When the loss function is based on the second differences, $\mathrm{M}=D'D'DD$.

OLS estimator and an equal distribution of the aggregate discrepancies. The other two cases have been analysed by Chow and Lin (1971) and by Fernandez (1981) and Litterman (1983). These are the most widely applied models in the literature: the Chow and Lin method, modified by Barbone et al. (1981) and successively revised, is currently employed by ISTAT in the construction of quarterly national accounts data (see ISTAT, 1992).

2.4.1 ε *distributed as a white noise*

When ε follows a white noise, the covariance matrix Ω_ε is given by $\sigma_\varepsilon^2 I$ and (5) may be rewritten as:

$$\begin{aligned} &\min_{\beta, z} (z - W\beta)' (z - W\beta) \\ &s.t. \quad y = Cz \end{aligned} \tag{17}$$

β may be estimated with OLS applied to the aggregate model without any efficiency loss; in this case, the smoothing matrix is equal to $(1/m)C'$ and the optimal solution is:

$$\hat{z} = W\hat{\beta} + \frac{1}{m} C' \hat{u} \tag{18}$$

It may occur that, as $\hat{u}$ is equally distributed to the high frequency data, the estimate of z is characterised by undesirable "jumps" between adjacent observations. This is due to the fact that, as with adjustment methods, it is assumed that there is no correlation among the disturbances. In the empirical work the solution is to assume much more complex stochastic processes, which allow to distribute the discrepancies in a more realistic way.

2.4.2 ε *distributed as an AR(1)*

An alternative to the restrictive white noise hypothesis is proposed by Chow and Lin (1971), who suppose that ε might be generated by an ARMA process, simplified in their application and in most of the following literature into the AR(1):

$$\varepsilon_i = \rho \varepsilon_{i-1} + \nu_i \tag{19}$$

where ν_i is white noise and $|\rho| < 1$ to guarantee stationarity of ε_i.

The idea is that this specification may capture dynamic elements ignored in the model (see Hendry and Mizon, 1978). In effects, when ε_i is generated as in (19), model (3) may be rewritten as:

$$z_i = \rho z_{i-1} + W_i \beta - W_{i-1}(\beta\rho) + \nu_i \tag{20}$$

This hypotheses is however quite arbitrary and should be tested (see section 4 on this point). The covariance matrix of ε_i is:

$$V_\varepsilon = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{mT-1} \\ \rho & 1 & \rho & \cdots & \rho^{mT-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{mT-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{mT-1} & \rho^{mT-2} & \rho^{mT-3} & \cdots & 1 \end{bmatrix} \tag{21}$$

and (5) becomes:

$$\begin{aligned} &\min (z_i - W_i\beta)' V_\varepsilon^{-1} (z_i - W_i\beta) \\ &s.t. \quad y = Cz \end{aligned} \tag{22}$$

The optimal solution is very similar to (6):

$$\begin{aligned} &\tilde{\beta} = \left[W'C'V_u^{-1}CW \right]^{-1} W'C'V_u^{-1} y \\ &L = V_\varepsilon C' V_u^{-1} \end{aligned} \tag{23}$$

where $V_u = CV_\varepsilon C'$..
Notice that the smoothing matrix distributes the discrepancies over high frequency data according to decreasing powers of ρ. As ρ is less than unity in absolute terms, more distant observations are only marginally affected by the updating of the series due to the availability of new aggregate values.

The estimate of the autoregressive parameter ρ is rather difficult; iterative methods *à la* Cochrane-Orcutt (1949) [56] cannot be applied because only the aggregate error is known. Under the assumption of normality, however, a possible solution

---

6 In general, when m is odd Cochrane-Orcutt procedure may be used: see Wei and Stram (1990).

is provided by the maximum likelihood estimation technique. β, σ_v^2 and ρ may therefore be jointly estimated by solving:

$$\max_{\beta,\sigma_\varepsilon^2,\rho} \max \left(2\pi\sigma_\varepsilon^2\right)^{-\frac{1}{2}n} |V_u|^{-\frac{1}{2}} \exp - \frac{(y - CW\beta)' V_u^{-1} (y - CW\beta)}{2\sigma_\varepsilon^2} \tag{24}$$

The computing of these estimates is not free of problems: when the high frequency data present sharp fluctuations, the estimated values may not be acceptable (see DiF, chapter 5).

To overcome these problems Barbone et al. (1981) propose a new method called "estimated generalised least squares" (Visco, 1983), based on the minimisation, with respect to β and ρ, of the following expression:

$$\left(y - CW\tilde{\beta}\right)' V_u^{-1} \left(y - CW\tilde{\beta}\right) \tag{25}$$

The estimate of ρ is obtained by scanning over a grid of values in the interval (-1,1) according to the procedure proposed by Hildreth and Lu (1960). Comparing (24) and (25) it may be observed that the estimate of ρ is independent of $|V_u|^{-\frac{1}{2}}$ so that it is not really a maximum likelihood estimate.

The extrapolation of n values, when the aggregate is not yet known, is easy to obtain by using the optimal solution (8). The proposal of Bournay and Laroque (1979) is equivalent to augmenting C with n vectors of zero as:

$$C^+ = (C \mid 0) \tag{26}$$

where 0 is a $(mT \times n)$ matrix. Further, V_ε is substituted by V_ε^+ where

$$V_\varepsilon^+ = E\begin{pmatrix} \varepsilon \\ \varepsilon^+ \end{pmatrix} \left(\varepsilon' \mid \varepsilon^{+\prime}\right) = \begin{pmatrix} V_\varepsilon & E\left(\varepsilon\varepsilon^{+\prime}\right) \\ E\left(\varepsilon^+\varepsilon'\right) & E\left(\varepsilon^+\varepsilon^{+\prime}\right) \end{pmatrix} \tag{27}$$

where $\varepsilon^+ = \left(\varepsilon_{Tm+1}, \ldots, \varepsilon_{Tm+n}\right)'$. The estimate of the unknown values is then given by:

$$\hat{y} = CW\hat{\beta} + V_\varepsilon^+ C^{+\prime} V_u^{-1} \hat{u} \tag{28}$$

2.4.3 ε distributed as a random walk

Fernandez (1981) reinterprets Denton's procedure (1971; section 2.3) in the framework of optimal methods. The crucial point is the hypothesis that ε is a random walk:

$$\varepsilon_i = \varepsilon_{i-1} + v_i \qquad i = 1, \ldots, mT \tag{29}$$

where v_i is a white noise with mean 0 and variance σ_v^2. In this case, the proposal of Fernandez is similar to that discussed in the previous section, with the additional assumption that $\rho = 1$.
This allows to simplify the computations because the autoregressive parameter has not to be estimated; given the initial conditions, the covariance matrices are:

$$V_\varepsilon = \sigma_v^2 (D'D)^{-1}$$
$$V_u = \sigma_v^2 C (D'D)^{-1} C' \tag{30}$$

Once a preliminary estimate of $\tilde{z}$ has been obtained, the optimisation problem is the same as Denton's and the solution for $\tilde{\beta}$ and L is similar to (23), with the matrices given in (30).

2.5 The missing data approach

In many empirical cases, it may occur that the high frequency data are available only for some sub period of the sample (i.e. yearly data to some point and quarterly data thereafter). If there exist one or more high frequency indicators for the whole period, then the missing data procedures may be used to construct the whole set of high frequency observations. It should be noticed that the missing data problems is somehow more difficult to solve in econometrics than in other disciplines, such as biometrics [7]. The principal reason for this it the intrinsically dynamic structure of economic systems.

A first attempt to estimate an econometric model in the presence of missing data is proposed by Sargan and Drettakis (1974). Essentially, this is an iterative maximum likelihood method where the missing data are seen as parameters to be estimated [8]. At first, the structural parameters of the model are estimated by employing the sample period for which the high frequency data are available. Missing data are then

computed by extrapolation from the model just estimated. The procedure is repeated by completing the information set with the extrapolated observations and the parameters are re-estimated [9]. If the model is correctly specified, the estimates are asymptotically consistent and efficient. In this context the estimate of missing data has to be considered as a by-product of the more general estimation problem. The practical application of this method is however quite complex; Drettakis (1973) explicitly considers the disaggregation problem with the imposition of constraint (1).

Gilbert (1977) concentrates only on the case of single equation estimators. He shows that in the case of flow variables, by substituting the disaggregation mean for the missing data and/or the aggregate data ($\frac{y}{m}$ for each of the m observations), the OLS [10] estimator is equivalent to the GLS on the observed data and to that of Theil-Goldberger (1961). The approach may be extended so that the aggregation constraint may be imposed [11].

Gilbert (1977) shows also that when in a static simultaneous system there is a problem of missing data for all the endogenous variables in the same period, the 2SLS estimator is asymptotically equivalent to the LIML estimator if the period with missing observations grows with the sample.

The dynamic case is tackled by Gilbert through an autoregressive model and the application of an iterative procedure similar to that advocated by Sargan and Drettakis (again the maximum likelihood estimate is obtained on convergence). It is interesting to observe that this estimator is equivalent to that of Chow and Lin (1971 and 1976), when the disturbances are distributed as an AR(1) [12].

Palm and Nijman (1984) analyse the more general dynamic case with ARMA disturbances. In particular, they show the superiority of maximum likelihood (ML) estimates without substitution of the missing data with some linear transforms of the aggregate values. However, they warn against possible problems of parameter identification, especially when residuals have MA components.

The estimation methods just seen may be correctly defined as ML only when the likelihood function is marginalised for the missing data (see Pena and Tiao, 1991). In dynamic models, however, the marginalisation is very difficult to perform so that iterative estimation techniques have been employed which cannot be defined as ML, even if this term is currently used.

A way to avoid explicit marginalisation is to evaluate the (log) likelihood function and its derivatives in prediction error decomposition form by means of the Kalman filter. In this context, (see Ansley and Kohn, 1983) the model has to be transformed into the state-space form and the (high frequency) missing data may be computed through the smoothed estimator of the state vector (see Harvey and McKenzie, 1983 for an application of Kalman filter technique to the model of Sargan and Drettakis). This approach has successively been generalised by Harvey and Pierse (1984), who show that by employing a state-space transformation of (3) is possible to modify and extend the Chow and Lin method to the case when the

---

7 See Afifi e Elashoff (1966) for a classical survey of missing data problems in biometrics.

8 The necessity of iterative methods stems from the impossibility in the dynamic case of partitioning the likelihood according to the data observed at different frequencies.

9 However, notice that there is in the likelihood a correcting term to take into account the estimate d nature of some observations.

10 The estimate must be corrected for the number of degrees of freedom, which are the same of a regres sion on the observed data.

11 See Tzerkezos (1993) for an application with distributed lags.

12 Acutally the equivalence is obtained through an approximation of the matrix V_ε.

residual ε is generated by a stationary ARMA(p,q) process1[13]. With this approach, using an extended Kalman filter, it is also possible to consider non linear transformations of the data, such as the logarithmic one (see Anderson and Moore, 1979).

Kalman filter based methods seem to be more easily applicable than ML ones; however, a deep knowledge of the properties of this estimator is requested for a better implementation. In particular, the non uniqueness of the state space representation needs a specific analysis in each case, while for instance, the Chow and Lin method does not.

It is useful to observe that with the missing data approach the estimates crucially depend on the hypotheses underlying the generating model of the data. As we have already seen, the estimates are asymptotically efficient and consistent only if the structural model is correctly specified. Here the point is the interpretation of the missing data as fixed parameters to be estimated. This seems a contradiction in terms: it would be much more natural to consider the missing data as random variables, with the same probabilistic structure of the already known observations (see Pena and Tiao, 1991).

3 Indirect estimate of several series with contemporaneous adding-up constraints

In many instances, the disaggregation problem refers to economic variables characterised by adding-up constraints. In the case of the resources and uses account:

$$GDP_t = C_t + \mathrm{I}_t + S_t + \mathrm{E}_t - \mathrm{M}_t \qquad (32)$$

where GDP is the gross domestic product, C is consumption, I investments, S the change in stocks, E exports and M imports. (32) is an identity valid every time period. In general, the application of some of the disaggregation methods discussed so far allows to satisfy the aggregation constraint (1), but not (32).

A direct, but quite arbitrary, solution to this problem, consists in disaggregating all the series but one to be determined as a residual. Needless to say, this series will be characterised by the errors of the disaggregation procedure applied to the other variables. Another possibility is to obtain the quarterly series for all the variables and then to balance them in order to satisfy the adding-up constraints (see Stone et al., 1942; Stone, 1990 for a general treatment of this problem in the context of national accounts).

A more efficient alternative should be based on an explicit consideration of adding-up constraints in the disaggregation procedures. In this case, there are 3 kinds of constraints:

a) Contemporaneous adding-up constraint.

Let l be a vector of S series, and let z be the sum of its components at each time period. The constraint may then be written as follows:

$$\sum_{j=1}^{s} l_{j,m(t-1)+k} = z_{m(t-1)+k} \qquad k = 1, \mathrm{K}, m \qquad (33)$$

b) Temporal aggregation constraint.

This is the usual constraint analysed up to now:

$$\sum_{k=1}^{m} l_{j,m(t-1)+k} = y_{j,t} \qquad j = 1, \mathrm{K}, S \qquad (34)$$

c) Adding-up constraint at the aggregate level.

For this case, it is assumed that as the aggregate value of the series is known, the adding-up constraint is automatically satisfied.

Once the constraints have been specified, the procedure is similar to the univariate case. DiF proposes a generalisation of the Chow and Lin method: the difficulty is how to estimate the covariance matrix of the residuals, given the hypotheses on the different error terms; for instance, in the situation where the disturbance follows a multivariate AR(1), with S series to disaggregate, $S(S+1)/2$ autoregressive parameters should be estimated with iterative techniques.

13 In this case, β should be part of the state vector. On this, see Harvey e Phillips (1979).

To avoid this kind of compications, Rossi (1982) suggests a two stage procedure, which is a reproposition of Denton's method (section 2.3), where the two phases consist of the preliminary estimate and the following adjustment. For the first one, Rossi proposes to compute the S disaggregated series with the univariate procedure of Chow and Lin (section 2.4.2)[14]; at this stage the single series do not satisfy the constraint in (34). In the second step a kind of smoothing matrix to adjust the series is computed. Di Fonzo (1990) describes Rossi's method and show that the adjustment phase is based on the hypothesis that the multivariate disturbance is a white noise, thus contradicting the assumption of an AR(1) for the single series.

The same logic of the univariate case applies for extrapolating the series when the aggregate values are not known. It is important to notice that every piece of information must be taken into account: for instance, when only one of the aggregate observations is available.

4 Inference

The results of the disaggregation procedure depend crucially on the model used for the temporally aggregated variables. In fact, the initial step in a Chow-Lin-like disaggregation procedure is the regression model

$$y_t = X_t\beta + u_t \quad (t = 1, \mathrm{K}, \mathrm{T}) \tag{35}$$

where y_t represents the variable to be disaggregated, and X_t is the t-th row of the matrix of temporally aggregated indicators, i.e. $X=CW$.

Because of the small number of indicators (we are not trying to "explain" the true DGP) and of the static nature of (35), the residuals u_t are generally autocorrelated and heteroskedastic. For example, assume that the true relation between the variable and the indicator is

$$z_i = \gamma + \psi(B)W_i + \zeta(B)\varepsilon_t \quad (i = 1, \mathrm{K}, m\mathrm{T}) \tag{36}$$

where $\psi(B) \equiv 1 + \sum_j \psi_j B^j$ and $\zeta(B) \equiv 1 + \sum_j \zeta_j B^j$ are lag polynomials whose roots are all outside the unit circle: then it is possible to show that the residuals of the equation (35) are in fact autocorrelated (Tiao and Wei, 1976).

If $\mathrm{E}(uu') = \sigma^2\Omega \neq \sigma^2\mathrm{I}$, then it is very well known that the OLS estimator of β in (35) is not efficient and inference based on OLS estimates is not correct in general[15]. Of course, it is possible to use Aitken's (1935) GLS or equivalent methods, but in this case the usual "goodness of fit statistics" have to be computed consistently (see e.g. Buse, 1973). The theory is well known and we will not treat it here [16]. However, the main issue is that the GLS estimator is unbiased and consistent if and only if the covariance matrix used in the estimate is the true one and if autocorrelation and/or heteroskedasticity are real features of the disturbances in the GDP. If residual autocorrelation and/or heteroskedasticity are induced by model misspecifications (as it is often the case in a temporal disaggregation framework), then the GLS estimator is no longer unbiased and consistent. Therefore, an empirical assessment of the nature of the autocorrelation using e.g. the COMFAC test (Hendry and Mizon, 1978) is recommended. A general problem, common to GLS and maximum likelihood methods, is that at least the structure of Ω must be specified a priori, unless a Kalman filter approach, which does not consider Ω explicitly (see Harvey and Phillips, 1979), is used. An alternative is offered by GMM estimators (Hansen, 1982), based on consistent estimates of the covariance matrix [17]. However, it

14 In Rossi's procedure the disturbances are generated by a process free of correlation. The approach may however be extended to the case analysed by Fernandez (paragraph 2.4.3).

15 See for example Nicholls and Pagan (1977) and Dufour (1988). There are particular cases in which OLS are efficient and/or consistent (see Fiebig et al., 1992). However, these cases do not seem to find natural applications in the present context.

16 Interested readers might refer to Amemiya (1973) for an extensive discussion of the properties of GLS estimator in the presence of an estimated variance-covariance matrix.

would be necessary to study further the properties of these estimators in small samples. Finally, solutions can be searched in the applications of bootstrap methods (Efron, 1979). In this case the main difficulty is given by the necessity of using consistent bootstrap techniques in the presence of dependent data. However, there are rather general conditions under which the autocorrelation-heteroskedasticity problems are alleviated, as we will see in Section 5.

Inference should also be used in order to assess the consistency of preliminary estimates of disaggregated variables arising from the aggregate regression model (35). Under stationarity, and in the case of only one indicator, Guerrero (1990) suggests a method based on the following hypotheses:

(i) the indicator and the variable have the same ARMA representation;
(ii) $E(z_i|\tilde{z}_i) = \tilde{z}_i$;
(iii) $E(z_i|\Im_{i-1})$ is independent of W, where $\Im_j$ is the information set as of time j.

If (i) holds, then the preliminary estimate $\tilde{z}_i$ follows the same ARMA process as the true series. This ARMA model can be written according to the Wold representation as

$$\tilde{z}_i = \theta(B)\varepsilon_i \tag{37}$$

and Guerrero derives the minimum variance estimator as

$$\begin{aligned} \hat{z} &= W + L(Y - CW) \\ &= W + \Theta\Theta' C' \left(C\Theta\Theta' C'\right)^{-1} (Y - CW) \end{aligned} \tag{38}$$

where Θ is a matrix whose elements are functions of the parameters of the Wold representation (37).

A consistency test is then derived under normality, computing the statistic

$$R = \frac{\left(y - CW\tilde{\beta}\right)' \left(C\hat{\vartheta}\hat{\vartheta}' C'\right)^{-1} \left(y - CW\tilde{\beta}\right)}{\tilde{\beta}^2 \hat{\sigma}_\varepsilon^2} \tag{39}$$

where $\hat{\vartheta}$ is a triangular matrix whose non zero elements are the estimates of the coefficients of the Wold representation (37). In practice, in the presence of one indicator, the estimate of σ_ε^2 can be derived from the ARMA model $\phi(B)W_i = \theta(B)\varepsilon_i$. Under the null that the preliminary aggregated estimate, $CW\tilde{\beta}$, is consistent with observed data, R is asymptotically distributed as a χ^2 with T degrees of freedom.

Unfortunately, this procedure is rather complex and becomes extremely complicated when more than one indicator is used. Furthermore, and this is perhaps the main issue at hand, the emphasis put on the stochastic structure of the residuals in the Chow-Lin method (and its generalisations), is posed directly on the relationship among the unknown disaggregated variable, its preliminary estimate and the indicators. There is a crucial problem: the preliminary estimate $\tilde{z}$ and the disaggregated series z cannot have the same ARIMA model if $\tilde{z}_i \sim I(1)$ and $z_i \sim I(1)$. In fact, assume

$$\phi(B)(1-B)z_i = \theta(B)\varepsilon_{1i} \tag{40}$$

$$\phi(B)(1-B)\tilde{z}_i = \theta(B)\varepsilon_{2i} \tag{41}$$

from which, $(1-B)\phi(B)(z_i - \tilde{z}_i) = \theta(B)\eta$. Therefore, in this case the differences between the preliminary estimate and the true series are non stationary. Alternatively, suppose that z_i and $\tilde{z}_i$ are generated by the simple bivariate cointegrated process

$$\begin{pmatrix} 1 & -\beta \\ 0 & (1-\beta) \end{pmatrix} \begin{pmatrix} z_i \\ \tilde{z}_i \end{pmatrix} = \begin{pmatrix} \theta_{11}(B) & 0 \\ 0 & \theta_{22}(B) \end{pmatrix} \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \tag{42}$$

It is easy to see that

$$\Delta z_i = \theta_{11}(B)(1-B)\varepsilon_{1i} + \beta\theta_{22}(B)\varepsilon_{2i} \tag{43}$$

The simplest case is when $\theta_{11}(B) \equiv 1$ and $\beta \equiv 1$. However, even in this rather special case

$$\Delta z_i = \Delta\varepsilon_{1i} + \theta_{22}(B)\varepsilon_{2i} \tag{44}$$

---

17 Consistent covariance estimators are studied, among others, by White (1984), Newey and West (1987), Andrews (1991), Andrews and Monahan (1992), Hansen (1992).

It follows that the true series and the preliminary estimate cannot be generated by the same ARIMA process.

5 Non-stationarity and temporal disaggregation

Cointegration is of paramount importance for estimation and inference in models with integrated time series. The intuition is very simple: if two (or more) variables are cointegrated, they cannot drift apart too much as time passes, and a stationary linear combination must exist. From the economic point of view, cointegration arises from the existence of a long-run equilibrium relationship.

In temporal disaggregation contexts, usually the indicator X_t measures substantially the same phenomenon as the variable y_t to be disaggregated. However, important differences can exist, for example in the way the two variables measure the same quantity. A typical case is when X_t is measured on a subsample with respect to the sample used to measure y_t. These differences notwithstanding, the general behaviour of the two variables should be very similar.

In what follows we will assume that $y_t \sim I(1)$. For X_t to be a valid indicator for y_t it must be that the two variables are cointegrated, i.e., $(y_t X_t) \sim CI(1,1)$. This of course implies that in (35) $u_t \sim I(0)$, in apparent contradiction with some disaggregation procedures, notably those by Fernandéz (1981) and Litterman (1983), but also with the general approach analysed by Stram and Wei (1986b). In other words, while according to Fernandéz (1981) random walk residuals should be considered as a realistic hypothesis, in our opinion it is a misleading one, meaning that the indicators used are not valid. This contradiction is partly explained by the fact that most studies on time disaggregation are purely statistical, and consider residuals as autonomous processes rather than, more correctly, model-based derived quantities.

In this section we will try to explain in much greater detail our viewpoint, using well known results concerning integrated variables. Remember that in the case of regressions between stationary variables, the asymptotic properties of the most common estimators are based on the assumptions (cf. White, 1984):

$$\begin{aligned} &p\lim \mathrm{T}^{-1}(X'X) = \mathrm{M} \\ &p\lim(X'u) = 0 \end{aligned} \qquad (45)$$

with M finite positive definite matrix. Most of the difficulties with integrated variables stem from the fact that second moments and cross-products, when appropriately normalised, do have non standard asymptotic distributions, which can be in general expressed as functionals of standard Wiener processes.

The most common cases are examined below.

a) $X_t \sim I(0)$

Actually, this is not a very interesting case in that it is self-evident that (35) is not well balanced in the sense that the dependent and the explanatory variables have different orders of integration. This implies that residuals are non stationary and the estimate of β tends to zero.

b) Linear trend used as indicator

(35) becomes in this case

$$y_t = \alpha + \beta_t + e_t \qquad (46)$$

Durlauf and Phillips (1988) treat this case extensively and demonstrate that

b.i) the distributions of $\hat{\alpha}, F_{\alpha=0}$ and t_α diverge;
b.ii) $\hat{\beta}$ tends to the true value of the drift of y_t but the distributions of $F_{\beta=0}$ and t_β diverge;
b.iii) the estimated R^2 has a non degenerate asymptotic distribution.

The empirical results from (46) could in general induce the investigator to assume that the trend is a "good" indicator for y_t having a significant t-statistic and being able to produce a relatively high R^2. However the Durbin-Watson (1950, 1951) DW statistic in this case gives an important indication that something must be wrong. In fact:

b.iv) DW is asymptotically zero.

The convergence to zero of the DW statistic is due to the nonstationarity of the equation residuals, and

explains Granger and Newbold's (1974) findings and their heuristic suggestion to be very cautious when a regression is characterised by $R^2 > DW$.

c) $X_t \sim I(1)$ **but independent of** y_t

This is the potentially most common case. The relevant results for this case are proved by Phillips (1986). In particular:

c.i) Even if X_t and y_t are independent processes, nevertheless $\hat{\beta}$ in (35) does not converge to its true value zero and possesses a non degenerate asymptotic distribution. Further, the asymptotic distribution of t_β diverges;

c.ii) the probability distributions of $\hat{\alpha}$ and t_α diverge;

c.iii) the R^2 statistic is asymptotically distributed according to a non degenerate distribution.

Results (c.i) – (c.iii) show that the problem of spurious regression is potentially very dangerous in a time disaggregation context. However, in the same way as (b.iv):

c.iv) DW converges asymptotically to zero.

Again, these results explain Granger and Newbold's (1974) findings and justify their heuristic caveat.

It is interesting to note that in both cases (b) and (c) above, the use of GLS is not justified and produces incorrect results. In particular, Durlauf and Phillips (1988) show that GLS estimates obtained via iterative algorithms *à la* Cochrane-Orcutt (1949) are such that:

c.v.) $T^{1/2}\hat{\alpha}_{GLS}$ and $T^{1/2}\hat{\beta}_{GLS}$ converge to non Gaussian random variables and $\hat{\alpha}_{GLS}$ and $\hat{\beta}_{GLS}$ are inconsistent;

c.vi) $F_{\beta=0}$ converges to a non$-\chi^2$ random variable;[18]

c.vii) *DW* tends in probability to zero.

Therefore, it should be clear that GLS methods cannot overcome nonstationarity problems.

Finally, the inclusion of a trend with a non-cointegrated *I*(1) indicator is not again a solution. Equation (35) can be explicitly written in this case as

$$y_t = \alpha + X_t\beta + \gamma t + u_t \qquad (47)$$

Durlauf and Phillips (1988) show that:

c.viii) $\hat{\beta}$ converges to a non-normal random variable;

c.ix) $\hat{\gamma}$ tends in probability to zero;

c.x) $\hat{\alpha}$, $\hat{\sigma}^2$, and $F_{\beta=0}$ diverge;

c.xi) *DW* converges in probability to zero.

It should be clear now that non-stationary residuals cannot be considered as an advantage as advocated e.g. by Fernandéz (1981), but rather is a serious warning against the reliability of the final disaggregated estimates.

For X_t to be considered a valid indicator of y_t the following minimal condition must hold:

d) y_t **and** X_t **are cointegrated**

The literature concerning estimates and inference in the presence of cointegrated variables is enormous. We will not try to give an exhaustive survey but rather we will concentrate on the main results which have a direct impact on the problem at hand.

If X_t and y_t in (35) are cointegrated[19], then

18 In (c.v), (c.vi), and (c.viii), convergence is defined in terms of convergence in probability measure (Billinglsey, 1968).

19 What follows is a valid if and only if regressors are not cointegrated among themselves. More general situations are discussed for example in Park and Phillips (1989), Sims et al. (1990), Phillips (1991). However, in most cases only one indicator is used in real time series disaggregation problems. For this reason we consider more convenient not to treat the more complicated cases directly and refer the interested readers to the relevant literature.

d.i) $\hat{\beta}$ is super-consistent (*T*-consistent).

This fundamental result is proved in different contexts by Andrews (1987), Phillips (1987a), Phillips and Durlauf (1986), Stock (1987), and highlight that OLS estimates among cointegrated variables. Furthermore, this results continues to hold even if the residuals are correlated with the regressors and also in the presence of measurement errors. OLS on stationary variables would have produced inconsistent estimates under the same circumstances [20]. However,

d.ii) $T(\hat{\beta}-\beta)$ is asymptotically distributed according to a mixture of normals centred on zero,

and implies that non standard distributions have to be used to draw inference (Park and Phillips, 1988). A further asymptotic advantage offered by the presence of a cointegrating relationship is given by the fact that (Phillips and Park, 1988)

d.iii) $T(\hat{\beta}-\beta)$ and $T(\hat{\beta}_{GLS}-\beta)$ have the same asymptotic distribution when $\hat{\beta}_{GLS}$ is Aitken's (1935) GLS estimator.

Therefore, in contrast with the stationary case, OLS are asymptotically efficient even in the presence of autocorrelation. Furthermore, Phillips and Park (1988) show that, if X_t is strictly exogenous (in the sense of Engle et al., 1983) [21], then

d.iv) the Wald test for *q* linear restrictions is asymptotically distributed, as in the stationary case, as a χ^2 with *q* degrees of freedom.

In most temporal disaggregation problems, the application of cointegration theory should require great care, since sample sizes are typically too small for the validity of asymptotic results; however, cointegration analysis should continue to play an important role as a "guide" also in finite samples [22]. It has been observed that, despite super consistency results, static regressions among cointegrated variables can give biased results in small samples. Two main solutions have been suggested. The first is to avoid static regressions and use dynamic models in the form of unrestricted ECM's (Banerjee *et al.*, 1986; Boswijk, 1992); the second is to use semiparametric corrections (Phillips and Hansen, 1990 [23]. It is known that greater biases correspond to smaller values of the R^2 statistic [24]. For this reason a very good fit may matter. Further, the bias does not converge to zero at the super-consistency rate in finite samples, and may persist even if the sample size is relatively large [25].

6 An example with real data

In order to be able to compare our results with preceding studies and different methodologies, we chose to examine the disaggregation of Italian National Accounts production using industrial

---

20 A classical reference is Haavelmo (1943)

21 Strict exogeneity is a rather peculiar feature. In general, if this condition is not satisfied, con ventional tests cannot be used. For more on this, see Phillips (1988, 1991), Park and Phillips (1988, 1989), Johans en (1992). Even lack of weak exogeneity can produce non trivial complications also in extremely simplified mod els (Hendry, 1993).

22 See e.g. Hendry (1973, 1982) on the role of asymptotic theory in finite samples.

23 There is Monte Carlo evidence that indicates that in many realistic cases the unrestricted ECM esti mator is to be preferred to OLS and to modified OLS with Phillips and Hansen's (1990) semiparametric correction . Details are contained in Inder (1993).

24 However, remember that here R^2 tends asymptotically to a non degenerate random variable.

25 This result has been found by Banerjee *et al*. (1986) using Monte Carlo experiments. It is justified theoretically by Abadir (1993a), who shows that if $Y_0/\sigma_\varepsilon \neq 0, Y_0$ being the initial condition and σ_ε the standard deviation of the stationary series of shocks, then in small samples the bias is inversely proportional to Y_0/σ_ε but converges to zero at a rate which itself is inversely proportional to the same quantity. Of course, the asymp totic convergence rate to zero is $O_p(T^{-1})$.

production and price indices as indicators. The same problem is examined, under a different viewpoint, in Gennari e Giovannini (1993) (GG, henceforth) [26].

Let $YC^{(k)}$ be the current prices production of branch k, and denote by $IPI^{(k)}$ and $POUT^{(k)}$ the industrial production index and the output price of the same branch. The annual equation is

$$YC_t^{(k)} = \alpha + \beta\left(IPI_t^{(k)} \times POUT_t^{(k)}\right) + e_t^{(k)} \qquad (48)$$

where $e_t^{(k)}$ is the equation residual. In order to have a better fit, estimates are often corrected introducing dummy variables and/or broken trends. However, the choice of dummies and trends is usually arbitrary, and pre-testing problems are likely to occur in practice.

The branches considered here are the same as those examined in GG, namely branches 21 and 41 of the NACE classifications, corresponding to "Agricultural and Industrial Machinery" and "Textiles and Clothing", respectively. Our OLS estimates, relative to the equations exposed in GG [27], are reported in table 1[28]. The diagnostic statistics indicate the presence of misspecification problems in both equations. GG's proposal is to represent the broken trends in the equations by means of changing parameters models. Their estimates are based on the Kalman filter [29], and seem to be decidedly better than OLS estimates. The authors give also an interesting structural interpretation of their estimates in terms of the different behaviour between large and small firms, and of the composition of the industrial production index.

In our approach, consistently with the discussion contained in the preceding section, the misspecification problems are treated as arising from the lack of cointegration between integrated variables. Table 2 lists the results of the Dickey-Fuller (1979, 1981) test for stationarity: they seem to indicate that both the variables $YC^{(21)}, YC^{(41)}$ and the indicators $\left(IPI^{(21)} \times POUT^{(21)}\right), \left(IPI^{(41)} \times POUT^{(41)}\right)$ are well approximated by $I(1)$ processes. Note that the Dickey-Fuller tests on the OLS residuals reported in table 1 do not reject the null hypothesis of no cointegration[30]. Of course, the two step Engle-Granger method applied here, is not necessarily the most efficient one to test for cointegration [31]. However, given that only one cointegrating vector can be present in each of our equations, and since our basic model is a static one, Engle and Granger two step method is naturally nested in our problem. Further, this choice allows us to use finite sample critical values (see MacKinnon, 1991).

The OLS estimates of the basic equations (48) are listed in table 3, from which other important pieces of information can be gathered. Note that, while it is important to obtain a very good fit at this stage of the disaggregtion process, nevertheless the use of broken trends and dummy variables may weaken the relationship between variable and indicator, as it can be clearly seen from comparing tables 1 and 3. In particular, note the highly significant value of the $W_{\beta=1}$ test.

---

26 Strictly speaking, our estimates are not exactly comparable with those in GG in that are based on a temporal sample which includes also the year 1992 (instead of just 1991) and embodies the annual revision of National Accounts 1990-91 data carried out in March 1993

27 In fact, our second equation is a reparameterization of that proposed in GG. Our parameterization is closer to that commonly used in the analysis of structural breaks (see for example Perron, 1989 and Zivot and Andrews, 1992).

28 The table reports the estimated models, the diagnostic statistics and the Dickey-Fuller tests for the residuals. The diagnostics of the Dickey-Fuller equations are also listed in the same table.

29 The precise algorithm is discussed in Carraro and Sartore (1987).

30 This occurs despite the fact that pre-testing problems are present in these equations and that the critical values should be greater in absolute value in order to take account of this. See for example Zivot and Andrews (1992) for similar considerations in the unit root tests framework.

31 On this problem see for example Boswijk and Franses (1992), Johansen (1988), Phillips (1991), Phillips and Loretan (1991) and Phillips and Ouliaris (1990).

A crucial aspect is the result of Ramey's RESET test (Ramsey, 1969; Ramsey and Schmidt, 1979) which is significant in both equations in tables 1 and 3. Indeed, it has been observed (Godfrey *et al.*, 1988) that the RESET test has good properties as far as the choice between linear and log-linear parameterizations are concerned. There are also theoretical reasons which make the log-linear alternative particularly appealing for economic variables. Consider for example the simple GDP:

$$\Delta y_t = \mu + \varepsilon_t,\ \mu > 0, \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right) \qquad (49)$$

In this case, $\sigma_\varepsilon^2 / Var(y_t) = O_p\left(T^{-1}\right)$, so that the series tends to be "asymptotically deterministic". Furthermore, $\mathrm{E}(\Delta y_t) = \mu \ \ \forall t$, and $\mathrm{E}(\Delta y_t) / y_{t-1} = O_p\left(T^{-1}\right)$ so that the $\Delta y_t's$ become less and less important. These are unappealing properties for economic time series. On the contrary, if the series were generated by the following process in logarithms

$$\Delta \log(y_t) = \gamma + \nu_t, \gamma > 0, \nu_t \sim NID\left(0, \sigma_\varepsilon^2\right) \qquad (50)$$

then, using the approximation $\log(1+\gamma) \cong \gamma$, we would have

$$y_t = (1+\gamma) y_{t-1} e_t,\ \log(e_t) \sim NID\left(0, \sigma_\varepsilon^2\right) \qquad (51)$$

Here, $\mathrm{E}(\Delta \log\ y_t) = \mathrm{E}\{\log(y_t / y_{t-1})\} = \gamma, \forall\ t$, so that the relative increments, and not the absolute ones, have a constant mean. Further, σ_e varies proportionally to y_t and the series does not become "asymptotically deterministic".

In general, integration and cointegration properties of time series are not invariant to logarithmic transforms of the data. In particular, while the existence of a cointegrating relationship in the levels implies the existence of an analogue relation in the logarithms, the converse is not generally true [32].

This justifies the analysis of the logs of our variables. Table 4 reports the Dickey-Fuller tests of the transformed variables. Once more, the tests cannot reject the null of no stationarity. The results of the OLS regressions for the variables expressed in logs are listed in table 5: the estimates are noteworthy better than those arising from the non-transformed OLS regressions, without having to use arbitrary dummy variables and broken trends.

Variables and indicators appear to be cointegrated and this ensures *T*-consistency of the relevant parameters. The bias should be moderate, given the high R^2 and the absence, at least in the second equation, of significant residual autocorrelation [33].

Though our results are very preliminary and partial, nevertheless they seem to indicate that more effort should be posed in the analysis of time disaggregation problems in the presence of non-lineary transformed data. It could be useful to examine further the possibilities offered in this context by the extended Kalman filter as suggested by Harvey and Pierse (1984).

7 Concluding remarks

The paper discusses the main temporal disaggregation procedures, specially focusing on the optimal methods, which, being based on indicators, exploit a larger information set. Barcellan and Di Fonzo (1994) provide preliminary evidence on the reliability of these methods in comparison to the univariate ones. In this context our effort is directed to give a critical appraisal of the several procedures in the light of recent econometric literature. This allows us to highlight the main problems of each method which deserve more analysis.

Some area for further research has been found: for instance, the application of the optimal methods in the presence of non linear transformations of variables and of general forms for the covariance matrix of the

32 These and related aspects are treated for example in Granger and Hallman (1991) and Ermini and Gra nger (1993).

33 However, remember that in the present framework R^2 is a non-degenerate random variable.

error term; the development of suitable inference tools in order to test the consistency of the disaggregate estimates with the observed, aggregate, series; the application of cointegration analysis. These issues are particular aspects of the more general problem of finding a well specified model to be employed in the disaggregation of economic time series. While in the applied economic analysis model specification exploits the information provided by economic theory, in this framework the same information should be of very limited use. Cointegration analysis should then be used mainly as a statistical device in order to find meaningful relationships among variables and indicators. This raises the crucial point of defining when to stop in the specification search of the "best" model.

The approach followed in the paper helps to show that the assumption of a random walk process for the high frequency disturbance is not appropriate even if it greatly simplifies the computations. In this case, the problem of spurious regression emerges quite clearly with the danger of obtaining very biased results.

Form the discussion above it is clear that the model to be employed for the disaggregation is aimed at providing a reliable measure of the phenomenon under analysis and not at explaining it. This may help understanding that when the aggregate is not yet known, the extrapolated values are not to be considered as efficient forecasts, but simply as anticipated, provisional measures.

Appendix

The statistical approach to temporal disaggregation

Let

$$z_i = W_i\beta + \varepsilon_i \quad (i = 1, \mathrm{K}, m\mathrm{T}) \tag{A.1}$$

be the true relationship at the disaggregated level, where W_i is the *i*-th row of the $(m\mathrm{T}\times k)$ indicators matrix *W*, and ε_i is a stationary component. Under the strong hypothesis that there is no omitted dynamics, (A.1) implies[34]

$$y_t = X_t\beta + u_t \quad (t = 1, \mathrm{K}, T) \tag{A.2}$$

when X_t is the *t*-th row of the indicators at the aggregated level. Further, let $\tilde{\beta}$ be a consistent estimate of β in (A.2) and define the preliminary estimate of the unknown disaggregated series $\{z_i\}_1^{mT}$ as

$$\tilde{z}_i = W_i\tilde{\beta} \tag{A.3}$$

In general, the difference $\{\tilde{u}_t\}_1^T$ between the preliminary estimate and the true series $\{y_t\}_1^T$ is non-null. Stram e Wei (1986b) assume $u_t \sim ARIMA(p,d,q)$ and show that aggregated discrepancies can be distributed solving the constrained minimisation problem

$$\min_{\varepsilon} r'\Omega_r^{-1}r \tag{A.4}$$

$$s.t. \sum_{j=(t-1)m+1}^{tm} \varepsilon_j = \tilde{u}_j \tag{A.5}$$

where $r_i = (1-B)^d \varepsilon_i$ and Ω_r is the $(mT-d)\times(mT-d)$ covariance matrix of r. Usually Ω_r specified *a priori* and this is equivalent to impose a particular stochastic structure to $\{\varepsilon_i\}$. Major differences among alternative optimal methods of temporal disaggregation are essentially based on the specification of the structure of Ω_r.

Following Stram and Wei (1986b) let us consider

$$r = D_{m\mathrm{T}}^d \varepsilon \tag{A.6}$$

$$s = D_{\mathrm{T}}^d \tilde{u} \tag{A.7}$$

Where $r = (r_{d+1}, \mathrm{K}, r_{mT})'$, $x = (x_1, \mathrm{K}, x_{mT})'$ and D_{mT}^d is $(mT-d)\times(mT-d)$ matrix of the kind

$$D_{MT}^d = \begin{pmatrix} \delta_0 & \delta_1 & \mathrm{K} & \delta_d & 0 & \mathrm{K} & 0 \\ 0 & \delta_0 & \mathrm{K} & \delta_{d-1} & \delta_d & \mathrm{K} & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0 & 0 & 0 & 0 & \mathrm{K} & \mathrm{K} & \delta_d \end{pmatrix} \tag{A.8}$$

Where δ_j is the coefficient of B^j in $(B-1)^d$. In practice, D_{MT}^d is the matrix difference operator of order d. For example, if $d = 1$

$$D_{mT}^1 = \begin{pmatrix} -1 & 1 & 0 & 0 & \mathrm{K} & 0 & 0 \\ 0 & -1 & 1 & 0 & \mathrm{K} & 0 & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0 & 0 & 0 & 0 & \mathrm{K} & -1 & 1 \end{pmatrix} \tag{A.9}$$

Further,

$$s = C^d r \tag{A.10}$$

where C^d is a $(T-d)\times(mT-d)$ matrix given by

$$C^d = \begin{pmatrix} c' & 0 & \mathrm{K} & 0 \\ 0_m' & c' & \mathrm{K} & 0 \\ 0_{2m}' & \mathrm{K} & \mathrm{K} & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0_{(T-d-1)m}' & \mathrm{K} & \mathrm{K} & c' \end{pmatrix} \tag{A.11}$$

when $0_j'$ is a $(1\times j)$ vector of zeros and $c' = (c_0, c_1, \mathrm{K}, c_{(d+1)(m-1)})$, where c_j is the coefficient associated to B^j in $(1+B+\mathrm{K}+B^{m-1})^{d+1}$. In practice, when $d = 1$ and $m = 4$, we have

34 A static relationship is necessary to write (A.2) ; see Tiao e Wei (1976).

$$C^1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & \mathrm{K} & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & \mathrm{K} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & \mathrm{K} & 0 \\ \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} & \mathrm{K} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathrm{K} & 1 \end{pmatrix} \tag{A.12}$$

The covariance matrix of (r, s) is

$$Cov(r,s) = \begin{pmatrix} \Omega_r & \Omega_r \left(C^d\right)' \\ \left(C^d\right)'\Omega_r & \Omega_s \end{pmatrix} \tag{A.13}$$

so that, if the r's are normal, the conditional expectation of r, given s, is

$$\bar{r} = \mathrm{E}\left(r|s\right) = \Omega_r \left(C^d\right)'\Omega_s^{-1} s = \Omega_r \left(C^d\right)'\Omega_s^{-1} D_T^d \tilde{u} \tag{A.14}$$

Let be $u^* = \left(u_{T-d-1}, \mathrm{K}, u_T\right)$, and denote the identity matrix of order d by I_d; if $J_m^{'}$ is a row vector of m zeros we have:

$$\begin{pmatrix} r \\ \tilde{u}^* \end{pmatrix} = \begin{pmatrix} D_{mT}^d \\ 0 \mathrm{I}_d \otimes J_m^{'} \end{pmatrix} \varepsilon \tag{A.15}$$

so that $\{\varepsilon\}$ can be derived from (A.15) as:

$$\varepsilon = \begin{pmatrix} D_{mT}^d \\ 0 \mathrm{I}_d \otimes J_m^{'} \end{pmatrix}^{-1} \begin{pmatrix} r \\ \tilde{u}^* \end{pmatrix} \tag{A.16}$$

Using (A.14), it is possible to obtain:

$$\varepsilon = \begin{pmatrix} D_{mT}^d \\ 0 \mathrm{I}_d \otimes J_m^{'} \end{pmatrix}^{-1} \begin{pmatrix} W_r \left(C^d\right)' W_s^{-1} D_T^d \\ 0 \mathrm{I}_d \end{pmatrix} \tilde{u} = L\tilde{u} \tag{A.17}$$

where L is the smoothing matrix. Stram and Wei (1986b) show that the matrix $\begin{pmatrix} D_{mT}^d \\ 0 \mathrm{I}_d \otimes J_m^{'} \end{pmatrix}$ exists and is invertible and that the solution of (A.4) – (A.5) for $\{\varepsilon_i\}_1^{mT}$ satisfies the aggregation constraint $\sum_{j=(t-1)m+1}^{tm} \varepsilon_j = \tilde{u}_t \left(t = 1, \mathrm{K}, T\right)$. Therefore the unknown series $\{z_i\}_1^{mT}$ can be computed as:

$$z = W\beta + \varepsilon \tag{A.18}$$

When $\{\tilde{u}_t\}$ is stationary, d = 0, and (A.16) becomes:

$$\varepsilon = \Omega_\varepsilon \left(C^0\right)'\Omega_u^{-1}\tilde{u} \tag{A.19}$$

where C^0 corresponds to C in section 2.1. Stationary residuals, ε, not only are theoretically more appropriate, but also offer advantages in terms of computing algorithms.

The smoothing matrix L in (A.17) can be used to derive, as particular cases, all the smoothing matrices examined in section 2. Furthermore, this method can be used also to disaggregate a series y without using indicators; in this case it is sufficient to substitute z to ε in (A.6) and y to $\tilde{u}$ in (A.7).

Table 1. OLS estimates with *trends* ad *dummy* variables.

$$YC_t^{(21)} = \underset{\substack{(872.60)\\[629.25]}}{-3241.4} + \underset{\substack{(16.176)\\[17.524]}}{477.97} * \left(IPI_t^{(21)}\ POUT_t^{(21)}\right) - \underset{\substack{(1912.0)\\[678.19]}}{3036.1} * D87 + \underset{\substack{(262.72)\\[322.35]}}{1548.3} * TR8492$$

$$\overline{YC}_t^{(21)} = \underset{\substack{(0.0241)\\[0.0308]}}{-0.108} + \underset{\substack{(0.0289)\\[0.0313]}}{0.859} * \overline{\left(IPI_t^{(21)}\ POUT_t^{(21)}\right)} - \underset{\substack{(0.0711)\\[0.0252]}}{0.113} * D87 + \underset{\substack{(0.00977)\\[0.0120]}}{0.0576} * TR8492$$

R^2=0.996	F(3,19)=1648.4 [0.000]	ó =1830.16 [5.224%]	DW=0.922	RSS=63.64*10^6
LM(2)=7.229 [0.005]	ARCH(1)=3.413 [0.082]	Norm.=1.444 [0.486]	Het.=1.307 [0.320]	RESET=4.540 [0.047]
ADF$_t$(1)=-2.717 [-3.253]	LM(2)=0.985 [0.395]	ARCH(1)=1.187 [0.292]	Norm.=1.600 [0.449]	Het.=1.709 [0.208]
$W_{b=1}$=26.106 [0.000]				

$$YC_t^{(41)} = \underset{\substack{(980.24)\\[541.25]}}{-494.30} + \underset{\substack{(61.728)\\[73.107]}}{416.96} * \left(IPI_t^{(41)}\ POUT_t^{(41)}\right) - \underset{\substack{(1574.5)\\[743.03]}}{3474.8} * D86 + \underset{\substack{(307.31)\\[315.32]}}{652.95} * t + \underset{\substack{(264.00)\\[314.93]}}{2433.2} * TR7992$$

$$\overline{YC}_t^{(41)} = \underset{\substack{(0.124)\\[0.142]}}{-0.563} + \underset{\substack{(0.0784)\\[0.0929]}}{0.530} * \overline{\left(IPI_t^{(41)}\ POUT_t^{(41)}\right)} - \underset{\substack{(0.0472)\\[0.0223]}}{0.104} * D86 + \underset{\substack{(0.00921)\\[0.00945]}}{0.0196} * t + \underset{\substack{(0.00791)\\[0.00944]}}{0.0729} * TR7992$$

R^2=0.996	F(4,18)=3045.9 [0.000]	ó =1448.69 [3.163%]	DW=1.320	RSS=37.8*10^6
LM(2)=0.992 [0.392]	ARCH(1)=0.124 [0.729]	Norm.=0.278 [0.870]	Het.=1.037 [0.463]	RESET=5.253 [0.035]
DF$_t$=-3.001 [-3.243]	LM(2)=0.300 [0.744]	ARCH(1)=0.915 [0.352]	Norm.=0.070 [0.966]	Het.=0.341 [0.716]
$W_{ß=1}$=35.933 [0.000]				

Notes:

Barred variables are standardised. "D86" and "D87" are two dummy variables for 1986 and 1987, respectively. "*t*" indicates a linear trend over the whole period while "TR8492" and "TR7992" represent two broken trends for the periods 1984-1992 and 1979-1992, respectively. Standard errors in parenthesis and heteroskedasticity consistent standard errors in brackets (White, 1980). Diagnostics include R^2, joint significance F test, equations standard errors (ó). Durbin-Watson statistics (DW [Durbin e Watson, 1950; 1951]), residuals sum of squares (RSS), residuals autocorrelation tests up to order two (LM(2) [Breusch e Pagan, 1980]), residuals first order ARCH test (ARCH(1), [Engle, 1982]), normality tests (Norm. [Jarque e Bera, 1980]), heteroskedasticity tests (Het. [White, 1980]), RESET tests (RESET [Ramsey, 1969; Ramsey e Schmidt, 1976]), Wald tests for the null ß=1 in the standardised equations ($W_{ß=1}$). In the table the Dickey-Fuller *t*-tests (DF [Dickey e Fuller, 1979; 1981]) and the augmented Dickey-Fuller *t*-tests with *k* lags (ADF(*k*) [Said e Dickey, 1984]) on equations residuals are also listed. Under each test, the probability levels are reported in square brackets; for the DF and ADF (*k*) tests the 10% critical value is given (MacKinnon, 1991]). For ó, the value expressed in percentage of the mean of the dependent variable is shown in brackets.

Table 2. Dickey-Fuller tests

Variable	*k*	ADF(*k*)	LM(2)	ARCH(1)	Norm.	Het.
$YC^{(21)}$	3	-2.986 [-3.276]	1.418 [0.283]	0.346 [0.568]	1.372 [0.504]	0.279 [0.933]
$YC^{(41)}$	1	-2.550 [-3.260]	1.453 [0.265]	2.557 [0.131]	0.375 [0.829]	1.281 [0.347]
$IPI^{(21)} * POUT^{(21)}$	2	2.490 [-3.268]	1.770 [0.209]	0.009 [0.925]	0.622 [0.733]	0.441 [0.859]
$IPI^{(41)} * POUT^{(41)}$	0	-2.420 [-3.253]	0.082 [0.921]	0.166 [0.689]	1.287 [0.525]	0.857 [0.513]

Notes:

Standard errors in parenthesis and heteroskedasticity consistent standard errors in brackets (White, 1980). Diagnostics of the Dickey-Fuller equations include R^2, joint significance F test, equations standard errors (ó). Durbin-Watson statistics (DW [Durbin e Watson, 1950; 1951]), residuals sum of squares (RSS), residuals autocorrelation tests up to order two (LM(2) [Breusch e Pagan, 1980]), residuals first order ARCH test (ARCH(1) [Engle, 1982]), normality tests (Norm. [Jarque e Bera, 1980]), heteroskedasticity tests (Het. [White, 1980]), RESET tests (RESET [Ramsey, 1969; Ramsey e Schmidt, 1976]) Wald tests for the null ß=1 in the standardised equations ($W_{ß=1}$). In the table the Dickey-Fuller *t*-tests (DF [Dickey e Fuller, 1979; 1981]) and the augmented Dickey-Fuller *t*-tests with *k* lags (ADF(*k*) [Said e Dickey, 1984]) on equations residuals are also listed. Under each test, the probability levels are reported in square brackets: for the DF and ADF(*k*) test the 10% critical value is given [MacKinnon, 1991]. For ó, the value expressed in percentage of the mean of the dependent variable is shown in brackets.

Table 3. OLS estimates

$$YC_t^{(21)} = \underset{\substack{(1158.9)\\[932.55]}}{-6221.1} + \underset{\substack{(13.141)\\[14.738]}}{557.38} * \left(IPI_t^{(21)}\ POUT_t^{(21)}\right)$$

$$\overline{YC_t^{(21)}} = \underset{\substack{(0.0234)\\[0.0234]}}{-1.640 * 10^{-16}} + \underset{\substack{(0.0234)\\[0.0263]}}{0.994} * \overline{\left(IPI_t^{(21)}\ POUT_t^{(21)}\right)}$$

R^2=0.998	F(1,21)=1799.1 [0.000]	ó =3022.48 [8.628%]	DW=0.519	RSS=191.8*10^6
LM(2)=13.244 [0.005]	ARCH(1)=6.262 [0.082]	Norm.=4.675 [0.486]	Het.=1.078 [0.320]	RESET=6.944 [0.047]
DF_t=-0.014 [-3.832]	LM(2)=2.632 [0.101]	ARCH(1)=0.025 [0.877]	Norm.=0.872 [0.647]	Het.=1.210 [0.350]
$W_{ß=1}$=0.0609 [0.807]				

$$YC_t^{(41)} = \underset{\substack{(1339.8)\\[966.37]}}{-5850.0} + \underset{\substack{(17.087)\\[18.109]}}{783.07} * \left(IPI_t^{(41)}\ POUT_t^{(41)}\right)$$

$$\overline{YC_t^{(41)}} = \underset{\substack{(0.0217)\\[0.0217]}}{2.039 * 10^{-16}} + \underset{\substack{(0.0217)\\[0.0230]}}{0.995} * \overline{\left(IPI_t^{(41)}\ POUT_t^{(41)}\right)}$$

R^2=0.990	F(1,21)=2100.3 [0.000]	ó =3474.42 [7.586%]	DW=0.493	RSS=253.5*10^6
LM(2)=13.33 [0.000]	ARCH(1)=1.004 [0.329]	Norm.=0.859 [0.651]	Het.=4.891 [0.020]	RESET=47.357 [0.000]
DF_t=-1.533 [-3.243]	LM(2)=1.240 [0.313]	ARCH(1)=0.385 [0.543]	Norm.=0.152 [0.927]	Het.=1.240 [0.314]
$W_{ß=1}$=0.0522 [0.8214]				

Notes:

Barred variables are standardised. "D86" and "D87" are two dummy variables for 1986 and 1987, respectively. "*t*" indicates a linear trend over the whole period while "TR8492" and "TR7992" represent two broken trends for the periods 1984-1992 and 1979-1992, respectively. Standard errors in parenthesis and heteroskedasticity consistent standard errors in brackets (White, 1980). Diagnostics include R^2, joint significance F test, equations standard errors (ó). Durbin-Watson statistics (DW [Durbin e Watson, 1950; 1951]), residuals sum of squares (RSS), residuals autocorrelation tests up to order two (LM(2) [Breusch e Pagan, 1980]), residuals first order ARCH test (ARCH(1), [Engle, 1982]), normality tests (Norm. [Jarque e Bera, 1980]), heteroskedasticity tests (Het. [White, 1980]), RESET tests (RESET [Ramsey, 1969; Ramsey e Schmidt, 1976]), Wald tests for the null ß=1 in the standardised equations $W_{ß=1}$. In the table the Dickey-Fuller *t*-tests (DF [Dickey e Fuller, 1979; 1981]) and the augmented Dickey-Fuller *t*-tests with k lags (ADF(*k*) [Said e Dickey, 1984]) on equations residuals are also listed. Under each test, the probability levels are reported in square brackets; for the DF and ADF (*k*) tests the 10% critical value is given (MacKinnon, 1991]). For ó, the value expressed in percentage of the mean of the dependent variable is shown in brackets.

Table 4. Dickey-Fuller tests on logarithms

Variable	*k*	ADF(*k*)	LM(2)	ARCH(1)	Norm.	Het.
$\log(YC^{(21)})$	2	-0.096 [-3.268]	1.437 [0.273]	1.379 [0.261]	0.984 [0.611]	0.784 [0.635]
$\log(YC^{(41)})$	2	0.904 [-3.268]	0.704 [0.0.513]	0.223 [0.645]	0.289 [0.865]	0.434 [0.864]
$\log(IPI^{(21)} * POUT^{(21)})$	2	-0.144 [-3.268]	1.588 [0.242]	0.673 [0.427]	0.722 [0.697]	1.013 [0.508]
$\log(IPI^{(41)} * POUT^{(41)})$	2	-2.092 [-3.268]	0.403 [0.676]	0.875 [0.367]	0.576 [0.750]	0.475 [0.837]

Notes:

Standard errors in parenthesis and heteroskedasticity consistent standard errors in brackets (White, 1980). Diagnostics of the Dickey-Fuller equations include R^2, joint significance *F* test, equations standard errors (ó). Durbin-Watson statistics (DW [Durbin e Watson, 1950; 1951]), residuals sum of squares (RSS), residuals autocorrelation tests up to order two (LM(2) [Breusch e Pagan, 1980]), residuals first order ARCH test (ARCH(1) [Engle, 1982]), normality tests (Norm. [Jarque e Bera, 1980]), heteroskedasticity tests (Het. [White, 1980]), RESET tests (RESET [Ramsey, 1969; Ramsey e Schmidt, 1976]) Wald tests for the null ß=1 in the standardised equations ($W_{ß=1}$). In the table the Dickey-Fuller *t*-tests (DF [Dickey e Fuller, 1979; 1981]) and the augmented Dickey-Fuller *t*-tests with *k* lags (ADF(k) [Said e Dickey, 1984]) on equations residuals are also listed. Under each test, the probability levels are reported in square brackets: for the DF and ADF(*k*) test the 10% critical value is given [MacKinnon, 1991]. For ó, the value expressed in percentage of the mean of the dependent variable is shown in brackets.

Table 5. OLS estimates of the logarithms

$$\log\left(YC_t^{(21)}\right) = \underset{\substack{(0.0599)\\[0.0605]}}{-4.882.1} + \underset{\substack{(0.0146)\\[0.0146]}}{1.279} * \log\left(IPI_t^{(21)}\ POUT_t^{(21)}\right)$$

$$\overline{\log\left(YC_t^{(21)}\right)} = -1.440 * 10^{-15} + \underset{\substack{(0.0114)\\[0.0114]}}{0.999} * \overline{\log\left(IPI_t^{(21)}\ POUT_t^{(21)}\right)}$$
(0.0234)
[0.0234]

R^2=0.997	F(1,21)=7668.5 [0.000]	ó =0.0589 [0.589%]	DW=1.320	RSS=0.0729
LM(2)=5.053 [0.0174]	ARCH(1)=4.936 [0.0386]	Norm.=2.106 [0.349]	Het.=0.0193 [0.981]	RESET=0.00371 [0.952]
ADF_t(3)=-4.686 [-3.276]	LM(2)=2.014 [0.176]	ARCH(1)=0.015 [0.904]	Norm.=0.388 [0.842]	Het.=0.707 [0.685]
$W_{ß=1}$=0.0144 [0.906]				

$$\log\left(YC_t^{(41)}\right) = \underset{\substack{(0.0442)\\[0.0339]}}{5.8119} + \underset{\substack{(0.0111)\\[0.0100]}}{1.1634} * \log\left(IPI_t^{(41)}\ POUT_t^{(41)}\right)$$

$$\overline{\log\left(YC_t^{(41)}\right)} = 2.199 * 10^{-16} + \underset{\substack{(0.00952)\\[0.00863]}}{0.999} * \overline{\log\left(IPI_t^{(41)}\ POUT_t^{(41)}\right)}$$
(0.00952)
[0.00952]

R^2=0.998	F(1,21)=1.1002 [0.000]	ó =0.0457 [0.437%]	DW=1.640	RSS=0.043
LM(2)=0.237 [0.791]	ARCH(1)=0.0499 [0.826]	Norm.=0.717 [0.699]	Het.=0.317 [0.732]	RESET=2.889 [0.104]
DF_t=-1.533 [-3.243]	LM(2)=1.240 [0.999]	ARCH(1)=0.385 [0.9402]	Norm.=0.152 [0.726]	Het.=1.240 [0.939]
$W_{ß=1}$ =0.0100 [0.921]				

Notes:

Barred variables are standardised. "D86" and "D87" are two dummy variables for 1986 and 1987, respectively. "*t*" indicates a linear trend over the whole period while "TR8492" and "TR7992" represent two broken trends for the periods 1984-1992 and 1979-1992, respectively. Standard errors in parenthesis and heteroskedasticity consistent standard errors in brackets (White, 1980). Diagnostics include R^2, joint significance F test, equations standard errors (ó). Durbin-Watson statistics (DW [Durbin e Watson, 1950; 1951]), residuals sum of squares (RSS), residuals autocorrelation tests up to order two (LM(2) [Breusch e Pagan, 1980]), residuals first order ARCH test (ARCH(1), [Engle, 1982]), normality tests (Norm. [Jarque e Bera, 1980]), heteroskedasticity tests (Het. [White, 1980]), RESET tests (RESET [Ramsey, 1969; Ramsey e Schmidt, 1976]), Wald tests for the null ß=1 in the standardised equations $W_{ß=1}$. In the table the Dickey-Fuller *t*-tests (DF [Dickey e Fuller, 1979; 1981]) and the augmented Dickey-Fuller *t*-tests with k lags (ADF(*k*) [Said e Dickey, 1984]) on equations residuals are also listed. Under each test, the probability levels are reported in square brackets; for the DF and ADF (*k*) tests the 10% critical value is given (MacKinnon, 1991]). For ó, the value expressed in percentage of the mean of the dependent variable is shown in brackets.

References

ABADIR, K.M., (1992), "A Distribution Generating Equation for Unit Root Statistics", Oxford Bulletin of Economics and Statistics, 54, 305-323.

ABADIR, K.M., (1993a), "OLS Bias in a Nonstationary Autoregression", Econometric Theory, 9, 81-93.

ABADIR, K.M., (1993b), "The Limiting Distribution of the Autocorrelation Coefficient Under a Unit Root", Annals of Statistics, 21, 1058-1070

AFIFI, A.A. – ELASHOFF, R.M. (1966), "Missing Observation in Multivariate Statistics, I: Review of the Literature", Journal of the American Statistical Association, 61, 595-604.

AITKEN, A.C., (1935), "On Least Squares and Linear Combinations of Observations", Proceedings of the Royal Society of Edinburgh, 55, 42-48.

AL-OSH, M., (1989), "A Dynamic Linear Model Approach for Disaggregating Time Series Data", Journal of Forecasting, 8, 85-96.

AMEMIYA, T., (1973), "Generalised Least Squares with an Estimated Autocovariance Matrix", Econometrica, 41, 723-732.

ANDERSON, B.D.A. – MOORE, J.B., (1979), Optimal Filtering. Englewood Cliffs, N.J., Prentice Hall

ANDREWS, D.W.K., (1987), "Least-Squares Regression with Integrated or Dynamic Regressors under Weak Error Assumptions", Econometric Theory, 3, 98-116.

ANDREWS, D.W.K., - MONAHAN, J.C., (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", Econometrica, 60, 953-966

ANSLEY, C.F. – KOHN, R. (1983), "Extract Likelihood of Vector Autoregressive Moving Average Process with Missing or Aggergated Data", Biometrika, 70, 275-278.

BANEJEE, A. – DOLADO, J.J. – HENDRY, D.F. – SMITH, G.W., (1986), "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence", Oxford Bulletin of Economics and Statistics, 48, 253-277

BARBONE, L. – BODO, G. – VISCO, I., (1981), "Costi e profitti in senso stretto: un'analisi su serie trimestrali, 1970-1980", Bollettino della Banca d'Italia, 36, 465-510.

BARCELLAN, R. – DI FONZO, T., (1994), "Disaggregazione Temporale e Modelli ARIMA", Atti della XXXVII Riunione Scientifica della SIS, 355-362

BASSIE, V.L., (1958), Economic Forecasting. New York, McGraw-Hill

BEVERIDGE, S. – NELSON, C.R., (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'", Journal of Monetary Economics, 7, 151-174

BILLINGLEY, P., (1968), Convergence of Probability Measures. New York, Wiley

BOOT, J.C.G. – FEIBES, W. – LISMAN J.H.C., (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data", Applied Statistics, 16, 65-75

BOSWIJK, H.P., (1992), "Efficient Inference on Cointegration Parameters in Structural Error Correction Models", University of Amsterdam, Econometrics Discussion Paper.

BOSWIJK, H.P. – FRANSES, P.H. (1992), "Dynamic Specification and Cointegration", Oxford Bulletin of Economics and Statistics, 54, 369-381.

BOURNAY, J. – LAROQUE, G., (1979), "Reflections sur la Méthode d'Elaborations des

Comptes Trimestriels", Annales de l'INIFE, 36, 3-30.

BOX, G.E.P. – JENKINS, G.M., (1976) , Time Series Analysis: Forecasting and Control (2nd Edition). San Francisco, Holden Day

BREUSCH, T.S. – PAGAN, A.R., (1980) , "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics", Review of Economic Studies, 47, 239-253

BUSE, A. (1973), "Goodness of Fit in Generalised Least Squares Estimation", The American Statistician, 27, 106-108

CARRARO, C. – SARTORE, D., (1987) , "Square Root Iterative Filter: Theory and Applications to Econometric Models", Annales d'Economie et Statistique, n. 6-7, 435-459

CHOW, G. – LIN, A.L., (1971) , "Best Linear Unbiased Interpolation Distribution and Extrapolation of Time Series by Related Series", Review of Economics and Statistics, 53, 372-375

CHOW, G. – LIN, A.L., (1976) , "Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series", Journal of the American Statistical Association, 71, 719-721

COCHRANE, D. – ORCUTT, G.H., (1949) , "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", Journal of the American Statistical Association, 44, 32-61

DAVIDSON, R. – HENDRY, D.F. – SRBA, F. – YEO, S., (1978), "Econometric Modelling of the Aggregate Relationship Between Consumers' Expenditure and Income in the United Kingdom", Economic Journal, 88, 661-692

DAVIDSON, R. – MACKINNON, J.G., (1993) , Estimation and Inference in Econometrics. Oxford, Oxford University Press

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 Editore.

DI FONZO, T., (1990) , "The Disaggregation of M Time Series When Contemporaneous and Temporal Aggregates are Known"" Review of Economics and Statistics, 72, 178-182

DICKEY, D.A. – FULLER, W.A., (1979) , "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, 74, 427-431

DICKEY, D.A. – FULLER, W.A., (1981) , "Likehood Ratio Statitics for Autoregressive Time Series with a Unit Root", Econometrica, 49, 1057-1072

DRETTAKIS, E.G., (1973) , "Missing Data in Econometric Estimation" Review of Economic Studies, 40, 537-552

DUFOUR, J.M., (1988) , "Estimators of the Disturbance Variance in Econometric Models: Small-sample Bias and the Existence of Moments", Journal of Econometrics, 37, 277-292

DURBIN, J. – WATSON, G.S., (1950) , "Testing for Serial Correlation in Least Squares Regression I", Biometrika, 37, 409-428

DURBIN, J. – WATSON, G.S., (1951) , "Testing for Serial Correlation in Least Squares Regression II", Biometrika, 38, 1-19

DURLAUF, S.N. – PHILLIPS, P.C.B., (1988) , "Trends versus Random Walks in Time Series Analysis", Econometrica, 56, 1333-1354

EFRON, B., (1979), "Bootstrapping Methods: Another Look at the Jackknife", Annals of Statistics, 7, 1-26

ENGLE, R.F., (1982), "Autoregressive Conditional Heteroscedasticity, with Estimates of the

Variance of United Kingdom Inflation", Econometrica, 50, 987-1008

ENGLE, R.F. – GANGER, C.W.J., (1987) , "Co-integration and Error-correction: Representation, Estimation and Testing", Econometrica, 55, 251-276

ENGLE, R.F. – HENDRY, D.F. – RICHARD, J-F., (1983), "Exogeneity", Econometrica, 55, 251-276

ERMINI, L. – GRANGER, C.W.J., (1993) , "Some Generalizations on the Algebra of I(1) Processes", Journal of Econometrics, 58; 369-384

FERNANDEZ, R.B., (1981) , "A Methodological Note on the Estimation of Time Series", Review of Economics and Statistics, 63, 471-478

FIEBIG, D.G. – MCALEER, M. – BARTELS, R., (1992), "Properties of Ordinary Least Squares in Regression Models with Nonspherical Disturbances", Journal of Econometrics, 54, 321-334

GENNARI, P. – GIOVANNINI, E., (1993) , "La stima trimestrale dei conti nazionali mediante modelli a parametri variabili", Quaderni di Ricerca Serie Economica e Ambiente, ISTAT, n.5/1993

GILBERT, C.L. (1977) , "Regression Using Mixed Annual and Quartely 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

GODFREY, L.G. – MCALEER, M. – MCKENZIE, C.R., (1988), "Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models", Review of Economics and Statistics, 70, 492-503

GRANGER, C.W.J. – HALLMAN, J., (1991) , "Nonlinear Transformations of Integrated Time Series", Journal of Time Series Analysis, 12, 207-224

GRANGER, C.W.J. – NEWBOLD, P., (1974) , "Spurious Regressions in Econometrics", Journal of Econometrics, 2, 111-120

GUERRERO, V.M., (1990) , "Temporal Disaggregation of Time Series: an ARIMA based Approach", International Statistical Review, 58, 29-36

HAAVELMO, T., (1943) , "The Statistical Implications of a System of Simultaneous Equations", Econometrica, 11, 1-12

HANSEN, B.E., (1992) , "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Procesifs", Econometrica, 60, 967-972

HANSEN, L.P., (1982) , "Large Sample Properties of Generalised Method of Moments Estimators", Econometrica, 50, 1029-1054

HARVEY, A.C., (1981) , The Econometric Analysis of Time Series. Oxford, Philip Allan

HARVEY, A.C. – MCKENZIE, C.R., (1984) , "Missing Observations in Dynamic Econometric Models: A Partial Synthesis", in E. Parzen (ed.) Time Series Analysis of Irregularly Observed Data. Berlin, Springer-Verlag

HARVEY, A.C. – PIERSE, R.G., (1984) , "Estimating Missing Observations in Economic Time Series", Journal of the American Statistical Association, 79, 125-131

HARVEY, A.C. – PHILLIPS, G.D.A., (1979) , "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances", Biometrika, 66, 49-58

HENDRY, D.F., (1973) , "On Asymptotic Theory and Finite Sample Experiments", Economica, 40, 210-217

HENDRY, D.F., (1982) , "A Reply to Professors Maasoumi and Philips", Journal of Econometrics, 19, 203-213

HENDRY, D.F., (1993), "On the Interactions of Unit Roots and Exogeneity. Nuffield College, mimeo

HENDRY, D.F. – MIZON, G.E., (1978), "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England", Economic Journal, 88, 549-563

HERRNDORF, N, (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", Annals of Probability, 12, 141-153

HILDRETH, C. – LU, J. Y., (1960), "Demand Relations with Autocorrelated Disturbances", Michigan State University, Research Bulletin 76

HYLLEBERG, S. – MIZON, G.E. (1989), "Cointegration and Error Correction Mechanisms", Economic Journal (Supplement), 99, 113-125

INDER, B., (1993), "Estimating Long-Run Relationships in Economics: A Comparison of Different Approaches", Journal of Econometrics, 57, 53-68

ISTAT, (1992), "I Conti Economici Trimestrali con baif 1980", Note e Relazion, 1

JARQUE, C.M. – BERA, A.K., (1980), "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", Economics Letters, 6, 255-259

JOHANSEN, S., (1988), "Statistical Analysis of Cointegration Vectors", Journal of Economic Dynamics and Control, 12, 231-254

JOHANSEN, S., (1992), "Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis", Journal of Econometrics, 52, 389-402

LITTERMAN, R.B., (1983), "A Random Walk, Markov Model for the Distribution of Time Series", Journal of Business and Economic Statistics, 1, 169-173

LUPI, C., (1992), Measures of Persistence, Trends, and Cycles in I(1) Time Series: A Rather Skeptical Viewpoint, University of Oxford, Institute of Economics and Statistics, Applied Economics Discussion Paper No. 137

MACKINNON, J.G., (1991), "Critical Values for Cointegration Tests", in R.F. Engle e C.W.J Granger (Eds.): Long-Run Economic Relationships: Readings in Cointegration. Oxford, Oxford University Press

MCLEISH, D.L., (1975a), "A Maximal Inequality and Dependent Strong Laws", Annals of Probability, 3, 829-839

MCLEISH, D.L., (1975b), "Invariance Principles for Dependant Variables". Z. Wahrscheinlichtstheorie und Verw. Gebiete, 32, 165-168

NELSON, C.R. – PLOSSER, C.I., (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", Journal of Monetary Economics, 10, 139-162

NEWEY, W.K. – WEST, K.D., (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica, 55, 703-708

NICHOLLS, D.F. – PAGAN, A.R., (1977), "Specification of the Disturbance for Efficient Estimation: An Extended Analysis", Econometrica, 45, 211-217

NICKELL, S.J., (1985), "Error Correction, Partial Adjustment and All That: An Expository Note", Oxford Bulletin of Economics and Statistics, 47, 119-129

ONOFRI, P., (1992), "Osservazione empirica e analisi economica: Esperienze di indagine sulle fluttuazioni cicliche", SIS Bollettino, n.26, 134-164

PALM, F.C. – NIJMAN, T.E., (1984), "Missing Observations in the Dynamic Regression Model", Econometrica, 52, 1415-1435

PARK, J.Y. – PHILLIPS, P.C.B., (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", Econometric Theory, 4, 468-497

PARK, J.Y. – PHILLIPS, P.C.B., (1988), "Statistical Inference in Regressions with Integrated Processes: Part 2", Econometric Theory, 5, 95-131

PENA, D. – TIAO, G.C., (1991), "A Note on Likelihood Estimation of Missing Values in Time Series", American Statistican, 45, 212-213

PHILLIPS, P.C.B., (1986), "Understanding Spurious Regressions in Econometrics", Journal of Econometrics, 33, 311-340

PHILLIPS, P.C.B., (1987a), "Time Series Regression with a Unit Root", Econometrica, 55, 277-301

PHILLIPS, P.C.B., (1987b), "Towards a Unified Asymptotic Theory for Autoregression", Biometrika, 74, 535-547

PHILLIPS, P.C.B., (1988), "Reflections on Econometric Methodology", Economic Record, 64, 544-559

PHILLIPS, P.C.B., (1991), "Optimal Inference in Cointegrated Systems", Econometrica, 59, 283-306

PHILLIPS, P.C.B. – DURLAUF, S.N., (1986), "Multiple Time Series Regression with Integrated Processes, Review of Economic Studies, 53, 473-495

PHILLIPS, P.C.B. – HANSEN, B.E., (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", Review of Economic Studies, 57, 99-125

PHILLIPS, P.C.B. – LORETAN, M., (1991), "Estimating Long-Run Economic Equilibria", Review of Economic Studies, 58, 407-436

PHILLIPS, P.C.B. – OULIARIS, S., (1990), "Asymptotic Properties of Residual Based Tests for Cointegration", Econometrica, 58, 165-193

PHILLIPS, P.C.B. – PARK, J.Y., (1988), "Asymptotic Equivalence of Ordinary Least Squares and Generalised Least Squares in Regressions with Integrated Variables", Journal of the American Statistical Association, 83, 111-115

RAMSEY, J.B., (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", Journal of the Royal Statistical Society, Series B, 31, 350-371

RAMSEY, J.B. – SCHMIDT, P., (1976), "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests", Journal of the American Statistical Association, 71, 389-390

ROSSI, N., (1982), "A Note on the Estimation of Disaggregated Time Series When the Aggregate is Known", Review of Economics and Statistics, 64, 695-696

SAID, S.E. – DICKEY, D.A., (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", Biometrika, 71, 599-607

SARGAN, J.D. – DRETTAKIS, E.G., (1974), "Missing Data in an Autoregressive Model", International Economic Review, 15, 39-58

SIMS, C.A. – STOCK, J.H. – WATSON, M.W., (1990), "Inference in Linear Time Series Models with Some Unit Roots", Econometrica, 58, 113-144

STOCK, J.H., (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", Econometrica, 55, 1035-1056

STONE, R., (1990), "Adjusting the National Accounts", Annali di Statistica, 9, 311-325

STONE, R. – CHAMPERNOWNE, D.G. – MEADE, J.E. (1942) , "The Precision of National Income Estimates", Review of Economic Studies, 9, 111-125

STRAM, D.O. – WEI, W.W.S., (1986a) , "Temporal Aggregation in the ARIMA Process", Journal of Time Series Analysis, 7, 279-292

STRAM, D.O. – WEI, W.W.S., (1986b) , "A Methodological Note on the Disaggregation of Time Series Totals", Journal of Time Series Analysis, 7, 293-302

TAYLOR, W.E., (1981) , "On the Efficiency of the Cochrane-Orcutt Estimator", Journal of Econometrics, 17, 67-82

THEIL, H. – GOLDBERGER, A.S., (1961) , "On Pure and Mixed Statistical Estimation in Economics", International Economic Review, 2, 65-78

TIAO, G.C. – WEI, W.W.S., (1976) , "Effect of Temporal Aggregation on the Dynamic Relationship of Two Time Series Variables", Biometrika, 63, 513-523

TSERKEZOS, D.E., (1993) , "Quarterly Disaggregation of the Annually Known Greek Gross Industrial Product Using Related Series", Rivista Internazionale di Scienze Economiche e Commerciali, 40, 457-474

VANGREVELINGHE, G., (1966) , "L'Evolution à Court Terme de la Consommation des Ménages: Connaissance, Analyse et Prévision", Etudes et Conjoncture, 9, 54-102

VISCO, I., (1983) , "Materiale per un Seminario ISTAT sulla Trimestralizzazione", mimeo, Roma

WEI, W.W.S. – STRAM, D.O., (1990) , "Disaggregation of Time Series Models", Journal of The Royal Statistical Society, Series B, 52, 453-467

WHITE, H., (1980) , "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", Econometrica, 48, 817-838

WHITE, H., (1984) , Asymptotic Theory for Econometricians. London, Academic Press

ZIVOT, E. – ANDREWS, D.W.K., (1992) , "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", Journal of Business and Economic Statistics, 10, 251-270.

Propositions pour une désagrégation temporelle basée sur des modèles dynamiques simples

Stéphane GREGOIR
INSEE

La construction de comptes nationaux trimestriels français se fait de manière indirecte. Elle repose sur l'utilisation de méthodes économétriques qui permettent de quantifier les relations entre des indicateurs infra-annuels et des postes comptables évalués annuellement à partir d'informations très fines. Dans le passé récent, les comptes trimestriels français n'ont pas su décrire avec précision l'amplitude des mouvements macro-économiques. Cette faiblesse est peut-être liée au fait que les relations économétriques utilisées dans ce processus sont statiques et présentent parfois des spécifications de faible qualité au vu des apports récents de la théorie des séries temporelles. Ce travail analyse quelques aspects de l'emploi de modèles dynamiques simples dans la construction de séries désagrégées en sous-périodes. Différentes spécifications avec peu de retards sont considérées, la construction de comptes trimestriels bruts semble être une approche possédant un certain nombre d'avantages.

1 Introduction

Les comptes trimestriels ont pour objectif de décrire les évolutions des agrégats économiques à un rythme infra-annuel en s'appuyant sur un cadre comptable cohérent afin de permettre d'observer les cycles conjoncturels, d'évaluer les délais entre événements et par là de mieux suivre, comprendre et prévoir la dynamique des mouvements économiques. Ils permettent aussi l'estimation de modèles macro-économiques. Enfin, ils cherchent à construire une évaluation de l'année en cours la plus robuste et la plus précise possible en fonction de l'information infra-annuelle disponible.

On peut séparer les instituts nationaux en charge de la construction des comptes trimestriels en deux catégories suivant la méthode qu'ils emploient pour construire ces derniers. La première approche repose sur un système d'enquêtes auprès des agents économiques et applique la même méthode pour construire les données annuelles et les données trimestrielles. La deuxième approche utilise des méthodes indirectes économétriques pour construire les évaluations trimestrielles qui sont recalées sur le passé sur les données annuelles obtenues par des méthodes de type enquêtes et/ou exploitation de sources fiscales (voir Di Fonzo, 1986, ou Di Fonzo et Filosa, 1987, pour un survol de la littérature). Ces deux approches ne sont pas si éloignées. Il est possible d'interpréter la seconde comme une méthode s'appuyant sur des enquêtes mais pour lesquelles les procédures de redressement reposeraient sur des modèles de régression. Une théorie plus générale des sondages reposant sur des modèles permet donc de réconcilier, si nécessaire, les deux approches.

En France, la deuxième approche est utilisée par l'INSEE pour construire les évaluations des comptes trimestriels. Lors des dernières années pendant lesquelles l'économie française a connu un mouvement cyclique marqué, un problème a pris une

ampleur remarquée. Il a été en effet observé que les révisions que les comptables trimestriels sont amenés à faire sur leurs premières estimations sont fortement corrélées avec la phase du cycle : révision à la hausse en phase de croissance, révision à la baisse en phase de décroissance. Parmi plusieurs explications possibles, deux ont retenu l'intérêt des comptables. La première porte sur le type d'indicateur utilisé pour construire l'évolution d'un agrégat. Lorsque cet indicateur est à champ constant, il ne permet pas de capturer les phénomènes de démographie des entreprises et peut donc être amené à sous-estimer la croissance en phase croissante et inversement. Une autre explication possible porte sur le fait que les modèles économétriques utilisés pour construire les évaluations courantes sont statiques et ne décrivent pas les mouvements dynamiques dans leurs enchaînements.

Le but de ce travail est de tenter de répondre à cette dernière critique en recherchant des spécifications dynamiques qui capturent à la fois les propriétés de court terme et de moyen et long terme. Pour ce faire, il est nécessaire de prendre en compte les développements récents de l'analyse des séries temporelles (intégration, cointégration) pour obtenir des spécifications appropriées. Il nous semble surtout important d'éviter d'utiliser des modèles où les erreurs sont intégrées et donc à variance croissante en fonction du temps. Après avoir proposé une spécification avec peu de retards assez générale pour tenir compte de ces propriétés, nous montrerons que plusieurs spécifications trimestrielles sont compatibles avec une spécification annuelle tout en satisfaisant certaines contraintes de cohérence inter temporelle (telle que la somme de quatre trimestres est égale à l'année pour des variables de flux). Dans ce travail, nous recherchons des méthodes simples, facilement "réplicables", ne demandant pas une connaissance statistique ou en algorithmique trop grande de façon à pouvoir être utilisées par beaucoup de personnes au niveau de connaissance hétérogène. Suivant que l'on veut construire des comptes corrigés des variations saisonnières ou des comptes bruts, les modèles simples que nous proposons ne présentent pas des propriétés semblables.

Le plan du papier est le suivant. Dans un premier temps, on rappelle les définitions d'intégration et de cointégration et l'on montre comment ces propriétés se propagent de séries annuelles à des séries trimestrielles (conservation de propriétés) mais aussi on souligne l'impossibilité de connaître des propriétés du même genre pour des fréquences différentes de 0. En d'autres termes, toute l'information sur la saisonnalité, lorsque que l'on veut construire des comptes bruts, ne peut être apportée que par les indicateurs. La méthode doit veiller à la transmettre de la façon la moins déformée possible aux variables de compte. Ensuite on montre à l'aide d'un exemple, les conséquences que peut avoir sur l'estimation et l'inférence le fait de négliger le propriétés d'intégration des variables. Puis on propose une spécification générale qui peut par des contraintes appropriées englober un grand nombre de cas.

Ensuite on s'intéresse à l'emploi d'un modèle dynamique simple compatible avec cette spécification permettant de construire des variables comptables trimestrielles (de flux et de stock) corrigées des variations saisonnières. La recherche d'une méthode simple d'estimation suivant un principe de statistique de moindres carrés ne donne pas de bonnes propriétés et se révèle coûteuse en temps machine. Quelques simulations viennent illustrer ce propos. Nous nous tournons ensuite vers la construction de comptes saisonniers pour lesquels nous obtenons de meilleures propriétés : simplicité d'usage et optimalité. Enfin nous effectuons quelques essais en vraie grandeur.

2 Quelques propriétés des séries temporelles

2.1 Définitions

L'approche économétrique de la construction des comptes trimestriels de la Nation repose sur un domaine de la statistique qui a connu un développement important durant les vingt dernières années, celui des séries chronologiques (ou séries temporelles). L'usage de ces techniques mathématiques dans l'analyse quantitative des phénomènes économiques a suscité l'introduction de nouveaux concepts et l'étude de classes de séries temporelles particulières. En effet, des travaux empiriques au début des années quatre-vingt (Nelson et Plosser (1982) *inter alia*) ont montré que nombre de séries macro-économiques pouvaient être correctement décrites en termes statistiques à l'aide de

processus non-stationnaires. L'étude de ces processus a entraîné l'introduction de nombreuses techniques de test et d'estimations. Plus précisément, la théorie statistique des processus temporels la plus communément utilisée s'appuie sur le concept de séries stationnaires du second ordre. Ces processus sont tels que leurs moments du premier et deuxième ordre vérifient des propriétés d'invariance temporelle qui permettent l'inférence usuelle. Un processus stationnaire du second ordre purement non déterministe admet ainsi une représentation de Wold unique de la forme:

$$x_t = c(B)\varepsilon_t$$

où B est un opérateur de décalage $Bx_t = x_{t-1}$, tel quel ε_t est un bruit blanc de moyenne nulle, $c(0) = 1, c(u) = \sum_{i=0}^{+\infty} c_i u^i$ satisfaisant $\sum_{i=1}^{+\infty} c_i^2 < +\infty$ et tel que les zéros de ce polynôme soient sur ou à l'extérieur du cercle d'unité. L'analyse usuelle d'un tel processus est effectuée à l'aide de sa structure d'autocorrélation et de son spectre. Beaucoup de séries économiques semblent ne pas satisfaire cette représentation, mais être mieux décrites par des processus intégrés d'ordre 1, voire 2, dont la définition est la suivante (cf. Engle et Granger (1987)):

Définition 1*: un processus stochastique purement non déterministe est dit intégré d'ordre h (h entier), ce que l'on note I(h), si* $(1-B)^h x_t$ *est un processus stationnaire du second ordre dont le polynôme c(u) associé à sa représentation de Wold vérifie* $c(1) \neq 0$.

Une série temporelle stationnaire du second ordre est parfois appelée I(0). Il existe d'autres formes d'intégration associées à des non-stationnarités liées à la saisonnalité. Ce type d'intégration saisonnière ne peut être pris en compte directement dans la modélisation de séries trimestrielles à partir de la seule observation de la série des quantités annuelles. En revanche, l'emploi de séries connexes observées de façon infra-annuelles permet d'obtenir de l'information sur ces propriétés. Aussi nous renvoyons le lecteur intéressé par ces aspects à d'autres travaux (Hylleberg, Engle, Granger et Yoo (1991) ou Gregoir (1994)).

Un autre concept a aussi pris de l'importance au cours de ces dernières années du fait de son interprétabilité en termes économiques. Il s'agit de la cointégration (Engle et Granger (1987)). Ce concept caractérise la situation dans laquelle il existe une combinaison linéaire d'au moins deux séries intégrées du même ordre qui soit intégrée d'un ordre inférieur. Mathématiquement, cela s'énonce de la façon suivante :

Définition 2: *un processus multivarié* x_t *de dimension n, intégré d'ordre h est dit cointégré de degré m (m entier), s'il existe une vecteur* $\alpha(1 \times n)$ *tel que* αx_t *est un processus intégré d'ordre h-m.*

La situation la plus communément observée dans les études empiriques correspond à une cointégration de degré 1 entre des variables intégrées d'ordre 1, obtenant ainsi des relations de cointégration stationnaires du second ordre. Nous nous limiterons à cette situation par la suite, mais une extension des propositions énoncées ci-dessous peut être facilement obtenue pour des cadres de travail plus généraux. Par ailleurs, il existe d'autres formes de cointégration liées à la saisonnalité dont nous ne traiterons pas explicitement dans ce travail mais auxquelles nous ferons parfois référence par la suite.

Dans ce paragraphe et par défaut dans la suite de ce texte, nous nous sommes placés dans le cadre usuel de l'analyse des séries temporelles. Un cadre plus large pourrait être utilisé. En effet, nous pourrions considérer le cadre des processus périodiques (Gladishev, 1961) ou plus généralement celui des processus harmonisables (Loève, 1963, Karhunen 1947) pour modéliser la structure des séries trimestrielles en l'absence d'information sur la nature saisonnière des séries. Mais l'emploi de ce cadre ne serait pas parcimonieux (multiplication des paramètres pour isoler les non-stationnarités) et réduirait la qualité des inférences.

2.2 Notations

En pratique, nous allons travailler à partir de séries de valeurs annuelles de la comptabilité nationale et de séries d'indicateurs observés de façon infra-annuelle pour construire des évaluations trimestrielles. Il importe donc de savoir s'il y a conservation des propriétés de non-stationnarité entre les deux séries. Nous devons distinguer deux situations selon que l'on

travaille sur des variables de flux ou des variables de stocks. Mais avant cela, il nous faut introduire quelques notations qui seront utilisées dans la suite de ce travail.

Dans ce qui suit nous dénoterons les quantités annuelles par des majuscules (X, Y,...) et les quantités trimestrielles correspondantes par des minuscules (x, y,...). "a" sera l'indice du temps annuel, "t" l'indice du temps trimestriel, et "i" le numéro du trimestre dans l'année (i = 1, 2, 3, 4). Ainsi par construction, pour toute valeur de "t", il existera une valeur unique de "a" et une valeur unique de "i" telles que t=4a+i.

Les variables de flux vérifient l'équation $Y_a = \sum_{i=1}^{4} y_{4(a-1)+i}$ alors que les variables de stock satisfont $Y_a = y_{4a}$.

2.3 Conservation des propriétés intégration et de cointégration entre des séries trimestrielles et des séries annuelles

Il importe avant toute chose de se poser la question de la conservation des propriétés précédemment décrites dans le processus d'agrégation temporelle. En fait, la réponse est relativement simple. Toutes les formes d'intégration et de cointégration non liées à la saisonnalité se transmettent des observations trimestrielles aux observations annuelles et réciproquement.

Proposition 1: *soient x_t une série univariée trimestrielle et X_a la série annuelle correspondante, alors si les deux séries suffisamment différenciées admettent une représentation de Wold,*
$x_t \approx \mathrm{I}(h) \Leftrightarrow X_a \approx \mathrm{I}(h)$

Preuve: voir annexe 1.

Cette propriété est satisfaite que l'on considère une variable de flux ou une variable de stock. En revanche, et de façon bien évidente, il n'y a pas conservation des propriétés d'intégration liées à la saisonnalité entre des observations trimestrielles et annuelles puisqu'il n'y a pas de conservation de l'information sur la structure d'intégration saisonnière.

Proposition 2: *soient x_t une série multivariée de dimension* n, I(1) et *X_a la série annuelle correspondante, alors si les deux séries suffisamment différenciées admettent une représentation de Wold.*
$\exists \alpha, \alpha x_t \approx \mathrm{I}(0) \Leftrightarrow \exists \alpha, \alpha X_a \approx \mathrm{I}(0)$

Preuve: voir annexe 1.

Là encore, cette propriété est vérifiée à la fois pour des variables de stock et de variables de flux.

Remarque 1: nous n'avons pas considéré dans cette partie des processus possédant une partie déterministe, mais les propriétés que nous venons d'énoncer sur les degrés d'intégration sont valables pour le degré des polynômes des parties déterministes des quantités trimestrielles et annuelles, ainsi que pour les propriétés de cointégration. De même, la non conservation de l'information sur la structure d'intégration liée à la saisonnalité demeure lorsque l'on considère les tendances déterministes associées, à savoir les polynômes combinaisons linéaires de monômes de la forme $\left(t^k \cos \omega t, t^k \sin \omega t\right), \omega = \frac{\pi}{2}, \pi$.

2.4 Conséquences de l'omission de ces propriétés

En pratique, il importe de tenir compte des propriétés d'intégration et de cointégration entre des variables pour pouvoir mener à bien une analyse économétrique. Dans le travail qui nous intéresse nous recherchons des situations où l'apport d'information par des séries économiques disponibles à un niveau infra-annuel peut permettre de construire des évaluations infra-annuelles pertinentes des postes comptables. En général, ceci se ramène à un régression des moindres carrés sur des indicateurs infra-annuels annualisés et des évaluations annuelles des postes comptables. Si, par exemple, *ces variables sont intégrées d'ordre 1 mais qu'elles ne sont pas cointégrées*, travailler sur le niveau des variables à des conséquences non négligeables sur la qualité de l'ajustement économétrique. Le résultat suivant donne une illustration de ces situations:

Exemple 1: *soit X_t variable I(1) satisfaisant la relation $(1-B)X_t = \beta(1-B)Y_t + \varepsilon_t$ pour $t = 1, \mathrm{K}, T$, où ε_t est i.i.d. de variance σ_ε^2, Y_t est défini par $(1-B)Y_t = \eta_t$ avec η_t i.i.d. de variance σ_η^2 et indépendant à toutes les dates de ε_t, et $X_0 = Y_0 = \varepsilon_0 = 0$, soit $\hat{\beta}_T$ l'estimateur des moindres*

carrés ordinaires de la régression de X_t sur Y_t, alors lorsque T tend vers l'infini :

$$\hat{\beta}_T = \beta + O_p(1)$$

$$\left(\frac{\hat{\beta}_{T+1}}{\hat{\beta}_T} - 1\right) = O_p\left(\frac{1}{T}\right)$$

En d'autres termes, si l'on effectue à tort une régression en niveau alors qu'il faudrait l'effectuer en différences premières, on peut obtenir un estimateur sans biais qui ne converge pas, pour lequel les règles usuelles de l'inférence normale ne s'applique pas. Les résidus de l'équation ne sont pas stationnaires et ont une variance qui croit linéairement avec le temps. Le deuxième résultat sur la convergence de l'estimateur du coefficient montre que si l'on veut analyser les sources de révisions des prévisions, le caractère intégré des termes d'erreur joue un rôle important sur le niveau de la révision. En effet, nous pouvons écrire :

$$x_{T+1} - \hat{x}_{T+1} = \frac{\hat{\beta}_{T+1} - \hat{\beta}_T}{\hat{\beta}_T} \hat{x}_{T+1} + \hat{\varepsilon}_{T+1}$$

et nous obtenons la convergence suivante :

$$\frac{x_{T+1} - \hat{x}_{T+1}}{\sqrt{T}} = O_p(1)$$

Ce qui signifie que l'erreur de révision croit en $O_p\left(\sqrt{T}\right)$.

2.5 Spécification générale d'une équation dynamique sur données annuelles

Nous venons d'illustrer le fait qu'une mauvaise spécification d'une équation entre des variables éventuellement intégrées peut avoir des conséquences non négligeables sur l'inférence que l'on peut pratiquer et sur les prévisions que l'on peut établir. Un grand nombre de séries macro-économiques semblent être assez bien décrites par une modélisation de variable intégrée, il faut donc pouvoir à travers une spécification assez souple tenir compte des différentes situations possibles. Par souci de simplicité, nous nous limiterons ici à des spécifications avec peu de retards, ce qui est relativement bien adapté à des modèles sur quantités annuelles, un peu moins pour des quantités trimestrielles. L'extension de cette écriture à des modèles linéaires plus complexes est directe.

La forme suivante :

$$Y_t = \alpha Y_{t-1} + \beta X_t + \gamma X_{t-1} + \varepsilon_t$$

englobe les différentes situations que l'on peut rencontrer avec des séries intégrées d'ordre 1. En effet, nous avons

- si $X_t \approx \mathrm{I}(1), Y_t \approx \mathrm{I}(1)$ et sont cointégrés alors nous devons utiliser une représentation à terme à correction d'erreur (théorème de représentation de Granger et Weiss (1983)). Le paramètrage $\alpha = 1 + a, \beta = c, \gamma = ab - c$ à cette situation. L'usage des moindres carrés ordinaires sous cette forme moyennant l'hypothèse d'orthogonalité du terme d'erreur par rapport aux variables présentes dans le membre de droite de l'équation donne des estimateurs convergents des paramètres α, β, γ.
- si $X_t \approx \mathrm{I}(1), Y_t \approx \mathrm{I}(1)$ et ne sont pas cointégrés alors la représentation doit faire intervenir des différences premières des variables. Le paramètrage qui convient est $\alpha = 1, \beta = a, \gamma = -a$. L'usage des moindres carrés ordinaires sous cette forme moyennant l'hypothèse d'orthogonalité du terme d'erreur par rapport aux variables présentes dans le membre de droite de l'équation donne des estimateurs super-convergent du paramètre α et convergents de β, γ.sSi $X_t \approx \mathrm{I}(0), Y_t \approx \mathrm{I}(1)$, la variable intégrée doit intervenir sous forme de différences premières $\alpha = 1, \beta = a, \gamma = b$. L'usage des moindres carrés ordinaires sous cette forme moyennant l'hypothèse d'orthogonalité du terme d'erreur par rapport aux variables présentes dans le membre de droite de l'équation donne des estimateurs super-convergent du paramètre α et convergents des paramètres β, γ.
- si $X_t \approx \mathrm{I}(1), Y_t \approx \mathrm{I}(0)$, la variable intégrée doit intervenir sous forme de différences premières $\alpha = a, \beta = b, \gamma = -b$. L'usage des moindres carrés ordinaires sous cette forme moyennant l'hypothèse d'orthogonalité du terme d'erreur par rapport aux variables présentes dans le membre de droite de l'équation donne des estimateurs convergents des paramètres α, β, γ.

- si $X_t \approx \mathrm{I}(0), Y_t \approx \mathrm{I}(0)$, il n'y a aucun problème dans la représentation, nous sommes sous une forme standard.

Pour des modèles avec des variables intégrées d'ordre deux, il suffit d'augmenter de 1 le nombre de retards dans l'écriture de l'équation. Nous ne pouvons en général rien affirmer sur la nature de ε_t. Il se peut qu'il existe une corrélation entre les endogènes et les termes d'erreur. Si tel est le cas, il faut modéliser la forme de la corrélation qui existe.

Les différents types de convergence énoncés ci-dessus impliquent que les règles usuelles d'inférence ne peuvent être utilisées de façon systématique. Suivant la nature des variables et l'existence ou non de relation de cointégration, les propriétés des statistiques usuelles ne sont pas les mêmes et les tables que l'on doit utiliser diffèrent. Pour décider de la situation dans laquelle on se trouve, il faut pratiquer un certain nombre de tests (test d'intégration, de cointégration) dont la puissance est faible. Néanmoins, des travaux récents (Kitamura et Phillips (1994), Kitamura (1994), Phillips (1993a,1993b)) permettent de construire des statistiques de tests de contraintes linéaires sur les coefficients dont la distribution asymptotique est un chi-deux quelles que soient les propriétés individuelles ou collectives des variables tant que l'on manipule des séries au plus intégrées d'ordre 1. Ces tests sont néanmoins légèrement conservatifs.

Les comptables trimestriels sont confrontés au problème de "créer" une dynamique trimestrielle cohérente avec une dynamique annuelle. Ceci signifie qu'il faut respecter les propriétés d'intégration et de cointégration. Nous pouvons partir d'une dynamique annuelle et rechercher une dynamique trimestrielle compatible. Cette voie est relativement compliquée. Dans certains cas nous serions amenés à considérer des dynamiques de type moyenne mobile infinie sur les variables trimestrielles pour obtenir des dynamiques autorégressives simples sur les quantités annuelles. Nous préférons effectuer la démarche dans le sens inverse, partir du trimestriel pour construire une dynamique annuelle et en déduire ce qui peut être inféré sur la dynamique infra-annuelle. De façon schématique, deux solutions sont envisageables. Nous pouvons construire des comptes trimestriels corrigés des variations saisonnières modélisés par une dynamique de type ARI infra-annuelle en utilisant des indicateurs désaisonnalisés et rechercher la dynamique agrégée compatible. Une autre approche consiste à construire des comptes bruts à partir d'une modélisation SARI des comptes trimestriels en utilisant des indicateurs bruts et étudier la dynamique annuelle compatible. Cette solution a l'avantage de permettre la construction de comptes trimestriels corrigés pour jours ouvrables comme une étape ultérieure du processus de fabrication. Nous proposons d'étudier ces deux solutions dans des spécifications parcimonieuses des dynamiques évoquées ci-dessus.

3 Construction d'un modèle trimestriel corrigé des variations saisonnières

3.1 Le modèle

3.1.1 Variable de flux

Nous considérons dans un premier temps des variables de flux. Nous postulons une dynamique simple de la forme suivante:

$$y_t = \alpha y_{t-1} + x_t\beta + \varepsilon_t, \qquad |\alpha| \le 1 \qquad (1)$$

où $\varepsilon_t \approx i.i.d.$ avec $\mathrm{E}\varepsilon_t | \underline{x}_t, \underline{y}_{t-1} = 0, V\varepsilon_t | \underline{x}_t, \underline{y}_{t-1} = \sigma_\varepsilon^2$, x_t est un vecteur d'indicateurs observés trimestriellement (il peut contenir des différences premières ou des variables décalées dans le temps) et $\underline{x}_t$ représente l'ensemble des valeurs passées prises par le processus stochastique x jusqu'à la date t. Nous en déduisons la dynamique annuelle :

$$Y_a = \alpha^4 Y_{a-1} + X_a\beta + \alpha X_{a,1}\beta + \alpha^2 X_{a,2}\beta + \alpha^3 X_{a,3}\beta + \mathrm{E}_a$$

avec $X_{a,j} = \sum_{i=1}^{4} x_{4(a-1)+i-j}$

et

$$\begin{aligned} E_a = {} & \varepsilon_{4a} + (\alpha+1)\varepsilon_{4a-1} + (\alpha^2+\alpha+1)\varepsilon_{4a-2} \\ & + (\alpha^3+\alpha^2+\alpha+1)\varepsilon_{4a-3} + (\alpha^3+\alpha^2+\alpha)\varepsilon_{4a-4} \\ & + (\alpha^3+\alpha^2)\varepsilon_{4a-5} + \alpha^3\varepsilon_{4a-6} \end{aligned}$$

Cette représentation n'est pas une représentation canonique du processus annuel, en particulier, nous n'avons pas isolé l'innovation du processus puisque $Cov(Y_{a-1}, \mathrm{E}_a) \neq 0$. Par ailleurs, la dynamique simple autorégressive sur la variable trimestrielle d'intérêt entraîne une modification notable de la dynamique de l'indicateur que nous n'avons pas explicitée jusqu'ici sous la forme annuelle déduite.

Nous ne pouvons pas utiliser de façon simple et directe la représentation annuelle pour estimer les paramètres d'intérêt de la dynamique trimestrielle. L'emploi de toute l'information disponible entraînerait l'usage de méthodes non-linéaires et plutôt de type variables instrumentales du fait de la corrélation entre le terme d'erreur et l'endogène retardée. Ne pas tenir compte des contraintes non-linéaires et estimer le modèle par une méthode des variables instrumentales entraîne un problème d'identifiabilité sur le signe de α[1].

3.2.1 Variable de stock

Nous considérons maintenant des variables de stock. Comme précédemment, nous postulons une dynamique simple de la forme suivante:

$$y_t = \alpha y_{t-1} + x_t \beta + \varepsilon_t, \qquad |\alpha| \leq 1 \qquad (2)$$

où $\varepsilon_t \approx i.i.d.$ avec $\mathrm{E}\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = 0, V\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = \sigma_\varepsilon^2$ et x_t est un vecteur d'indicateurs observés trimestriellement. Nous en déduisons la dynamique annuelle :

$$Y_a = \alpha^4 Y_{a-1} + x_{4a}\beta + \alpha x_{4a-1}\beta + \alpha^2 x_{4a-2}\beta + \alpha^3 x_{4a-3}\beta + \mathrm{E}_a$$

avec $\mathrm{E}_a = \varepsilon_{4a} + \alpha\varepsilon_{4a-1} + \alpha^2\varepsilon_{4a-2} + \alpha^3\varepsilon_{4a-3}$.

Cette représentation est une représentation canonique du processus annuel, nous avons isolé l'innovation du processus puisque $Cov(\underline{Y}_{a-1}, \mathrm{E}_a) = 0$. La dynamique simple autorégressive sur la variable trimestrielle d'intérêt entraîne une dynamique annuelle de l'indicateur où intervient chacune des observations effectuées au cours de l'année.

Comme pour les variables de flux, nous ne pouvons pas utiliser de façon simple et directe la représentation annuelle pour estimer les paramètres d'intérêt de la dynamique trimestrielle. L'emploi de toute l'information disponible entraînerait l'usage de méthodes non-linéaires. Ne pas tenir compte des contraintes non-linéaires et estimer le modèle par une méthode des moindres carrés entraîne ici encore un problème d'identifiabilité sur le signe de α.

3.2 Méthode d'estimation et ses propriétés

Nous ne voulons pas considérer des méthodes sophistiquées pour estimer ces modèles. Nous avons besoin de méthodes simples, robustes, employables par toutes les personnes impliquées dans la construction de comptes trimestriels. De plus des méthodes trop coûteuse en temps de calcul informatique seraient pénalisantes. Il doit être possible de réestimer rapidement les quelques milliers d'équations qui servent à la construction d'un compte. Nous nous tournons vers des méthodes du type moindres carrés ordinaires (du moins dans l'écriture du problème).

Nous proposons donc de travailler directement à partir de la représentation trimestrielle et de minimiser la somme des carrés des termes d'erreur sous contrainte de cohérence entre les évaluations trimestrielles et les comptes annuels. Nous ne tenons pas compte des contraintes de positivité qui existent sur les variables de comptabilité dans la mesure où si la contrainte devenait active, cela signifierait qu'une valeur comptable doit être mise égale à zéro, ce qui a peu de pertinence si nous ne disposons pas d'informations exogènes qui nous permettent d'asseoir ce fait. De même, nous n'imposons pas la contrainte $|\alpha| \leq 1$.

De façon pratique, si la valeur absolue de l'estimateur prend une valeur supérieure à 1, cela peut être une information sur le degré d'intégration de la série (série

---

1 Un parallèle peut être tiré à ce propos entre le modèle autorégressif et le modèle avec autocorréla tion des erreurs tel qu'envisagé par Bournay et Laroque (1979). Sous cette spécification, il existe aussi un problème d'identifiabilité du fait qu'un rapport de produit croisé de résidus puisse permettre l'estimation du coefficient d'autocorrélation des résidus mais élevé à la puissance 4.

intégrée d'ordre 2) et inciter à introduire un retard supplémentaire dans la représentation autorégressive considérée. En revanche, si la valeur est faiblement supérieure à 1, il faudrait être en mesure de tester son égalité afin d'imposer la valeur 1 pour éviter des processus explosifs. La fonctionnelle que nous devons minimiser est quadratique en les paramètres $\{y,\alpha,\beta\}$ où est le vecteur des variables trimestrielles inobservées. Nous considérons donc pour des variables de flux le problème :

$$(\mathrm{I})\begin{cases} \underset{y,\alpha,\beta}{Min}\sum_{t=2}^{T}\left(y_t-\alpha y_{t-1}-x_t\beta\right)^2 \\ \mathrm{I}_A\otimes(1111)\,y=Y \end{cases}$$

et pour des variables de stock, le problème suivant:

$$(\mathrm{II})\begin{cases} \underset{y,\alpha,\beta}{Min}\sum_{t=2}^{T}\left(y_t-\alpha y_{t-1}-x_t\beta\right)^2 \\ \mathrm{I}_A\otimes(0001)\,y=Y \end{cases}$$

où $\mathrm{I}_A$est la matrice identité de dimension $A=T\div 4$. La résolution numérique de ces problèmes est relativement simple et peut être obtenue en résolvant de façon itérative les conditions du premier ordre déduites du Lagrangien du problème. Ceci s'écrit sous la forme suivante où l'on a séparé la condition initiale y_1 du reste du vecteur y:

étape 1(i+1) :

$$\begin{pmatrix} {}^{i+1}\hat{\alpha} \\ {}^{i+1}\hat{\beta} \end{pmatrix} = \left({}^{i}\hat{V}',\ {}^{i}\hat{V}\right)^{-1}\left({}^{i}\hat{V}',\ {}^{i}\hat{y}\right)$$

où ${}^{i}\hat{V}=\left({}^{i}\hat{y}_{-1}\quad x\right)$ et ${}^{i}\hat{y}_{-1}={}^{i}\hat{y}_1 e_t+J\,{}^{i}\hat{y}$, **J** est la matrice de Jordan de taille (4A-1,4A-1) (avec des 1 sous la diagonale principale et des 0 ailleurs) et e_t est un vecteur de taille 4A-1 avec un 1 en première position et des 0 ailleurs,

étape 2(i+1) :

(a)

$${}^{i+1}\hat{y}_1=\frac{V\left({}^{i+1}\hat{\alpha}\right)'\left(\Phi\left({}^{i+1}\hat{\alpha}\right)\Phi\left({}^{i+1}\hat{\alpha}\right)'\right)^{-1}\left(\Phi\left({}^{i+1}\hat{\alpha}\right)x^{i+1}\hat{\beta}-Y\right)}{V\left({}^{i+1}\hat{\alpha}\right)'\left(\Phi\left({}^{i+1}\hat{\alpha}\right)\Phi\left({}^{i+1}\hat{\alpha}\right)'\right)^{-1}v\left({}^{i+1}\hat{\alpha}\right)}$$

(b)

$${}^{i+1}\hat{A}=\left(\Phi\left({}^{i+1}\hat{\alpha}\right)\Phi\left({}^{i+1}\hat{\alpha}\right)'\right)^{-1}\left(Y-\Phi\left({}^{i+1}\hat{\alpha}\right)x^{i+1}\hat{\beta}-v\left({}^{i+1}\hat{\alpha}\right)^{i+1}\hat{y}_0\right)$$

(c)

$${}^{i+1}\hat{y}=\left[\left(\mathrm{I}-{}^{i+1}\hat{\alpha}J\right)'\left(\mathrm{I}-{}^{i+1}\hat{\alpha}J\right)\right]^{-1}\left[\left(\mathrm{I}-{}^{i+1}\hat{\alpha}J\right)'\left({}^{i+1}\hat{\alpha}e_t^{\,i+1}\hat{y}_0+x^{i+1}\hat{\beta}\right)+\Phi'^{i+1}\hat{A}\right]$$

avec $v(\alpha)=\alpha\Phi(\alpha)e_t+e_a 1_{\{y \text{ est une variable de flux}\}}$, $\Phi(\alpha)=\Phi(\mathrm{I}-\alpha J)^{-1}$, où Φ est telle que pour des variables de flux après concaténation à e_a, un vecteur de taille A avec un 1 en première position et des 0 ensuite, on obtient la matrice d'annualisation $\left((e_a\quad\Phi)=\mathrm{I}_a\otimes(1\quad 1\quad 1\quad 1)\right)$ et pour des variables de stocks, après concaténation à un vecteur nul de dimension A, on obtient la matrice d'annualisation appropriée $\mathrm{I}_a\otimes(0\quad 0\quad 0\quad 1)$.

Le fait que les fonctions que l'on minimise sont quadratiques en les paramètres et que les contraintes sont linéaires en les paramètres assure que ce type élémentaire d'algorithme converge (plus ou moins vite). Néanmoins les propriétés statistiques des estimateurs que l'on obtient sont mauvaises. Il est simple de remarquer les estimateurs $\hat{y}$ ne sont pas convergents dans la mesure où le nombre de paramètres augmente avec le nombre d'observations. Ceci est toujours vérifié dans ce type de problème. En revanche l'étude des biais apporte de l'information sur la qualité des estimations.

Proposition 3: *les estimateurs* $\left\{\hat{y},\hat{\alpha},\hat{\beta}\right\}$ *solutions du programme (I) ou (II) sont des estimateurs biaisés de* $\{y,\alpha,\beta\}$.

Preuve : voir annexe 2

Les défauts de cette approche sont relativement nombreux. En plus du biais possible des estimateurs, il faut remarquer qu'au cours d'une nouvelle année, lorsque l'on utilise le modèle sans information sur la valeur annuelle, on utilise généralement des indicateurs susceptibles de révisions au fur et à mesure que les réponses aux enquêtes peuvent être exploitées.

Il s'ensuit que la construction du compte serait fragilisée par l'accumulation des erreurs faites sur l'évaluation de chaque trimestre du fait du caractère auto régressif du modèle.

3.3 Expériences de simulations

Nous avons procédé à quelques exercices de simulations pour illustrer cette méthode et ses défauts. Pour ce faire, nous avons simuler cinq différents modèles dynamiques en trimestriel, puis nous avons annualisé ces observations et effectué une estimation à l'aide de la méthode décrite ci-dessus. Les modèles simulés sont cohérents avec la nulle de la méthode d'estimation proposée, ils correspondent simplement à différentes configurations de stationnarité ou d'intégration des variables intervenant dans la relation. Les cinq modèles simulés sont les suivants :

H1: $y_t = 0.8 y_{t-1} + x_t + \varepsilon_t$, $x \approx I(0)$, i.i.d.N(0,10), ε i.i.d.N (0,1)

H2: $y_t = y_{t-1} + x_t + \varepsilon_t$, $x \approx I(0)$, i.i.d.N(0,10), ε i.i.d.N (0,1)

H3: $y_t = 0.8 y_{t-1} + x_t + \varepsilon_t$, $x \approx I(1)$, $(1\text{-}B)x_t$ i.i.d.N (0,10), ε i.i.d.N(0,1)

H4: $y_t = x_t + \varepsilon_t$, $x \approx I(0)$, i.i.d.N(0,10), ε i.i.d.N (0,1)

H5: $y_t = x_t + \varepsilon_t$, $x \approx I(1)$, $(1\text{-}B)x_t$ i.i.d.N (0,10), ε i.i.d.N(0,1)

Dans H1 et H4, la série simulée est stationnaire. Dans H2, elle est intégrée d'ordre 1 mais non cointégrée avec x. Sous H3 et H5, les deux variables sont intégrées et cointégrées entre elles.

Nous avons effectué 1000 simulations d'échantillons de 100 observations et représenté la distribution des paramètres estimés à l'aide d'un estimateur à noyau gaussien.

Les distributions que nous obtenons sont asymétriques, la masse à gauche du mode est plus importante que la masse à sa droite pour les hypothèses H1,2,3,5 et inversement pour H4. Même si le mode semble bien positionné, un biais peut subsister. La valeur du biais empirique que nous obtenons sur ces échantillons est illustré dans le tableau 1.

Les biais empiriques sont relativement faibles et d'autant plus faibles que les variables sont intégrées d'ordre 1.

Ceci est compréhensible dans la mesure où la divergence des moments croisés est plus rapide pour ces séries.

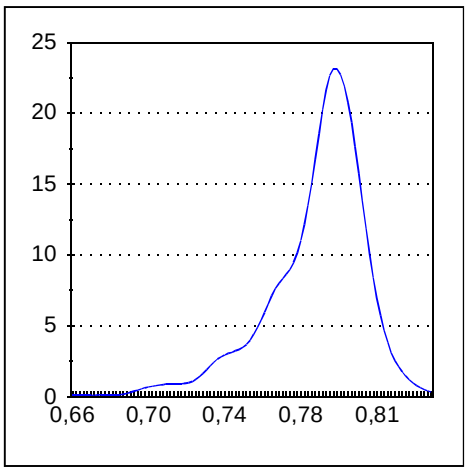
Les distributions que nous obtenons pour le paramètre de l'indicateur présentent une asymétrie inverse à celle des distributions des α, la masse à droite du mode est plus importante que la masse à sa gauche pour les hypothèses H1,2,3,5 et inversement pour H4. Le mode semble correspondre à la valeur sous l'hypothèse nulle. Le biais empirique que nous obtenons sur ces échantillons est illustré dans le tableau 2.

Tableau 1: bias empirique

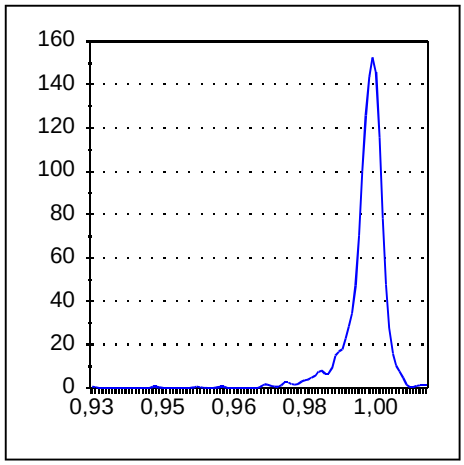
	H1	H2	H3	H4	H5
biais sur α	-0.012	-0.003	-0.002	0.038	-0.003

Tableau 2: bias empirique

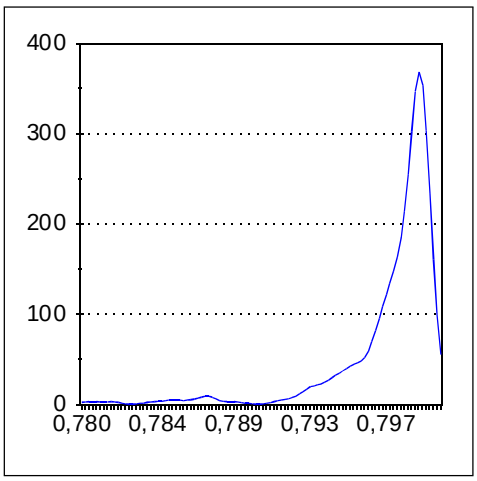
	H1	H2	H3	H4	H5
biais sur β	0.033	0.020	0.010	-0.027	0.003



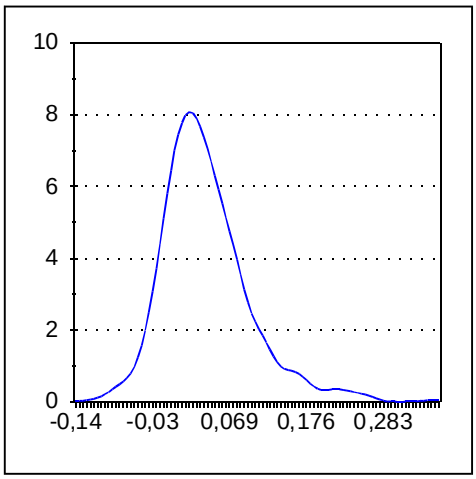
Distribution de alpha sous H1



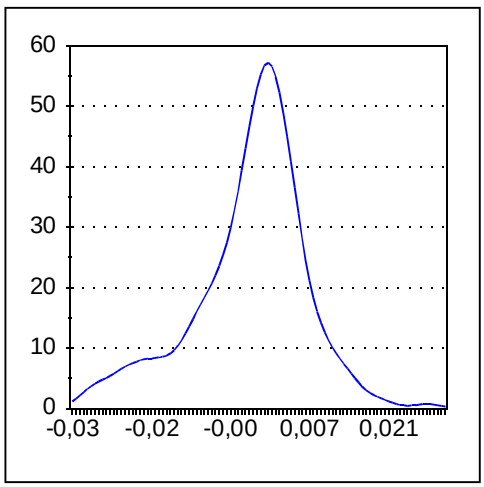
Distribution de alpha sous H2



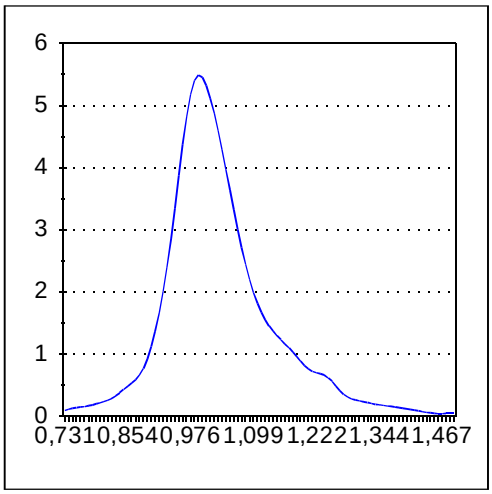
Distribution de alpha sous H3



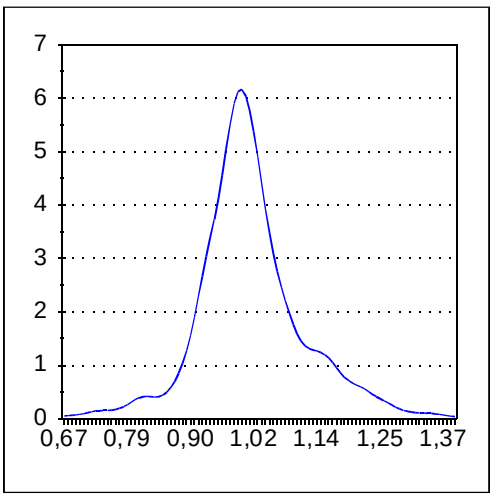
Distribution de alpha sous H4



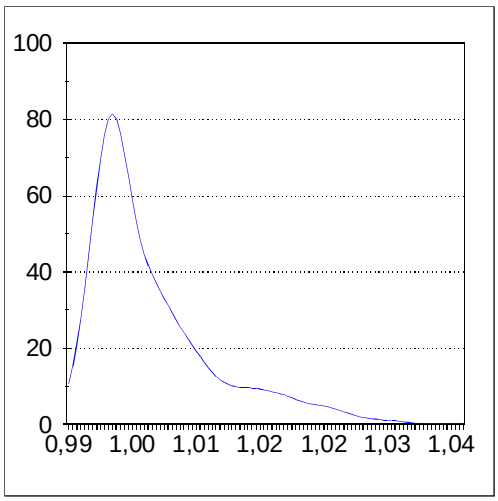
Distribution de alpha sous H5



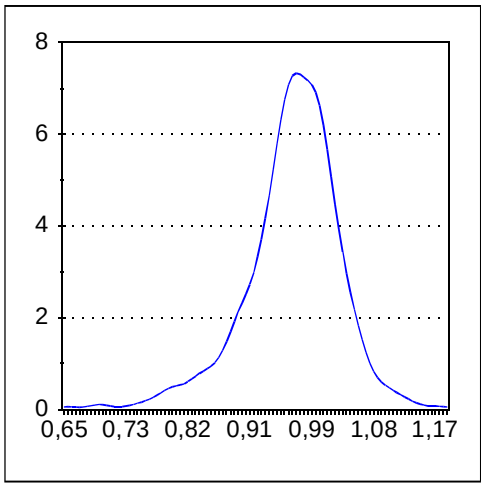
Distribution de bêta sous H1



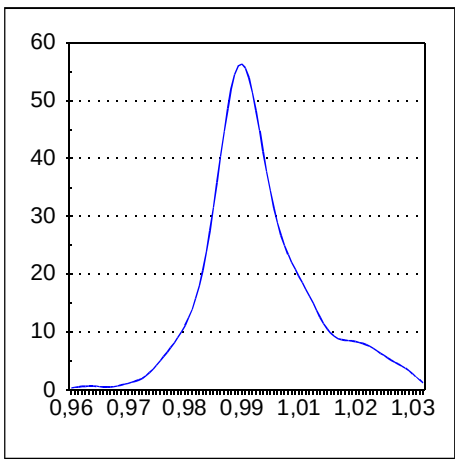
Distribution de bêta sous H2



Distribution de bêta sous H3



Distribution de bêta sous H4



Distribution de bêta sous H5

Les biais empiriques sont plus importants sur le coefficient de la variable qui joue le rôle de l'indicateur. Il est plus faible (et opposé en signe) lorsque les variables sont intégrées d'ordre 1 et non cointégrées entre elles. La cointégration introduit des termes stationnaires qui polluent la vitesse de convergence pour tous les coefficients. Les biais observés sur les β sont de signe opposé à celui observé sur les α, ce biais provient de l'interaction des contraintes annuelles qui sont imposées dans l'estimation. Il y a donc une sorte de compensation entre les deux erreurs. Ceci entraîne que l'erreur faite sur les évaluations trimestrielles peut être plus faible.

4 Construction d'un modèle trimestriel saisonnier

4.1 Le modèle

L'écriture d'un modèle saisonnier est ici entendue dans le sens où le comportement des variables trimestrielles peut être relativement bien décrit par une représentation SAR avec peu de retards. Dans cette écriture, la variable d'intérêt présente une forte autocorrélation d'ordre 4. Si nous considérons des variables intégrées à des fréquences saisonnières, il nous faut alors travailler avec les différences d'ordre 4 des variables.

Ce qui a été dit a propos d'une écriture qui permette d'englober différentes configurations de cointégration entre les variables demeurent valables a l'exception près que par cette écriture nous ne pouvons pas prendre en compte les formes de cointégration saisonnière.

Il faudrait en effet dans ce cas faire intervenir les variables en niveau décalées de 1,2 et 3 dates (cf. théorème de représentation énoncé dans Gregoir (1994)). Mais pour le travail que nous voulons effectuer, ceci n'a pas beaucoup de pertinence car cette structure de cointégration saisonnière ne peut être induite à partir de l'information accessible au niveau annuel.

En d'autres termes, elle n'est pas identifiable. En effet, dans le cas le plus simple de cointégration saisonnières nous avons, lorsque nous considérons un vecteur de variables endogènes :

$$\begin{aligned} y_t - y_{t-4} &= \alpha_0\beta_0\left(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}\right) \\ &+\alpha_\pi\beta_\pi\left(y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}\right) \\ &+\alpha_{\frac{\pi}{2}}^1\left(\beta_{\frac{\pi}{2}}^1\left(y_{t-1} - y_{t-3}\right) + \beta_{\frac{\pi}{2}}^2\left(y_{t-2} - y_{t-4}\right)\right) \\ &+\alpha_{\frac{\pi}{2}}^2\left(\beta_{\frac{\pi}{2}}^1\left(y_{t-2} - y_{t-4}\right) + \beta_{\frac{\pi}{2}}^2\left(y_{t-3} - y_{t-1}\right)\right) + \varepsilon_t \end{aligned}$$

mais la seule relation de cointégration que nous pouvons observer est celle décrite par β_0 (associée à la fréquence 0) qui demeure sur les quantités annuelles (la preuve de cette affirmation est relativement longue, elle n'est pas donnée pour ne pas alourdir le texte). Il semble en fait plus important de construire des séries comptables qui reproduisent les propriétés de cointégration saisonnière rencontrées sur les indicateurs.

Si deux indicateurs bruts sont cointégrés, il est souhaitable que les variables comptables construites à partir de leur observation satisfassent la même propriété. Qualitativement, on voit que cette conservation des propriétés de cointégration saisonnière sera satisfaite si les variables comptables sont construites à partir de mêmes filtrages linéaires "stationnaires" mono-directionels des indicateurs. Par linéarité, les propriétés seront conservées.

4.1.1 Variable de flux

Nous proposons donc de considérer la forme suivante de modèles dynamiques :

$$y_t = \alpha y_{t-4} + x_t\beta + \varepsilon_t, \quad |\alpha| \le 1 \qquad (3)$$

où $\varepsilon_t \approx i.i.d.$ avec $\mathrm{E}\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = 0, V\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = \sigma_\varepsilon^2$ et x_t est un vecteur d'indicateurs observés trimestriellement de données non désaisonnalisées (il peut contenir des différences quatrièmes ou des variables décalées dans le temps). Nous en déduisons la dynamique annuelle :

$$Y_a = \alpha Y_{a-1} + X_a\beta + \mathrm{E}_a$$

La grande simplicité du passage du modèle trimestriel au modèle annuel permet de considérer des modèles avec de l'autocorrélation des résidus saisonnière (elle demeure conservée au niveau annuel) comme l'ont proposé Bournay et Laroque (1979). L'autocorrélation

des résidus est alors directement mesurée sur les résidus annuels.

4.1.2 Variable de stock

Pour les variables de stock, nous proposons de considérer des modèles dynamiques de la forme suivante :

$$y_t = \alpha y_{t-4} + x_t\beta + \varepsilon_t, \quad |\alpha| \le 1 \qquad (4)$$

où $\varepsilon_t \approx i.i.d.$ avec $E\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = 0, V\varepsilon_t \left| \underline{x}_t, \underline{y}_{t-1} \right. = \sigma_\varepsilon^2$ et x_t est un vecteur d'indicateurs observés trimestriellement de données non désaisonnalisées. Nous en déduisons la dynamique annuelle:

$$Y_a = \alpha Y_{a-1} + x_{4a}\beta + \varepsilon_{4a}$$

Là encore, la simplicité du passage du modèle trimestriel au modèle annuel permet de considérer des modèles avec de l'autocorrélation des résidus saisonnière qui est mesurée sur les résidus annuels.

4.2 Identifiabilité de la série trimestrielle

Néanmoins l'usage d'une telle modélisation n'est pas sans quelques inconvénients. Le plus important porte sur l'identifiabilité de la série trimestrielle à partir de l'information annuelle. En effet, il existe un nombre infini de séries trimestrielles qui vérifient à la fois une telle écriture autorégressive et la contrainte comptables liant les quantités trimestrielles à la quantité annuelle. Ceci est particulièrement évident pour les variables de stock, puisque nous n'imposons aucune contrainte sur les valeurs des trois premiers trimestres de l'année. Mais ceci est aussi vrai pour le modèle le plus général de variables de flux.

Proposition 4 *: si y_t vérifie la dynamique autorégressive $\varphi(B^4) y_t = z_t$ où $\varphi(u)$ est de degré n, et satisfait les contraintes $\forall a, \sum_{i=1}^{4} y_{4(a-1)+i} = Y_a$, alors $y^*_{4(a-1)+i} = y_{4(a-1)+i} + \sum_{j=1}^{n} \lambda_j^a k_{j,i}$ satisfait aussi les contraintes et vérifie la dynamique autorégressive lorsque $(\lambda_j)_{j=1,K,n}$ sont les racines du polynôme $\varphi(u)$ et les constantes réelles $(k_{j,i})\ j = 1, K, n, i = 1,2,3,4$ sont telles que $\forall j, \sum_{i=1}^{4} k_{j,i} = 0$.*

Exemple 2 : *Dans le cas du modèle (3), les processus de la forme $y^*_{4(a-1)+i} = y_{4(a-1)+i} + \sum_{j=1}^{n} \lambda_j^a k_{j,i}$ satisfont à la fois la dynamique autorégressive et les contraintes sur les sommes annuelles. L'espace des solutions possibles est donc de dimension trois, mais ces degrés de liberté jouent un rôle de plus en plus faible sur l'évolution de au fur et à mesure que l'on s'éloigne de la date initiale.*

Ceci signifie que nous devons imposer une information supplémentaire pour pouvoir estimer une série trimestrielle. Différents critères peuvent être envisagés. Nous nous limiterons à des critères associé à la variabilité de la série.

Dans la mesure où nous n'observons pas la vraie série, il nous semble prudent de rechercher des séries à faible variabilité ou du moins pour lesquelles les conditions initiales contribuent le moins possible à la variabilité totale de la série.

4.3 Méthode d'estimation et ses propriétés

Nous recherchons des méthodes simples, facilement mises en oeuvre, peu coûteuses en temps de calcul, satisfaisant les approches standard de la statistique. Le paragraphe précédent nous a montré qu'une information supplémentaire doit être introduite pour obtenir l'estimation d'une série trimestrielle. De façon générale, nous posons le programme pour des variables de flux sous la forme suivante sans prendre en compte la contrainte sur la valeur absolue de α:

$$(\mathrm{III}) \begin{cases} \begin{cases} \underset{y_0}{Min} Critère(\hat{y}(y_0)) \\ (1111)\, y_0 = Y_0 \end{cases} \\ \hat{y}(y_0) \in \arg \begin{cases} \underset{y,\alpha,\beta}{Min} \sum_{t=5}^{T} (y_t - \alpha y_{t-4} - x_t\beta)^2 \\ I_A \otimes (1111)\, y = Y \end{cases} \end{cases}$$

où y_0 représente les conditions initiales sur lesquelles on reporte l'indétermination. Le problème pour des variables de stock s'écrit de la même façon aux contraintes près :

$$(IV)\begin{cases} \begin{cases} \underset{y_0}{Min}\, Critère(\hat{y}(y0)) \\ (0001)\, y_0 = Y_0 \end{cases} \\ \hat{y}(y_0) \in \arg \begin{cases} \underset{y,\alpha,\beta}{Min} \sum_{t=5}^{T} (y_t - \alpha y_{t-4} - x_t \beta)^2 \\ I_A \otimes (0001)\, y = Y \end{cases} \end{cases}$$

Les propriétés des solutions du programme "primaire" de minimisation de la somme des carrés des résidus sont proches de celles trouvées par Chow et Lin (1971) en ce qui concernent les estimateurs.

Proposition 5: *les estimateurs de* (α,β) *issus du programme "primaire" sont les estimateurs des moindres carrés du modèle annuel.*

Preuve : *voir annexe 2*

Par ailleurs, le caractère quadratique des fonctions que l'on minimise assure que les familles de solutions $\hat{y}(y_0)$ sont des fonctions linéaires des conditions initiales :

Proposition 6: *les solutions* $y(y_0)$ *du programme "primaire" sont données par les équations suivantes:*

$$\hat{y}(y_0) = (I_{A-1} - \alpha J_{A-1})^{-1} \otimes I_4 (\alpha e_{A-1} \otimes I_4 y_0 + x\beta) + (I_{A-1} - \alpha J_{A-1})^{-1} \otimes \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \hat{\varepsilon} / 4$$

pour des variables de flux et

$$\hat{y}(y_0) = (I_{A-1} - \alpha J_{A-1})^{-1} \otimes I_4 (\alpha e_{A-1} \otimes I_4 y_0 + x\beta) + (I_{A-1} - \alpha J_{A-1})^{-1} \otimes \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \hat{\varepsilon}$$

pour des variables de stock, où e_{A-1} *est un vecteur de taille A-1 dont toutes les valeurs sont égales à zéro excepté la première qui vaut 1,* J_{A-1} *est la première matrice de Jordan dont la première sous diagonale est égale à 1 et les autres paramètres sont nuls et* $\hat{\varepsilon}$ *est le vecteur des résidus estimés du modèle annuel correspondant.*

Preuve : *voir annexe 2*

Ces propriétés sont conservées si nous considérons des modèles avec des spécifications légèrement plus compliquées, i.e. avec plus de retards ou avec de l'autocorrélation d'ordre quatre des termes d'erreurs. Il faut simplement dans ce dernier cas procéder à une estimation en plusieurs étapes pour construire un estimateur des moindres carrés généralisés des paramètres du modèles sur quantités annuelles.

Il nous faut maintenant regarder les problèmes liés au programme complet que nous considérons. Le premier critère naturel que nous pouvons considérer est celui qui fait choix de la trajectoire qui présente la variabilité la plus faible[2]. Le critère est donc :

$$V\begin{pmatrix} y_0 \\ \hat{y}(y_0) \end{pmatrix}$$

Les solutions de ce programme sont simples (cf. annexe 2), néanmoins les interactions qui peuvent exister entre les conditions initiales et les séries d'indicateurs peuvent entraîner des révisions sur le

2 Nous faisons abstraction ici des problèmes de non stationarité de la solution obtenue, la variance de la variable I(1) en niveau est un critère peu satisfaisant dans la mesure où il diverge en T (où T est la taille de l'échantillon) et où proprement redéfini en le divisant par T (ce qui est sans conséquence à distance finie), il converge vers une variable aléatoire. Ce critère n'a donc pas toujours de pertinence statistique. Il faudrait pouvoir proposer un critère satisfaisant qui converge vers une constante quelle que soit la nature de la série, intégrée ou non. Il faut toutefois noter que la divergence de la variance ne peut être due aux conditions initiales, mais à la présence de variables intégrées parmi les indicateurs.

début de période lors des réestimations nécessaires après l'introduction de nouvelles observations.

Le fait que la solution du problème "primaire" soit linéaire en les conditions initiales permet aussi de considérer comme critère la contribution des conditions initiales à la variance de la série [3].

Le critère devient :

$$\frac{Cov\left(\begin{pmatrix} y_0 \\ \hat{y}(y_0) \end{pmatrix}\left(\begin{matrix} y_0 \\ (I_{A-1} - J_{A-1})^{-1} \varepsilon_{A-1} \otimes I_4 \, y_0 \end{matrix}\right)\right)}{V\begin{pmatrix} y_0 \\ \hat{y}(y_0) \end{pmatrix}}$$

Les conditions du premier ordre issues de ce programme qualifient en général deux extrema, un minimum et un maximum (cf. annexe 2). Le calcul de la fonction objectif permet de faire le choix des conditions initiales appropriées. Remarquons que la résolution de ces programmes avec un modèle auto régressif avec plus de retard ne pose pas de problèmes techniques majeurs.

Par ailleurs, une autre méthode simple, facilement mise en oeuvre, consisterait à construire des valeurs initiales vérifiant la contrainte annuelle et reproduisant le profil de l'indicateur brut utilisé.

Enfin, un soucis de lisibilité des comptes peut inciter les comptables trimestriels à construire les séries qui présentent le moins de modifications possible entre deux publications [4]. Le critère devient alors la minimisation de la distance entre les deux dernières estimations sur une période sélectionnée, ce qui s'écrit:

$$\left(\hat{y}_1 - \hat{y}_2(y_0)\right)' P \left(\hat{y}_1 - \hat{y}_2(y_0)\right)$$

où $\hat{y}_1$ est issu des données de l'année précédente , $\hat{y}_2(y_0)$ est calculé avec les données révisées et restreint à la période d'observation correspondante, et **P** sélectionne la période sur laquelle on désire rendre les observation les plus proches possible avec éventuellement des poids pour qualifier les périodes d'intérêt. Le problème de cette approche réside dans son initialisation.

Lorsque le nombres de retards présents dans l'équation dynamique est supérieur à un, les inconvénients d'une telle méthode résident dans le fait que l'on travaille avant tout sur des évolutions. Les évaluations successives des années deviennent sensibles aux taux de croissance des années antérieures. Ceci correspond bien aussi à l'idée de travailler sur des variables stationnaires. En revanche, lorsque le nombre de retard est de un et le coefficient de retard a une valeur proche de 1, en révisant la valeur de l'année précédente, on décale la valeur courante d'autant et on préserve au premier ordre le taux de croissance (sans révisions des indicateurs).

5 Etudes des résultats empiriques

5.1 Choix des critères d'évaluation

Il est difficile de définir des critères qui au niveau trimestriel permettent de juger de la qualité d'un ajustement économétrique parmi plusieurs modèles disponibles. Seules les "performances" du modèle au niveau annuel peuvent être évaluées dans la mesure où seules les quantités annuelles de la variable dépendante sont observées. Néanmoins, en termes de comptabilité nationale, il faut quatre ans pour connaître la valeur "définitive" d'un agrégat et les évaluations de comptes trimestriels ne peuvent prétendre à partir d'indicateurs relativement simples correspondant à des niveaux agrégés de la nomenclature décrire avec précision des évolutions qui seront obtenues à l'issue d'un processus de quatre années basé sur les comptes des entreprises à un niveau fin. Il me semble donc utile de comparer les performances des méthodes en compétition à la première évaluation disponible des comptes annuels (le compte provisoire) issue d'un travail de confrontation des données des comptables annuels et trimestriels. Néanmoins il ne faut pas attribuer trop de significativité à ces comparaisons dans la mesure où le

3 Une remarque semblable à celle faite pour le critère précédent reste pertinente.

4 Cette approche a été suggérée par Jacques Bournay .

poids relatif de chacune des parties prenant part à la concertation varie suivant les variables et les années.

5.2 Le cas de la production de textile sur données françaises

Pour la période 1970-1992, l'étalonnage retenu dans les comptes trimestriels est le suivant :

$$y = 25.57x + 1283.88t$$

la racine carré de la moyenne des erreurs au carré (RMSE) sur données annuelles est de 3769.9 et la statistique de Durbin-Watson s'élève à 0.44, illustrant une forte corrélation des résidus annuels. La recherche d'un modèle auto régressif d'ordre 4 sur les mêmes séries aboutit au modèle suivant (une indicatrice a été introduite en 1986 année de départ de la rétropolation de la base 1980) :

$$y = 1.04\, y_{-4} + 35.2x - 35.6x_{-4}$$

le RMSE est de 2121.7 et la statistique de Durbin-Watson vaut 1.82. Les statistiques obtenues pour les deux modèles nous permettent de dire que le modèle dynamique domine d'un point de vue statistique le modèle statique.

Dans le tableau 3 nous avons simulé à l'horizon un, les deux modèles retenus dans les deux approches, pour différentes versions des comptes annuels. Nous avons comparé les taux que nous donnaient ces équations à la valeur publiée après la concertation des comptables trimestriels et annuels. Le modèle dynamique calé donne des résultats convenables excepté en 1992 et 1993.

Le modèle statique calé quant à lui donne des taux plus satisfaisants pour ces deux années, mais est dominé par le modèle dynamique calé pour les trois premières années.

Il faut néanmoins être prudent quant à la signification de ces résultats, ils dépendent du poids de chacun des intervenants dans la phase de concertation.

Nous avons utilisé les résultats présentés ci-dessus pour calculer les séries trimestrielles correspondantes en imposant la contrainte $|\alpha| \leq 1$ dans l'estimation.

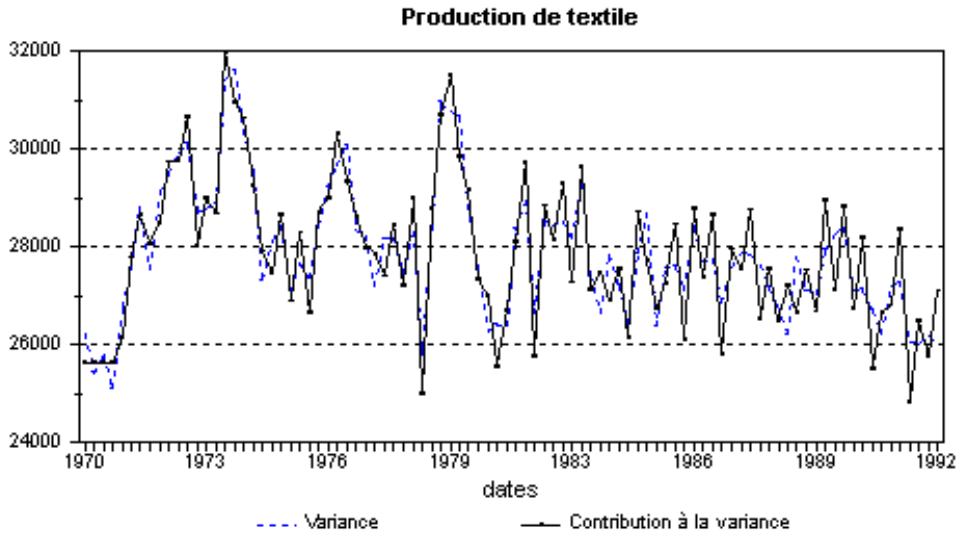
Dans le graphique 1, les séries obtenues pour les deux critères retenus (minimisation de la variance et minimisation de la contribution des conditions initiales à la variance) ont été représentées. On observe que la série obtenue en minimisant la contribution à la variance des conditions initiales présente une première années pratiquement plate.

En l'absence d'indicateur utilisé implicitement dans l'estimation de la dynamique infra-annuelle de cette première année, il semble naturel que la minimisation de la contribution à la variance corresponde à un profil relativement plat lors de l'année d'initialisation. Le profil obtenu pour les autres années diffère entre les deux méthodes dans la mesure où le coefficient auto régressif sur la variable comptable est égal à 1, ce qui entraîne une persistance de l'influence du profil de la première année sur le reste de la série. Ainsi la série obtenue en minimisant la variance apparaît beaucoup moins variable que la série obtenue en minimisant la contribution à la variance des conditions initiales.
Nous comparons ensuite la série obtenue par minimisation de la contribution à la variance à la série publiée par les comptes trimestriels lors des résultats détaillés du quatrième trimestre 1992.
La série obtenue à partir du modèle dynamique joue le rôle d'une série brute, puisqu'étalonnée sur un indicateur brut, elle présente donc des fluctuations plus marquées.

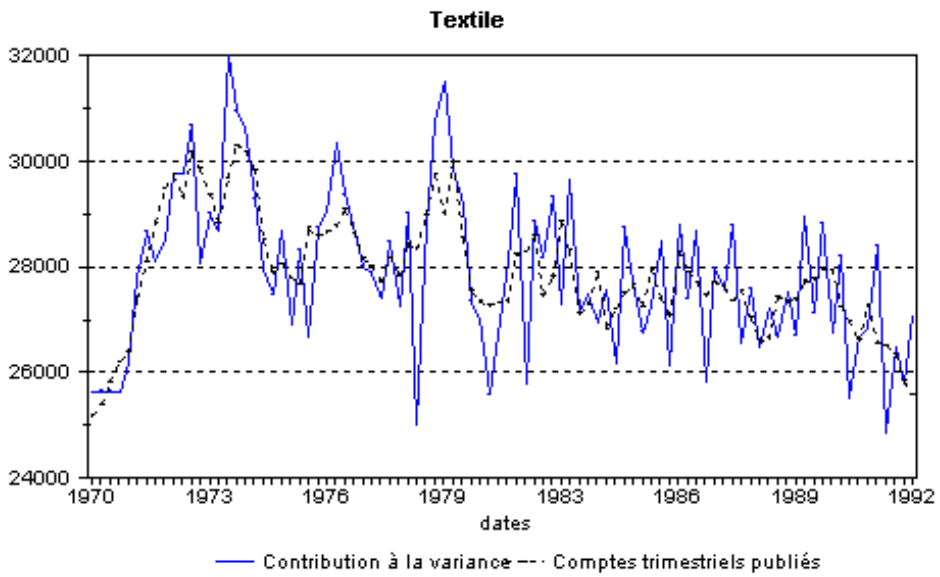
Tableau 3: Simulation à l'horizon 1

année	modèle statique non calé	modèle statique calé	modèle dynamique non calé	modèle dynamique calé	publication au provisoire
1989	-2.7	-1.2	-2.6	-0.2	-0.3
1990	-5.5	-2.1	-2.6	-1.9	-1.1
1991	-7.0	-3.1	-3.6	-2.6	-1.9
1992	-6.1	-2.2	-2.3	-0.8	-2.9
1993	-7.2	-3.6	-4.1	-5.3	-1.7

Graphique 1: Production textile



Graphique 2: Comparation avec les comptes trimestriels



5.3 Le cas de la production de minerais et métaux non-ferreux sur données françaises

Pour la période 1970-1992, l'étalonnage retenu dans les comptes trimestriels est le suivant :

$$y = 10.24x + 730.26t$$

la racine carré de la moyenne des erreurs au carré (RMSE) sur données annuelles est de 2901.3 et la statistique de Durbin-Watson s'élève à 0.73. La recherche d'un modèle auto régressif simple d'ordre 4 sur les mêmes séries aboutit au modèle suivant :

$$y = 0.89\,y_{-4} + 13.68x - 12.29x_{-4}$$

le RMSE est de 2114.3 et la statistique de Durbin-Watson vaut 2.18. Le modèle dynamique obtenu est proche du modèle suivant :

$$\left(1 - 0.89B^4\right)\left(y - 13.74x\right) = \varepsilon$$

qui indique une possibilité de relation de cointégration entre les variables, relation mal capturée dans le modèle statique du fait de la présence d'une tendance déterministe. Là encore les statistiques obtenues pour le modèle dynamique sont beaucoup plus satisfaisantes que celles obtenues pour le modèle statique.

Dans le tableau 4 nous avons simulé à l'horizon un les deux modèles retenus dans les deux approches, pour différentes versions des comptes annuels. Nous avons comparé comme précédemment les taux que nous donnaient ces équations à la valeur publiée après la concertation des comptables trimestriels et annuels. Aucun modèle semble d'un qualité supérieure, compliquer légèrement le modèle dynamique permettrait peut-être d'améliorer ses résultats.

Nous avons utilisé les résultats présentés ci-dessus pour calculer les séries trimestrielles correspondantes. Dans le graphique ci-dessous, les séries obtenues pour les deux critères retenus (minimisation de la variance et minimisation de la contribution des conditions initiales à la variance) ont été représentées. Dès la troisième année, les profils obtenus par les deux méthodes sont très proches. L'influence des conditions initiales disparaît avec la rapide convergence vers 0 des puissances de 0.89.

Nous comparons ensuite la série obtenue par minimisation de la contribution à la variance à la série publiée par les comptes trimestriels lors des résultat détaillés du quatrième trimestre 1992 (graphique 4). Il faut noter que le profil des comptes trimestriels publiés est particulièrement heurté en fin de période soulignant un problème de traitement de la saisonnalité.

5.4 Le cas de la production de la fonderie et travail des métaux sur données françaises

Pour la période 1970-1992, l'étalonnage retenu dans les comptes trimestriels est le suivant :

$$y = 32.78x$$

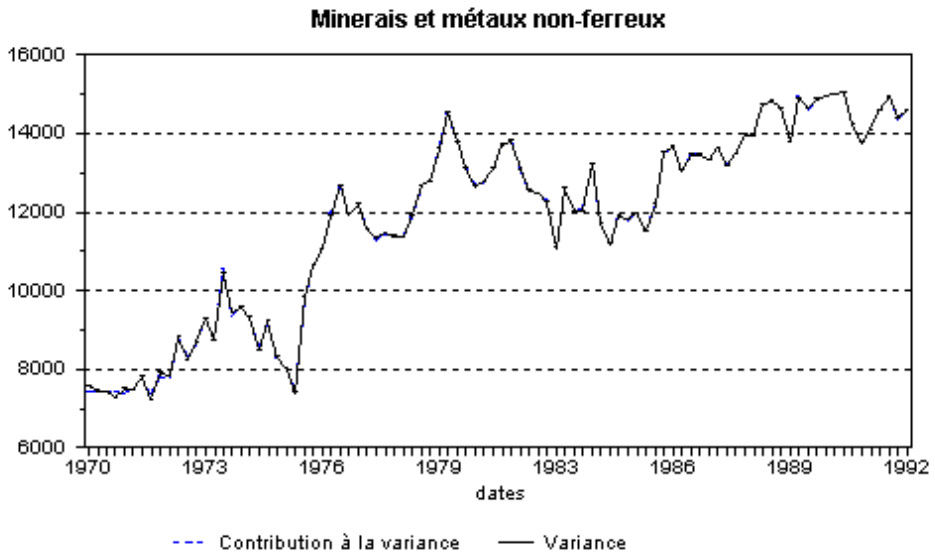
la racine carré de la moyenne des erreurs au carré (RMSE) sur données annuelles est de 4412.7 et la statistique de Durbin-Watson s'élève à 0.46, illustrant une forte corrélation des résidus annuels. La recherche d'un modèle autorégressif simple d'ordre 4 sur les mêmes séries aboutit au modèle suivant :

$$y = 0.82\,y_{-4} + 29.27x - 23.24x_{-4}$$

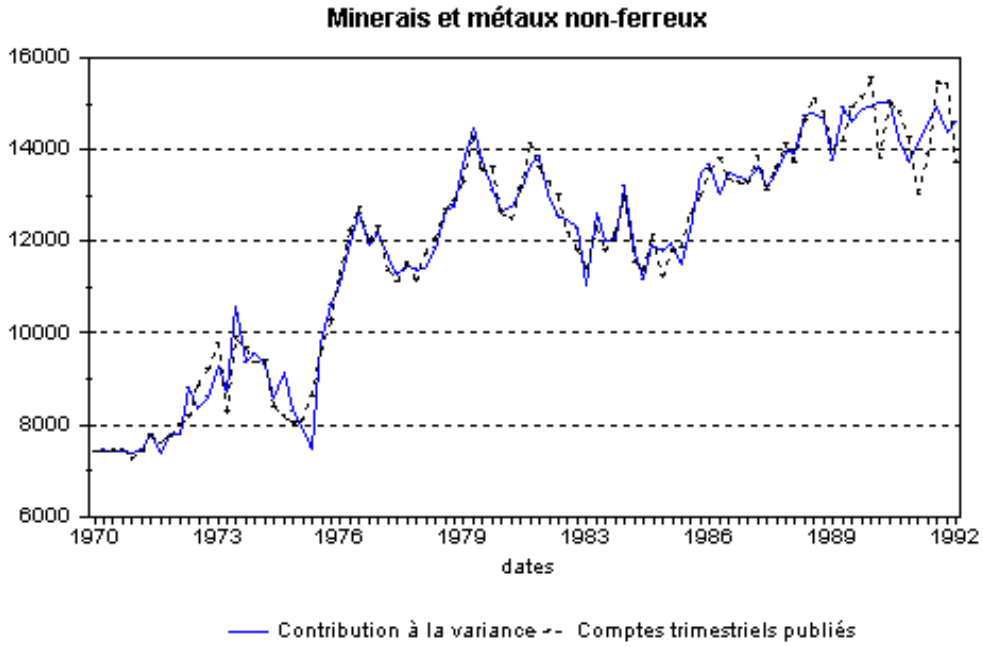
Tableau 4: simulation à l'horizon 1

année	modèle statique non calé	modèle statique calé	modèle dynamique non calé	modèle dynamique calé	publication au provisoire
1989	6.9	3.5	2.4	2.7	6.9
1990	3.6	2.3	0.8	1.6	0.9
1991	3.9	-0.1	-2.5	-5.7	-3.3
1992	5.7	2.6	1.4	2.9	-3.1
1993	-1.0	-4.4	-7.8	-7.0	-1.9

Graphique 3: Minerais et métaux non-ferreux



Graphique 4: Comparaison avec les comptes trimestriels



le RMSE est de 2958.2 et la statistique de Durbin-Watson vaut 1.15. Nous remarquons que le modèle dynamique est proche d'un modèle du type

$$\left(1-0.82B^4\right)\left(y-28.90x\right)=\varepsilon$$

qui fait penser à la possibilité d'une relation de cointégration entre le poste comptable et l'indicateur, ce qui correspondait au choix effectué dans la méthode statique. Là encore les statistiques obtenues pour le modèle dynamique sont beaucoup plus satisfaisantes que celles obtenues pour le modèle statique. Néanmoins, la forte corrélation des résidus du modèle dynamique peut inciter à rechercher un modèle avec autocorrélation des résidus. Ceci est possible avec un légère modification des algorithmes envisagés, mais ne sera pas effectué ici pour ne pas surcharger le propos.

Dans le tableau 5 nous avons simulé à l'horizon un les deux modèles retenus dans les deux approches, pour différentes versions des comptes annuels. Nous avons comparé comme précédemment les taux que nous donnaient ces équations à la valeur publiée après la concertation des comptables trimestriels et annuels. La décision prise pour l'évolution de la dernière année semble peu en accord avec les deux types de modèle.

Comme précédemment, nous avons utilisé les résultats présentés ci-dessus pour calculer les séries trimestrielles correspondantes. Dans le graphique ci-dessous, les séries obtenues pour les deux critères retenus (minimisation de la variance et minimisation de la contribution des conditions initiales à la variance) ont été représentées.

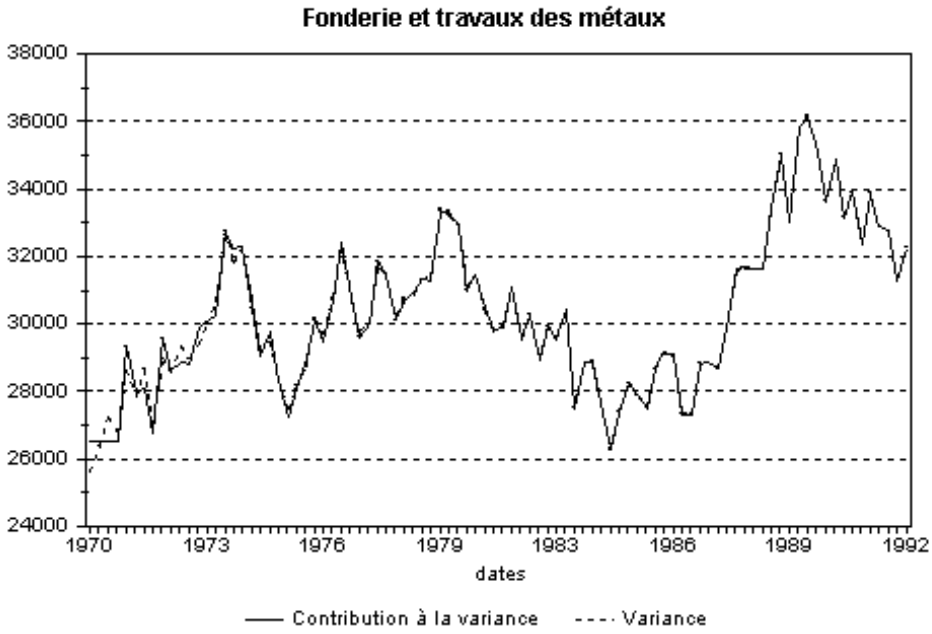
Comme nous l'observons, les deux méthodes donnent des séries très voisines dès la cinquième année, le profil des conditions initiales de la série obtenue en minimisant la variance est assez heurté et joue un rôle plus persistant dans le profil du reste de la série.

Nous comparons enfin la série obtenue par minimisation de la contribution à la variance à la série publiée par les comptes trimestriels lors des résultat détaillés du quatrième trimestre 1992 (graphique 6). La série issue du modèle dynamique présente un profil saisonnier qui oscille autour de la série publiée par les comptes trimestriels.

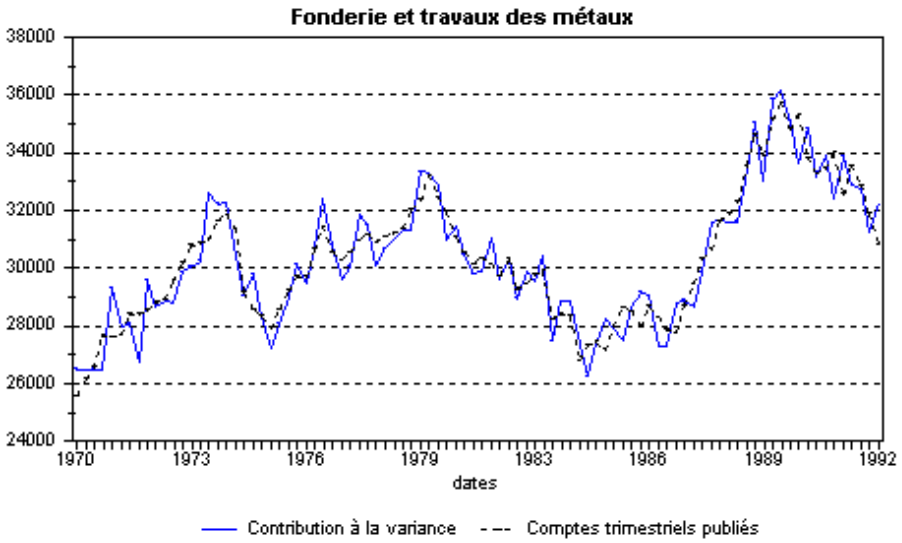
Tableau 5: Simulation à l'horozon 1

année	modèle statique non calé	modèle statique calé	modèle dynamique non calé	modèle dynamique calé	publication au provisoire
1989	10.6	7.1	5.9	7.1	6.8
1990	2.4	0.6	0.9	-1.5	0.9
1991	-6.0	-4.4	-3.9	-3.5	-3.4
1992	-7.6	-2.9	-2.8	-2.3	-3.1
1993	-10.8	-8.4	-7.7	-7.9	-2.2

Graphique 5: Fonderie et traveaux des métaux



Graphique 6: Comparaison avec les comptes trimestriels



ANNEXE 1

Proposition 1 :

Les conséquences de l'agrégation temporelle sur les propriétés de stationnarité des séries ont fait l'objet de nombreux papiers. Dans la mesure où nous nous intéressons à des variables agrégées sur des périodes de temps finies les propriétés sont conservées.

Pour plus de détails et des éléments bibliographiques, nous renvoyons le lecteur à un papier récent de Pierse R.G et A.J.Snell (1995).

Proposition 2 :

La proposition 2 est une conséquence directe de la proposition 1.

ANNEXE 2

Preuve de la proposition 3 :

Nous utilisons les notations introduites dans le paragraphe 2.2. Des calculs directs à partir des c onditions du premier ordre permettent d'écrire :

$$\hat{y} - y = (I - \hat{\alpha} J)^{-1} \left\{ \left(\hat{\alpha} e_t - \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} v(\hat{\alpha}) \right) (\hat{y}_0 - y_0) \right.$$
$$\left. \left(I - \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) \right) \left(x \left(\hat{\beta} - \beta \right) + (\hat{\alpha} - \alpha)(e_t y_0 + Jy) - \varepsilon \right)$$

ce qui illustre l'absence de convergence des estimateurs des valeurs des trimestres, mais aussi le fait que le biais sur $\hat{y}$ est fonction des biais sur les paramètres (α, β, y_0). A l'aide des mêmes conditions d'orthogonalité, nous obtenons:

$$M \begin{pmatrix} \hat{y}_0 - y_0 \\ \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} = \begin{pmatrix} v(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) \varepsilon \\ \left(J\hat{y} + e_t \hat{y}_0 \right)' \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) \varepsilon \\ x' \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) \varepsilon \end{pmatrix}$$

où

$$M = \begin{pmatrix} v(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} v(\hat{\alpha}) & v(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} ' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) (e_t \hat{y}_0 + J\hat{y} \; x) \\ (e_t \hat{y}_0 + J\hat{y} \; x)' \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} v(\hat{\alpha}) & (e_t \hat{y}_0 + J\hat{y} \; x)' \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) (e_t \hat{y}_0 + J\hat{y} \; x) \end{pmatrix}$$

apparaît donc que même en l'absence de corrélation à toutes les dates entre les indicateurs et les termes d'erreur, le deuxième élément du vecteur de droite, à savoir

$$\left(J\hat{y} + e_t \hat{y}_0 \right)' \Phi(\hat{\alpha})' \left(\Phi(\hat{\alpha}) \Phi(\hat{\alpha})' \right)^{-1} \Phi(\hat{\alpha}) \varepsilon$$

introduit un biais qui se propage à tous les termes. En effet ce terme n'est pas toujours d'espéran ce nulle du fait de l'action de l'opérateur $\Phi(\hat{\alpha})$ qui mélange les observations en annualisant des sommes pondérées.

Preuve de la proposition 4 :

La preuve est laissée au lecteur, il suffit de vérifier que la solution proposée satisfait l'équati on.

Preuve de la proposition 5 :

Le modèle s'écrit vectoriellement sous la forme suivante :

$$y_1 = \alpha\left(J^4 y_1 + e_{A-1} \otimes I_4 y_0\right) + x\beta + \varepsilon$$

où y_1 est le vecteur des valeurs trimestrielles des dates 5 à 4A et y_0 le vecteur des conditions initiales. Par la suite , Y_1 représente le vecteur des valeurs annuelles correspondant à y_1. Le lagrangien du système est :

$$L = \left(y_1 - \alpha\left(J^4 y_1 + e_{A-1} \otimes I_4 y_0\right) - x\beta\right)'\left(y_1 - \alpha\left(J^4 y_1 + e_{A-1} \otimes I_4 y_0\right) - x\beta\right) + \lambda'(\psi y_1 - Y_1)$$

où ψ est la matrice d'annualisation appropriée suivant que l'on travaille avec des variables de flux ou des variables de stock. Les conditions du premier ordre sont simplement dérivées du lagrangien et donnent :

(1) $\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right)'\left(\hat{y}_1 - \hat{\alpha}\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right) - x\hat{\beta}\right) = 0$

(2) $x'\left(\hat{y}_1 - \hat{\alpha}\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right) - x\hat{\beta}\right) = 0$

(3) $\psi\hat{y}_1 - Y_1 = 0$

(4) $\left(I - \hat{\alpha}J^4\right)'\left(\hat{y}_1 - \hat{\alpha}\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right) - x\hat{\beta}\right) + \psi'\hat{\lambda} = 0$

fe (4) reporté dans (1) et (2), nous tirons :

(5) $x'\left(I - \hat{\alpha}J^4\right)^{-1}\psi'\hat{\lambda} = 0$

Et

(6) $\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right)'\left(I - \hat{\alpha}J^4\right)^{-1}\psi'\hat{\lambda} = 0$

et de (3) et (4), nous obtenons l'expression de $\hat{\lambda}$:

(7) $\hat{\lambda} = -\frac{1}{e'e}\left(I_{A-1} - \hat{\alpha}J_{A-1}\right)'\left(I_{A-1} - \hat{\alpha}J_{A-1}\right)\left(Y_1 - \left(I_{A-1} - \hat{\alpha}J_{A-1}\right)^{-1}\left(\hat{\alpha}e_{A-1}Y_0 + X\beta\right)\right)$

où **e** est égal à (1111) ou $(0\,0\,0\,1)$ suivant les cas. Le résultat en est déduit en remarquant que les égalités suivantes sont satisfaites :

$$\psi\left(I - \hat{\alpha}J^4\right)^{-1}x = \left(I_{A-1} - \hat{\alpha}J_{A-1}\right)^{-1}X$$

et

$$\psi\left(I - \hat{\alpha}J^4\right)^{-1}\left(J^4 \hat{y}_1 + e_{A-1} \otimes I_4 y_0\right) = \left(I_{A-1} - \hat{\alpha}J_{A-1}\right)^{-1}\left(J_{A-1}Y_1 + e_{A-1}Y_0\right)$$

Ce résultat peut être étendu à des modèles avec des dynamiques plus longues sans aucun problème.

La solution de ce problème est simple, le lagrangien s'écrit sous la forme suivante :

$$L = y_0' y_0 + \hat{y}_1(y_0)' \hat{y}_1(y_0) - \overline{Y}^2 + \lambda(ey_0 - Y_0)$$

où $\overline{Y}$ représente la moyenne sur l'ensemble des observations des quantité de référence (annuelle ou quatri ème trimestre) et **e** est le vecteur d'annualisation approprié. Si l'on note $c(\alpha)$ la quantité suivante :

$$c(\alpha) = 1 + \alpha^2 \left(e_{A-1}' (I_{A-1} - \alpha J_{A-1})'^{-1} (I_{A-1} - \alpha J_{A-1})^{-1} e_{A-1} \right)$$

les conditions du premier ordre issues du lagrangien s'écrivent :
$c(\hat{\alpha})\hat{y}_0 + \hat{Z} + e'\hat{\lambda}_1 = 0$

où
$$\hat{Z} = \hat{\alpha} e_{A-1}' \otimes I_4 \left(I - \hat{\alpha} J^4 \right)'^{-1} \left\{ \left(I - \hat{\alpha} J^4 \right)^{-1} x\hat{\beta} + \left(I_{A-1} - \hat{\alpha} J_{A-1} \right)^{-1} \otimes e' \hat{E} / ee' \right\}$$
d'où nous tirons l'expression des valeurs trimestrielles de l'année initiale à l'aide de la contrai nte comptable sur l'année

$$\hat{y}_0 = (e' Y_0 / ee') - \left\{ (I_4 - e'e / ee') \hat{Z} / c(\hat{\alpha}) \right\}$$

Solution du problème de minimisation de la contribution à la variance :

Nous réécrivons l'expression des solutions du programme primaire données par la proposition 6 sous la forme suivante :

$$\hat{y}(y_0) = Py_0 + Q$$

L'expression de la fonction objectif, à savoir la contribution à la variance, s'écrit sous la forme suivante:

$$(y_0' T y_0 + S y_0 + y_0' S' + U)^{-1} (y_0' T y_0 + S y_0)$$

où

$$T = (I_4 \quad P') \left(I_{4A} - \frac{e_{4A} e_{4A}'}{4A} \right) \begin{pmatrix} I4 \\ P \end{pmatrix}$$
$$S = (0 \quad Q') \left(I_{4A} - \frac{e_{4A} e_{4A}'}{4A} \right) \begin{pmatrix} I_4 \\ P \end{pmatrix}$$
$$U = (0 \quad Q') \left(I_{4A} - \frac{e_{4A} e_{4A}'}{4A} \right) \begin{pmatrix} 0 \\ Q \end{pmatrix}$$

Les conditions du premier ordre donnent les équation suivantes :

$$\begin{cases} 2(S\hat{y}_0 + U) T\hat{y}_0 + (U - \hat{y}_0' T\hat{y}_0) S' - (\hat{y}_0' T\hat{y}_0 + 2S\hat{y}_0 + U)^2 e' \lambda_0 = 0 \\ e\hat{y}_0 = Y_0 \end{cases}$$

d'où

(1) Si (S', e') est une famille de vecteurs indépendants, on pose $\$_0 = vT^{-1}S + \mu T^{-1}e'$ et l'on tire des expressions $S\$_0, \$_0 'T\$_0, e\$_0$ les équations suivantes :

$$\begin{cases} v^2 ss + v\mu se + (2v+1)U - \mu Y_0 = 0 \\ ves + \mu ee = Y_0 \end{cases}$$

où $ss = ST^{-1}S', se = ST^{-1}e', es = eT^{-1}S', ee = eT^{-1}e'$. La solution de ce système d'équations consiste en deux couples de solutions. La fonction objectif étant un rapport de formes quadratiques dont les termes dominants sont égaux et dont le dénominateur est une forme définie positive, elle possède deux extrema : un minimu m et un maximum. Le calcul des valeurs de la fonction en les solutions permet de faire choix du minimum.

(2) Si (S', e') est une famille de vecteurs dépendants, alors :

$$\$_0 = \frac{Y_0}{eT^{-1}e'} T^{-1}e'$$

La dérivée seconde calculée en ce point est égale à **2UT** qui est définie positive (**T** est une variance). La solution correspond donc à un minimum.

Références

Bournay J. et G.Laroque (1979) "Réflexions sur la méthode d'élaboration des comptes trimestriels" *Annales de L'INSEE*, 36, p3-30

Chow G.C. et A.L.Lin (1971) "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series" *The Review of Economics and Statistics*, 53, p372-375

Di Fonzo T. (1986) "La stima indiretta di serie economiche trimestrali", Thèse, Université de Padoue, décembre

Di Fonzo T. et R.Filosa (1987) "Methods of estimation of quarterly national accounts series: a comparison", papier présenté aux "Journées franco-italiennes de comptabilité nationale" Lausanne 18-20 mai 1987

Engle R.F et C.W. Granger (1987) "Co-Integration and Error Correction : Representation, Estimation and Testing", *Econometrica*, 55, p45-62

Gladishev E.G. (1961) "Periodically correlated random sequences" *Soviet Math.* 2, p383-388

Granger C.W.J et A.A.Weiss (1983) "Time Series Analysis of Error Correcting Models" in *Studies in Econometrics, Time Series and Multivariate Statistics*, NewYork, Academic Press, p225-278

Gregoir S. (1994) "Multivariate Time Series with Various Unit Roots : Part I : Integral Operator Algebra and Representation Theorem", miméo

Hylleberg S., R.F.Engle , C.W.J. Granger et B.S.Yoo (1991) "Seasonal Integration and Cointegration" *Journal of Econometrics*, 44, p215-238

Karhunen K. (1947) "Über lineare Methoden in des Wahrscheinlichkeitsrechnung" *Ann. Acad. Sci. Fenn. Ser. AI, Math,* 37, p3-79

Kitamura Y. (1994) "Hypothesis testing in models with possibly non-stationary processes", working paper , department of Economics, University of Minnesota

Kitamura Y. et P.C.B. Phillips (1994) "Fully Modified IV, GIVE and GMM Estimation with Possibly non-Stationary Regressors and Instruments" Cowles foundation discussion paper #1082

Loève M. (1963) Probability Theory, Van Nostrand, N.Y.

Nelson C.R. et C.I. Plosser (1982) "Trends and Random Walks in Macro-economic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10,p139-162

Phillips P.C.B. (1993a) "Fully Modified Least Squares and Vector Autoregression" Cowles foundation discusson paper #1047

Phillips P.C.B. (1993b)

Pierse R.G et A.J.Snell (1995) "Temporal Aggregation and the power of tests for a Unit Root" *Journal of Econometrics*, 65, p333-345.

SECTION 3 -
SURVEY-BASED APPROACH I

Quarterly National Accounts of the Federal Statistical Office of Germany

Heinrich LÜTZEL
Statistiches Bundesamt

Since 1978 the Federal Statistical Office (FSO) has published quarterly main aggregates of German national accounts. The time series start with 1968. The time-lag of the publication is about nine weeks. These data are of greatest interest to specialists as well as to the general public in Germany.

Quarterly data are mainly used for short-term economic analysis, the analysis of business cycles, and as the basis for economic forecasts. The internal use of quarterly data is also important especially as a basis for timely annual estimates. Compared with these purposes the use of econometric models may be considered less important.

At the FSO the quarterly national accounts aggregates are prepared by the same persons who also prepare annual estimates. In some cases, quarterly and annual methods are the same up to the final estimates. In other cases different indicator methods for quarterly and first annual estimates are used. In some cases only special rough estimates are possible. Preparing quarterly and annual data "in one hand" is in many respects preferable. The original quarterly estimates should be prepared seasonally unadjusted.

T*here is no "best" method for seasonal adjustment. No method can meet all requirements at one time. It must be decided whether accounting rules (additivity over time and of aggregates) or the quality of the adjustment should be given priority. Taking the main purposes of seasonally adjusted quarterly national accounts into account, the FSO gives priority to the quality and stability of seasonal adjustment.*

A handbook on quarterly national accounts should discuss typical conceptual problems of quarterly accounting and should give advice on how to handle them. Special problems referring to the time of recording transactions like income, taxes or fixed capital formation have to be dealt with. Other questions concern the treatment of typical seasonal changes in prices and problems of estimating the natural growth of crops or of holding gains in stocks of inventories.

1 Introduction

In 1978, first quarterly national accounts data for the Federal Republic of Germany (FRG) were published by the Federal Statistical Office (FSO).

In November 1994 the time series cover the period from the first quarter of 1968 to the second quarter of 1994 for the former territory of the fourth quarter of 1993.

From 1995 onwards the quarterly accounts will be calculated only for the unified Germany.

Before 1978 quarterly estimates were published by several German economic research institutes, for instance by the DIW in Berlin. At that time the FSO prepared quarterly estimates only for internal purposes, while half-yearly data were published.

2 Main purposes of quarterly accounts

Three main purposes of quarterly data can be distinguished (in Germany)

- Economic analysis and policies
- Econometric models
- Internal uses

2.1 Economic analysis

The use of quarterly data for economic analysis and, more specifically, analysis of business cycles is the most important purpose of quarterly national accounts in Germany. The data should be meet the following requirements:

– timely estimates must be available;
– highly aggregated main indicators are needed ;
– original and seasonally adjusted time series should be provided ;
– reliable estimates with small revisions are important

The FSO publishes the quarterly accounts data about nine weeks after the end of the latest quarter. At that time monthly basic statistics for all three months of the quarter are available. The DIW publishes first estimates about two weeks earlier.

The FSO provides quarterly estimate for the production side of GDP (value added of 5 industries), the expenditure side (7 main aggregates and 14 additional aggregates), employment (5 industries), and on unemployment. Analysts and the general public are interested above all in the expenditure side (at constant prices) and employment data. The annual growth rate of real GDP often is (mis) used as “the” indicator for the success of the economic policy of the government.

2.2 Econometric models

Many model builders prefer quarterly accounts because the number of data in the time series is bigger and thus the t-statistics of their estimators are better. It is less important to them that the accuracy of annual results is better than the accuracy of quarterly results. Long time series and a complete system of quarterly national accounts are demanded by model builders. Compared with internal uses and with the use of quarterly data for short-term economic policy, the use for econometric models is of minor importance.

2.3 Internal uses

Quarterly national accounts are very important also for internal purposes because they allow to prepare very timely annual data using monthly and quarterly statistics. The FSO publishes first annual estimates just ten days after the end of the year. It is possible to produce these estimates only on the basis of quarterly accounts.

However, no quarterly figures are presented at that time, because only one month of the fourth quarter has been covered by statistics. At the press conferences, the journalists urgently demand information on the fourth quarter every year. This seems to be the most important information from national accounts.

Another important internal use of quarterly data regards the preparation of short-term economic forecasts. The FSO does not publish any forecasts. But it is an important task to provide a picture of the economy at the moment when the government prepares its forecasts.

Timeliness of national accounts is most important. To achieve this goal, quarterly national accounts are necessary.

3 Preparation of quarterly national accounts

The way and the method used in quarterly accounts depend above all on the availability of timely monthly and quarterly statistics. There is no "best method" of calculating quarterly accounts data.

At the FSO the quarterly estimates are prepared by the same staff members who produce annual estimates. This is preferable because these persons

- know the data and methods of "their" aggregate or industry;
- know what is going on in "their" economic sector;
- prepare quarterly data which are always consistent with the annual data provided at the end of the year;
- use quarterly results for the first annual estimates.

In addition some annual results are originally calculated on the basis of quarterly statistics. In these cases it is obvious, that quarterly and annual calculations should be prepared by the same persons.
The two tables of "Quarterly National Accounts in Germany" show in the column "Dominating method" which aggregates are originally calculated on a quarterly basis in Germany.

In most cases quarterly data of the previous year are extrapolated with indices. For example, the value added at constant prices of he third quarter of 1994 is estimated by multiplying the value added at constant prices of the third quarter of 1993 by the growth rates of production index for the same periods. But this is not done in a strictly mathematical way.

This relation can be modified "manually" taking into account that in some industries, or in special phases of business cycles, the indicator normally shows a faster (or slower) change than the national accounts aggregate. When preparing national accounts, experience often is more important than highly sophisticated mathematical methods.

For some aggregates (industries) the statistical basis for the quarterly estimates is very poor. Here weak information (employment data from the Federal Institute for Employment), plausibility checks (change in stocks) or quarterly breakdown of (estimated) annual data (consumption of fixed capital) are used, too. For transforming annual data into quarterly data without using any indicator, the following formula is applied:

1. $Q_t = (12A_t + 5A_{t-1} - 1A_{t+1}):64$
2. $Q_t = (20A_t - 1A_{t-1} - 3A_{t+1}):64$
3. $Q_t = (20A_t - 3A_{t-1} - A_{t+1}):64$
4. $Q_t = (12A_t - 1A_{t-1} + 5A_{t+1}):64$

Q: quarterly results
A: annual results
T: year of recording

This formula has a goods smoothing effect and makes sure that the four quarters add up to the annual result. But at the end of the time series the annual data A_t and A_{t+1} must be estimated.

Other principles for quarterly estimates at the FSO are:

- They are prepared on an original basis and not seasonally adjusted.
- They always add up to annual data by integrating the quarters into the annual data, if available.
- They always add up to the GDP. Adjustments are made in such a way that no residuals have to be published.
- Timeliness and accuracy are given priority over comprehensiveness of the system. The FSO does not present a full quarterly system of accounts.

The strong demand of economic analysts for timely quarterly data, the increasing interest of journalists in our press conferences as well as the positive reaction of the general public show that at least quarterly main aggregates of national accounts are extremely important.

4 Seasonal adjustment

In Germany two main methods of seasonal adjustment are used:

- "Berlin Method", fourth variance

– "Census X-11 Method", with and without working day adjustments

The FSO publishes seasonally adjusted quarterly data in three steps:

- In the harmonised press releases of the FSO and the Ministry of Economic Affairs, a uniform growth rate is published for the real GDP. A working day adjustment is included. The adjustment is made by the Deutsche Bundesbank using the original data of FSO.
- In the subject-matter series quarterly national accounts, seasonally adjusted data based on the Berlin Method are published. All series are adjusted with a uniform version of this method. No adjustments are made for extremes or working days.
- In the generally available working document "Materialien zur Zeitreihenzerlegung", the FSO adjusts the series with the Berlin Method and the Census X-11 Method in a special uniform version for quarterly accounts. All components are presented here, as well as growth rates versus the previous year (original data) and the previous quarter (adjusted data) also as an "annual rate".

This praxis of publication meets the special needs of the different users:

- The general public may be confused if different results (according to the method used) are published.
- The "ordinary" economic analyst wants to obtain seasonally adjusted data to find out about the latest trend of the economy or the phase in the business cycle.
- The specialist wants to have all the details. He knows that different methods or different versions of the same method will result in different adjusted data. These differences will show him that in some quarters it is dangerous to rely on seasonally adjusted data.

At the beginning of seasonal adjustment exercises the FSO tried to find out which is the "best" adjustment method for quarterly national accounts. An ideal method should meet the following requirements:

– the seasonal component should not be too stable and not too flexible over time;
– the trend should be smooth but true;
– the first estimates of the seasonal component should not be corrected excessively (stability in adjustment);
– the adjusted quarters should add up to the original result (additivity over time);
– the adjusted aggregates should add up to the adjusted sum (additivity of aggregates);
– adjusted nominal aggregates divided by adjusted deflators should equal adjusted real aggregates.

Hundreds of comparable calculations were made. The findings are clear: there is no "best" method. It is impossible to meet only three of these requirements at the same time. Stability in and quality of adjustment was given priority over other requirements.

The main purpose of seasonally adjusted time series is the analysis of economic trends and business cycles at the end of the time series. Thus, the quality of should be given priority over the accounting rules in the system (additivity over time and of aggregates). In addition, other adjustments should be made and published separately. Working day adjustment and weather adjustment are typical examples.

A search for the "best" method is more or less useless. The differences in results caused by employing different versions of the same method are often greater than the differences between the results of different methods. A very smooth trend component and a clear theoretical foundation are the advantages of the Berlin Method. It is based on the theory of spectral analysis.

5 General conceptual problems

In general, the same definitions and accounting rules should be used in quarterly national accounts as well as in the annual accounts of the SNA 1993 and the ESA 1995, respectively. But there are some special points which should be discussed and explained in detail in a handbook of quarterly national accounts. The problems concern the time at which transactions are recorded and special seasonal effects within a year.

This paper raises some of these problems. Other issue should be included and discussed.

Problems:

» Should typical seasonal price movements over the year (mostly prices of agricultural products) be treated as changes in prices or in volume? According to the SNA 1993 and the ESA 1995 these are changes in volume (different products in different months). What is the consequence for quarterly accounts, especially for changes in stocks?

» The SNA 1993 and the ESA 1995 recommend that in agriculture the natural growth of crops should be included in output. How can this be estimated in quarterly accounts? Is there any experience in national accounts of other countries?

» For intermediate consumption, usually only annual information is available. Is it acceptable to assume that there is no seasonal influence on the intermediate consumption ratio by quarters, except, of course, for crop production in agriculture?

» Special tinning problems of recording income flows should be explained:

A. Wages and salaries often paid before (after) the month for which they are paid
B. Normal annual payments like salaries for a 13th month at the end of the year
C. Quarterly estimates of interest income
D. Time of recording dividends paid on shares.

» Special tinning problems of recording taxes:

A. Are advance payments of taxes (in Germany typical of VAT and income tax) an acceptable proxy for the accrual basis?
B. Recording the results of annual tax declarations in the following year
C. Recording the results of tax inspections some years later
D. Practical advice on how to adjust cash data on taxes to an accrual basis.

» Special tinning problems of recording fixed capital formation:

A. New buildings and other construction have to be recorded as fixed capital formation in the period when they are produced. The commodity-flow approach provides good estimates.
B. New machinery and equipment have to be recorded as fixed capital formation at the time when the change in ownership takes place. The commodity-flow estimates must be corrected for changes in stocks of such investment goods. What information can be used for such corrections?

» Is there any experience in calculating quarterly changes in stocks? Which methods are applied in other countries for adjusting holding gains and losses?

These are some examples for typical problems, which should be explained in the handbook on quarterly national accounts.

Quarterly National Accounts in Germany (former FRG, t+70 days estimates)

1. Gross Value Added (GVA) at Market Prices

Industry	Dominating method	Sources, indicators, m: monthly, q: quarterly	Breakdown	d.: deflation, i.: inflation
Agriculture	Original (valuation of quantities)	Animals slaughtered (m), stocks of animals, crops harvested (m) sales prices (basic year)	34 types of commodities	i. with the price index of agricultural products
Forestry, fishing	Estimates	Felled timber (m), catches of fish (m), sales prices (basic year)	2 industries	i. with the price index of timber (m)
Manufacturing, mining, electricity	Indicators (extrapolation of GVA)	Production indices (m)	3 industries	i. with the producers' price index
Construction	Indicators (extrapolation of GVA)	Hours worked (m) Estimates of change in productivity	2 industries	i. with the construction price index (q)
Trade	Indicators	Turnover (m)	2 industries	d. with retail and wholesale price index (m)
Transport	Indicators	Information from railways, post offices, road transport (m) Balance of payments (m)	9 industries	Partly i. With CPI (Consumer price index), (m)
Other market services	Indicators	Information from banks, hotels, insurance's (m) Employment (m)	7 industries	i. with CPI (consumer price index), (m)
General government	Indicators	Finance statistics (q)	3 cost components 4 levels of government	i. with compensation index

Quarterly National Accounts in Germany (former FRG, t+70 days estimates)

2. Expenditure of GDP

Aggregate	Dominating method	Sources, indicators, m: monthly, q: quarterly	Breakdown	d.: deflation, i.: inflation
Household final consumption	Original, indicators (supplier approach)	Statistics on turnover of trade, restaurants, energy (m)	120 branches of suppliers, 300 purposes	d. with CPI (consumer price index, m) 300 commodities by purpose
Final consumption of general government	indicators	Finance statistics (q)	6 cost components, 1 sale	d. indices for cost components
Fixed capital formation -in machinery and equipment -in building and construction	Original (commodity flow method) Indicators (commodity flow method)	Production statistics (q) foreign trade statistics (m) Hours worked (m) plus estimated change in productivity	1600 commodities 8 types of construction	d. with producer's price index (m), foreign trade price index (m), 180 commodities i. with price index of buildings (q) 8 types of building
Change in stocks	Estimates	Difference between production and expenditure side of GDP		
Exports and imports of goods and services	original	Foreign trade statistics (m) Balance of payments statistics (m)	1 800 commodities	d. with foreign trade price index by 1 800 commodities

Quarterly National Accounts in Germany (former FRG, t+70 days estimates)

3. Distribution of income

Aggregate	Dominating method	Sources, indicators, m: monthly, q: quarterly	Breakdown	d.: deflation, i.: inflation
Consumption of fixed capital	Estimates	Annual perpetual inventory estimates and formula	2 commodities	i. with deflators of gross fixed capital formation
Taxes/subsidies	Original/indicators	Tax receipts (m)	3 items	
Compensation of employees	Indicators	Statistics on employment (m), wage sale index	23 industries	
Social contributions	Original	Information from social insurance institutions (m)	11 social insurance schemes	

4. Employment

Aggregate	Dominating method	Sources, indicators, m: monthly, q: quarterly	Breakdown	d.: deflation, i.: inflation
Employment	Indicators	Statistics on employment (m), estimates for self employed persons	31 industries	
Unemployed persons	Original	Information from the Federal Institute for Employment (m)		

1 $1°\ Q_t = \left(12D_t + 5D_{t-1} - 1D^{t+1}\right):64;\ 2°\ Q_t = \left(20D_t - 1D_{t-1} - 3D_{t+1}\right):64;\ 3°\ Q_t = \left(20D_t - 3D_{t-1} - 1D_{t+1}\right):64;\ 4°\ Q_t = \left(12D_t - 1D_{t-1} + 5D_{t+1}\right):64$

Q: quarterly results; D: annual data; y: year of recording

Special aspects of the Quarterly GDP Accounts in East Germany in the initial years after German Unification

Ralf HAIN
Statistisches Bundesamt

1 Introduction

With the unification of Germany on Oct. 3, 1990 the statistical system of the Federal Republic of Germany was adopted in the new Laender. This meant the adoption of the SNA calculations for the national accounts at short notice. Owing to the differences in the economic level and the varying database and quality in the new Laender as compared with the former federal territory, separate accounting for West and East Germany are carried out for some years. From the very beginning efforts have been made to reach a possibly complete adaptation of the methods and depth of calculation and periodicity to those applied for the former federal territory. As from 1995 the transition to an all-German presentation is envisaged. Then, still only a few facts will be calculated separately for West and East Germany.

The present article is aimed at outlining the methods of the quarterly calculation of the gross domestic product (GDP) for East Germany and describing some special aspects that arise from the economic situation.

2 Sequence and Basis of Quarterly Accounts

For the first time, the results achieved in the 2 nd half-year of 1990 in East Germany were published in April 1991. The currency conversion effected on July 1, 1990 did not allow providing an account for the whole year. Actual annual and semi-annual results have, in principle, been calculated or revised simultaneously with the West German results from September 1991 on. As from 1992 quarterly results have been published once a year in September for the preceding year. The uncertainties of the database have not yet allowed to present actual quarterly results at the same time with those for the former federal territory, always approximately two months after the end of the quarter, though, from the very beginning, the methods of calculation have been envisaged for quarters, analogously to those applied for West Germany. The determination of seasonally adjusted rows for East Germany was not possible either.

Annual accounts of the German national accounting are, for the most part, based on statistical surveys carried out once a year or at intervals of a few years only for full calendar years. The database for quarterly accounts is essentially smaller, for the most part these are the results of short-term (monthly and quarterly) economic statistics. The fact that they do not cover the whole domain of the economy or all components of the facts to be proved, corresponding in general less to the definitions and deliminations of the national accounts than the annual statistics is to the disadvantage of the national accounts. Owing to the higher uncertainties thus arising the publication of quarterly results of the national accounts is restricted to a few facts.

The calculations are carried out separately for the production and the use of the GDP. The co-ordination of the overall aggregates is made proceeding from qualitative criteria.

Problems for the national accounts in East Germany resulted, in particular, from the fundamental changes proceeding in economy that started directly with the adoption of the economic and currency union. Carrying out surveys according to the Federal

Statistics Act meant a complete radical change in the statistical database. To avoid data gaps or an unjustifiable delay in the supply of important results, provisional regulations for carrying out various surveys on the territory of the new Laender were adopted with the unification treaty and in a special statistics adaptation ordinance.

3 Survey of the Methods of Production and Expenditure Calculation

The quarterly calculation of the production side of the GDP is, in principle, based on the allocation of the final annual results or in the case of actual calculations on updating the results with the aid of the actual economic indicators available for quarters or months. As a rule, the same indicators are applied in allocating the final annual results to quarters and for actual calculations. Statistical data on changes of stocks, self-constructed equipment and, in particular, the cost structure which, in general, are recorded only for the whole years, are lacking.

The gross value added of the manufacturing industry is determined for the new Laender proceeding from the turnover trend of the enterprises with 20 and more employees and the handicraft enterprises. On the former federal territory the net production index is applied to update the real gross value added.

As for West Germany the gross value added of the building industry is effected on the basis of the data on hours worked, and, additionally, for preliminary calculations on the basis of the estimates of the development of labour productivity and an estimation factor for newly established companies. As to the wholesale trade the gross value added is distributed quarterly with the aid of the turnover data.

It was not possible to determine the gross value added of the retail trade on the basis of the monthly business statistics which through the lack of an actual trade census was based on an outdated register from 1990. That is why the turnover volume was indicated as being much too low. Therefore, it was necessary to make rough estimations proceeding from the Private consumption in the first years after the unification.

For other service enterprises only very few statistical data were available for individual branches, e.g. transport, credit and insurance companies and health care. It is only possible to make a rough quarterly estimate of the gross value added of most of the service branches through the development of the employment figure. In 1991 and 1992 quarterly service statistics were additionally complied (see item 3).

As regards the method, the quarterly calculations of the GDP-expenditure side, do not differ essentially from the annual account as the biggest part of the statistical data is quarterly available.

Until 1993 the Private consumption in the new Laender was mainly determined on the basis of the household surveys carried out quarterly. After evaluating the 1993 trade census a complete account for the supply sector will be possible – as it is common practice for the old Laender

Investments in fixed assets are calculated proceeding from the production, import and export data according to the commodity flow method as for the former federal territory. Also the quarterly export and import account that is mainly based on the monthly data of the foreign trade statistics and service transactions data corresponds to that for West Germany.

4 Special Problems

4.1 Effects of the Economic Situation on the DataBase and Methods of Calculation

Immediately after the currency union was established, the production in East Germany declined strongly. A big number of enterprises was closed down. Many others worked at a loss.

Altogether, an enormous structural break proceeded in economy.

The net production index was not applied for quarterly accounts in the manufacturing industry. The structures of economic sectors and costs existing in the 2 nd half-year of 1990 applied to determine the weighting scheme were already outdated in 1991. That is why the

calculations for the new Laender are made proceeding from the determined turnover and supplementary estimates made for the lacking facts.

Also the assumption of constant input ration for all quarters was not appropriate as they varied by more than ten percentage points in individual sectors from year to year. That is why "curves" were estimated for the quarterly input ration proceeding from the available yearly information.

Problems have been arising in evaluating the surveys carried out in the manufacturing industry. For the turnover of companies with 20 and more employees determined on the basis of monthly and yearly surveys and cost structure statistics very differing results were achieved. That is why an extrapolation was made on the basis of all companies included in the surveys. Thus, an additional correction of level by more than 10 per cent had to be made for the gross value added of the manufacturing industry for 1991.

Due to the strong structural changes it was not possible to make an estimate in one sum analogously to the West German account made for the manufacturing industry. That is why until 1994 the calculations have been made subdivided into 35 sectors of the manufacturing industry. The data of handicraft enterprises with a turnover rising much faster than in the remaining manufacturing industry during the last few years were separately considered.

The calculation in constant prices was also effected subdivided in a more detailed way than it has been common practice for West Germany, mainly to exclude the effects of the outdated weighting scheme for the producer's price index of 1989 applied in the first years. Owing to the strong quarterly variations of the production value level when going over to the 1991 basis of quotation the price index was calculated as a weighted average.

Many methods of calculation involved uncertainties based on the updating of the data of a base year or the same period of the preceding year. This refers notably to sectors with random sampling (e.g. in building industry, handicraft and retail trade). For instance, for handicraft enterprises of the manufacturing industry a corrective element was estimated for individual quarters proceeding from the number of newly established companies with the aid of which the turnover was estimated.

The suitability of indicators for updating had to be also examined thoroughly. Thus, the number of gainfully employed persons was not applied as an indicator for the development of the gross value added of service enterprises. Owing to the strong backlog of demand a very strong increase in production, partly with saturation setting in slowly subsequently, was to be stated in the individual sectors. In such cases it was scarcely possible to assess the development of productivity.

4.2 Additional Surveys in the Transitional Period

Service statistics and household budget surveys were of special importance among the surveys carried out additionally. The surveys of cost structures carried out in all sectors supplied very important information on the input ration, yet only as annual data.

In 1991 and 1992 the calculations of the gross value added in the service sectors were based on the results of the random sampling effected in this sector quarterly. As the random sampling plan and the extrapolation range of the survey were primarily based on the survey of gainfully employed persons carried out in 1990 many newly established enterprises or subsidiaries of West German companies were considered only incompletely. Thus, strong corrections of the results of the random sampling were required. Owing to the additional information on individual sectors the gross value added of other service companies was, in principle, doubled in preliminary estimates. The first presentation of the results of the 1992 turnover tax statistics showed, however, that a remarkable increase in the gross value added by 50 per cent was once more necessary. Given the conditions of fast economic changes it would have been only possible to achieve better results if the survey was based on a constantly updated register.

The results of the quarterly budget survey carried out for 3500 households in connection with the population structure obtained through the micro census proved to be useful in estimating the Private consumption.

5 Conclusions

The methods of calculation applied for quarterly national accounts depend essentially on the database available, yet also on the special features of the economic development. Under the radical change taking place in East Germany it was necessary to examine the methods of calculation applied especially thoroughly. The statistical results to be applied as basic data or indicators in the event of a “normal” development proceeding proved to be partly useless or incomplete.

Owing to the strong structural changes going on in the economy accounting had to be performed with a more detailed subdivision and additional estimates of changes of the operating sector had to be made. As the necessary information was frequently not available it had, to a large degree, to be supplemented by estimates.

The determination of the absolute level and the development of the gross valued added in the service branches was connected with great difficulties. The random sampling carried out in retail trade, transport and other service branches supplied results which being based on outdated registers were only conditionally useful. It would have been necessary to keep the registers up-to-date and to adapt random sampling in each case.

SURVEY-BASED APPROACH II

An outline of the Swedish Quarterly National Accounts

Helena RUNNQUIST
Eddie KARLSSON

Statistics Sweden

1 Introduction

This outline of the Swedish quarterly National Accounts (NA) deals with the scope of the accounts and the methods used in balancing and analysing data at an aggregated level. It is not a document of sources of data. Swedish quarterly NA have been compiled since 1972. The time series runs from 1970 but present series are consistent from 1980.

Calculations are made from the production and the expenditure side. The extension of the system is increased in the way that estimations now are made also in current prices (at the start only in constant prices). Variables are in principle the same as from the start, which means, GDP by expenditure and kind of activity, disposable income for households and number of hours worked and persons employed (for more details on labour statistics, see appendix I).

Number of activities, expenditures and other details are much more comprehensive now compared to the first calculations. Value added is computed for 58 activities of industries, producers of government services, private non-profit institutions serving households and domestic services. For expenditures there are a great amount of details, in households consumption, for example, we have more than 100 items. Government final consumption expenditures are calculated for central and local government and by type of cost. Figures on fixed gross capital formation are achieved for different types of capital goods and sector and, nine groups of activities of industries.

Foreign trade is divided on goods and services in accordance with the HS classification system. Changes in stocks are calculated for a wide range of branches in mining and manufacturing. There are also stock data for retail and wholesale, agriculture and forestry (concerning classification levels for calculations, see appendix II).

2 Calculation

For expenditures like private final consumption expenditure, changes in stocks and foreign trade of goods, the sources are quarterly and used for the annual estimates as well. The other parts of the expenditure estimates are covered quarterly by surveys measuring data at a more aggregated level than the annual ones. But for an essential part of the expenditures we have information on kind of use. For the production estimates information on intermediate consumption is mostly not available. Measures of the production (volume or value) are therefore used at indicators for values added. With the assumptions that input coefficients in volume are unchanged compared to previous yearly estimate and also that coefficients are constant during all four quarters.

In the process of calculating the quarterly NA the same staff is involved as in the annual accounts. One of the advantages with this we think is the thorough knowledge in their special fields that each person can develop and make use of in the quarterly calculations. We want to keep a very close relation between the annual and the quarterly accounts and quarterly

accounts are always adjusted to yearly level is a satisfying solution. This is however not always the case with the estimates of value added at constant prices. The result is that the time series shows discontinuities at the turns of year. To avoid this unwanted effect Statistics Sweden have developed a method called MIND4. This least square-method minimises D4 (see below) combined with the conditions that the yearly total is unchanged and that the first adjusted quarter is linked to the last quarter in the previously adjusted period.

$$D4 = \sum \left(Y_i / X_i - Y_{i-1} / X_{i-1} \right)^2$$

where X_i = the original quarterly series; and Y_i = the adjusted quarterly series. This method of adjusting the quarterly series to a yearly total is used once a year in autumn when the annual accounts are completed. It is applied to a period of two years at a time. Each year is adjusted twice. The practice shows however that the problem in identifying the « best » quarterly pattern is the most important to overcome in calculating quarterly time-series.

3 Reconciliation and analysing results

Some years ago we adopted an input-output system functioning at a half year basis. In recent years it has been somewhat transformed to operate also quarterly. Balances are tuned quarterly in constant prices and in both current and constant prices at a half year basis. (Comments below refers to constant price calculations). Supply and demand are divided on 58 commodity groups (see appendix II).

On supply side that means, gross output of industries, government sales, sales of non-profit institutions (NPI) serving households, import and import duties, trade margins and indirect taxes and subsidies.

On demand side there are intermediate consumption, government intermediate consumption, private consumption expenditure, gross fixed capital formation, change in stocks and exports.

The information available quarterly is limited and can not offer the total range of information required for making input-output calculation of previous or present year.

The key between activities and commodities are taken from the annual system. Quarterly, only production estimates for activities are available. As mentioned before input coefficients are also unchanged compared to the annual system and consequently inputs are spread out on various commodities in consistence with the annual key. Exports and imports are easily distributed on goods, and by using some common sense also on services. Import duties are calculated by adopting a percent figure on imports.

Government sales and intermediate consumption of governments are divided on commodities in the same way in previous or present yearly calculation.

For gross capital formation there are fairly reliable information on type and by kind of activity at a 9 digit level. Stock figures must be divided on goods by using keys both yearly and quarterly. The information is like in most countries of poor quality and consequently the variable is much useful in input-output handling.

Private consumption expenditures are calculated at purpose and are therefore easily distributed.

Trade margins are calculated in the system at the outcome of the balancing process. For different kind of use and kind of commodity there are constant rates on trade margins. For example, we have the commodity car and the uses of private consumption, gross fixed capital formation and exports. By applying the fixed trade margins, varying among different kind of use, it is possible to get a total value for commodity as well as total economy.

Indirect taxes and subsidies are calculated in the same way as trade margins as shares for different commodities and kind of use.

In varying degree residuals will occur on the 58 groups of commodities. A problem in handling the residuals in the quarterly estimate is the appearance of seasonal patterns on divergences. The IO system was introduced only a few years ago and then the patterns were disclosed. The aim in the analyse is of course to minimise residuals at commodity level and hereby also at total level.

Current price estimates are produced at a half-year basis.

On demand side current price figures are directly calculated and distributed.

On supply side output price indices are used for getting commodity figures. Total supply in current and constant prices yields an implicit price index which is used for reflating intermediate consumption at commodity level. The outcomes are figures in current prices for both intermediate consumption and value added at activity level.

At yearly level GDP is determined from the expenditures. The positive or negative residual is hence on production side. For a single quarter corrections are a bit more complicated. For the first and third quarter GDP is determined by using the average growth rate between expenditures and production hereby obtaining a GDP level.

If there is a complete half year, GDP for both quarters is determined from production side by taking the weights for each quarter within it's half year and then applying that on the half year GDP.

Consequently, at the quarterly basis there are discrepancies on both sides.

Principles mentioned in table 1 are valid for both current and constant price estimates. In practise the method is a bit different in current prices as the production estimates are available only at half-year basis. The solution in getting current price estimates is to construct a relevant price index for reflation of the correction item on the expenditure side. GDP at current prices is the sum of expenditures and the correction item.

To facilitate analysis of the quarterly time series we calculate series adjusted for calendar effects and for seasonal variations.
The adjustments for calendar effects is done for production data at constant prices and number of hours worked only. For each activity a quota is calculated. It is based on the actual number of working days for the present quarter compared to a "normal" quarter; an average of the latest twelve years. Number of working days varies for different "working schedules". We use four different kinds, five days a week, six days a week; shifts 29 days a month and 30 days a month. The quota is also weighted with the information of what kind of schedules are used in each activity. Several kinds can be applied for one activity. This information is only obtained by telephoning branch-organisations etc. and it is not systematised. This is probably a weak link in the calculations. By tradition no yearly totals for the time-series adjusted for number of working days are published or presented since they differ from the original ones.

Seasonally adjusted series are calculated for a period of the latest twelve years (for year 1983 to 1994 at present). We recalculate the adjustments for the whole period when we introduce a new quarter. Since we have found that seasonal patterns can be hard to distinguish in many of the time-series due to relatively big irregular components, we adjust the series at a rather aggregated level (for levels of adjustment, see appendix III). This may be a problem specific for relatively small economies like Sweden, compared to the United States for example.

Seasonal adjustments are calculated for values added on series adjusted for calendar effects and for series of expenditure of GDP. We calculate readjustments for extreme irregular values that we know of, like for example the big conflict on the labour market that we had in 1980. The problem is of course to estimate the impact of an event like that on values added. These readjustments are made since we do not want the seasonal adjustment factors to be influenced by so called extreme values. The adjusted series contains the irregularities. Estimation of the factors for seasonal adjustment is made with a computer programme developed by Statistics Canada; X-11 ARIMA\88. We use a function for correcting extreme values (other than the ones mentioned above) and the ARIMA

Table 1: Rules on quarterly tuning

Yearly GDP = Expenditures = Production +(-) Discrepancy
Half yearly GDP = Expenditures = Production +(-) Discrepancy
Quarterly GDP = Expenditures +(-) Correction +(-) Discrepancy

forecasting model in the programme. Each data series is adjusted separately which means that no form of additivity exists in figures. Both additive and multiplicative methods are used depending on what is most appropriate for each series. The lack of additivity does not seem to have been a problem for the users so far. The yearly totals are not affected by the seasonal adjustments.

4 Revisions and publishing

GDP figures are presented in current and constant prices (1991 prices from this autumn) from first quarter of 1980. Expenditures are presented quarterly in both current and constant prices while production data are shown quarterly in constant prices and in current prices at a half year basis. GDP expenditures and value added are also presented in a seasonally adjusted version. GDP and value added are also available in a version corrected for number of working days as well as seasonal patterns.

In tables both levels and changes in volume are noted. In the seasonal adjusted series the volume changes are noted with regard to previous quarter without any upgrading to yearly levels or changes.

There are some fixed rules for revisions in the quarterly series. Data are published about 2.5 months after the quarter in question. The estimating process starts approximately one month earlier for the early estimates like employment and households expenditures. Recently a project has started aiming at quicker estimates.

In June figures for the first quarter are published. There are no revisions for preceding years.
In September second quarter is completed and first quarter revised.

In October/November the annually accounts for the previous two years are completed and the corresponding quarterly estimates are adjusted to the new levels. When the third quarter is published in middle of December the revised series are also published.

In March the four quarters of the preceding year are published. The first three-quarters are revised. An exception from the scheme are quarterly data on disposable income of households which are compiled and published only two times a year, in September and March.

The main results are always presented in a press release. A short version of the press release is also introduced in a data base (Key Economic Indicators) containing a wide range of economic statistics. Clients are mostly banks, stockbrokers and companies on the financial market. There is also a more comprehensive database for more common use (TSDB).

The division is also compiling and publishing a comprehensive quarterly publication. From Statistics Sweden there are two more publications (monthly) containing all kinds of economic statistics including NA.

APPENDIX I

Labour statistics used in the National Accounts

The main source in estimating number of persons employed and number of hours worked, quarterly as well as yearly, is the Labour Force Survey (LFS).

The LFS is a monthly inquiry of the total labour force with an representative sample of 18 000 persons. The inquiry tells both number of persons employed as well as number of hours worked for all of the activities and sectors in the Swedish system of labour accounting.

The LFS also determines the total level of employment.

For some activities where more reliable sources are at hand, the LFS is replaced with other statistics available. That concerns employment within manufacturing and local governments. The source for manufacturing is two separate monthly inquires which, for example, measures absence and salaries for employees, but also number of persons employed and hours worked. Information for local governments is received from monthly statistics produced by their associated organisations. This information is the only one used in labour accounting which is not produced by Statistics Sweden.

APPENDIX II

Classification of households goods and services

Function

1000 Food, beverages and tobacco

1100 Food

1110 Bread and cereals
1120 Meat
1130 Fish
1140 Milk, cheese and eggs
1150 Oils and fats
1160 Fruits and vegetables other than potatoes and similar tubers
1170 Potatoes, manioc and other tubers
1180 Sugar
1191 Coffee, tea, cocoa
1192 Confectionery, etc.

1200 Non-alcoholic beverages

1300 Alcoholic beverages

1310 Spirits, wine and export beer
1311 Spirits
1312 Wine
1313 Export beer
1320 Beer n.e.c.

1400 Tobacco

2000 Clothing and footwear

2100 Clothing including repairs

2110 Clothing, fabrics and yarn
2120 Repairs to clothing

2200 Footwear including repairs

2210 Footwear
2220 Repairs to footwear

3000 Gross rent fuel and power

3100 Gross rent and water charges

3110 Gross rent
3120 Imputed rents of owner occupied dwellings
3130 Imputed rents of secondary dwellings
3140 Tenants repair costs

3200 Fuel and power
3210 Electricity
3220 Gas
3230 Liquid fuels
3240 Other fuels
3250 Purchased heat

4000 Furniture, furnishings, household equipment and operation

4100 Furniture, carpets and repairs

4110 Furniture, carpets and lamps
4120 Repairs to furniture

4200 Household textiles and other furnishings incl. Repairs

4300 Heating and cooking appliances, refrigerators, washing machines and similar major household appliances including fittings and repairs

4310 Heating and cooking appliances, refrigerators, washing machines and similar major household appliances including fittings
4320 Repairs to major household appliances

4400 Glassware, tableware and household utensils

4500 Household operation except domestic services
4510 Non-durable household goods
4520 Household services excluding domestic service
4600 Domestic services

4610 Private welfare services, children
4620 Local government services, children

4630 Welfare services elderly

5000 Medical care and health expenses

5100 Medical and pharmaceutical products

5110 Medicines
5120 Other products

5200 Therapeutic appliances and equipment

5300 Fees paid to physicians, dentist and related practitioners

6000 Transport and communication

6100 Personal transport equipment

6110 Purchases of motor cars
6120 Other transport equipment
6121 Motorcycles and bicycles
6122 Caravan

6200 Operation of personal transport equipment

6210 Repair charges, parts and accessories
6220 Gasoline, oils and greases
6230 Other expenditures on cars
6231 Compulsory tests of cars
6232 Driving lesson
6233 Garaging
6234 Parking
6235 Car leasing

6300 Purchased transport

6310 Railways
6320 Bus and local traffic
6340 Cabs
6350 Ships
6360 Airlines
6370 Services of travel agencies and air charter
6380 Removal
6400 Communication

6411 Postal services
6412 Telephone services

7000 Recreation, entertainment, education and cultural services

7100 Equipment and accessories, including repairs

7110 Radio and television
7120 Photographic equipment, musical instruments, boats and other major durables
7121 Photographic equipment
7122 Boats
7130 Other recreational goods
7140 Repairs to recreational goods, etc.
7141 Parts and accessories for and repairs to recreational goods
7142 Port services

7200 Entertainment, recreational and cultural services excluding hotels, restaurants and cafés

7211 Entertainment and photo services
7212 Television licences
7213 Gambling, lotteries etc.
7214 Veterinary services

7300 Books, newspapers and magazines

7310 Books
7320 Newspapers and magazines
7330 Other printed matter

7400 Education

7410 Music school
7420 Education

8000 Miscellaneous goods and services

8100 Services of barber and beauty shops etc.

8110 Services of barber and beauty shops etc.
8120 Goods for personal care

8200 Goods n.e.c.
8210 Jewellery, watches; rings and precious stones
8220 Other personal goods
8230 Writing and drawing equipment and supplies

8300 Expenditure in restaurants, cafés and hotels

8310	Expenditure in restaurants and cafés
8320	Expenditure for hotels and similar lodging services
8500	Financial services
8600	Services n.e.c.
8611	Undertaking
8612	Other services
Sum	
Direct purchases abroad by resident households	
Total private final consumption expenditure by households	
Private non-profit organisations	
Total private final consumption expenditure	

By durability

1100	Goods
1110	Durable goods
1111	Of which: purchases of motor cars
1112	Other durable goods
1120	Semi-durable goods
1130	Non-durable goods
1131	Of which: food and beverages
1132	Other non-durable goods
1200	Services
1210	Of which: gross rent
1220	Other services

Classification of industries by kind of economic activity

SNR-REV[1]		SNR[2]	SNI[3]
1000	Agriculture, forestry, fishing	1000	1
1100	Agriculture	1100	11
1200	Forestry and logging	1200	12
1300	Fishing	1300	13
2000	Mining and quarrying	2000	2
2100	Iron ore mining	2100	2301
2200	Non-ferrous ore mining	2200	2302
2900	Other mining	2300	21,22,29
3000	Manufacturing	3000	3
3100	Manufacture of food, beverages and tobacco	3100	31
3110	Protected food manufacturing	3111	3111/2, 3116/8
3120	Import-competing food manufacturing	3112	3113/5, 3119, 312
3130	Beverage and tobacco manufactures	3120	313/4
3200	Textile, wearing apparel and leather industries	3200	32
3300	Manufacture of wood and wood products, incl. furniture	3410	33
3310	Saw mills and planing mills	3411	33111
3320	Other wood industry	3412	33 exkl 331111
3400	Manufacture of paper and paper products, printing and publishing		34
3410	Manufacture of pulp	3421	34111
3420	Manufacture of paper and paperboard	3422	34112
3430	Manufacture of pulp, paper and paperboard products	3423	34113, 3412/9

SNR-REV[1]		SNR[2]	SNI[3]
3440	Printing, publishing and allied industries	3430	342
3500	Manufacture of chemicals, plastic products and petroleum	3500	35
3510	Manufacture of industrial chemicals, incl. Plastic materials	3521	351
3520	Manufacture of other chemical products	3522	352
3530	Petroleum refineries and manufacture of products of petroleum and coal	3530	353/4
3550	Manufacture of rubber products	3510	355
3560	Manufacture of plastic products	3523	356
3600	Manufacture of non-metallic mineral products, except products of petroleum and coal	3600	36
3700	Basic metal industries	3700	37
3710	Iron and steel basic industries	3710	371
3720	Non-ferrous metal basic industries	3720	372
3800	Manufacture of fabricated metal products, machinery and equipment	3800	38
3810	Manufacture of fabricated metal products, except machinery and equipment	3811	381
3820	Manufacture of machinery and equipment	3812	382
3830	Manufacture of electrical machinery, apparatus, appliances and supplies	3830	383
3840	Manufacture of transport equipment, except ship building	3813	384 exkl 3841
3850	Manufacture of professional, scientific, measuring and controlling equipment, and of photographic and optical goods	3814	385
3860	Ship building and repairing	3843	3841
3900	Other manufacturing	3900	39
4000	Electricity, gas and water, incl. Steam and hot water supply and sewage disposal	4000	4
4100	Electric light and power, steam and hot water supply	4100	4101, 4103
4200	Gas manufacture and distribution	4200	4102
4300	Water works and supply, incl. sewage disposal	4410, 9200 del	42, 92001
5000	Contstruction	5000	5
6000	Wholesale and retail trade, hotels and restaurants	6000	6
6100	Wholesale and retail trade	6100	61/62
6300	Restaurants and hotels	6300	63
7000	Transport, storage and communication	7000	7
7100	Transport and storage	7100	71
7110	Land transport		711
7111	Railway transport	7110	7111

SNR-REV[1]		SNR[2]	SNI[3]
7112	Urban, suburban and interurban highway passenger transport	7120	7112
7113	Other passenger land transport	7130	7113
7114	Freight transport by road	7140	7114
7116	Supporting services to land transport	7180 del	7116
7120	Water transport and allied services		712/9
7121	Ocean, coastal and inland water transport	7150	7121/2
7123	Supporting services to water transport	7160	7123
7130	Air transport	7170	713
7190	Services allied to transport	7180 del	719
7200	Communication	7200	72
7210	Postal services	7210	72001
7220	Telecommunication	7220	72002
8000	Financing, insurance, real estate and business services	8000	8
8100	Financial institutions	8110	81
8200	Insurance	8210	82
8300	Real estate and business services		83
8310	Real estate		831
8311	One- and two-family houses and leisurehouses	83, 84	831 a
8312	Other real estate	83, 84	831 b
8320	Business services	8500	832/3
9000	Other personal services	9000	9
9200	Sanitary and similar services, except sewage disposal	9200 del	92 exkl 92001
9300	Education and health services	9300	93
9400	Recreational and cultural services	9400	94
9500	Personal and household services	9500	95
9510	Repair services not elsewhere classified	9510	951
9520	Other services	9520	952/9

1 Revised code of classification by kind of activity in the Swedish national accounts (SNR).

2 The former SNR classification.

3 Swedish standard industrial classification of all economic activities (see Reports on statistical c o-ordination 1969:8 and 1977:9). SNI is a Swedish version of the ISIC and is equal to that on a four digit level .

Classification of Gross Fixed Capital Formation

Kind of activity:

ISIC 1 – 9
Local Government
Central Government

By type of capital good **:**

Residential buildings
Other buildings and constructions
Machinery and other

APPENDIX III

Levels of calculating Seasonal adjustment

Value added

Activity 1-9 ISIC
Unallocated banking services
Indirect commodity taxes net
Central government
Local government
NPI
Aggregated levels for example the sum of ISIC, GDP etc.

Hours worked

Activity 1-9 ISIC
Unallocated banking services
Central government
Local government
NPI

Aggregated levels for example the sum of ISIC; Total number of hours worked etc.

Expenditure of GDP

Private final consumption expenditure, total by durability (see Appendix I)

General government consumption, total
- Central government consumption
- Local government consumption

Gross fixed capital formation, total
- Gross fixed capital formation, buildings
- Gross fixed capital formation, machinery and other

Changes in stocks

Exports, total
- Exports of goods
- Exports of services

Imports, total
- Imports of goods
- Imports of goods

GDP

Data flows in Dutch Quarterly National Accounts [1]

Ronald JANSSEN
Peter OOMENS
Nico van STOKROM
Statistics Netherlands

This paper describes the dataflows in the Dutch Quarterly National Accounts. It gives insight in methods, organisation and publications. Also, new developments and future fields of study are discussed.

1 Introduction

The subject of this paper is to describe the transformation of basic short-term statistics into a consistent set of Quarterly National Accounts (QNA). In terms of a flow-chart of data: from non-integrated or single statistics (basic information) to fully integrated and consistent National Accounts. An outline of the philosophy of the Dutch system of national accounting will be given. Also remarks will be made on the organization, the publications and new developments and new fields of study for the Quarterly National Accounts in the Netherlands.

In the process of compiling QNA in the Netherlands there are three main phases. The first phase is that of data collection and adjustment. In the second phase all data are brought together in an input output table. In the Dutch system of (Q) NA and input output table (or a make use table) is an instrument for statistical integration. The result of the second phase is a fully balanced input output table from which the relevant macro economic totals can be derived. The third phase concerns publications.

The organization of the Dutch QNA reflects the dataflows described. There are three project groups working on the QNA. These are: basic information ; integration and publication.

The output in terms of publications of the Dutch QNA is quite varied. There are press reports highlighting the changes of important macro economic variables. More detailed analysis follows in bulletins and other printed publications. There are also electronic publications available to the public.

A recent development is the preliminary (flash) estimate of GDP. Furthermore, the influence of variations in numbers of working days on output is being studied. Also, seasonal adjustment procedures are under scrutiny. In the field of publications, the electronic publication will be renewed and Statistics

1 The views expressed in this paper are those of the authors and do not necessarily reflect the views of Statistics Netherlands.

Netherlands is investigating alternative ways, such as electronic mail, to distribute the electronic publications.

2 Some remarks on macro economic statistics

The many different statistics produced by Statistics Netherlands can be thought of as a coherent system. On one hand there is long-term (mainly annual) statistical information which describes the economic structure or another phenomenon with great detail. This information is characterized as reliable and detailed but available only after a relatively long period of time. On the other extreme there is short-term information. This usually is rough and aggregated but relatively fast. There are of course lots of statistics which can be put in categories between these extremes. Specifically, the system of macro-economic statistics contains annual and quarterly national accounts and monthly indicators. The annual national accounts go in the first (detailed but slow) category. The monthly indicators are of the rough and ready kind . In between, we have the QNA.

One of the main macro-variables in national accounting is total production (Gross Domestic Product, GDP). Production (value added of production) is generated in the various branches of industry. This approach to the macro-economic aggregate of GDP is called the output approach. Calculating the same macro-economic aggregate by way of the process of income distribution and redistribution, where all individual incomes are added up, is known as the income approach. The third sub-process is the expenditure process. Here the variables of the confrontation of supply and demand are measured separately. GDP can be determined as a balancing item if information is available about total consumption (by households and government), total investments (by government and enterprises, including changes in stocks and work in progress), total exports (of goods, services and primary income) and total imports (ditto). This estimation of GDP is the expenditure approach. In the Dutch National Accounts the input output table is traditionally an important tool for integrating most of the information on the economic process. The three above mentioned approaches come together very efficiently in the input output table. The output approach is applied by aggregating gross value added in the columns of the table. The breakdown of primary costs in the table is a result of the income approach and the breakdown in destination of goods and services produced by industry reflects the expenditure approach. In practice, it is important to keep the table as simple as possible, which means that most breakdowns necessary for the three methods do not appear in full detail in the input output table. For example, the income approach requires a breakdown of incomes into their elements. The expenditure approach requires a division of household consumption by distribution channel or a breakdown of investments by type of asset. No such details are shown in the input output table, although they do play a role in the process of statistical integration.

The extrapolation of statistical structures to more recent periods is a kind of short-term statistical analysis. Changes in (economic) structures cannot be quantified very reliably in this way. Of course, it is possible to obtain indications, for example about the nature and the direction of structural changes. The results of such short-term statistical analysis depend on, among other things, a judgement about the set of used indicators; both separately and in mutual relationships. Is is clear that this kind of short-term statistical analysis takes more time than producing separate indicators.

In the Dutch system of QNA, an extrapolation of the structure of the Dutch economy is the result of the exercise. It is done to make consistent all available statistical information about changes in production, income and expenditure. The input output table of a base period can be seen as a two dimensional weighting scheme for integrating all short term information. In other words: the input output table is an instrument for the integration of short-term statistics. It is not a goal in itself.

It goes without saying that the input output table can be compiled on a more disaggregated level and in more detail as more time passes after the end of the reporting period. Integration of very short-term (monthly) statistics is considered as not possible (yet). The monthly figures describe aspects of one specific row or column of the input-output table. Generally speaking monthly indicators are like a thermometer ;

they do not describe the state of health, but they spot any significant changes in it.

In the last decade the emphasis in the development of Dutch macro economic statistics has been on the short-term statistics . In the early eighties this led, among other things, to the resumption of the QNA. There were QNA for 1948 to the second quarter of 1953. These were discontinued. The statisticians in those days also had to work on the annual national accounts. It was found that the differences between annual national accounts and the sum of four quarters were very significant. The reliability of the quarterly data was considered insufficient and the work on QNA was discontinued. The budget restraints at that time were very strict. Priority was given to the annual national accounts. The motto was : first get your annual statistical structure in order and then you can extend and extrapolate to quarters.

The designers of the new QNA were confronted with a demand for optimal timeliness. On the one hand the availability of short-term statistical information was important, while on the other the timeliness of the process of statistical integration was also an issue. With respect to the former, initially the necessary basic statistical information was complete about 16 to 17 weeks after the end of the quarter under report. At that moment the process of integration could start. This process – building up the input output table, confronting statistically calculated supply and demand, deriving statistical discrepancies, deliberating these differences and consulting specialists and, last but not least, taking decisions – takes about two weeks. Later the timeliness of necessary statistical information came to be somewhat shorter, so that the regular QNA could be published in full detail at a fairly steady 17 weeks after the end of the quarter.

This delay, however, is quite large. Several users of QNA have asked for figures with better timeliness. Also, other countries produce their QNA faster. The timeliness of the underlying statistics is often shorter. It is our impression that the compilation of QNA in many other countries is usually based on one of the three approaches, or on a combination of the approaches in which the integration takes place on the macro-level. For the Netherlands CBS the relatively large delay led to reconsideration of this problem. A research project was started to try to find a new optimum between the availability of short-term indicators, the possibility of making additional estimates and the wish to shorten the time lag. The possibility to use the extrapolation method of the regular QNA in the fast QNA was also investigated.

The research led to the fast quarterly GDP estimate. Generally, it appears that some of the most important monthly indicators were available about 6 weeks after the end of the month or quarter under report. Obviously, the timeliness for the statistical basic data for the fast QNA was established at 6 weeks as well. It is hard to cut down on the 2 weeks needed for the integrating and balancing process. The process was even becoming more complex because of the larger statistical margins. Some time was cut by performing the balancing process on a more aggregated level (a smaller input output table). All in all the starting point was that the first results of the fast QNA should be published 7 to 8 weeks after the end of the quarter under review. The so called flash estimate gives the result of economic growth only. We present GDP-volume and some information on the production side of the economy.

3 Data collection and adjustment in the QNA

The purpose of the QNA is to create from unstructured and sometimes inconsistent information a comprehensive, complete and consistent synopsis of production, consumption and of income generation from productive activities in the economy (scheme 1). Therefore we first need data from economic actors. Basic data is derived from a diverse number of sources (e.g. production, turnover and investment statistics per industry, inquiries within households, foreign trade statistics, government accounts). In practice it is common that the data is not completely consistent. These inconsistencies between supply and demand of commodities must be balanced.

The system of extrapolation of Quarterly National Accounts provides quarterly input output tables in both current and constant prices. The coherence between these tables and the method of extrapolation is outlined in scheme 2. The same method of extrapolation is applied for the fast QNA, although there is less

statistical information and the balancing process is less disaggregated.

Scheme 2 shows that there is a relation between the table for the base-quarter (t-1) in current prices and the table for the reporting quarter t (this is the same quarter one year ahead) in current prices. This relationship is represented by a dotted line. It is usual to decompose information about value changes into a price and a quantity (volume) component. With these two components the continuous lines are followed. Furthermore, the scheme shows that the balancing process of the two tables (in current and in constant prices) is a simultaneous process. Balancing decisions are based on volumes as well as prices and values.

Branche specialists in the quarterly national accounts are responsible for the data collection. Most data are gathered by other divisions of Statistics Netherlands. An important task of the specialist is to transfer the basic data into data suitable as input in the quarterly input output table. A major element in this task is to check the data for continuity. Important is: do year-to-year changes in the data (quarter under report compared to the corresponding quarter a year before) represent real growth rates or are they perhaps caused partly by changes in the classifications of statistical units?

The branch specialists also complete the data : often small firms are not included in the statistical surveys of basic statistics. Since the QNA require, just as the annual National accounts, complete estimates for an industry, data from basic statistics must be grossed up. Also some estimates for hidden transactions are best made with a view on the total values in the input-output tables.

Further on data are subjected to some plausibility checks. E.g. the year-to-year quantity changes of total output, total input, total value added and the number of employees are compared. If necessary data are corrected.

Scheme 3 sums up the activities by the branch specialists in the Netherlands carried out before the results from basic statistics can be included in the integration process.

4 Statistical integration

The integration of data in compiling QNA takes place in a input output framework. The quarterly economic structure of a base quarter is updated. This is done simultaneously in current prices and in constant prices (average prices of the previous year). The first tasks of the integration specialist is to build up an unbalanced input output table. The input output table in the QNA is an industry-industry type table. It consists of 86 rows, which are branches of industry. On the input side 63 columns are distinguished, which correspond with 63 branches of industry (at a slightly higher level of aggregation).

The tables are compiled in current as well as in constant prices. To extrapolate the economic structure a production function with fixed coefficients is assumed. The change in gross production is used to calculate the change in intermediate consumption of the branches of industry. Basicly, to produce 10% more cars, 10% more metals, glass, rubber products, etc is needed. The next step is to add data on household consumption, capital formation, exports and imports and, if available, data or estimates on the change in stocks and work in progress. The result is an estimation of total demand. This is confronted with the estimation of gross production (= total supply). Subsequently the various output and expenditure indicators at constant and current prices are used to calculate a price index for intermediate demand (row). The final result is a column of statistical discrepancies at current and constant prices. That column regrettably hardly ever contains cells with a value of zero.

Now the real work of statistical integration can start. The statistical discrepancies are to be resolved. Much depends on the quality of the basic data and the stubbornness or flexibility of the branch and integration specialists. The process is one of negotiation. The real questions in this process are : what is the quality of the indicator of production; how were changes in stocks and other final expenditure variables estimated; and nowadays a very relevant point, what about the level of imports and exports. If gross production is changed because demand is found to be higher than production, then the whole column (input) of this branch must be recalculated. That of course leads to new statistical discrepancies. This is called the process of iteration. Finally, the statistical

differences are brought to zero. The input output table in its final stage may not contain residual statistical differences.

The advantage of this method of working is that the consequences of a decision concerning a statistical difference immediately become clear. If intermediate demand was considered to be too low and was subsequently raised, then value added has been decreased and the operating surplus goes down correspondingly.

Fully integrated and well-structured data are the outcome of this statistical process. Of great importance for improving the quality of the relevant basic statistics is the « feed-back » of the adjusted data to the surveying divisions. If production for a branch of industry is consistently too high or too low relative to other indicators, that is reported to the statistical departments responsible for gathering the data. Regretfully, there is much feed bak on the foreign trade statistics these days.

The input output table or an aggregated version of it does not give all the needed information. Capital formation by type of asset cannot be derived from the input output table. Additional calculations are necessary. For example a matrix of capital formation could have rows showing branches of industry where the capital assets are produced (or imported) and columns which show the type of capital good.

Changes in capital formation resulting from the integration process lead automatically to a change in capital formation by branch of industry. Matrix calculations can be made for the consequences to the values of investment of various types of capital goods. Household consumption and other expenditure variables are handled in the same way.

The input output table is the common core of a set of matrices relating to various economic variables. In the end the integration specialist declares himself satisfied with the data in the balanced input output table. The results are then examined and discussed. This sometimes leads to further corrections. Finally, the aggregated dataset is made and presented to the publications group.

5 Publications of QNA

The above mentioned dataset is the starting point for various calculations.

The first calculations are about time series. The time series at constant and current prices must be updated with the most recent quarterly data. Calculations in the input output table are done at current prices and at constant (average) prices of the previous year. For the QNA publication time series with a fixed base year are wanted. The base year at this moment is 1990. From the values at constant prices in the input output tables the volume changes of all variable are calculated. These changes are then chained on the values at constant prices in the base year. The outcome is not a time series with constant weights as in a Laspeyres index, it is a time series of chained volume changes. This is done for all variables for which values at constant prices are calculated. These problems of course do not exist for the values at current prices. These are derived directly from the input output tables.

There are also several variables which are not found in the input output tables. Data on primary incomes and unrequited income transfers to and from the rest of the world is obtained from the Netherlands Central Bank. These income flows are part of the balance of payments. Data is adjusted to national accounts definitions and then national product, national income and the surplus of the nation on current transactions is calculated. The third part in the additional calculations done by the publications group is the seasonal adjustment of the time series. The method used by Statistics Netherlands is CENSUS X-11. Each variable is adjusted separately. Statistical discrepancies between seasonally adjusted aggregates and the sum of partial series will occur.

Once the whole set of data in all its dimensions has been completed, work on the publications can begin.

The first publication is always a press release. The press report has two functions. The first is of course to inform the general public. The second purpose is to signal to our customers that the QNA data are available.

The press report typically consists of about three pages. It gives a one page abstract of recent economic developments in the Dutch economy and two pages with some tables and notes (name of spokesman, telephone numbers, etc.).

There usually is quite a lot of media attention for these press reports. We have interviews on radio, articles in newspapers and occasionally (Holland going into recession) a television performance. In the case of our fully based estimates the press report highlights the expenditure approach to GDP and extra details are given on capital formation. The press report in the case of the flash estimate of GDP gives fewer details. The change in GDP volume (seasonally adjusted and original) and some details on the output approach to GDP are presented. The volume changes of value added are given for three groups of industries. These are the producers of goods, commercial services and non commercial services. Government value added is included in the last category.

The second outlet for the QNA is on Videotext. Videotext is a kind of databank with pages of information accessible by computer modem. The tables of the printed publication (full details) are presented. Each table can be downloaded but there is no possibility to get other data from the QNA. There is no publication delay for Videotext. Both Videotext and the press report are published on the same day at 10 o'clock in the morning. The exact publication dates are known to press and other customers well in advance.

The immediate needs of the clients can be fulfilled with the press report and Videotext. The aim is to have the full set of QNA data with the customers within days of the press release. The electronic publication is the means to do so. At this moment QNA has an electronic publication which consists a lots of large tables with time series of all published variables. There are 28 (ascii) files with the tables of the printed publication and nearly 50 files with times series on specific variables. The time series typically go back to 1977. Values at constant prices of 1990, values in current prices and (implicit) price indices are presented. For all these there are original and seasonally adjusted time series. Also the changes of values, etc. are (electronically) printed. The present electronic publication of QNA in its ascii format is not very customer friendly. So the publication format is being changed to a different standard. The Netherlands CBS has recently updated its software for electronic publications. This is CBSview, release 2.0. This software is presented free of charge to customers and must be installed on their computer system. It uses menu bars and has a tree structure to select variables from a database.

The electronic publication QNA will be presented in its new format in the near future. It will enable customers to select data from the database of QNA with relative ease. The selection can furthermore be put in a format preferred by the customer. For example : a spreadsheet in WK1 format, a DBASE file, a SPSS file, DIF files, plain ascii format, etc.

In the new version of our electronic publication many new features have been added to the existing possibilities. There are three partial publications : time series, publication tables and investment matrices.

The time series publication has been extended. Many previously unpublished time series have been added. The second part of the electronic publication is based on the tables in the printed publication. The format of those tables has been adapted so as to make it look right on a computer screen. One of the simplest changes is putting the newest figures in the first datacolumn. In a printed publication one would put the newest figure in the last column. On a computer screen that just does not work.

The investment matrices are published electronically for the first time. These are matrices where the capital formation of one quarter is given in further detail. There are two sets of matrices. One gives capital formation by type of capital good and by industry of origin. The other gives capital formation by type of capital good and industry of destination. Each table is presented in current as well as in constant prices. Data are available from 1987 to the present.

The new electronic publication not only holds data. It also has text on the quarterly national accounts in general and the variables and tables in particular. Explanations can be called on screen at almost all levels in the tree structure.

A test version of the new electronic publication has been completed. Some of the most favoured customers (central bank, various commercial banks, ministry of economic affairs) have been asked study and criticize the result. Their comments will be used for the final publication. That will be introduced to all Dutch customers in early 1995. An English version will be made immediately afterwards.

Lots of questions about the economic interpretation of the figures are posed immediately after the press release. An extended version of the press release would be useful. Statistics Netherlands publishes a weekly bulletin. In this bulletin there is room for text and tables with data. The production of the bulletin and distribution to the customers takes about one week. It is therefore fast enough for our purpose. Every quarter we fill two pages in this bulletin with the main results and comments on the QNA. Publication takes place some days after the press report.

The last publication is the printed publication. This 50 page magazine has all information necessary for a proper interpretation of the figures. The publication opens with a special section called 'Topics'. In this section to main macro-economic trends are highlighted. Next, we have a section with explanatory notes on the results of the quarter. In this section economic growth is analyzed in many ways. Which expenditure category contributes most to GDP growth and why, for example. The contribution of branches of industry to GDP growth is calculated and analyzed. Details of gross fixed capital formation and consumption expenditure are highlighted and relationships with other trends in the economy are shown. Sometimes methodological comments are made.

In this publication there also is a section about the international trends. Developments in the economies of the other countries of the European Union are important to the Dutch economy. Recently, special attention was given to developments in the German economy. Also in the printed publication there is room for articles. Subjects have included a monthly production index of construction, economic developments in agriculture and economic relations with the rest of the world.

The last part of the printed publication is the tables section. These tables show the variables of the expenditure approach to GDP and details of expenditure, the variables of the output approach, variables of the income approach and a breakdown of total wages. Also there is a table giving details about the economic relations with the rest of the world.

6 Future fields of study

There are several issues which need further study.

The first calculations for the Dutch QNA were made for 1977. So, for many important variables the time series go back to 1977. Up to the mid eighties the quarterly input output tables were only compiled in current prices. The resulting values were then deflated to get at volume changes.

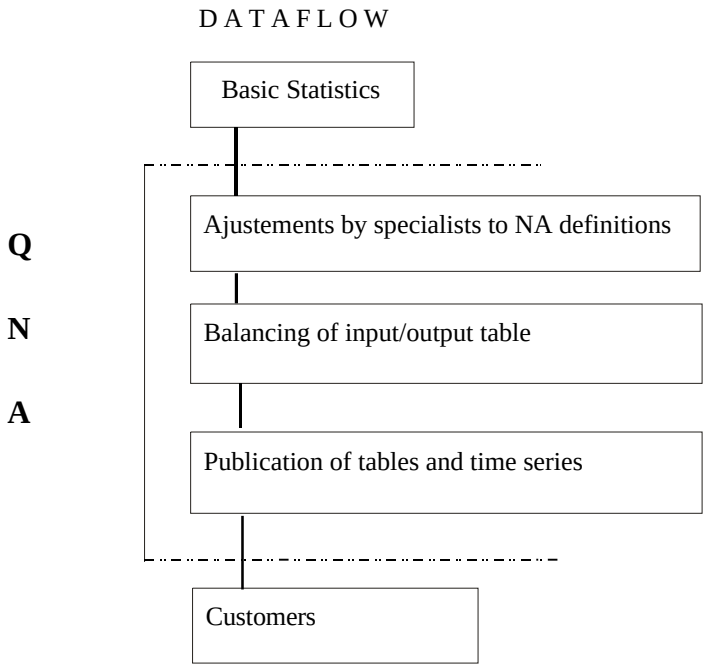
A major change in working practice led to a system where input output tables were compiled in current as well as in constant prices. For some variables the time series in constant prices start only in 1987. Work has started to extend the coverage of the time series back to 1977. Specifically, work is done on value added at constant prices by branch of industry. Now this is available only in current prices for the quarters of 1977 to 1986. The data on capital formation nowadays is fully consistent whith the QNA. In the years before 1987 this was not the case. The changes of capital formation were calculated independently from QNA. This so called investment index was subsequently used as and input in the QNA. The value in millions of guilders from the investment index however did not correspond with the levels published in the QNA. From 1987 onwards the investment index was integrated in the QNA. Inconsistencies no longer exist. What is lacking however, is a time series of the details of capital formation by type of capital good and branch of industry of destination for the quarter of 1977 to 1986. Work is in progress to recontruct from old data a consistent set of data on these subjects.

Another subject which needs more attention is the matter of working days or more precisely production days in a quarter. The number of production days changes between quarters. In the first quarter of 1992 economic growth in the Netherlands was 3.2%.

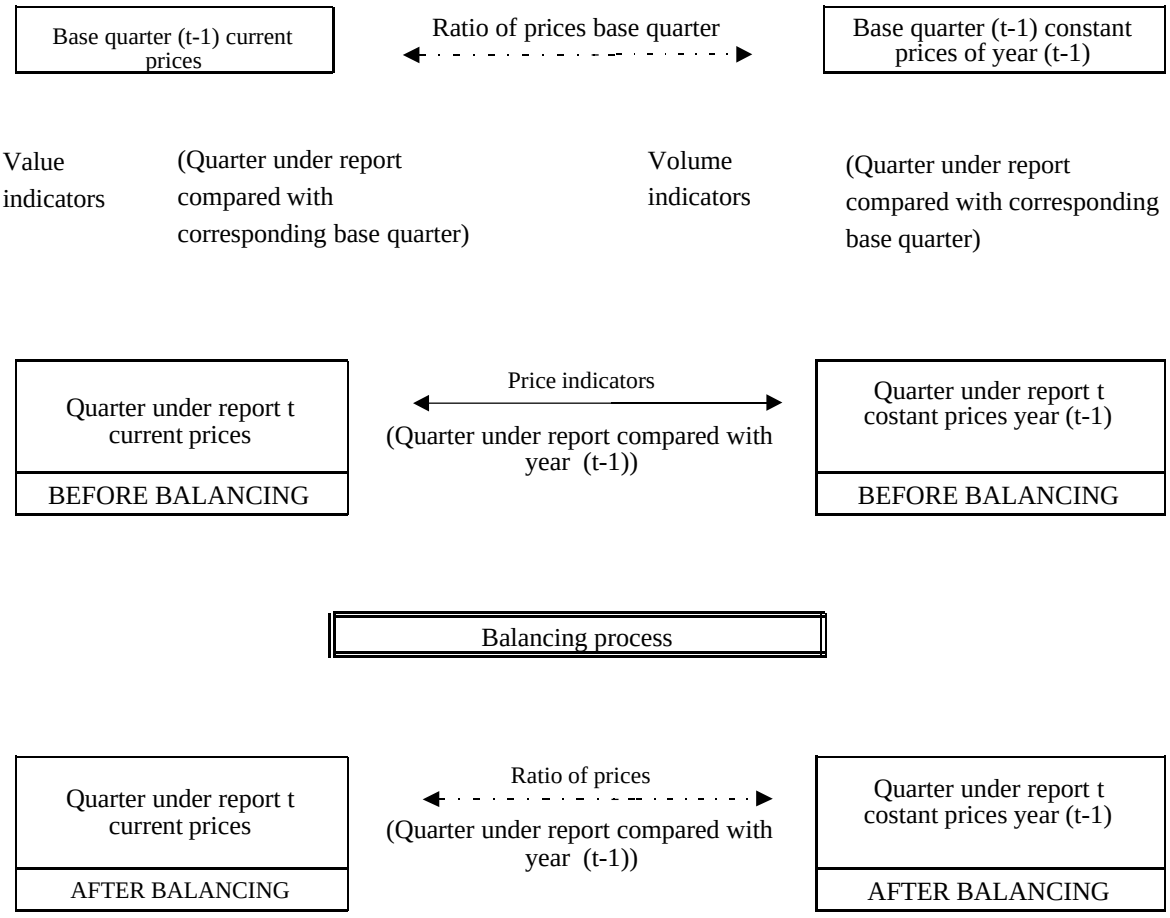
Roughly half of this growth had nothing to do with the business cycle but was caused by the leap day in combination with an extra production day in that quarter. These factors must be measured more exactly for a correct interpretation of the changes in GDP. The solution to this problem probably needs calculations of the effect of changes in the number of working days on value added for each branch of industry. That is being studied. In fact, we think a GDP-figure really describing the business cycle will result. This will make interpretation of the QNA in an economic sense more meaningful. To the end the seasonal adjustment procedures used in the Dutch QNA are being scrutinized. A more sophisticated seasonal adjustment procedure would for example keep track of production days in value added and shopping days in case of household consumption. These effects are not seasonal in nature. The point is to subtract (or add) from the time series the special effects and only then adjust the time series for seasonal influences. This probably would make the interpretation of seasonally adjusted data easier. The annual national accounts in the Netherlands also use input output tables. Recently however, the usual input output table has been replaced by use and make matrices. The use of the matrices in the QNA is to be investigated.

And last but not least : distribution of publications. The national and international use of Internet is growing very fast. Statistics Netherlands has E-mail facilities which are being used more and more. It is worth investigating the possibilities of distributing electronic publications by E-mail. Another even more promising possibility is to allow customers to access a computer (for example a GOPHER-like system) of Statistics Netherlands. The customer would of course need a username and password to get in. Subsequently, the customer can download or update the publication files. This service to customers certainly deserves further study.

Schema 1: Flow-chart of data



Scheme 2: Relations between quarterly input output tables



Scheme 3: Functions of specialists in NA and in QNA department

Functions of the specialists in I/O

for
- Industries
- Final expenditure

- Continuity of figures over time
- Adaptation to definitions of NA
- Completness for
  - cut off statistics
  - hidden economy
- Plausibility of figures
- Participation in the balancing

SECTION 4 -

SEASONAL ADJUSTMENT

Cycle tendance contre correction de variations saisonnières dans les Comptes nationaux trimestriels

Alfredo CRISTÓBAL CRISTÓBAL
Enrique Martín QUILIS
Instituto Nacional de Estadística, Madrid

1 Introduction[1]

L'emploi de la composante du cycle-tendance des indicateurs à court terme dans les Comptes Nationaux Trimestriels espagnols (CNTR) est une de leurs caractéristiques essentielles (INE, 1992, 1993). En effet, il y a une différence notable avec la pratique courante d'autres bureaux statistiques, qui corrigent les séries des variations saisonnières seulement.

La relative nouveauté des méthodes d'obtention de signaux du cycle-tendance dans le cadre économique [2] et la diffusion des méthodes du type "boite noire" comme X11 ou X11-ARIMA, très faciles d'employer mais difficiles de comprendre leur technique, à mené la production de statistiques vers un produit hybride, les séries corrigées des variations saisonnières, peu utiles pour les analystes et avec des principes statistiques améliorables.

Il s'agit de mettre en question les limitations de la méthode X11 dans la procédure d'extraction du signal de cycle-tendance. Cette méthode a eté très importante dans le développement de l'analyse des séries chronologiques, mais aujourd'hui, elle doit être changée par d'autres techniques plus efficaces et dépurées[3]. Dans ce rôle, on va présenter une méthode d'estimation du cycle-tendance employée à l'INE. Il s'agit d'une technique qui combine le dessin de filtres sur mesure (Melis 1983, 1986, 1988, 1989, 1991 et 1992) et les procédures modèles-basées (Hillmer et Tiao, 1982; Maravall, 1987; Bell et al., 1983).

Le filtre qu'on propose, appelé "Lignes Aériennes (Airlines) Modifié" (LAM) est robuste par rapport à des spécifications alternatives du modèle qui engendre les données, a un coût d'information réduit et une capacité formidable d'obtenir le signal du cycle-tendance.

L'organisation de l'article est articulée de la façon suivante: d'abord, on parle des raisons qui justifient la correction de la composante saisonnière et la procédure la plus habituelle employée (la méthode X11); après, on expose les fondements de l'obtention du signal de cycle-tendance, on critique la technique utilisée par la X11 (surtout les moyennes mobiles de Henderson) et on propose quelques solutions. En particulier, on présente une alternative choisie par l'INE. A la fin, on offre des conclusions.

---

1 Les avis manifestés dans l'article appartienent aux auteurs, non nécessairement à l'INE.

2 Cet événement est différent au procédé routinier de ces techniques dans d'autres domaines scientifi ques, par exemple dans la génie de télécommunication.

3 Il convient de souligner le prochain remplacement de la méthode X11 par une autre appelée "X12-ARI MA", réalisée par l'U.S. Bureau of the Census.

2 Pourquoi corriger des variations saisonnières ?

La correction des variations saisonnières est une pratique habituelle dans l'analyse à court terme et même dans les comptes trimestriels. Dans cette section, on examine les raisons économiques et statistiques qui permettent l'usage de cette technique. On va rapporter tous les résultats à des séries mensuelles, mais la généralisation aux séries trimestrielles est directe.

L'hypothèse des composantes dans le domaine du temps permet d'établir une série X_t comme la somme de quatre composantes orthogonales: la tendance, le cycle, la composante saisonnière et l'irrégularité:

$$[1] \qquad X_t = T_t + C_t + S_t + I_t$$

L'expression [1] est aussi valide pour un schéma multiplicatif $X_t = T_t^* + C_t^* + S_t^* I_t$ si on applique des logarithmes sur la série originale. Chaque composante est associée à une bande de fréquence:

(a) *la tendance*: est associée aux basses fréquences, c'est à dire, aux mouvements de longue durée, oscillations dont la période est plus grande que 60 mois (cinq ans). Cette composante peut être associée aux déterminants de la croissance économique: le développement technique; l'évolution du stock de capital physique; niveau, composition et qualification de la force du travail. En somme, tout ce qu'on appelle "Théorie de la Croissance" dans la Théorie Economique. Alors, on parlera de la tendance pour les oscillations dont la fréquence, exprimée en radians, soit entre 0 et $2\pi / 60$[4]. Il convient de souligner, dans cette bande de fréquence, sa limite inférieure, w=0, qui est associée aux oscillations dont la période est infinie, ou bien, dans les échantillons finis, avec des périodes dont la durée est plus grande que le nombre de données de la série. C'est la fréquence de la tendance absolue.

(b) *cycle:* la composante cyclique d'une série chronologique est caractérisée par des oscillations dont la durée est comprise entre deux et cinq ans, et d'une façon précise, entre vingt et soixante mois. C'est une composante associée aux basses fréquences, ainsi que la tendance, mais à l'origine est causée par des facteurs différents, où apparaîssent des aspects propres du court terme.

D'habitude, on peut caracteriser ces mouvements par la réponse optimum des agents rationnels à shocks exogènes différents, en prenant prix et/ou quantités comme instruments. C'est le domaine essentiel de la macroéconomie et de l'analyse à court terme. Il convient de souligner l'abondance d'explications théoriques pour ce genre de phénomènes. Par conséquent, on va considérer le cycle, toutes les oscillations dont la fréquence, en radians, soit compris entre $2\pi / 60$ et $2\pi / 20$.

Plusieur fois, c'est difficile de discriminer entre tendance et cycle, à cause des oscillations de période très longue. La plupart des séries macroéconomiques qu'on analyse sont très courtes et il peut être impossible de trouver des cycles longs. C'est pour cela que le dessin de filtres idéaux exclusifs pour la tendance ou le cycle est un problème très difficile à résoudre. D'un côté, d'un point de vue théorique, on accepte que plusieurs facteurs qui touchent la tendance sont aussi responsables de la conduite cyclique [5], ce qui rend peu convenable d'imposer une distinction très abrupte. On travaillera alors avec une composante mixte de cycle et tendance $P_t = T_t + C_t$ et comme ça [1] devient:

$$[1'] \qquad X_t = P_t + S_t + I_t$$

(c) *composante saisonnière:* il s'agit d'un mouvement périodique dont la durée est d'une année. Elle est causé, essentiellement, par facteurs institutionnels, climatiques et techniques qui évoluent doucement. La variation très faible de tels facteurs déterminent que cette composante ne soit pas la plus remarquable dans l'analyse à court

4 $w = 2\pi / p$ dont w est la fréquence (en radians) et p la période.

5 Surtout ceux qui touchent au stock de capital et à la productivité des facteurs.

terme, mais elle peut être importante dans un analyse structurel. Alors, on appelera composante saisonnière les oscillations dont la fréquence en radians soit $2\pi/12$, période d'une année (sans oublier les harmoniques $2k\pi/12, k = 2..6$).

(d) *irrégularité:* mouvements errants et très difficiles à prédire. Ils constituent généralement une source de bruit dans l'analyse à court terme, et c'est pourquoi on ne doit pas en tenir compte. Alors, on attribue à l'irrégularité aux oscillations dont la durée est inférieure à douze mois (une année), c'est à dire, inférieure à $2\pi/12$ radians, les harmoniques saisonniers exceptés.

Les composantes les plus importantes pour l'analyse à court terme sont la tendance et le cycle. Les variations saisonnières sont liées à des caractéristiques structurelles de l'économie, qui changent lentement. Le terme irrégulier ne transmet aucune information et c'est un élément de perturbation qui dénature un rapport stable.

Au contraire, d'un point de vue économique la tendance et le cycle captent ce qui est essentiel dans l'évolution d'une série. Ces composantes reprennent les effets des décisions des agents économiques sur l'accumulation de capital (physique et humain) ainsi que le rhytme et la manière de s'approcher à l'évolution à long terme.

Par rapport aux comptes trimestriels, on doit mettre en relief:

(a) les variations saisonnières des estimations trimestrielles obtenues par la méthode de Chow-Lin sont déterminées par celles de l'indicateur employé dans la régression. La plupart des fois, ces variations saisonnières ne seront pas les vraies puisque l'indicateur ne contiendra pas toute l'information de la variable annuelle.

(b) De plus, on ne peut pas vérifier de telles variations puisque la méthode de Chow-Lin met en rapport l'indicateur avec une série qui a l'information sur la tendance et le cycle seulement, dû à sa fréquence annuelle.

La manière la plus développée pour résoudre ces problèmes consiste à éliminer les variations saisonnières de la série originale, c'est à dire [6]:

$$[2] \qquad xs_t = X_t - s_t$$

Cette procédure est la "correction des variations saisonnières" et la méthode la plus populaire qui l'automatise c'est la X11 du U.S. Bureau of the Census et sa variante X11-ARIMA de Statistics Canada.

Cette méthode apporte un filtre simple, valable pour un grand nombre de séries et très facile d'interpréter dans le domaine du temps. On peut voir le tableau numéro 2 pour avoir une description complète de la méthode. La procédure basique de filtrage consiste dans un ensemble de filtres de moyennes mobiles symétriques tels que:

$$[3] \qquad H_1(B) = (1/2)\left(B^{1/2} + B^{-1/2}\right)(1/12) \cdot \left[\sum_{j=1K6} (B^{j-1/2} + B^{1/2-j})\right]$$

Après avoir appliqué [3] à la série originelle, on a une première estimation de la composante de cycle-tendance. On peut trouver la fonction de puissance de ce filtre dans le graphique 1, où on peut voir l'annulation des variations saisonnières, la relative atténuation du signal cyclique et la perceptible réduction de l'apport de la composante irrégulière.

La tendance reste invariable. Le filtre complémentaire de $H_1(B), H_2(B) = 1 - H_1(B)$, permet d'avoir une première estimation des composantes saisonnière et irrégulière. Sa fonction de puissance est dans le graphique 2.

On peut éliminer l'irrégularité par l'intermédiaire de filtres de moyennes mobiles saisonnières:

$$[4] \qquad H_3(B) = \left[(1/3)\left(\sum_{j=-1K1} B^{12j}\right)\right]^2$$

6 En general z_t est l'estimateur de Z_t.

[5]
$$H_5(B) = \left[(1/5)\left(\sum_{j=-2K2} B^{12j}\right)\right]\left[(1/3)\left(\sum_{j=-1K1} B^{12j}\right)\right]$$

La X11 a une fonction de décision qui permet de choisir entre [4] et [5] (Dagum, 1988). En avant, on va employer [5] sans perte de généralité.

On trouve dans le graphique 3 la fonction de puissance de [5]. Après avoir appliqué le filtre, il reste presque seulement la tendance et les variations saisonnières. Avec $H_2(B)$ y $H_5(B)$ et la restriction de que le filtre soit inoffensif pour les séries sans variations saisonnières (que la somme des termes intra-annuels soit nulle), on aura une première estimation de cette composante saisonnière:

[6] $H_6(B) = H_2(B)H_5(B)H_2(B)$

Alors, le filtre qui corrige les variations saisonnières sera:

[7] $H_7(B) = I - H_6(B)$

La figure 4 contient la fonction de puissance du filtre. On peut voir que après [7], les oscillations dont la période est supérieure à vingt mois restent intactes (puissance unitaire). L'irrégularité reste intacte aussi et les variations saisonnières s'évaporent.

La méthode X11 effectue une deuxième réitération dont la plus importante contribution est une estimation plus fine du cycle-tendance, fondée sur l'emploi comme input de la série corrigée de variations saisonnières. Pour cela, on emploie un filtre de passage bas (de basse fréquence) appelé "moyennes mobiles de Henderson", symétrique, dont la formule est:

[8] $$H_4(B) = \sum_{j=-mKm} c_j B^j$$

On va l'analyser [8] en détail dans la prochaine section. La X11 considère trois grandeurs pour [8]: m=4, 6 et 11. Le choix dépend du quotient signal/bruit, de telle sorte que pour les séries plus (moins) irrégulières, le filtre est plus grand (petit) (Dagum, 1988). Le tableau 1 expose les coefficients de [8] pour toutes les grandeurs de m.

Tableau 1 : Coefficients du filtre de Henderson

	m		
j	**4**	**6**	**11**
0	.32144	.23381	.13994
1	.26297	.21038	13472
2	.12540	.14854	.11977
3	-.00296	.07059	.09716
4	-.04612	.00422	.07003
5	—	-.02869	.04206
6	—	-.02195	.01697
7	—	—	-.00217
8	—	—	-.01346
9	—	—	-.01661
10	—	—	-.01302
11	—	—	-.00542

Tableau 2 : Equations basiques de la X11

A: PREMIERE ESTIMATION

Série	Formule
Cycle-tendance:	$p^{(1)}_t = H_1(B)\, X_t$
Comp. saison. + irrég.:	$si_t = X_t - p^{(1)}_t = [I - H_1(B)]\, X_t = H_2(B)\, X_t$
Comp. saisonnière:	$s^{(1)}_t = H_5(B)\, si_t = H_5(B)\, H_2(B)\, X_t$
	$s^{(2)}_t = H_2(B)\, s^{(1)}_t = H_2(B)\, H_5(B)\, H_2(B)\, X_t = H_6(B)\, X_t$
Irrégularité:	$i^{(1)}_t = X_t - p^{(1)}_t - s^{(2)}_t = [I - H_1(B) - H_6(B)]\, X_t = [H_2(B) - H_6(B)]\, X_t$
Corrigée de var. saison.:	$xs^{(1)}_t = X_t - s^{(2)}_t = [I - H_6(B)]\, X_t = H_7(B)\, X_t$

B: ESTIMATION DEFINITIVE

Série	Formule
Cycle-tendance:	$p^{(2)}_t = H_4(B)\, xs^{(1)}_t = H_4(B)\, H_7(B)\, X_t = H_8(B)\, X_t$
Comp. saisonnière:	$s^{(3)}_t = H_5(B)\, [X_t - p^{(2)}_t] = H_5(B)\, [I - H_8(B)]\, X_t$
	$s^{(4)}_t = H_2(B)\, s^{(3)}_t = H_2(B)\, H_5(B)\, [I - H_8(B)]\, X_t = H_9(B)\, X_t$
Irrégularité:	$i^{(2)}_t = X_t - p^{(2)}_t - s^{(4)}_t = [I - H_8(B) - H_9(B)]\, X_t$
Corrigée de var. saison.:	$xs^{(2)}_t = x_t - s^{(4)}_t = [I - H_9(B)]\, X_t = H_{11}(B)\, X_t$

L'estimation définitive de la composante saisonnière apparaît de la même façon qu'au début; on doit changer seulement H_1 par H_8 et alors:

[9] $H_9(B) = H_2(B)H_5(B)\,[I - H_8(B)]$

où $H_8(B) = H_4(B)\,[I - H_6(B)]$

Le filtre définitif pour corriger les variations saisonnières sera:

[10] $H_{11}(B) = I - H_9(B)$

L'image 5 montre la fonction de puissance du filtre [10]. Elle est très semblable à celle de [7] (graphique 4) sauf que:

(a) elle a une bande de rejet plus étroite

(b) elle offre une bande de passage plus ample dans la fréquence cyclique et

(c) les lobes de la bande irrégulière sont plus petits.

3 Pourquoi extraire le signal de cycle tendance?

L'emploi systématique des séries corrigées des variations saisonnières est contraire au faible intérêt que montrent celles du cycle-tendance, spécialement si on considère que les raisons exposées pour la correction des variations saisonnières sont les mêmes qui conseillent l'élimination de l'irrégularité. Se servir de séries corrigées des variations saisonnières revient à travailler avec le cycle-tendance contaminé par l'irrégularité. Puisque cette dernière composante n'a aucune information importante pour l'analyse à court terme, ce traitement n'est pas cohérent car on élimine seulement une partie de l'information hors ligne de l'analyse (les variations saisonnières).

Le terme irrégulier ne doit pas être confondu avec les innovations, surprises ou valeurs non anticipés. Une série chronologique peut être considerée comme l'agrégation d'un nombre infini de shocks non correlationnés, de moyenne nulle et variance finie (bruit blanc), selon l'expression:

[11] $X_t = \sum_{j=0K\infty} \psi_j a_{t-j} = \psi(B)a_t$

où $a_t \sim Niéd(0,\sigma_a)$ sont les shocks ou innovations de la série qui, estimés, représentent les erreurs de prédiction (Box et Jenkins, 1970).

L'hypothèse des composantes subjacentes dans le domaine du temps établit que la série observée peut être exprimée aussi selon [1]. De [1] et [11] on obtient:

[12] $\psi(B)a_t = T_t + C_t + S_t + I_t$

Selon [12] on peut déduire que, seulement sur des conditions restrictives à l'extrême et assez peu improbables, $I_t=a_t$. Par conséquent, dans le cas général, une innovation sera distribuée parmi toutes les composantes, non seulement dans l'irrégulière.

D'ordinaire, la correction de variations saisonnières est suivie d'un ajustage des effets du calendrier (cycle hebdomadaire ou correction par l'effet des jours ouvrables et aussi pour les fêtes mobiles, par exemple Pâques). Le cycle hebdomadaire produit un effet alias dans la série mensuelle. On peut voir une pointe dans son spectre à la fréquence de 2,873 mois, au domaine irrégulier. Corriger cet effet consiste à introduire un zéro dans la fonction de puissance du filtre (graphique 6). Maintenant, la question est, pourquoi éliminer cette irrégularité et laisser passer la restante question? Le motif de la correction du premier effet est connu et l'origine des autres inconnue mais cela n'empêche pas que les deux effets soient néfastes sur la clarté du signal à court terme estimé.

Conséquemment, on peut ajouter seulement des raisons statistiques pour justifier ce désintéressement pour les séries de cycle-tendance. On va analyser au fond le filtrage de la X11 pour l'estimation de cette composante.

Dans la première estimation on emploie une moyenne mobile centrée de douze termes [3] (graphique 1 pour sa puissance). Ce filtre a un pauvre rendement pour l'extraction du signal de basse fréquence, sauf pour la tendance pure, car il devient zéro très rapidement. La puissance du domaine cyclique est atténuée presque à un 70 pourcent. Cet estimateur est un outil défectueux pour l'extraction du cycle-tendance.

On estime de façon définitive cette composante avec le filtre:

[13] $H_8(B) = H_4(B)\,[I - H_6(B)]$

Le filtre $[I - H_6(B)]$ est l'estimateur préliminaire des variations saisonnières et $H_4(B)$ est la moyenne mobile de Henderson, dont la formule est :

[8] $$H_4(B) = \sum_{j=-m \mathrm{K} m} c_j B^j$$

On parle d'un filtre symétrique, de basse fréquence et grandeur variable qui dépend du quotient signal/bruit. On peut voir dans le graphique 7 la puissance pour les filtres de grandeurs m=6 (le médiane) et m=11 (le plus long).

Ce filtre est mieux que le préliminaire [3] et l'atténuance aux basses fréquences est moins accusée. Pourtant, il souffre de deux graves défauts:

(a) la fonction de puissance n'a pas de racines aux fréquences saisonnières. Par conséquent, il peut être seulement employé après avoir fait la correction des variations saisonnières, dans ce cas, $[I - H_6(B)]$. Si on applique le filtre de Henderson sur une série avec variations saisonnières, la série résultante sera assez douce mais elle les aura encore. La dépendance d'un filtrage préalable limite son application.

(b) la bande de passage du filtre est très ample car elle n'atténue pas suffisamment les oscillations dont sa période est entre 6 et 12 mois si m=6. Dans ce cas, le résultat ne sera pas très doux. Si m=11 cet effet disparaît, mais avec un coût supplémentaire.

Le graphique 8 montre la puissance du filtre définitif de cycle-tendance de la X11 avec des moyennes mobiles de Henderson de grandeur 13 et 23. Toutes les deux sont zéro aux fréquences harmomiques saisonnières. Le filtrage avec la moyenne de 23 termes donne un résultat plus doux que l'autre, mais attenue plus les oscillations du domaine cyclique. Il convient de souligner le schéma de fonctionnement de la X11: si la série est très irrégulière, le choix de la méthode est de prendre une moyenne de 23 termes. L'irrégularité deviendra nulle dans la série filtrée et les oscillations du domaine cyclique seront assez atténuées. Mais si la série n'est pas très irrégulière, le choix sera de prendre une moyenne de 13 termes. La série a des oscillations entre 6 et 12 mois, et le filtrage ne les élimine pas, seulement ils s'attenuent un peu. Au contraire, les oscillations du domaine cyclique restent presque invariables. Le résultat sera une série de cycle-tendance peu doux. Par conséquent, si la série est très irrégulière, l'output a une information tendencielle presque seulement et si elle est moins irrégulière, l'information du domaine cyclique reste invariable mais cachée derrière une irrégularité résiduelle non éliminée.

De plus, on peut faire une autre critique au filtre de Henderson. Celui-ci a une formule $H_4(B) = \sum_{j=-m \mathrm{K} m} c_j B^j$, où c_j sont ses coefficients, qu'on calcule avec une procédure d'optimum avec restrictions. Le filtre dépend d'une spécification très particulière de la composante du cycle-tendance:

[14] $X_t = P_t + U_t$

où: $P_t = a + bt + ct^2 + dt^3$ (tendance cubique déterministe) $U_t \sim iéd(0, \sigma_U)$ (perturbation genre bruit blanc).

On impose deux conditions:

(i) symmétrie : $c_j = c_{-j}\ \forall j$
(ii) invariabilité : $H_4(B) P_t = P_t\ \forall t$

La fonction objective de la procédure de minimisation est la norme de la série des troisièmes différences dusignal de cycle-tendance, P_t restreinte par (i) e (ii).

Toute la procédure est très arbitraire pour ces deux raisons:

(1) le choix de la fonction objective est très particulière, car prendre une différence troisième d'une série est assez rare dans l'analyse des séries chronologiques et de plus, elle a une interprétation statistique difficile.

(2) le modèle du cycle-tendance est très restrictif, arbitraire et peu consistant avec l'analyse moderne, où il est habituel d'employer des modèles ARIMA avec tendances stochastiques ou mixtes.

Il convient de souligner un autre élément très important: le coût du filtrage. C'est le nombre de

prédictions à l'extrême qu'il faut pour employer le filtre d'une façon que l'input et l'output soient simultanés, c'est à dire, le déphasage du filtre. On déduit du tableau 2 que l'estimation définitive du cycle-tendance a besoin de 54 prédictions, si on emploie la moyenne de 13 termes (m=6) et cinq de plus si on emploie celle de 23 termes (m=11). De celles-ci, 48 prédictions (presque le 90 pourcent) ont eté causées par la procédure (obligatoire) de correction des variations saisonnières qu'on doit faire avant d'appliquer le filtre de Henderson. La série définitive corrigée de variations saisonnières a besoin de 96 prédictions (m=6), 1.7 fois plus que l'estimation du cycle-tendance.

Cette différence devrait mener à l'emploi de cette dernière, surtout parce que l'usage de filtres très longs (193 termes dans la série corrigée de variations saisonnières si m=6) implique un grand nombre de révisions, spécialement aux extrêmes de la série. Par conséquent, l'évolution la plus récente ne peut être perçue avec clarté et elle devient incroyable et peu utile pour l'analyse à court terme.
Après ça, la X11 semble un outil inadéquat pour l'obtention du signal de cycle-tendance d'une série, dû aux déficences de ses filtres, surtout celui de Henderson et leurs coûts informatifs (en termes de prédictions à l'extrême). C'est pour cela que l'INE a choisi de travailler avec d'autres techniques d'extraction du signal pour essayer la correction de ces défauts-là.

On va exposer la méthode d'estimation du signal employée dans le système espagnol de comptes trimestriels.

Soit une série chronologique X_t construite par une somme de trois composantes orthogonales non observables: cycle-tendance (P_t), variations saisonnières (S_t) y bruit (I_t)

[15] $X_t = P_t + S_t + I_t$

Toutes les composantes peuvent être formulées avec des modèles ARIMA.

$$[16]\quad \begin{aligned} \Phi p(B)P_t &= \theta p(B)b_t & b_t &: Niéd(0,\sigma b)\\ \Phi s(B)S_t &= \theta s(B)c_t & c_t &: Niéd(0,\sigma c)\\ \Phi n(B)I_t &= \theta n(B)d_t & d_t &: Niéd(0,\sigma d) \end{aligned}$$

Toutes les répresentations n'ont pas des facteurs communs et les racines des polynômes en B sont sûr ou hors du cercle unitaire. Les termes b_t, c_t y d_t sont des bruits blancs indépendants.

Après [15] et [16] on obtient l'expression connue ARIMA de la série observée (forme réduite):

$$[17]\quad \Phi(B)X_t = \theta(B)a_t \qquad a_t : Niéd(0,\sigma a)$$

La relation entre [15] et [17] sera donnée par:

$$[18]\quad \begin{aligned} \Phi(B) &= \Phi_p(B)\Phi_s(B)\Phi_n(B)\\ \theta(B)a_t &= \left[\Phi_s(B)\Phi_n(B)\right]\theta_p(B)b_t\\ &\quad + \left[\Phi_p(B)\Phi_n(B)\right]\theta_s(B)c_t\\ &\quad + \left[\Phi_p(B)\Phi_s(B)\right]\theta_n(B)d_t \end{aligned}$$

Il y a un problème d'identification évident, car il est possible trouver un nombre infini de modèles en accord avec [15] et [16], compatibles avec la structure [17]. Alors, il convient d'introduire des spécifications a priori sur les modèles des composantes.
On va considérer que la composante de cycle-tendance et les variations saisonnières doivent suivre des schémas aléatoires qui respectent les renseignements statistiques habituels qu'on attribue dans l'analyse empirique d'extration des signaux.
Sur le domaine spectral, on peut associer les basses fréquences avec P_t (oscillations dont la durée est supérieure à vingt mois), et les pointes des fréquences saisonnières $(2k\pi/12, k = 1..6)$ avec S_t (oscillations dont la durée est de douze mois).

Un modèle adéquat pour le cycle-tendance sera un IMA(d, q_p), où $q_p \leq d$, car le spectre tend à l'infini à la fréquence zéro et après, il descend d'une façon monotone vers zéro aux hautes fréquences. La formulation de ce modèle est:

$$[19]\quad (1-B)^d P_t = \theta_p(B)b_t$$

Pareillement, un modèle ARMA$(11, q_s)$ où $q_s \leq 11$, réflète d'une manière adéquate les propriétés fréquentielles des variations saisonnières parce qu'il n'a aucune information à la fréquence zéro et tend à l'infini aux harmoniques saisonniers. La valeur moyenne, l'espérance mathématique est nulle. C'est:

[20] $U(B)s_t = \theta_s(B)c_t$

avec $U(B) = 1 + B + B^2 + .. + B^{11}$

après considérer [18] [19] et [20]

[21] $\theta(B) = (1-B)^d U(B)\Phi_n(B).$

L'expression [21] détermine que la partie autoregrésive du modèle de la série observée doit contenir le terme U(B) pour qu'il ait une décomposition valable.

Pour restreindre un peu plus le cadre d'analyse, on va considérer que la série X_t évolue en accord à un modèle du genre "lignes aériennes" (Airlines).

[22] $(1-B)(1-B^{12})X_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t$

$|\theta_1| < 1, \quad \theta_{12} > 0.$

La condition $\theta_{12} > 0$ s'établit pour que la série ait l'information dans les harmoniques saisonniers (Burman, 1980).
Après [21]:

[23]
$(1-B)(1-B^{12}) = (1-B)(1-B)U(B) = (1-B)^2 U(B)$

Et, par conséquent

[24] $\Phi_p(B) = (1-B)^2 \quad \Phi_s(B) = U(B) \quad \Phi_n(B) = 1$

Si on considère [18] on a:

[25]
$\theta(B)a_t = U(B)\theta_p(B)b_t + (1-B)^2\theta_s(B)c_t +$
$U(B)(1-B)^2\theta_n(B)d_t.$

Puisque $\theta(B)a_t$ est une moyenne mobile de treize termes, on doit remplir les restrictions suivantes :

[26] ordre $\theta_p(B) \le 2$
ordre $\theta_s(B) \le 11$
ordre $\theta_n(B) = 0$, c'est-à-dire, $\theta_n(B) = 1$

Alors le modèle général de la composante du cycle-tendance est un IMA (2,2) :

[27] $(1-B)^2 p_t = (1-\alpha_1 B - \alpha_2 B^2)b_t$

L'imposition du requis canonique (Bell, Hilmer et Tiao, 1983; Hilmer et Tiao, 1981) détermine que le spectre de P_t touche à l'axe d'abscisses à la fréquence $w = \pi$. Alors :

[28] $(1-\alpha_1 B - \alpha_2 B^2) = (1-h_1 B)(1-h_2 B)$

où $h_i \ i = 1,2$ sont les racines du polynôme MA(2). Par conséquent, la fonction de puissance du filtre qui origine P_t sera :

[29]
$g_p(w) = \left(\frac{\sigma_a}{\sigma_b}\right)(1-h_1\cos w)(1-h_2\cos w)/(1-\cos w)^2$

Doit au requis canonique :

[30] $g_p(w=\pi) = 0$

ce qui détermine que $h_1 = -1$.

Le modèle canonique de la composante de cycle tendance d'une série "ligne aériennes" est donné par :

[31] $(1-B)^2 P_t = (1+B)(1-\alpha B)b_t$

L'estimateur du cycle-tendance que minimise l'erreur quadratique moyenne est (Maravall, 1987) :

[32] $p_t = kV(B)V(F)X_t = kV(B,F)X_t$

où

$V(B) = \left[(1-\alpha_1 B - \alpha_2 B^2)U(B)\right] / \left[(1-\theta_1 B)(1-\theta_{12}B^{12})\right]$
$V(F) = \left[(1-\alpha_1 F - \alpha_2 F^2)U(F)\right] / \left[(1-\theta_1 F)(1-\theta_{12}F^{12})\right]$
$k = \sigma_b / \sigma_a$

Le filtre extracteur du cycle-tendance est symétrique, centré et avec des queues infinies. C'est pourquoi on a besoin d'un grand nombre de prédictions (en avant et en arrière) si θ_1 et/ou θ_{12} sont près de la limite d'inversibilité. De la même manière, pour calculer les paramètres α_1 *et* α_2 du filtre, il faur résoudre un sistème d'équations originé à partir de la méthode des moments (Maravall, 1987). Ce système a aussi des restrictions ajoutées, dû au principe de décomposition canonique.

Récemment, on a proposé une variation du filtre [32] appelée "lignes aériennes modifié" (LAM) (Melis et Gómez, 1989; Melis, 1991, 1992). Ce filtre évite la décomposition à fractions partielles proposée par Burman et Maravall et par conséquent son emploi est très simple.

Les coefficients $\alpha_i \; i=1..2$ dépendent des autres $\theta_i \; i=1..2$:

[35] $\alpha_i = \alpha_i(\theta_1, \theta_{12}) \qquad i=1,2$

On applique deux conditions au filtre :

a. Requis canonique : la puissance doit-être zéro à $w=\pi$ (absence de bruit). Alors, si $h_1 = -1$ on aura : $\left(1-\alpha_1(B=-1)-\alpha_2(B=-1)^2\right)=0$ c'est-à-dire :

[36] $\alpha_2 - \alpha_1 = 1$

b. Condition de tangence élevée à l'origine (Melis, 1992) : il faut que la deuxième dérivée de la fonction de puissance soit nulle à $w=0$ (de cette manière, on obtient une similitude maximum avec le filtre idéal de passage bas) :

[37] $d^2 g(w)/dw^2 = 0 \text{ À } w=0$
avec $g(w) = ||V(e^{-iw})||^2$

Dû aux conditions (a) et (b) on peut calculer les paramètres $\alpha_i \; i=1..2$, fonctions de $\theta_1 \; y \; \theta_{12}$ selon les expressions suivantes (Melis, 1992) :

[38] $C = -\left\{\left[\theta_1/(1-\theta_1)^2\right]+\left[s^2\theta_{12}/(1-\theta_{12})^2\right]\right\}$
avec s=12

[39] $\alpha_1 = \left[2-2\sqrt{(2-4C)}\right]/(4C-1)$

[40] $\alpha_2 = 1+\alpha_1$ (Requis canonique)

[41] $k = \left[(1-\theta_1)(1-\theta_{12})\right]/\left[s(1-\alpha_1-\alpha_2)\right],$
avec s=12

L'expression [41] découle de la condition de normalisation de la fonction de puissance, si on fait l'unité sa valeur à l'origine. Les avantages principaux du filtre V(B,F) sont deux:

(a) il s'adapte aux caractéristiques de la série, de manière qu'on amplifie (contracte) la bande de passage si l'input est peu (très) irrégulier.

(b) on connaît la modélisation de la composante du cycle-tendance et on peut réaliser des tests sur l'adéquation de l'extraction faite (p.e., analyse de similitudes entre les fonctions d'autocorrélation théorique et celle du signal estimé).

La désavantage principal de ce filtre apparaît quand la série a une valeur θ_1 positive (situation habituelle aux séries avec irrégularité moderée ou élevée). Dans ce cas, la fonction de puissance ferme trop la bande de passage et attenue beaucoup l'information du domaine cyclique (graphique 9). Alors, il devient un bon estimateur de la tendance mais non du cycle.

Les paramètres du filtre dépendent de ceux de la série modélisée, pourtant, on peut fixer θ_1 et θ_{12} et construire un filtre fixe de manière que la fonction de puissance soit appropiée pour l'extraction du cycle-tendance et le coût, nombre de prédictions à l'extrême, soit bas.

Au point de vue du dessinateur de filtres, ce n'est pas nécessaire que ceux-là soient originés par un bon modèle de prédiction; il sera suffisant que le filtre dessiné ait une fonction de puissance adéquate pour l'extraction de la signal et que le déphasage (coût à l'extrême) soit minimum. On ne va pas chercher une liaison directe entre le modèle de la série et le filtre extracteur. Pourtant, on va employer le filtre V(B,F) dessiné, mais avec quelques modifications:

(1)- on va maintenir $\theta_1 < 0$, fixe dans le filtre. De cette manière, on assure un bon choix du signal cyclique.
On fixera aussi θ_{12} dans le filtre. La procédure automatique assume $\theta_1 = -0{,}80$ et $\theta_{12} = 0{,}85$.

(2)- après V(B,F), et pour éliminer l'irrégularité qui contient la série filtrée, on va appliquer un autre filtre de passage bas de la famille Butterworth, autorégressif d'ordre quatre, avec une puissance moitié à seize mois dessiné sur mesure selon:

[42] $\Omega(B) = \Omega_0/\left(1+\Omega_1 B+\Omega_2 B^2+\Omega_3 B^3+\Omega_4 B^4\right)$

où $\Omega_0 = 0{,}0139$ $\Omega_2 = -3{,}4456$
$\Omega_1 = 2{,}9889$ $\Omega_3 = 1{,}0829$
$\Omega_4 = -0{,}3598$

On peut voir les fonctions de puissance et déphasage aux graphiques 10a et 10b.

(3)- en dernier lieu, il faut observer qu'on a un filtre non symétrique et par conséquent le nombre de prédictions dont on a besoin aux extrêmes de la série est différent. Au début, il faut prédire 17 observations, 13 du filtre V(B) (ARMA(13,13)) et 4 de l'adoucissant AR(4). A la fin, il faut prédire 18 observations, 13 dues au filtre V(F) et 5 dues au déphasage du AR(4) au domaine cyclique. Si on applique seulement $V(B)\Omega(B)$, c'est-à-dire, ignorant V(F), le coût informatif à la fin de la série sera la somme des déphasages des filtres V(B) et $\Omega(B)$ au domaine cyclique, qui sont, respectivement, un et cinq mois, c'st-à-dire, six mois, coût très bas.

Le filtre proposé, qu'on emploie dans la CNTR espagnole est :

$$[43] \qquad H(B) = kV(B)\Omega(B)$$

Les graphiques 11a et 11b montrent les fonctions de puissance et déphasage de H(B). On a remarqué le domaine cyclique. Il est nécessaire d'avancer six mois l'input afin de corriger le déphasage que le filtre induit sur l'output. L'expression complète du processus est :

$$[44] \qquad p_t = kV(B)\Omega(B)F^6 X_t$$

On a écrit une procédure SAS pour automatiser le filtre et pouvoir l'appliquer à n'importe qu'elle série. Celle-ci est accesible à toute personne qui la demande.

4 Conclusion

Il n'y a pas de raisons consistantes pour l'emploi de séries corrigées de variations saisonnières dans l'analyse à court terme au lieu de celles de cycle-tendance. Les filtres qui corrigent ces variations-là sont pires que celles qui extraient le cycle-tendance car ils laissent toute irrégularité (on ne peut pas voir avec clarté l'évolution de la série et alors, ce sera difficile de faire un diagnostic à court terme) et ils ont besoin d'un grand nombre de prédictions à l'extrême et, par conséquent, il faudra réviser les dernières données filtrées, dans quelques cas, presque cinq ans.

De toutes les méthodes analysées pour l'extration du cycle-tendance, le filtre proposé est le plus adéquat puisqu'il a la meilleure conduite (puissance) au domaine cyclique et en outre, son coût informatif à l'extrême est le plus bas.

En effet, le graphique 12a montre les fonctions de puissance des trois filtres de cycle-tendance analysés dans l'article: les deux correspondants à la X11 avec les moyennes mobiles de Henderson de 13 (m=6) et 23 (m=11) termes et le filtre proposé qu'on emploi dans la CNTR espagnole. Tous les trois éliminent les variations saisonnières et presque toute l'irrégularité produite par des oscillations dont la durée est plus petite que six mois.

Le graphique 12a et avec plus de détail le graphique 12b montrent les différences entre les puissances des trois filtres au domaine des oscillations d'une période plus grande de six mois. Le filtre X11 avec la moyenne de 13 termes est le meilleur estimateur du cycle, mais n'attenue pas beaucoup les oscillations de période entre six et douze mois, surtout celles de huit mois de durée. Alors, le signal produit sera assez irrégulier.

Les autres deux rendent nulles de telles oscillations. Après avoir analysé le domaine cyclique, on peut voir que le filtre proposé est toujours plus puissant que celui de la X11 construit avec la moyenne de 23 termes. C'est pour cela que le filtre proposé est préférable a celui de la X11.

En outre, ce filtre a un coût informatif très inférieur a celui-là: 6 observations par rapport à 59 de celui de la X11. C'est à dire, le filtre de la X11, produit une variabilité dans les 59 dernières observations du signal (cinq ans !). Cet aspect est contraire à l'analyse à court terme, plus intéressée aux observations les plus récentes, lesquelles sont les plus révisées. Cela démontre une fiabilité assez faible du signal estimée et par conséquent introduit un risque élevé de faire un mauvais diagnostic économique à court terme.

Références

Bell W.R., Hillmer S.C. et Tiao G.C. (1983) *Modeling considerations in the seasonal adjustment of economic time series* , en Zellner, A. (Ed.) *Applied Time Series Analysis of Economic Data*, U.S. Department of Commerce, Bureau of the Census, Washington, U.S.A.

Box G.E.P. et Jenkins G.M . (1970) *Time series analysis, forecasting and control* , Holden Day, San Francisco, U.S.A.

Burman J.P. (1980) *Seasonal adjustment by signal extraction*, Journal of the Royal Statisticial Society, series A, n. 143, p. 321-337.

Dagum E.B. (1988) *The X11-ARIMA seasonal adjustment method* , Statistics Canada, Ottawa.

Hillmer S.C. et Tiao G.C. (1982) *An ARIMA model-based approach to seasonal adjustment* , Journal of the American Statistical Society, vol. 77, n. 377, p. 63-70.

INE (1992) *Nota metodológica sobre la Contabilidad Nacional Trimestral* , Boletín Trimestral de Coyuntura n. 44, Instituto Nacional de Estadística.

INE (1993a) *Contabilidad Nacional Trimestral de España. Metodología y serie trimestral 1970-1992* , Instituto Nacional de Estadística.

Maravall A. (1987) *Descomposición de series temporales. Especificación, estimación e inferencia* , Estadística Española, vol. 29, n. 114.

Melis F. (1983) *Construcción de indicadores cíclicos mediante ecuaciones en diferencias* , Estadística Española, n. 98, p. 45-89.

Melis F. (1986) *Apuntes de Series Temporales* , Documento Interno, Instituto Nacional de Estadística.

Melis F. (1988) *La extracción del componente cíclico mediante filtros de paso bajo* , Documento Interno, Instituto Nacional de Estadística.

Melis F. (1989) *Sobre la hipótesis de componentes y la extracción de la señal de coyuntura sin previa desestacionalización* , Revista Española de Economía, vol. 6, ns. 1 y 2.

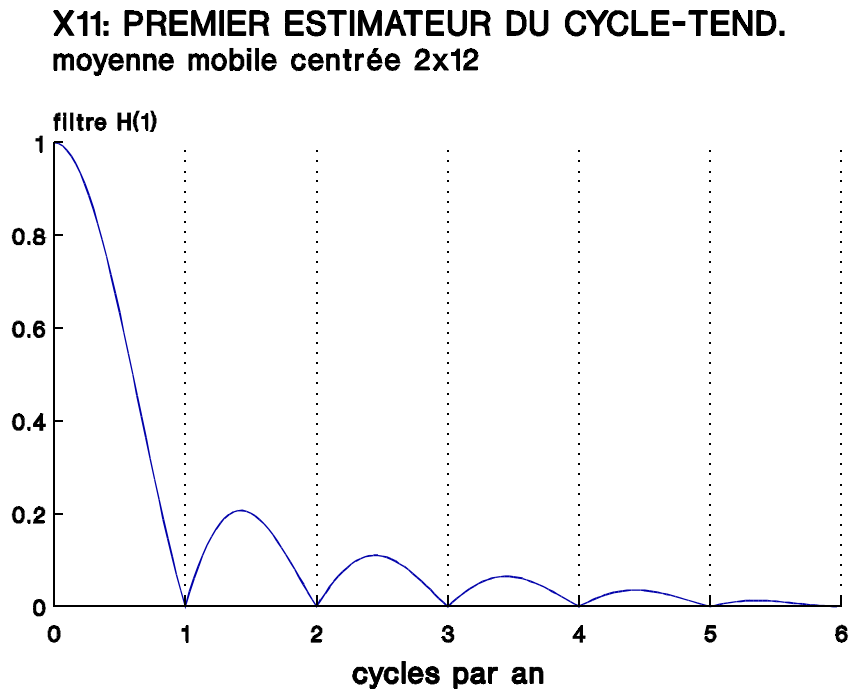
Melis F. (1991) *La estimación del ritmo de variación en series económicas* , Estadística Española, vol. 33, n. 126.

Melis F . (1992) *Agregación temporal y solapamiento o "aliasing"* , Estadística Española, n. 130, p. 309-346.

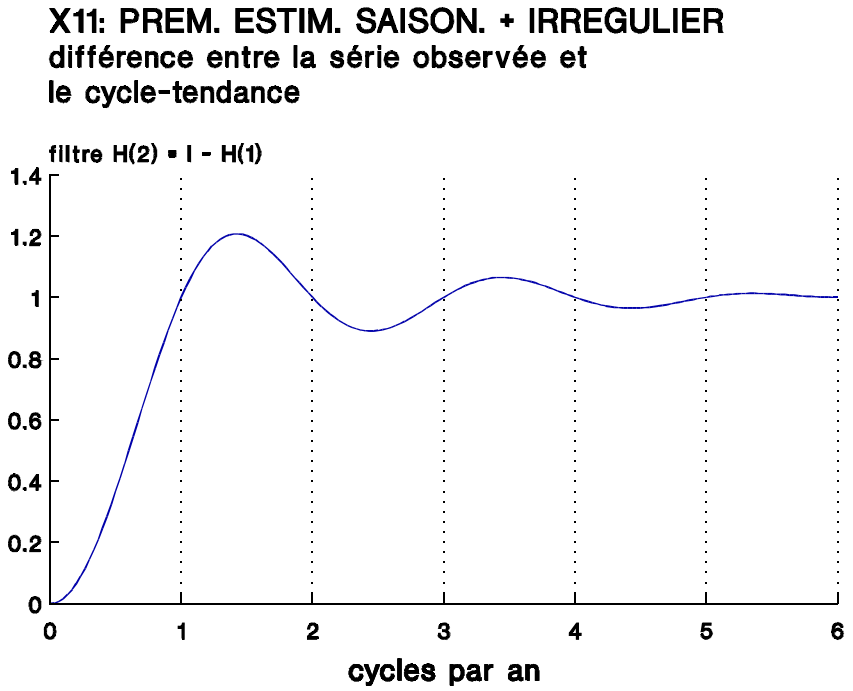
Melis F. et Gómez V. (1989) *Sobre los filtros de ciclo-tendencia obtenidos por descomposición de modelos ARIMA* , Documento Interno, Instituto Nacional de Estadística.

Appendix : graphiques

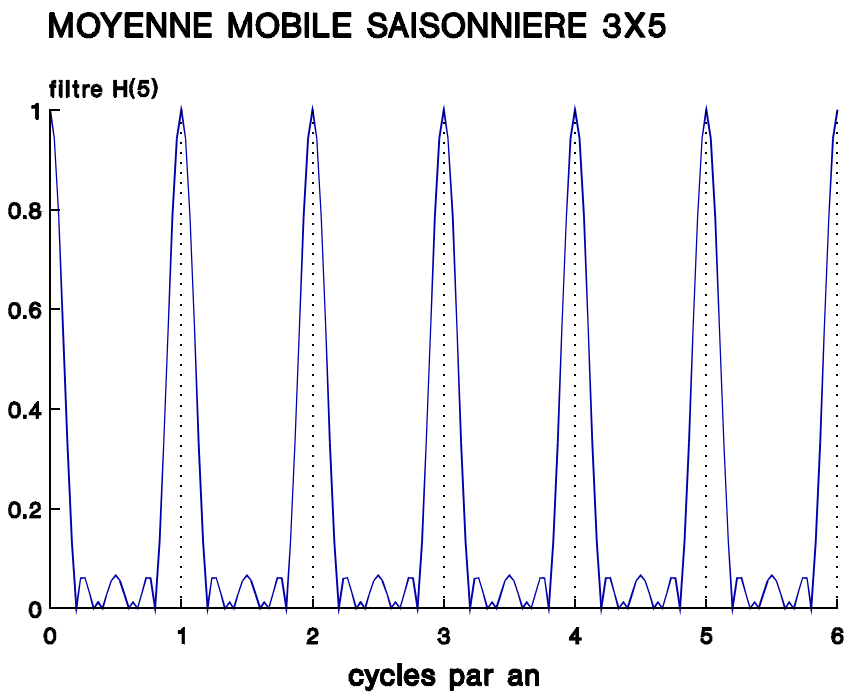
Graphique 1



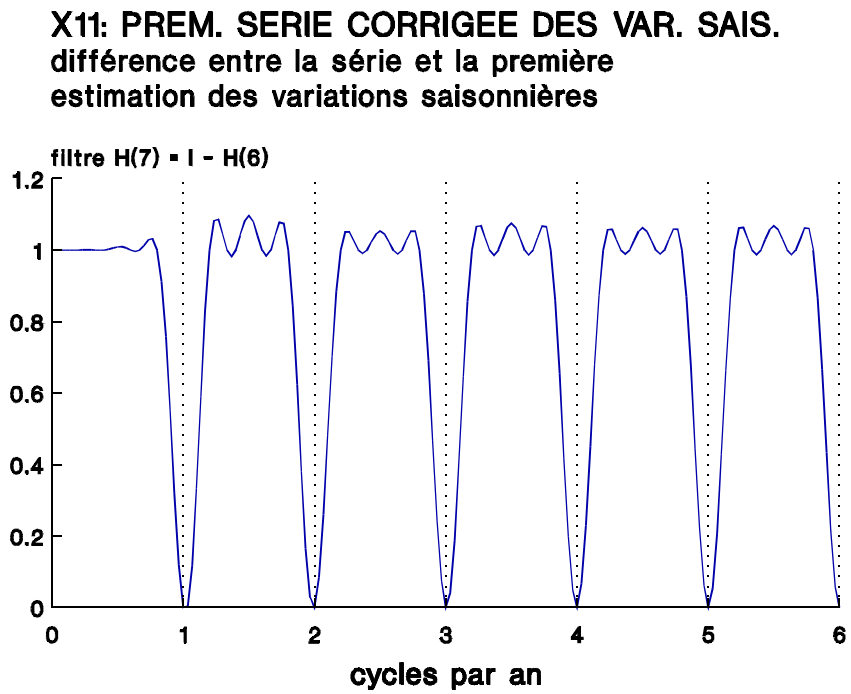
Graphique 2



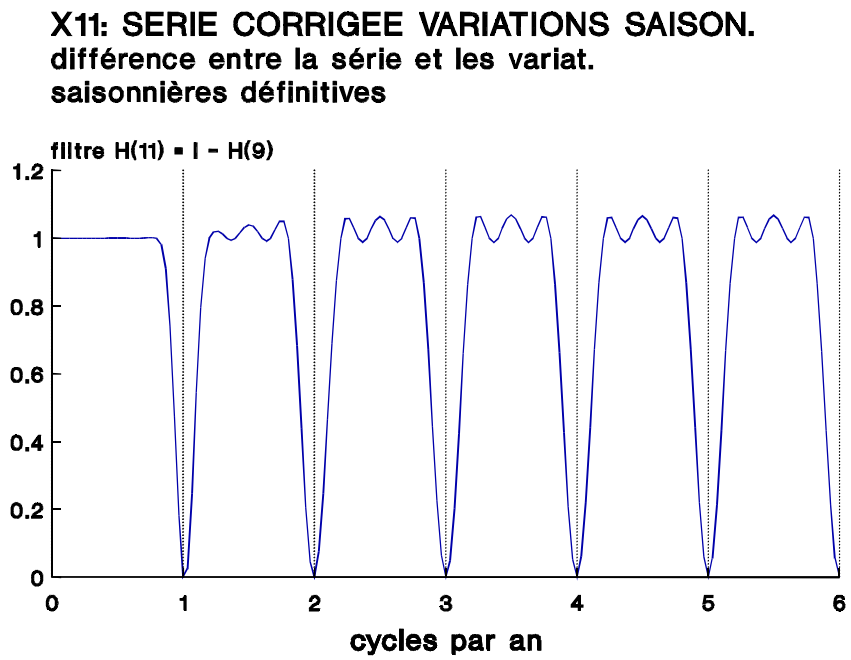
Graphique 3



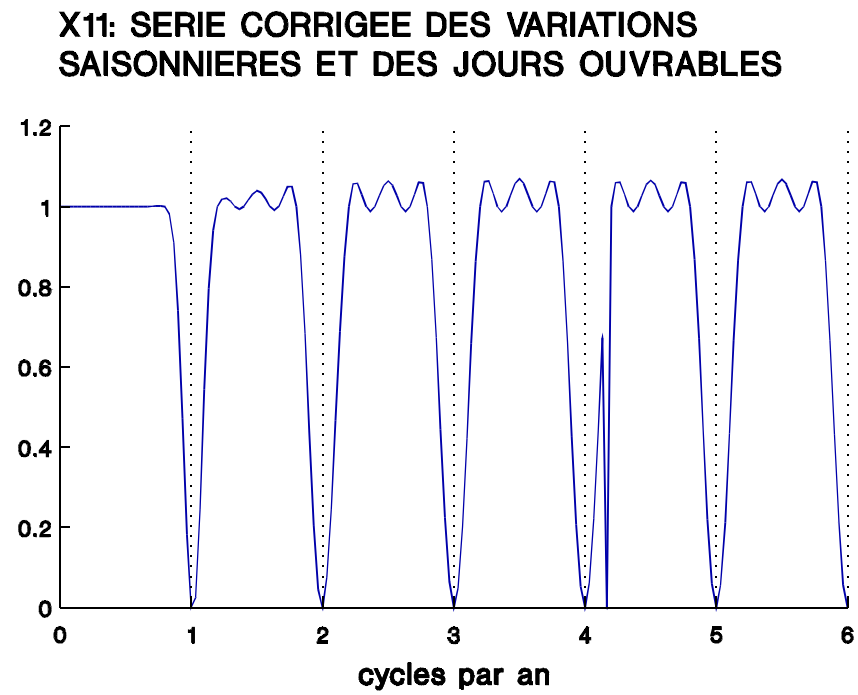
Graphique 4



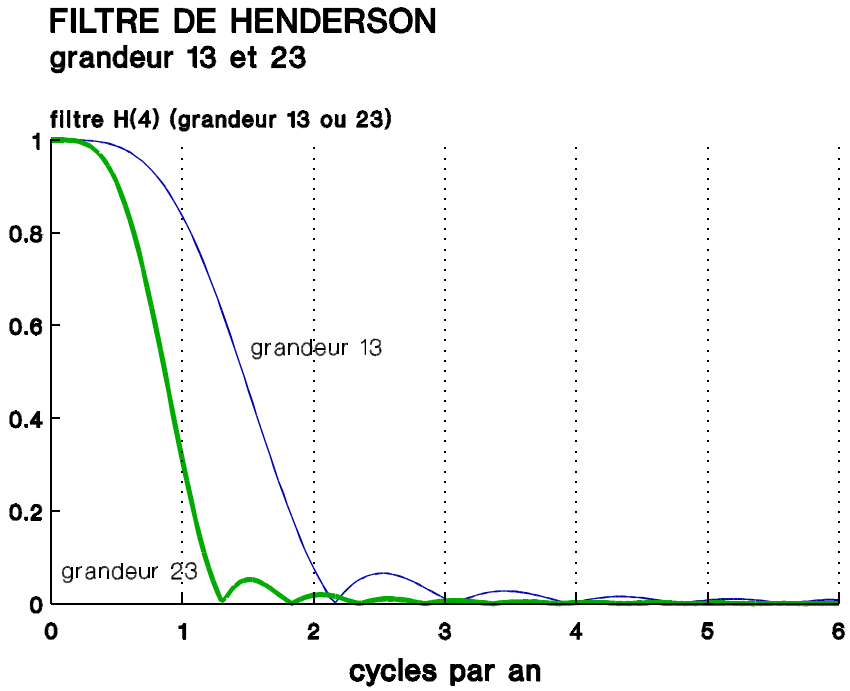
Graphique 5



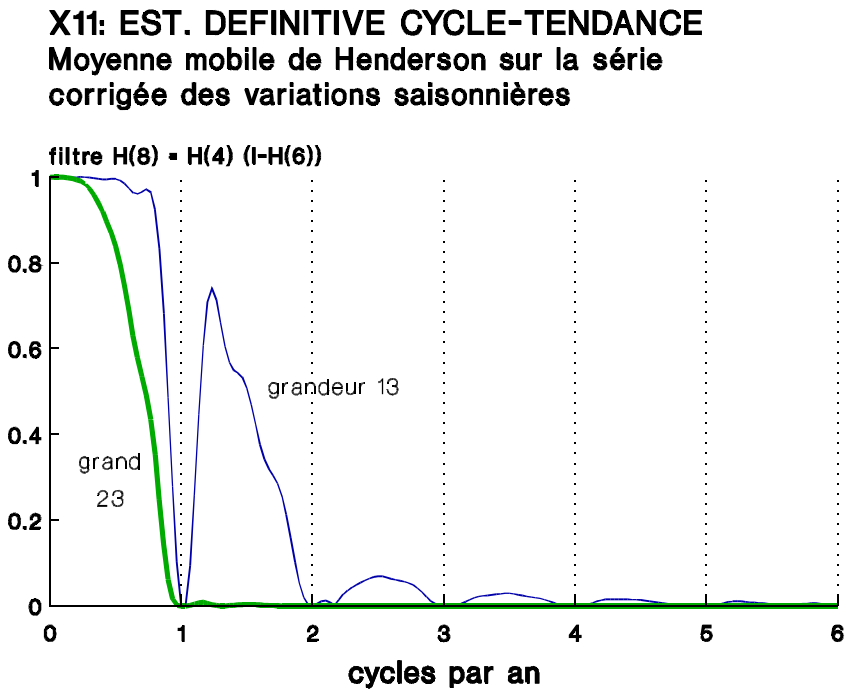
Graphique 6



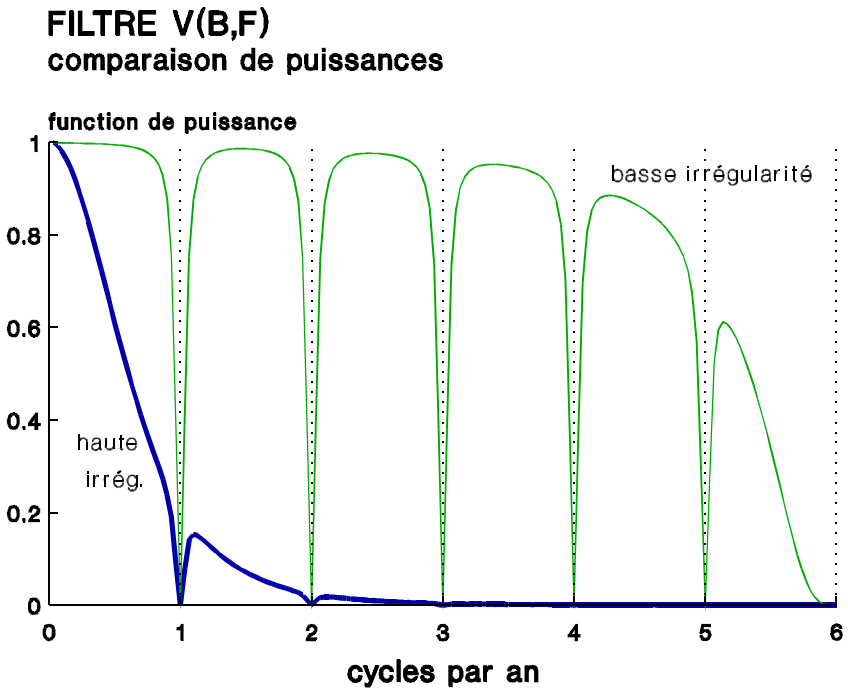
Graphique 7



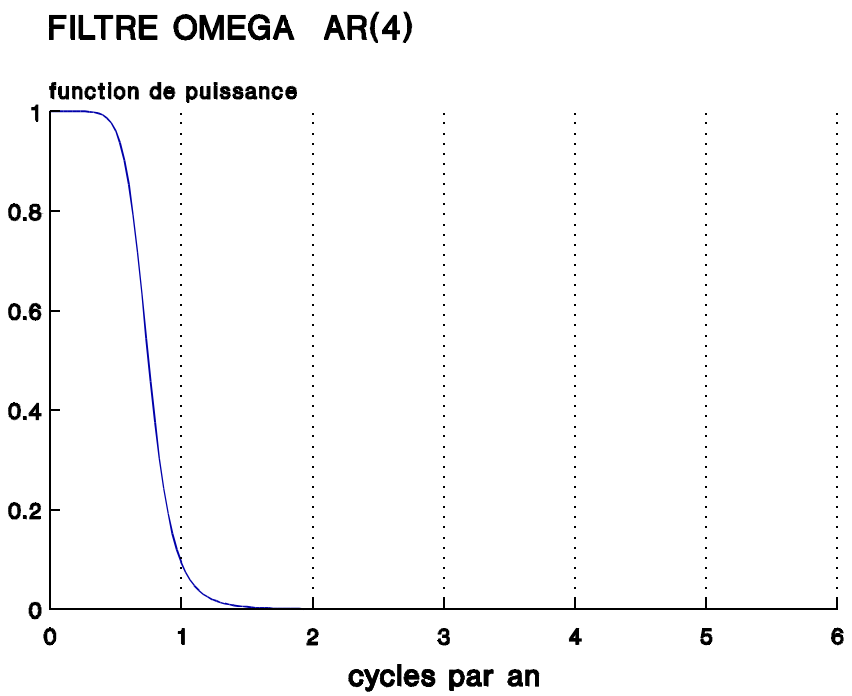
Graphique 8



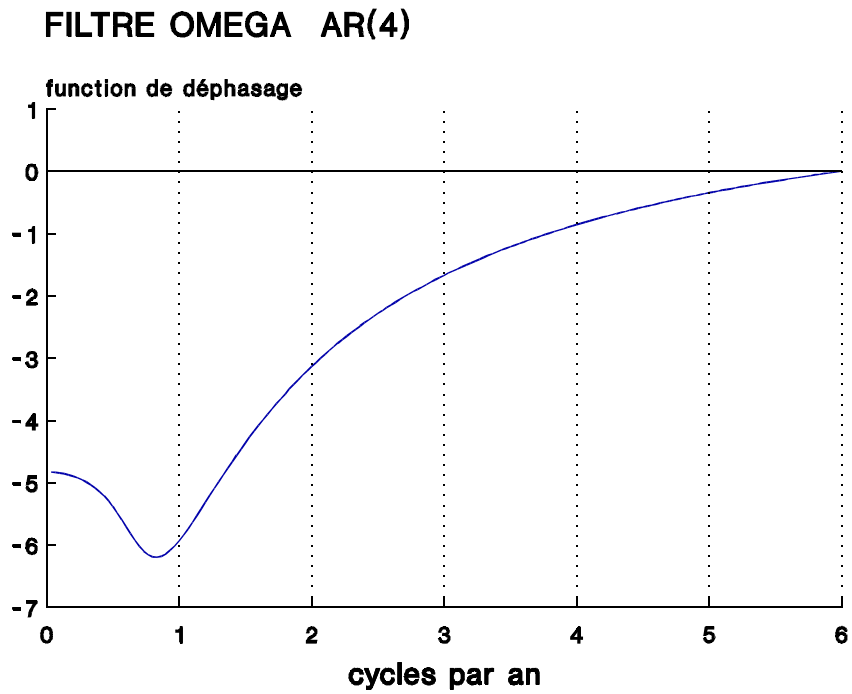
Graphique 9


FILTRE V(B,F)
comparaison de puissances
function de puissance
1
0.8
0.6
0.4
0.2
0
haute
irrég.
basse irrégularité
0
1
2
3
4
5
6
cycles par an

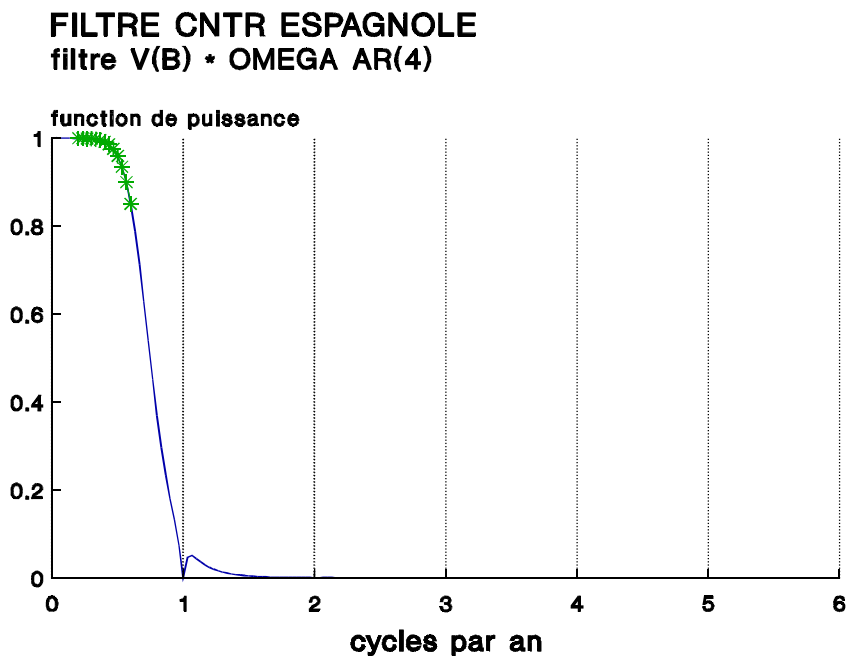
Graphique 10a



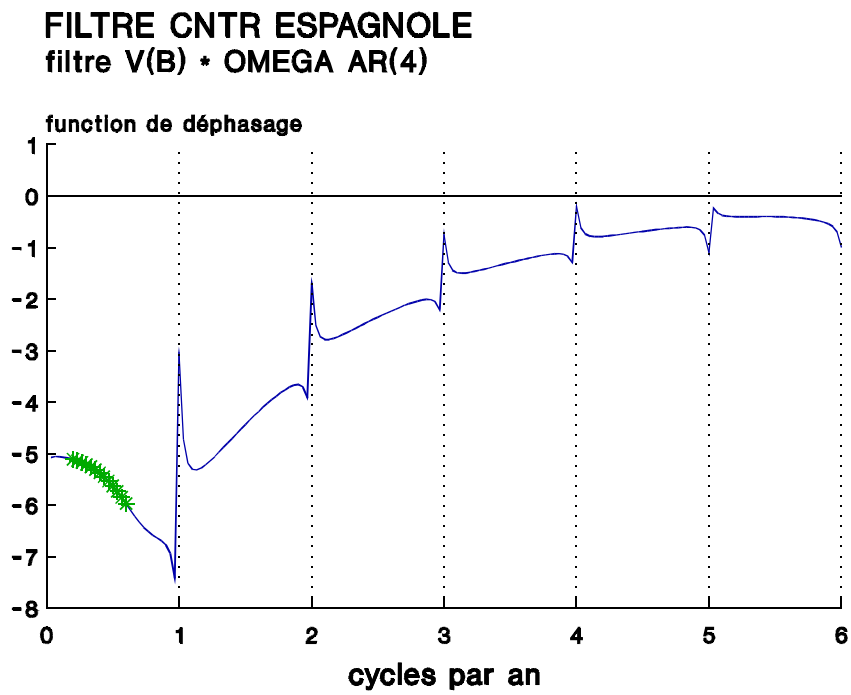
Graphique 10b



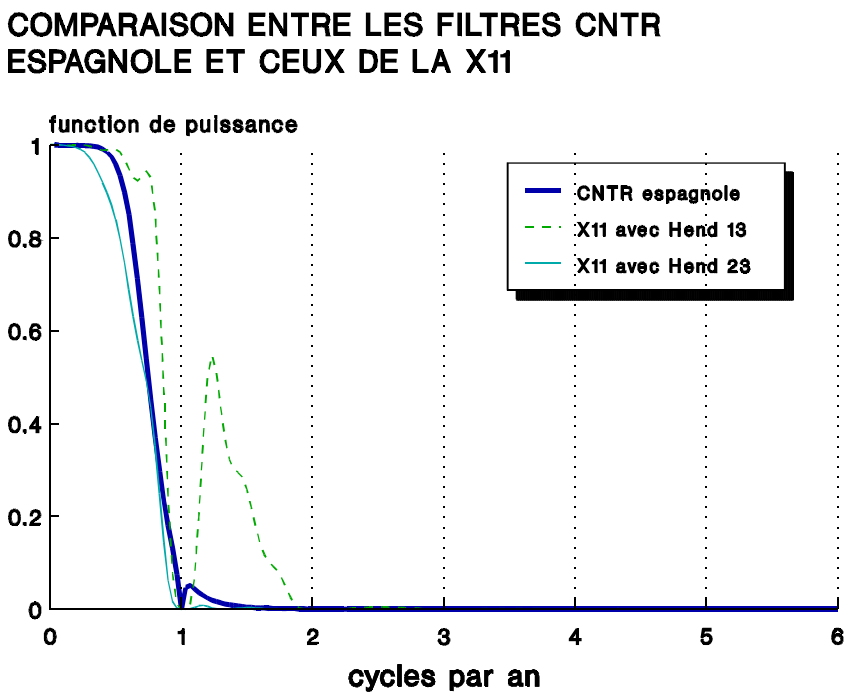
Graphique 11a



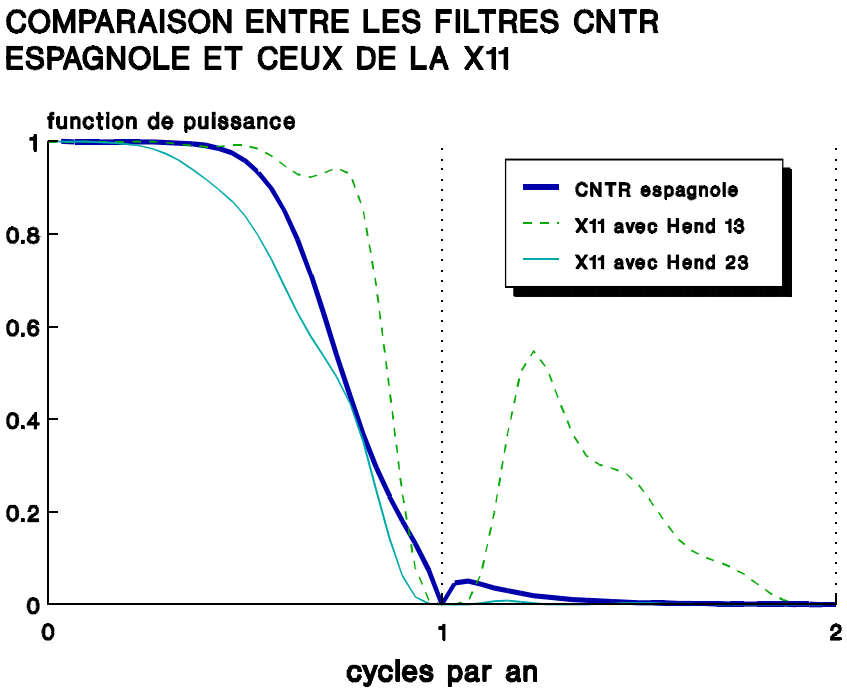
Graphique 11b



Graphique 12a



Graphique 12b



SECTION 5 -

INDIRECT METHODS

The monthly GDP indicator

Erkki LÄÄKÄRI
Statistics Finland

1 The purpose and structure of the monthly indicator

The monthly indicator of total output is a quick way of indicating the trend in overall economic development. It is based mainly on monthly statistics, speed being one of its key objectives in comparison with the quarterly accounts. The monthly indicator is completed within two months from the end of the month concerned.

The other objectives are simultaneousness and a high degree of correlation with the reference series. The quarterly GDP series, the widest and most familiar index describing total output, has been selected as the reference series. Linear interpolation is used to convert the series to a monthly one.

The structure of the monthly index is based on the principle of representativeness under which single indicators, or indicators composed of a few series, represent the output of the different sectors of the economy. This method was selected in order to make it possible to follow not only total output but also the trend in the output of the most important economic sectors.

In addition, the production method used resembles the method by which the GDP series is compiled in the quarterly accounts, thus facilitating the use of this series for reference.

2 Sectoral review

Primary production

Primary production, the first main sector of the economy, consists of agriculture and forestry. Its contribution to total output in 1993 was 6.3 per cent. Milk and meat production, which accounts for approx. 60 per cent of agricultural output, is used to represent agricultural output. The output of crop cultivation is taken into account during the harvest season, i.e. in August, September and October. Market fellings, which cover 80 to 90 per cent of the value added in forestry, are used to represent the output of forestry.

Manufacturing

The monthly volume index of industrial output is used as the indicator of industrial output. Manufacturing accounts for approx. 27.6 per cent of total output.
To enable output comparisons between counterpart months in different years, the monthly volume index of industrial output is calculated per working day.

Construction

The construction industry accounts for approx. 6.5 per cent of total output. The index of construction material inputs 1990=100 is used as the indicator of the output of the construction industry. It is modified with the help of the construction industry series on employment and on the hours worked which are derived from Statistics Finland's Labour Force Survey.

Trade

Trade accounts for approximately. 10.7 per cent of GDP. The volume indices of wholesale and retail sales are used as the indicator of the output of trade.
The monthly volume index of retail sales is calculated per trading day.

Transport and communications

Transport and communications accounted for 8.7 per cent of total output last year. The monthly indicator of transport and communications consists of three indicators: the real income of the Finnish P&T; the gross ton kilometres of goods transport on the State Railways; and the consumption of diesel oil, which mainly describes the development of professional road transport. Railway and road transport account for approximately 65 per cent of the output of transport and storage and for approx. 40 per cent of the output of the whole sector. Communications account for just over a quarter of the output of the whole sector.

Public and other services

The sector accounts for approx. 40.2 per cent of total output. The linear trend calculated from quarterly accounts data has turned out to be the best indicator of the sector for the purposes of the monthly indicator.

3 Calculation of the monthly indicator

The series of indicators describing an economic sector are weighted together to serve as the indicator of the sector. The indicators of the six economic sectors are made commensurable by indexing them according to a base year, currently the year 1990 (1990=100).

The monthly indicator of total output is formed by weighting the sectoral indicators together by means of their relative contributions to GDP at fixed prices in the previous year and by calculating the regression estimate of the weighted indicator with the help of the GDP series serving as the reference series.

4 The trend-cycle component of the monthly indicator

The monthly indicator is calculated in two versions: the one as based on the original series and the other, the so-called trend-cycle component, as adjusted for seasonal and random variation. The calculation of the trend-cycle component is analogous to the version based on the original series in that each series is separately adjusted for seasonal variation and, after the series have been weighted together, the regression estimate is calculated with the help of the seasonally adjusted quarterly series serving as the reference series.

The method of seasonal adjustment used is the one developed by the Bank of Finland and which is a slight modification of the second version, X-11, of the method developed by the US Bureau of the Census. In october 1994 we have changed the method to X-11-ARIMA.

4 Publication

The key results of the monthly indicator of total output are published as a four-page monthly report appearing two months after the end of the month in question.

5 The monthly indicator and quarterly national accounts

The quarterly national accounts are compiled indepently from GDP-indicator, but because of the data sources are partly the same, there is a close connection between these two estimates of the GDP.

The quarterly accounts are published about two months and three weeks after the end of the quarter. The data available at that moment is much larger than is available for monthly indicator.

The results of the indicator are used as a comparison data for quarterly accounts. It is also useful in estimating the economic development,because of the production lag in quarterly accounts.

The indicator measures GDP, or its changes in basic values. Therefore we have to compare the change in GDP-indicator with the change of quarterly GDP in basic values. Table 1 shows the annual volume changes of the indicator and the quarterly GDP.

Because the figures accounted by both ways are revised when new data is available, the comparison is made from the first release day figures.

When comparing the first estimates given by the monthly indicator with the first estimates given by quartely national accounts we can observe that the indicator is overestimating the development in 1992 and in the first half of 1993.

With closer examination we have observed that the monthly indicator has failed to forecast the development in two sectors of economy; namely Wholesale and retail trade and public and other services.

During depression the linear reggression in estimating the production of the public sector is not good anymore.

The problem with the wholesale and retail sales statistics was merely a problem of timing. The sales statistics is released one and a half week later than the monthly indicator is published.At the time the indicator was compiled the actual sales figures were not known.

These two problems we have solved by changing the method of estimating the public sectors development by receiving information of the salaries paid by the public sector. The problem with salesstatistics is also solved. We are receiving informatin from sales statistics as soon as they receive it from the firms.

As can be seen the development given by the monthly indicator after the first half of 1993 is quite similar to the development given by the quarterly national accounts.

Table 1: Annual volume changes of the monthly indicator and quarterly GDP

Year	Quarter	Monthly GDP indicator	GDP in basic values
		First release day	First release day of quarterly accounts
1992	I	-2.2	-2.4
	II	-0.5	-1.4
	III	-0.1	-1.8
	IV	-0.3	-1.5
1993	I	-0.5	-1.9
	II	-1.6	-2.9
	III	-1.7	0.0
	IV	0.7	0.0
1994	I	1.5	2.0
	II	5.2	5.2

The flash estimate of the Italian Real Gross Domestic Product

Giancarlo BRUNO ; Eurostat
Gianluca CUBADDA ; Università di Roma "La Sapienza"
Enrico GIOVANNINI ; ISTAT
Claudio LUPI ; ISTAT

1 Introduction

In Italy is the Istituto Nazionale di Statitica (ISTAT, the Italian Central Statistical Office) which releases the official estimates of the quarterly national accounts. As in many other European countries, in Italy, the quarterly national accounts are estimated resorting to indirect disaggregation methods. These methods, that can be traced back to Chow and Lin (1971), are based on the idea that it is possible to temporally disaggregate a low-frequency variable by using high-frequency related series. The literature on this topic is vast and we will not try to review it here. A critical appraisal is contained in Lupi and Parigi (1995) in this volume. However, the main idea is simple and can be loosely described as follows. Let $\{Y_t\}_1^{T_a}$ be an annual time series to be disaggregated at quarterly frequency and let $\{x_t\}_1^{T}$ (where $T = 4T_a$) be a quarterly time series whose annual value is related to according to the relation:

$$Y_t = \beta \sum_{j=4t-3}^{4t} x_j + \varepsilon_t \qquad (t = 1, \ldots, T_a \quad j = 1, \ldots, T)$$

Then the quarterly estimate of Y_t can be derived imposing a similar relationship among the quarterly variables.

It is evident that the timely availability of high-frequency indicators is of crucial importance for the estimation of the quarterly national accounts. In Italy, the problem of having quick access to the high frequency indicators is complicated by the number of the indicators themselves. Indeed, the present version of the quarterly national accounts covers estimates of Resources and Uses accounts, value added, standard labour units various measures of labour cost and productivity, all evaluated according to the detail of 44 branches of economic activity. Furthermore, household consumption is estimated for 50 consumption goods and services [1]. The amount of information needed to estimate the quarterly figures is therefore impressive. Unfortunately, this implies also that many of the indicators are available only with a considerable delay form the end of the reference quarter, so that the quarterly national accounts are currently released with a mean lag of 95 days form the end of the reference quarter. However, economic analysis calls for a more rapid evaluation at least or the main aggregates.

In this paper we describe the methodology followed by ISTAT for producing anticipated quarterly estimates of the main aggregates of the Resources and Uses Account. In section 2 we review various aspects

1 A complete description of the Italian quarterly national accounts can be found in Istat (1992)

concerning the compilation of quarterly national accounts in several European countries. Section 3 deals with the timing of the information needed in the final estimation of the Italian quarterly accounts. Section 4 is devoted to the discussion of the main methodological issues. Section 5 illustrates some experimental results relative to the flash estimate of the Italian GDP for the first quarter 1995. Section 6 concludes.

2 Situation of quarterly national accounts in European countries

The situation of quarterly accounts is quite different among the various European countries. Some differences relate to concepts that are common to the annual accounts, while some others arise from the specific nature of quarterly accounts. Here, we will present some of the main discrepancies in comparison to the Italian situation, how these influence the comparability of the figures among different countries, and how they affect the opportunity and feasibility of a flash estimate. We are particular interested in the time release of quarterly accounts in the different countries and in the revision process.

The main differences among the countries deal with the following subjects:

- the calculation method
- types of series produced
- the timing of publication
- the calendar and the extend of the revisions
- the extend of disaggregation

2.1 The calculation method

Quarterly accounts may be compiled using mainly two approaches : the direct approach and the indirect one. The former provide for a set of surveys so as to have "directly" the quarterly account aggregate; usually the information provided by these surveys. However there are countries like the United Kingdom which do not build quarterly accounts with a separate procedure from the annual ones; that is the quarterly accounts represent the main object of national accounts, while annual data are produced aggregating the quarterly figures.

An alternative approach is the indirect one: the annual figures are disaggregated by means of various techniques, which can use some indicators related to the variable we want to disaggregate; in the last years many enhancements have been made in this field (Di Fonzo, 1987).

In table 1 we can see the methods used in European countries. One should be aware that many countries that actually use direct methods, integrate them with the indirect ones for the items where they lack survey based quarterly information.

Table 1: Quarterly national accounts in the main European countries

Country	Methodology	Starting date of time series
Austria	direct	1970
Denmark	direct	1977
Finland	Mixed	1970
France	Indirect	1970
Germany	direct	1968
Italy	indirect	1970
Netherlands	direct (I/O)	1977
Norway	direct	1978
Portugal	indirect	1977
Spain	indirect	1970
Sweden	direct	1980
United Kingdom	direct	1955

Some countries which base quarterly national accounts on direct methods, use input-output annual tables as a check of the consistency of quarterly figures; a peculiar case is represented by the Netherlands, where a complete system of quarterly input-output tables, both at current and constant prices, is built. Concerning Finland, the reason why we indicated "mixed" in table 1, depends on the fact that in this case is difficult to find a prevailing method.

Considering only EU countries, we note that Belgium, Ireland, Greece and Luxembourg do not produce quarterly accounts. Nevertheless, some of these countries have already accomplished preliminary

studies to compile quarterly accounts, possibly using methods.

The choice between direct and indirect methods is due more to the statistical tradition and to empirical considerations that to theoretical ones. Anyway it has some consequences on flash estimation. In fact indirect methods already provide for relations of quarterly aggregates with some related series, so the problem of flash estimation consists of selecting some of these related series, on the basis of there early availability and/or the possibility to forecast them. In the case of direct methods, one must first find such relationships. In addition, once they have been identified, they are likely ot be less stable than the corresponding relations used with the indirect methods, given the ad-hoc subjective adjustments more frequently made by countries that use direct methods.

2.2 Types of produced series

Quarterly data, unlike annual figures, are affected by the "problem" of seasonality. Given the use of quarterly data for short-term analysis, statistical agencies are often required to produce, in addition to the raw figures, also seasonally adjusted ones. This raises some comparability problems among the adjusted figures of the different countries, due to the use of different seasonal adjustment methods. Moreover, even when the methods are the same, they can use different options (sometimes the possible choice is among lots of options, like in the X11 method) leading to quite different resulting series. Some countries publish also cycle-trend data, i.e. data from which the irregular component has been also removed.

In table 2 we observe the different types of series produced by the European countries and relative seasonal adjustment methods.

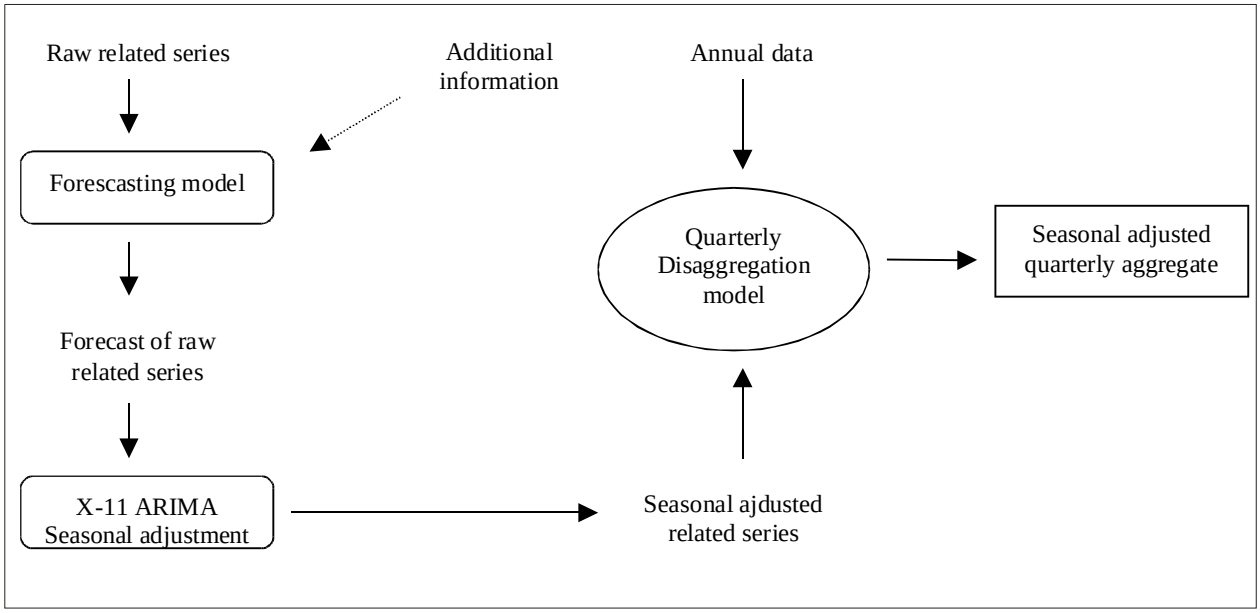
The main objective of flash estimates is to calculate the seasonally adjusted quarterly figures. The problem then arise, whether it is better to seasonally adjust the forecasted raw quarterly figures, or directly the seasonally adjusted ones. This heavily depends on the method already in use in each country. In our case the procedure to produce adjusted data consists of disaggregating annual figures by means of seasonally adjusted related series. So, when the adjusted indicator is available, we could directly estimate the seasonally adjusted quarterly series. Nevertheless in case we have to forecast the related series, it is worth noting that a forecasting model can take efficiently into account the seasonal pattern of series so as to improve the forecast itself. So the "best" procedure in this case can be sketched as in figure 1.

Table 2: Types of quarterly series supplied by the countries and method of seasonal adjustment used

Country	Raw data	Seasonal adjusted data	Cycle-trend data	Method
Austria	O	O*		SEATS
Denmark	O	O		X-11 ARIMA
Finland	O	O*		X-11 ARIMA
France		O		X-11 ARIMA
Germany	O	O*	O	BV, X-11
Italy	O	O		X-11 ARIMA
Netherlands	O	O		X-11
Norway	O	O	O	X-11 ARIMA
Portugal	O	O		
Spain			O	Ad hoc
Sweden	O	O*		X-11 ARIMA
United Kingdom	O	O		X-11 (CSO version)

** Not available for all quarterly series*

Figure 1: "Ideal" process of production of flash estimate of seasonally adjusted quarterly aggregat e

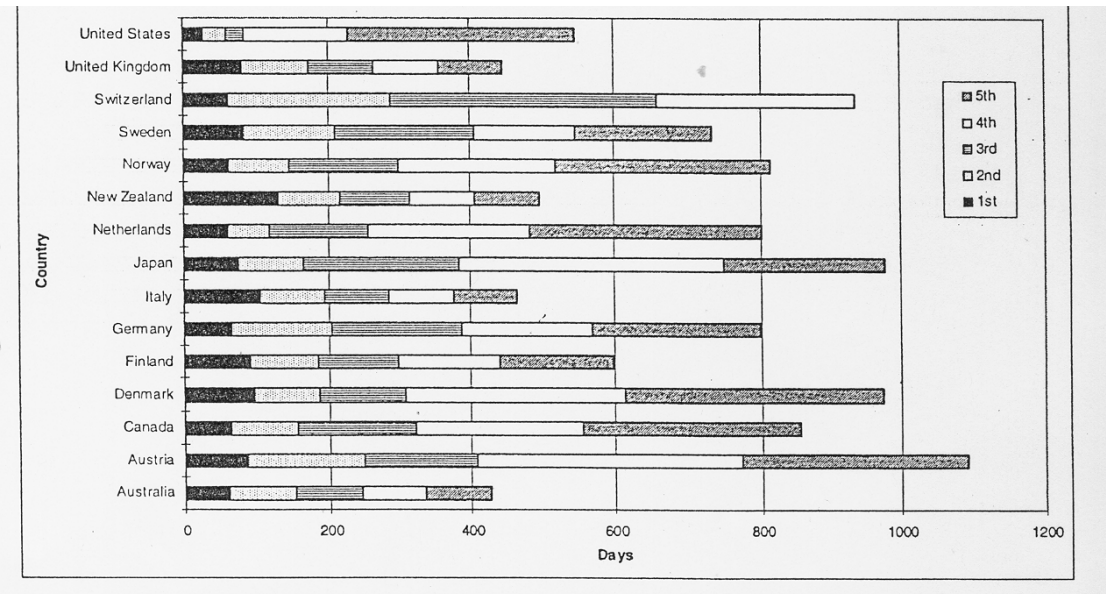


2.3 Timing of publication and process of revision

Also the timing of publication is quite variable among the different countries, due to the different choices between early availability of quarterly data and extent of the revisions. We cannot give her a complete comparative study on the timing of publication and the extend of the revisions in the different countries, but we can show some figures that give an idea of the main differences between the various countries (figure 2).

We can see from figure 2 that some countries, like USA, already provide very early estimates of quarterly accounts. In these cases the revisions are larger and can affect substantially the reliability of the first released figures. This effect can have a larger extend taking into account also the revisions due to the seasonal adjustment. Among European countries, UK provides the earliest estimate of quarterly GDP, three weeks after the end of the reference period (the figures provided are index numbers of output-based real GDP and its main components).

Figure 2: Timeliness of publication of the first five revisions of quarterly accounts



Source: Smith (1995). Reference year: 1991

ISTAT releases its estimates about 95 days after the end of the reference period. Although the long delay of publication, these data are subject to revisions too, due to the incomplete set of information used at that delay and to that delay and to the subsequent revisions of annual figures (Di Fonzo et al.1995).

If a systematic pattern could be found in subsequent revision , they could be efficiently used to improve the first release and consequently the flash estimates of quarterly accounts (Patterson, 1995).

2.4 The extend of disaggregation

While ISTAT produces anticipated quarterly accounts series at disaggregated level, the flash estimate is released at a fairly aggregated level, so that errors deriving from independent sources should compensate to a certain extent, leading to a more accurate estimate.

3 The situation of basic information in the Italian quarterly national accounts

Table 3 shows the current situation of the basic statistical information concerning the Italian quarterly national accounts at different delays after the end of the reference period, expressed as percentage of total information needed to calculate seasonally adjusted value added at constant prices for each branch value added at constant prices for each branch. The total is calculated as a weighted average of the branch values, where each branch weight is the same as the weighted average of the branch values, where each branch weight is the same as the corresponding branch share on total real value added in 1990.

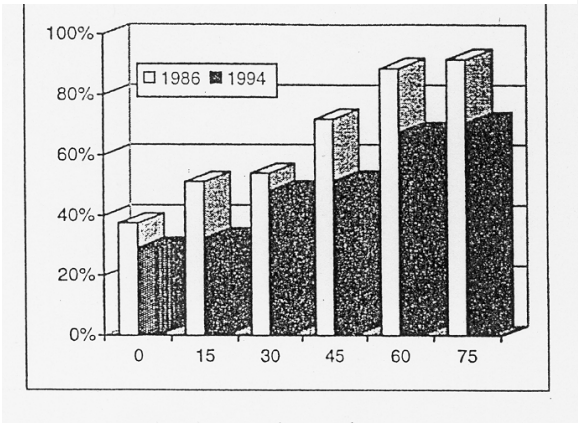
We can see that between 45 and 50 days there is a substantial increase in the available information, mainly due to the fact that there is a complete set of information for manufacturing and energy industry, consisting of industrial production indices.

In addition we can compare the difference in the available information between 1986 and 1994. In figure 3 we observe that there is an apparent worsening in the delay (Giovannini, 1988); while this result is due to the substantial increase in the set of base information used, nevertheless this makes more difficult to obtain a flash estimate of GDP. In figure 3 we have another evidence that the availability of the base information concerning the industry value added shows two “steps”, corresponding to the availability of

Table 3 : Available information at different time delays from the end of the reference period

days branch	0	15	30	45	60	75	90
Agriculture, forestry, fishing	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Industry, strict sense	33.3%	33.3%	66.7%	66.7%	100.0%	100.0%	100.0%
Building and construction	8.3%	8.3%	16.7%	16.7%	25.0%	25.0%	50.0%
Recovered goods and reparation	16.7%	33.3%	50.0%	66.7%	83.3%	100.0%	100.0%
Trade, lodging and catering	16.7%	33.3%	50.0%	66.7%	83.3%	100.0%	100.0%
Inland transport	0.0%	0.0%	33.3%	33.3%	66.7%	66.7%	100.0%
Air and water transport	8.3%	16.7%	16.7%	41.7%	41.7%	58.3%	66.7%
Supporting services to transport	1.4%	2.7%	30.6%	34.7%	62.6%	65.3%	94.6%
Communication	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%
Banking, finance and insurance services	0.0%	0.0%	16.7%	16.7%	33.3%	33.3%	100.0%
Business services	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Building lease	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Health, education and recreational services	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	55.6%
Non-market services	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Total	29.3%	32.5%	48.1%	51.4%	67.6%	70.9%	92.7%

Figure 3: Available information at different time delays in 1986 and 1994



the industrial production index, thirty and sixty days after the reference period. This means that if we were able to produce a reliable estimation of the third month industrial production index, we could have a flash estimate of GDP with 30/45 days delay form the end of the reference period.One way to do that is by means of business surveys results, which are compiled monthly by ISCO. They are quickly available, about 30 days after the end of the reference period. A way to use them to forecast the industrial production index is shown in Gennari (1991).

The strategy of the flash estimate exercise is then to forecast the industry value added by means of business surveys. The same applies to the building sector. Part of market and most of non-market services are forecasted using ARIMA models. This strategy is consistent with the observation that the turning points in GDP are well identified by the turning points of the industrial production index (Parigi and Schiltzer 1995).

4 Anticipating the quarterly national accounts estimates

Since the quarterly national accounts series are estimated by indirect methods, two possible forecasting routes are conceivable:

i) direct forecast of the quarterly series;

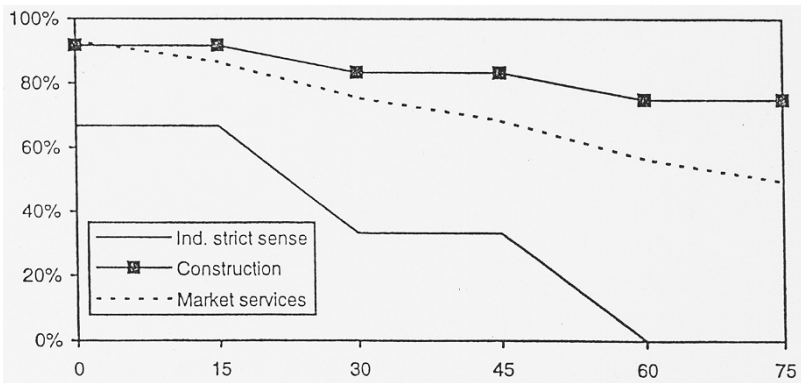
ii) forecast of the indicator series and time disaggregation along the standard procedure.

In general there is no solid theoretical foundation to assert if either (*i*) or (*ii*) is to be preferred. In fact, let $\{y_t\}_1^T$ be the quarterly series, $\{x_t\}_1^T$ the vector of indicator series, $\{u_t\}_1^T$ the quarterly residuals, β the vector of coefficients which relates y_t to x_t, and $\{z_t\}_1^T$ a vector of exogenous variables; further, let $F_T = \left(\{x_t\}_1^T, \{u_t\}_1^T, \{z_t\}_1^{T+1}\right)$ be the information set and $\mathrm{E}(\cdot|F_T)$ the conditional expectation as of time T, then

$$\mathrm{E}(y_{T+1}|F_T) = \beta \mathrm{E}(x_{T+1}|F_T) + \mathrm{E}(u_{T+1}|F_T)$$

so that, if $\mathrm{E}\left(u_{T+1}\middle|\{z_t\}_1^{T+1}\right) = 0$, then strategy (*i*) and strategy (*ii*) are equivalent.

Figure 4: Percentage "ignorance" at different time lags for base statistics infomation used to estimate value added at constant prices



However, the fact that the final estimates are revised quarterly constitutes a good reason why (*ii*) can be preferred to (*i*). Revisions are motivated both by revisions of the seasonally unadjusted indicator series and by seasonal adjustment carried out on the indicators themselves. Strategy (*ii*) has the advantage of exploiting the information conveyed by the revisions of the indicator series. Consistency of the flash estimate with the final quarterly one also suggests the opportunity of forecasting the missing indicators and use the same procedure to derive both the anticipated estimate and the final one. Therefore, we decided to follow this route in the present paper.

A particular situation arises when the indicator series are themselves the result of a preceding disaggregation step. Given the limited number of high-frequency indicators, this is not an infrequent case in the estimation of the Italian quarterly national accounts and in the present study we have been forced to consider these disaggregated series as truly observed ones. However, it should be highlighted that doing so can cause problems in estimation, inference, and forecasting both because of the "generated regressors" problem (Pagan, 1985) and because of noticeable data revisions[2].

Let's consider now strategy (*ii*) in greater detail. Two different settings can be distinguished:

(*ii-a*) a vector of exogenous variables $\{z_t\}_1^{T+1}$ is available which is useful in forecasting x_{T+1};

(*ii-b*) no exogenous information is available to forecast x_{T+1}.

Note that by "exogenous information" we mean essentially a proxy of the variable to be forecasted. Therefore, consistently with current quarterly national accounts estimation, we don't use behavioural models that rely on particular economic theories. In this sense our exercise differs considerably from a typical economic forecasting problem. In all those cases in which (*ii-a*) applies, we estimate a single-equation dynamic models of the kind

$$\alpha_0(L)x_t = \alpha_1(L)z_t + v_t$$

where the $\alpha_i(L)'s$ are finite polynomials in the lag operator L with all the roots not less than 1 and v_t is a white noise process. We identify these models using a general-to-specific approach (Hendry, 1995). In particular the model is estimated by OLS usually using eight lags for the lag polynomials $\alpha_i(L)'s$. Then the usual misspecification tests are carried out over each reduction of the model. In particular we test for residual serial correlation using the LM test (in the F form; see Breusch and Pagan, 1980 and Harvey, 1981) over various lags (usually from 4 to 12). We then check for ARCH structure in the residuals (Engle, 1982) and for residuals normality (Jarque and Bera, 1980). The null of linearity is tested against polynomial alternatives, while stability is checked using recursive Chow tests. In general the selected models satisfy all these tests at the usual confidence levels. Those cases in which normality, stability, and sometimes autocorrelation statistics are significant are considered as possibly affected by the presence of outliers or structural breaks. Impulse or step dummy variables are then included according to the indications arising from the recursive Chow tests and the graphical analysis of the residuals. Finally, in order to test the forecasting ability of the final (restricted) model, the forecasting performance of this is checked against the one-step ahead forecasts of the unrestricted model over the last eight quarters. These comparisons generally highlight a clear superiority of the restricted model.

Up to this point we did not address two major issues. The first is related to the choice of using single-equation instead of multi-equation models; the second is concerned with the possible presence of unit roots in the series under examination. As is well known, weak exogeneity in the sense of Engle *et al*. (1983) is the necessary condition to have correct inference in conditional (single-equation) models. In the present case, parameter stability is regarded as indirect evidence of weak exogeneity (Favero and Hendry, 1992). However, it should be stressed that the

2 Note that this is true also for all the models which use disaggregated time series. However, this f act does not appear to have been fully realized by models propietors .

variable-indicator setting can induce nontrivial complications as far as exogeneity is concerned (Lupi, 1995). For what concerns the unit root problem, various aspects should be considered. First, unrestricted (in the sense of not having imposed unit roots) models do not necessarily forecast worse than restricted ones in the presence of $I(1)$ variables (see e.g. Clements and Hendry, 1994). Secondly, our dynamic equation can be interpreted as a single row of a VAR model and the parameters estimates are consistent (Sims *et al.*, 1990)[3]. Finally, it is well known that standard unit roots tests have low power in the presence of small temporal samples, typical of the applications considered here [4].

When situation (*ii-b*) is relevant, there is not much to choose about the methods to be used and the standard ARIMA (*p,d,q*) model

$$\phi(L)\delta(L)x_t = \mu + \theta(L)u_t$$

is utilised. In this model $\phi(L)$ and $\theta(L)$ are polynomials in L of order p and q respectively, having all the roots outside the unit circle; $\delta(L)$ is polynomial of order d in L with all the roots equal to one; μ is a constant term.

The forecasting performance of ARIMA models depends crucially not only on the correct identification of the polynomials orders (*p,d,q*) but also on the preliminary correction of the outliers potentially present in the series. Correct identification and correction of the outliers would presume the use of a vast number of statistical and graphical tools and the use of qualitative information on the single economic phenomena. However, the flash estimate of the quarterly national accounts needs a large amount of forecasts to be obtained in a relatively short period of time so that this route would be impractical. In order to render outlier identification as much automatic as possible we decided to resort to the software TRAMO (Gomez and Maravall, 1994b). This software allows in fact the automatic selection of the orders p, d, q of the ARIMA model and the correction of four types of outliers. More specifically, additive and innovative outliers are considered along with transitory changes and level shifts. Loosely speaking the automatic procedure operates in three phases. In the first, the order d necessary to make the series stationary is selected: this is done avoiding over differencing, i.e. introducing unit roots in the MA polynomial $\theta(L)$. In the second phase the orders p and q are chosen by means of the Bayes Information Criterion (BIC). Finally, the kind and position of the outliers are identified using LM tests for each observation. When outliers are found the procedure starts again from the first step. Further details can be found in Gomez and Maravall (1994a, 1994b).

5 Some results

In order to motivate further our choice of forecasting the indicators instead of the final estimate, in table 5 we report an example based on the forecasts obtained by using a simple dynamic model for the value added of the building sector. The model exploits the information deriving from the ISCO monthly business survey. In particular, value added is forecasted using the level of the orders of the building sector. From the statistics reported in table 4 it is not possible to detect any clear sign of misspecification.

Table 4: Building Sector Model

Diagnostic Test	value	p-value
R2	0.9614	
F(4,48)	298.64	0.0000
DW	2.3000	
AR 1-4 F(4,44)	1.8237	0.1413
ARCH 4 F(4,40)	0.0211	0.9991
Normality Chi2(2)	1.6415	0.4401
Xi2 F(7,40)	0.8054	0.5878
RESET F(1,47)	0.0421	0.8384

3 However, we are aware of the fact that in the presence of unit roots nontrivial complications for i nference may arise.

4 The typical sample in this study is 1980q1-1994q4

The one-step ahead forecasting performance of the model, taking the value added series as fixed, is good (see table 5). However, the real value added series, as well as all the other quarterly national accounts series, is revised quarterly and is not fixed [5].

When the true series is used for estimating and forecasting the model, much worse predictions arise (table 5). This fact clearly reinforce our decision of forecasting the indicator instead of the final aggregate.

In table 6 the model diagnostics relative to the indicator series of the value added of the building sector are reported. Again, these tests do not suggest any misspecification of the model. Its forecasting performance is satisfactory apart, perhaps, the result associated to the second quarter 1993 (see table 7).

Similar situations arise in the other cases for which exogenous information is available to forecast the indicator variables, such as manufacturing, energy and gas, agriculture, and trade. These sectors in 1994

Table 5: One step ahead forecasts of the building sector value added changes

Date	Var% [a]	Var% [b]	Var% [c]	Var% [d]
93.1	-2.1	-2.1	-0.2	-1.7
93.2	-1.7	-1.7	-0.3	-1.2
93.3	-1.3	-0.3	-0.5	-2.5
93.4	-1.2	-1.2	-0.4	-2.4
94.1	-0.9	-1.1	-0.7	-1.7
94.2	-1.2	-1.9	-1.0	-1.2
94.3	-1.3	-2.2	-1.2	-0.4
94.4	-0.8	-0.5	-0.8	-0.3

Notes:

(a) as of at 1994.4

(b) one step ahead forecast treating the series as fixed as of at 1994.4

(c) as of at the reference date

(d) one step ahead forecast using the series as of at the reference date

Table 6: Building Sector Indicator Model

Diagnostic Test	value	p-value
R2	0.9477	
F(8,46)	104.24	0.0000
DW	2.3400	
AR 1-4 F(4,42)	1.1908	0.3289
ARCH 4 F(4,38)	0.1107	0.9780
Normality Chi2 (2)	4.1824	0.1235
Xi2	0.5163	0.9114
RESET	0.7536	0.3899

represented about 67% of the market goods and services value added. Where no exogenous information could be used in forecasting the indicator series, we used univariate forecasts of the indicators themselves. On this basis, in order to produce an anticipated estimate of the supply side growth of the economy, we decided to adopt the strategy of using the indicator series augmented by the one-step ahead predictions (where necessary) in the disaggregation procedure actually used to produce the final estimate.

Unfortunately, there are cases in which a disaggregated variable is used as an indicator of a further disaggregation step. This happens in particular

Table 7: One-step ahead forecasts of the building sector indicator changes

Date	True	Forecast
93.1	-2.1	-3.6
93.2	-1.8	-0.2
93.3	-1.4	-2.9
93.4	-1.4	-1.2
94.1	-1.2	-1.7
94.2	-1.4	-1.5
94.3	-1.5	-0.8
94.4	-0.5	-0.7

5 The last two years of the quarterly national accounts are revised quarterly. In addition, when the fourth quarter of a specific year is estimated, quarterly national accounts series are revised since five years before. A study of the revisions of the Italian quarterly national accounts is in Di Fonzo *et al.* (1995)

Table 8: Value added % changes of quarterly national accounts and flash estimate 1995.1

Sector	QNA	Flash
Agriculture, forestry and fishering	6.0	6.0
Manufacturing	3.0	3.2
Energy and gas	1.6	1.6
Building	-0.7	-0.6
Market Services	0.6	0.5
Market Goods & Services	1.5	1.5

in some cases related to the service sector, where there are more problems in finding relevant exogenous information. In these cases, at least until we have not completed the flash estimates for what concerns the demand side, we are forced to treat the disaggregated indicator as a truly observed one, and the *caveats* exposed before apply.

As a final example we provide the results relative to the flash estimate of sectoral value added growth rates for the first quarter 1995. These are reported in table 8. At the time in which the experiment was performed, the available indicators were essentially the industrial production index and most of the price indices along with some minor indicators. On the contrary, service and building sectors indicators had to be forecasted together with the remaining price indices. A particular case is represented by the agricultural sector, where an indicator is usually readily available but is revised very frequently. The information set just described is typically available 60 days after the end of the reference quarter. This would allow us to produce a first estimate of the market economy value added growth rate 35 days in advance with respect to the publication of the complete set of the quarterly national accounts.

6 Conclusions

This paper describes the methodology adopted by ISTAT for the flash estimate of the Italian real GDP. As it stands, the estimate is derived mainly from the supply side, given that the reliable information related to production is more readily available. The study highlights the opportunity of using forecasts of the indicator series in the standard disaggregation procedures instead of producing direct forecasts of sectoral value added series. The forecasts of the indicator series are obtained by utilising dynamic single-equation models, which exploit the information embodied in the business survey data provided by ISCO. When there is no genuine exogenous information, we rely on ARIMA-based procedures with outlier correction. Following this approach, we are able to provide an early estimate of constant prices GDP growth rates 60 days after the reference quarter, anticipating the release of the ordinary estimate of about 35 days. The reported results suggest that this methodology is able to produce accurate anticipations. This is encouraging towards the direction of implementing a flash estimate of the whole resources and uses account.

References

Breusch, T.S., and Pagan, A.R. (1980) : "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics", *Review of Economic Studies* , 47, 239-253.

Chow, G., and Lin, A.L. (1971) : "Best Linear Unbiased Interpolation Distribution and Extrapolation of Time Series by Related Series", *Review of Economics and Statistics* , 53, 372-375.

Clements, M., and Hendry, D.F. (1994) : "Towards a Theory of Economic Forecasting", *Economic Journal*.

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

Di Fonzo, T., Pisani, S., and Savio, G. (1995) : "Revisions to Italian Quarterly National Accounts", in this volume.

Engle, R.F. (1982): "Autoregressive Conditional Heteroscedasticity, with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1008.

Engle, R.F., Hendry, D.F., and Richard J-F. (1983) : "Exogeneity", *Econometrica*, 51, 277-304.

Favero, C., and Hendry, D.F., (1992) : "Testing the Lucas Critique: A Review", *Econometric Reviews*, 11, 256-306.

Gennari, P. (1991): "L'uso delle Indagini Congiunturali ISCO per la Previsione degli Indici della Produzione Industriale", *Rassegna dei Lavori dell'ISCO* , n.13.

Giovannini, E. (1988): "A Methodology for an Early Estimate of Quarterly National Accounts", *Economia Internazionale* , 41, 197-215.

Gomez, V., and Maravall, A. (1994a) : "Estimation, Prediction, and Interpolation for Nonstationary Series with the Kalman Filter", *Journal of the American Statistical Association* , 89, 611-624.

Gomez, V., and Maravall, A. (1994b) : "Program TRAMO: Time Series Regression with ARIMA Noise, Missing Observations, and Outliers", European University Institute, Florence, mimeo.

Harvey, A.C. (1981): *The Econometric Analysis of Time Series*. London: Philip Allan

Hendry, D.F. (1995): *Dynamic Econometrics* . Oxford: Oxford University Press.

Istat (1992): "I conti economici trimestrali con base 1980", *Note e Relazioni*, 1992, n.1.

Jarque, C.M., and Bera, A.K. (1980) : "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", *Economics Letters* , 6, 255-259.

Lupi, C. (1995): "Stationary Measurement Errors Among Cointegrating Variables", ISTAT, mimeo.

Lupi, C., and Parigi, G. (1995): "Time Series Disaggregation of Economic Variables: Some Econometric Issues", this volume.

Pagan, A.R. (1985): "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review* , 25, 221-247.

Parigi, G., and Schiltzer, G. (1995) : "Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicators", *Journal of Forecasting*, 14, 117-141.

Patterson, K.D. (1995) : "An Integrated Model of the Data Measurement and Data Generation Process with an Application to Consumers' Expenditure", *Economic Journal* , 105, 54-76.

Sims, C.A., Stock, J.H., and Watson, M.W. (1990) : "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica* , 58, 113-144.

Smith, P. (1995): "The Timeliness of Quarterly Income and Expenditure Accounts: An International Comparison", Joint OECD-Hungarian Central Statistical Office Workshop on quarterly national accounts, Budapest 21-24 February

Using data of qualitative business surveys for the estimation of Quarterly National Accounts

Carlos Coimbra [1]
Gabinete de estudos económicos do INE & Inst. superior ciências do trabalho e da empresa

The data from qualitative business survey is usually available first than the data from conventional quantitative statistical sources. the qualitative data may be useful for the quarterly accounts estimation. This problem of how to use the data from qualitative surveys for the estimation of quarterly national accounts is very relevant for the portuguese case. Often, we do not have all the intra-annual quantitative data usually available in the others countries of the European Union or we have it with a relative delay. In this paper we intend to present one of the research directions we are developing about this problem, which does not follow a kind of Carlson & Parkin approach. In general terms, we study the possibility of obtaining quantitative indicators, using an econometric friendly method, from qualitative results of Business Surveys in a way that preserves all the aggregated information obtained by them.

1 Introduction

The data from qualitative business survey is usually available before data from conventional quantitative statistical sources. Some of this qualitative data may be used to produce quantitative related indicators for the estimation of quarterly accounts. The usual form of quantification, the balance, is rather unsatisfactory. However, its simplicity probably explains why it is so popular. Maybe it is an impossible task to find another form of quantification maintaining that characteristic of simplicity and that, at the same time, does not have the problem of insufficiency affecting the balance quantification.

In this paper we develop another way of quantification which does not have the same degree of simplicity of the Balance, but that intends to correct its insufficiency problem.

In particular, in this paper we are concerned with how to use to data from qualitative surveys for the estimation of quarterly national accounts, or more precisely, for the first (fast) estimations of quarterly accounts. This problem is very relevant for the portuguese case because we do not have all the intra-annual quantitative data usually available in the others countries of the European Union or we have quantitative information with a considerable delay. In this paper we intend to present one of the research directions we are developing, which is not in the line of the Carlson & Parkin approach.

1 The author wish to thank J. Santos Silva for his assistance in developing the final version of the econometric model presented in this paper.

In the next section we discuss an alternative way of quantification of the business surveys information. In the last section we provide an illustration for the case of portuguese exports.

2 One approach less naive than the Balance

In general, the qualitative business surveys pose questions about the signal of the variation of some important variables for short term economic analysis. As result, the responses are aggregated in three complementary percentages. Consider P_1 as the percentage of positive responses, P_2 the percentage of neutral responses, i.e., the percentage of responses that report no variation, and P_3 the percentage of negative responses.

The variation V of the variable questioned can be seen as a weight average of the average variations inside the three groups of answers. For example, if the question is about the production evolution, V corresponds to weigh average of production variation rate of the concerned economic sector, being V_1, the average variation rate inside the positive responders (firms), V_2 the average variation rate inside the "neutral" responders, V_3 the average variation inside the negative responders (firms), in a given month or quarter, we can write the following identity equation:

$$V = V_1 \cdot P_1 + V_2 \cdot P_2 + V_3 \cdot P_3 \qquad (1)$$

Supposing that, like the percentages, V is observable, it is impossible to apply OLS to estimate the implicit average variations V_1,V_2 and V_3 because of the regressors perfect collinearity.

One easy way to surpass this problem, without losing information, is transforming (1) into:

$$V = \beta_0 + \beta_1 S + \beta_2 P_2 \qquad (2)$$

where $S = P_1 - P_3$, $\beta_0 = 1/2(V_1 + V_3)$, $\beta_1 = 1/2(V_1 - V_3)$ and $\beta_2 = V_2 - 1/2(V_1 + V_3)$.

If P_2 is not zero, the use of the Balance S will not be misleading only if $V_2 = 1/2(V_1 + V_3)$. In fact, even if $V_2 = 0$, P_2 is relevant to define V, unless V_1 and V_3 are symmetric.

So, in order to monitor V, one should pay attention both to the balance and to the P_2 answers. The question is how to combine this two indicators in a proper manner.

An obvious way, is to use equation (2) to estimate a composite indicator, taking as proxy regressors the surveys extrapolated proportions, which has not the problem of perfect collinearity present in (1).

However, the β coefficients are linear combinations of variation rates implicit in each group of responders. Those rates will not be constant. In fact, they probably are affected by the business cycle. Consequently, the same should apply to the β coefficients. But, also the P_2 answers and S are affected by business cycle. Thus, one possible way to take into accounts the variability with time of the β is to consider, for a given month or quarter, the three following stocastic equations:

$$\beta_1 = \alpha_{0,1} + \alpha_{1,i} S + \alpha_{2,i} P_2 + \varepsilon_i, \quad (i = 0,1,2) \qquad (3)$$

where, in each equation there is residual term ε_1.

After substituting the β_1 in (2) by (3) we obtain:

$$\begin{aligned} V = {} & \alpha_{0,0} + (\alpha_{1,0} + \alpha_{0,1}) S + (\alpha_{2,0} + \alpha_{0,2}) P_2 \\ & + \alpha_{1,1} S^2 + (\alpha_{2,1} + \alpha_{1,2}) S P_2 + \alpha_{2,2} P_2^2 + v \end{aligned} \qquad (4)$$

where $v = \varepsilon_0 + \varepsilon_1 S + \varepsilon_2 P_2$

Even if it was possible to run a regression in (4) there would be an identification problem on some of the α parameters.

Fortunately, this problem is not relevant in practical terms, since the objective is not to estimate the α parameters but simply to approximate V as a function of the available information. Moreover, we do not observe the true proportions, i.e., we ignore the effective values of S and P_2. We only obtain from the qualitative surveys proxy proportions, the extrapolated proportions form their reference samples. That is, we only obtain $\tilde{S}$ and $\tilde{P}_2$.

So we can rewrite (4) as:

$$V = \gamma_0 + \gamma_1 \tilde{S} + \gamma_2 \tilde{P}_2 + \gamma_3 \tilde{S}^2 + \gamma_4 \tilde{S}\tilde{P}_2 + \gamma_5 \tilde{P}_2^2 + \eta \qquad (5)$$

Supposing η an homoscedastic and not serial correlated residual variable[2], one may apply OLS to estimate V. This estimate can be seen as a quantitative indicator potentially useful for quarterly accounts estimation.

Next, taking as example the portuguese exports, the first we will show that one should not ignore the P_2 answers and, after, we use the equation (5) to estimate the exports with satisfactory results.

3 An empirical illustration

In the manufacturing industry monthly business survey there is question about the new orders from abroad variation. We have the extrapolated proportions of answers since 1987 to that question. Unfortunately we only have the exports, at constant prices, in a quarterly frequency. For that reason, we have aggregated the monthly data into quarterly data. After, studying the appropriated lags we have estimated three models using TSP package.

First we have estimated the following model by OLS;

$$V_t = \delta_0 + \delta_1 \tilde{S}_t + \mu_t \qquad (6)$$

The estimates for this model are (t-ratios in brackets):

$$\hat{\delta}_0 = \underset{(73.0)}{1.11}; \quad \hat{\delta}_1 = \underset{(4.7)}{0.15}$$

$\bar{R}^2 = 0.46$;
White's heteroscedasticity test=0.85;
DW=1.08.

To assess the importance of neutral answers proportions, we estimated the new model:

$$V_t = \delta_0 + \delta_1 \tilde{S}_t + \delta_2 \tilde{P}_{2t}^2 + \varepsilon_t \qquad (7)$$

The estimates for this model are (t-ratios in brackets):

$$\hat{\delta}_0 = \underset{(2.4)}{0.61}; \quad \hat{\delta}_1 = \underset{(-0.4)}{-0.04}; \quad \hat{\delta}_2 = \underset{(1.9)}{0.35}.$$

$\bar{R}^2 = 0.52$;
White's heteroscedasticity test=5.5;
DW=1.07.

Even if it is not clear that we do not have a serial correlation problem, the above results suggest that P_2 is a variable that should be considered. However this model is not satisfactory because there are some evidence of collinearity between the regressors and we are supposing the parameters constant. So we have estimated the model (5), and obtained the following estimates:

$$\hat{\gamma}_0 = \underset{(-2.0)}{-10.4}; \quad \hat{\gamma}_1 = \underset{(-2.2)}{-8.0}; \quad \hat{\gamma}_2 = \underset{(2.1)}{15.5};$$
$$\hat{\gamma}_3 = \underset{(-2.1)}{-2.9}; \quad \hat{\gamma}_4 = \underset{(2.2)}{11.0}; \quad \hat{\gamma}_5 = \underset{(-2.0)}{-10.5}$$

$\bar{R}^2 = 0.55$;
White's heteroscedasticity test=13.5;
DW=1.40.

Apparently, this model is a step in the right direction. The non-linearities seem to be rather significant[3], suggesting that one should account for the time variation in the coefficients of equation (2).

Naturally, the approach adopted here to model the coefficients variation is just one among other alternatives. The main points are that one should not waste the neutral answers proportion to get proper related indicators to the quarterly accounts estimation, and that equation (2) may be an interesting starting point.

---

2 These are not an indispensable hypotheses. In fact, they should be tested case by case and if they do not hold in a particular case, then the conventional alternatives should be applied.

3 Due to the possibility of serial correlated errors, we have estimated a covariance matrix that is r obust to this type of problem. However the results obtained are qualitatively equivalent.

References

Batchelor, R.A., 1982, "Expectations output and inflation: The European Experience", *European Economic Review*, 17.

Batchelor, R.A. and A.B. Orr, 1987, "Inflation Expectations Revisited", *Economica*, 55.

Carlson, J.A. and M. Parkin, 1975, "Inflation Expectations", *Economica*, 42

Dasgupta, S. and K. Lahiri, 1992, "A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data", *Journal of Business & Economic Statistics*, vol 10.

Person, B.A., 1977, "Balance Series as true and false indicators of qualitative change: Theory and Practice", CIRET, 13th Conference.

Person, B.A., 1985, "On the relationship between qualitative and quantitative indicators: Outline of an explanatory model"" CIRET, 17th Conference.

Soares, M.C., 1993, "Indicadores quantitativos a partir de respostas qualitativas: evolução das exportações em volume", *Boletim Trimestral do Banco de Portugal*, first quarter of 1993

Estimating Quarterly Regional Accounts for southern Italy

Margherita CARLUCCI and Roberto ZELLI [1]
University of Rome "La Sapienza"

1 Introduction

The availability of sub-national data in EU Countries is crucial to analyse regional disparities and to identify regional-specific policies. This is especially true in Italy where the South has experienced different patterns of growth with respect to the Northern part of the Country. National accounting data issued by the National Statistical Institute (Istat) at regional level concern only annual data, hence no attempt has been made so far to capture the conjunctural differences between North and South Italy.

To study Italian dualism at the edge of 1992 Europe unification, we made up a first attempt of disaggregation of regional annual data by deriving quarterly figures of the domestic product account and of value added by kind of activity of the Northern and Southern economies. As far as we know, no other regional disaggregation of the kind has been performed by either institutional or private researchers, at least for Italy. Therefore, this attempt may give helpful suggestions for users who are interested in building quarterly regional data for conjunctural purposes. Moreover, since the official regional data are released with considerable delay, the construction of a quarterly model allows one to provide forecasts of annual data.

The paper is set out as follows. In the second section a brief review of the methodology used is given. The third section concerns the choice of quarterly indicators available at sub-national level and the estimation procedure to obtain quarterly data. The fourth section describes the results obtained both in terms of historical quarterly profile and in terms of extrapolation of annual data.

2 Methodology

The methodology used is based on Barbone, Bodo and Visco (1981, BBV from now on) who modified the Chow and Lin (1971) method. This method is consistent with the one currently used by Istat to derive quarterly data of national accounts (Istat, 1992).

As it is well known, this method is based on a linear relation between unknown quarterly data of interest and known quarterly indicators. While other "smoothing" methods also based on indicators imply two subsequent steps, first the estimate of quarterly values and then their correction according to the annual constraint, this one automatically takes into account the annual constraint in the estimation procedure.

1 This paper is the outcome of joint work of the Authors; however, Roberto Zelli wrote sections 1 and 3, Margherita Carlucci wrote sections 2 and 4.

Denoting with Y the unknown $(4n\times 1)$ vector of observations, with X the $(4n\times k)$ matrix of related indicators and with β the $(k\times 1)$ vector of parameters, the following relation applies:

$$Y = X\beta + u$$

where the $(4n\times 1)$ random vector u has mean zero and covariance matrix V.

Defining the aggregation matrix B, we obtain the estimable relation on the annual data y:

$$y = BY = BX\beta + Bu$$

In this case the Best Linear Unbiased Estimator consistent with the aggregate time series is given by:

$$Y^{*} = Xb + VB'\left(BVB'\right)^{-1}\left(y - BXb\right)$$

where

$$b = \left(X'B'\left(BVB'\right)^{-1}BX\right)^{-1}X'B'\left(BVB'\right)^{-1}y.$$

Therefore, the estimated Y* are obtained as sum of a systematic component, deriving from the estimated regression, and an "error" one, which for each year distributes the bias between the sum of the regression quarterly data and the observed annual aggregate among the four estimated values according to the values of V. In other terms, higher variances suggest less reliability of the regression quarterly values, thus leading to an higher correction.

In most cases, the relationship between short-run movements in Y and $X\beta$ is fairly stable, but the levels of Y and $X\beta$ may vary over time: the residuals will then exhibit serial correlation (Litterman, 1983). Thus, random errors are supposed to be generated by a first-order Markov process (Chow and Lin, 1971), a random walk (Fernandez, 1981), an ARIMA (1,1,0) process (Litterman, 1983), a generalised ARIMA(p,d,q) process (Stram and Wei, 1986). Here, to avoid step discontinuities of quarterly estimates between years, the first order autoregressive process is adopted, with:

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \approx i.i.d.$$

Since ρ is generally unknown, it has to be estimated. Chow and Lin (1971) suggested a maximum likelihood procedure, whereas BBV adopted a feasible generalised least squares procedure. It consists of an iterative scanning on the values of , ranging in the interval (-1, 1), to build punctually the likelihood function. Here, we started with a step size of .01; in correspondence of the maximum likelihood value observed, we squeezed the step size to .001. The BBV procedure is shown to be preferable when the maximum likelihood quarterly time series are, for instance, characterised by dramatic fluctuations of uncertain economic meaning.

Moreover, since the assumption of autocorrelation of order one, seasonally adjusted quarterly data have been estimated.

"Optimal" methods, such as the Chow-Lin one, permit the extrapolation of the quarterly series outside the sample period. Extrapolated estimates are considered temporary data, to be revised when the annual constraint is available. We extend the extrapolation period to no more than 4 quarters, to avoid the summing up of prevision error.

Our series must satisfy two kinds of constraint: a temporal identity (adding up of quarterly values to the annual one) and a contemporaneous one (between resources and use aggregates). The balancing of aggregates was then achieved with a double-proportional algorithm.

3 Available information at sub national level: sources and timeliness

Concrete application have shown that when annual profiles of the indicators and of the series that have to be disaggregated are similar, then quarterly estimates obtained by alternative methods tend to be very close.

Therefore the main effort of the research has been the identification of reliable indicators. This problem is particularly evident when the purpose is temporal disaggregation of regional series since the availability of quarterly variables is in this case extremely poor. Moreover, "official" annual data that represent the constraint of the model are released with considerable delay, whereas first unofficial annual data of Southern

economy are given in July of the next year. Since the availability of quarterly indicators anticipates the estimation of annual data - the indicators are available within 5 months - therefore modelling quarterly data allows one to obtain provisional estimates of annual figures by the aggregation of quarterly figures.

At sub national level, it is possible to obtain information from the following Istat surveys:

- monthly survey on household consumption;
- monthly survey on consumer prices;
- quarterly survey on labour force.

Additional sources of information for Southern economy is the monthly survey on industrial electricity consumption led by national electric authority (Enel) and the monthly survey on entrepreneurs' expectation on economic activity conducted by the institute for short-term economic analysis (ISCO).

It should be noted that each survey has, at regional level, a different degree of reliability due to the different coverage of regional samples.

Unpublished elaboration of regional consumption are given for the following groups of goods and services: food, drink and tobacco; clothing, housing; fuel; transport and communication; other goods and services.

Survey on consumer prices gives provincial data. Therefore they have to be aggregated in order to obtain regional figures.

Labour force survey provides, at regional level, figures on workforce in employment by kind of activity (agriculture, industry, private and public services) and by condition (employees in employment and self-employed), and on unemployment, divided in first employment claimants and unemployed.

Choice of indicators is also linked to the timeliness with which they are released. At this regard, it is possible to distinguish three orders of delay:

- delay on annual data: 7 months;
- delay on Istat and Enel indicators: between 3 and 7 months;
- delay on qualitative Isco indicators: 1 month.

The structure of delays described above suggested an outline for the procedure of estimation of quarterly accounts in three steps:

- temporary estimation of a restricted number of aggregates based on Isco indicators;
- first revision of the estimates when all the indicators are available;
- final revision when the annual data is available

Utilisation of information gathered by Isco leads to some interpretative difficulties due to qualitative nature of the data. Expectation on variables as production, inventories, orders, prices and so on is given according to modalities as "normal", "more than normal", "less than normal". The procedure of quantification is the percentage of answers relating to each modality and eventually the balance between percentage of answers "more than normal" and answers "less than normal".

The problem is that each entrepreneur could give a different meaning to these modalities. However Isco has implemented a procedure in order to make the results coherent taking into account the long run average of the balance.

Isco results can be used in two different ways: as leading indicators of independent variables of the model or directly as indicators of the model itself. Since the coverage of the sample to Southern regions is reliable enough only starting from 1986, the second way of utilisation has been preferred [2].

The estimation spans the period 1980-1991. In the following outline (Table 1) the quarterly aggregates estimated, their main indicators, and the average delay these indicators are released with, are presented.

Given the hypothesis of the model adopted autoregressive process of order one in the error term indicators have been seasonally adjusted. For monthly indicators, we applied the X-11 method: quarterly data are obtained by sum of seasonally adjusted monthly ones. Household expenditure data have first been expressed in 5-terms moving averages, then interpolated with a third degree polynomial form ("spline" method, see Kendall and Stuart, 1976).

Table 1: Outline of quarterly aggregates estimated, main indicators and average delay

Aggregates	Indicators	Delay
Value added at factor cost:		
• Agriculture	• Workforce in employment National indicator	5 months 4 months
• Industry	• Industrial electricity consumption	3 months
• Private Services	• Workforce in employment National indicator	5 months 4 months
• Public Services	• Workforce in employment National indicator	5 months 4 months
GDP at market prices	Values added by kind of activity	5 months
Net Imports	GDP at market prices	5 months
Consumers' expenditure	Household consumption survey	7 months
General government final consumption	Workforce in employment	5 months
Gross domestic fixed capital formation	National indicator	4 months
Consumers' expenditure deflator	Consumer prices survey	6 months
GDP deflator	National GDP deflator	4 months

4 Short-term behaviour of Southern Economy

The availability of quarterly series of main regional economic aggregates allowed us to analyse the conjunctural differences between North and South Italy immediately before the European unification. A brief sketch of the main differences is reported in the following[3].

Due to different structural composition of the two economies, with an higher incidence of non-competitive sectors (agriculture, public services) in Southern one, international cyclic fluctuations seem to arrive in South Italy "smoothed" and with a certain delay, with some cases of counter-cycle in North and South (see Table 2).

At the beginning of the Eighties, Southern economy already presented negative growth rate in real terms; thus, neither it suffered the slowing down of the second part of the 1980, nor it fully benefited of the recovery of 1981.

After the slump of the third quarter of 1982, the Southern economy seemed to grow faster than the Northern one in the following three years. But, with the spreading of more favourable international conditions, the stronger economy of industrialised North Italy started an accelerating growth path, thus leading to an increasing gap between the two areas until the second half of 1989.

While Northern economy experienced a stable GDP growth rate pattern during this period, quarterly

2 On the basis of a sample of 74 monthly observations, a linear relation between electricity consumpt ion and Isco indicator has been estimated, where error term is assumed to follow a seasonal moving average proce ss. Results seem satisfying enough both in term of fitting and forcasting.

However the direct utilization of ISCO subjective indicators seems more interesting on perspective and will be adopted as soon as the time series will be long enough. As a matter of fact, utilization of industr ial electricity consumption can not be extended to further regional disaggregation since data are collected at "compartimental" level, according to the sources of electricity supply, which is not coherent with regional disaggregation.

3 From a different point of view, a discussion of more detailed economic results was first presented in: Centro Europa Ricerche, "Ciclo economico ed attivita' creditizia nel Mezzogiorno" ("Business cycle and ban king activity in South Italy"), *Rapporto n.6*, 1992.

variations in Southern GDP growth showed wider fluctuations and shorter persistence of positive swings.

Growth disparities persisted during the whole turndown period of the end of the decade; in 1991, due to an exceptional crop and a better performance of Southern industry than Northern one (the last one more affected by a fall of international investment demand), for the first time growth rates are higher in the South.

It is interesting to note that our model allowed us to forecast this unexpected gain in 1991 Southern GDP as compared to Northern one, which was confirmed only 3 months later by the annual data.

Consumption cycles are quite similar in the two areas even though Southern one seemed less sensitive to international cycle fluctuations and the same behaviour characterise investment cycle.

This piece of work should be interpreted as a first attempt in deriving quarterly regional accounts. However, results obtained seem interesting enough for future investigations.

Table 1: Main economic aggregates of South Italy Quarterly series 1980-1991

(billions of lire 1985)

Quarters	GDP	Net imports	Gross fixed Investments	Household Consumption	Collective Consumption	Changes in inventories
80.1	46.955	8.766	12.434	31.730	10.072	1.484
80.2	46.411	8.462	12.383	32.130	10.352	8
80.3	46.417	8.502	12.294	32.207	10.360	58
80.4	46.681	8.744	12.256	32.296	10.476	398
81.1	46.363	8.783	12.253	32.312	10.613	-32
81.2	46.281	8.883	12.108	32.482	10.535	39
81.3	46.285	8.993	11.911	32.783	10.573	11
81.4	46.602	9.246	11.737	32.966	10.643	502
82.1	47.073	9.580	11.682	32.987	10.658	1.326
82.2	46.895	9.520	11.629	33.083	10.882	819
82.3	46.358	9.219	11.662	33.417	10.898	-401
82.4	46.475	9.184	11.840	33.620	10.960	-760
83.1	47.137	9.345	12.032	33.632	11.245	-428
83.2	47.582	9.439	12.272	33.740	11.224	-215
83.3	48.505	9.796	12.477	34.036	11.368	420
83.4	49.027	9.952	12.636	34.344	11.312	688
84.1	49.159	9.893	12.906	34.640	11.507	-1
84.2	49.573	10.003	12.939	34.823	11.251	563
84.3	50.012	10.134	12.962	34.854	11.593	738
84.4	50.129	10.094	12.914	35.047	11.941	322
85.1	49.964	9.888	12.557	35.458	11.836	0
85.2	50.832	10.302	12.553	35.996	11.770	815
85.3	51.509	10.714	12.592	36.252	12.286	1.092
85.4	50.965	10.613	12.461	37.019	12.206	-107
86.1	50.746	10.814	12.425	37.198	12.113	-176
86.2	51.494	11.387	12.607	37.348	12.104	822
86.3	51.806	11.655	12.722	37.469	12.309	961
86.4	51.941	11.770	12.831	38.168	12.707	4
87.1	51.780	11.679	12.994	38.639	12.636	-811
87.2	52.966	12.297	13.265	39.034	12.606	359
87.3	53.703	12.718	13.211	39.642	12.751	818
87.4	54.172	13.044	13.310	40.006	12.947	952
88.1	54.047	13.123	13.548	40.380	12.994	247
88.2	54.564	13.431	13.595	40.786	13.198	416
88.3	54.892	13.555	13.655	41.284	13.215	293
88.4	55.327	13.638	13.719	41.689	12.988	570
89.1	55.555	13.498	13.763	41.922	13.048	320
89.2	55.925	13.580	13.746	42.233	13.302	225
89.3	56.164	13.729	13.882	42.626	13.362	23
89.4	56.496	14.063	14.247	42.577	13.222	513
90.1	56.526	14.410	14.439	42.613	13.225	659
90,2	56.811	14.687	14.511	43.299	13.386	301
90.3	57.648	15.066	14.534	44.050	13.615	515
90.4	56.981	14.500	14.401	44.238	13.506	-663
91.1	57.711	14.416	14.213	43.972	13.594	348
91.2	58.259	14.434	14.247	44.371	13.683	392
91.3	58.758	14.575	14.474	45.208	13.700	-49
91.4	58.922	14.675	14.560	44.944	13.749	344

References

Barbone L., Bodo G., Visco I. (1981), "Costi e profitti in senso stretto: un'analisi su serie trimestrali, 1970-80", *Bollettino della Banca d'Italia* , 36.

Chow G., Lin A. (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series", *The Review of Economics and Statistics*, 53.

Di Fonzo T. (1990), "The Estimation of M Disaggregate Time Series when Contemporaneous and Temporal Aggregates Are Known", *The Review of Economics and Statistics* , 72.

Fernàndez R.B. (1981), "A Methodological Note onthe Estimation of Time Series", *The Review of Economics and Statistics* , 63.

Istat (1992), "I conti economici trimestrali con base 1980", *Note e Relazioni*, 1.

Kendall M.G., Stuart A. (1976), *The Advanced Theory of Statistics*, vol. III, London, Griffin.

Litterman R.B. (1983), "A Random Walk, Markov Model for the Distribution of Time Series", *Journal of Business and Economic Statistics* , 1.

Stram D.O., Wei W.W.S. (1986), "A Methodological Note on the Disaggregation of Time Series Totals", *Journal of Time Series Analysis* , 7.