ISSN 2315-0807

eurostat
Statistical working papers

Overview of GDP flash estimation methods

2016 edition

eurostat

eurostat
Statistical working papers

Overview of GDP flash estimation methods

2016 edition

eurostat

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

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2016

ISBN 978-92-79-46829-2
ISSN 2315-0807
doi: 10.2785/51658
Cat. No: KS-TC-16-006-EN-N

Preface

Gross domestic product (GDP) is generally considered to be one of the most important economic indicators. In 2003 Eurostat published for the first time quarterly GDP flash estimates that were released 45 days after the end of the reference quarter. However, economic, financial and technological developments in recent years have made it increasingly important to have economic indicators that enable economic activity to be analysed at the earliest possible stage. Therefore, further advancement of the quarterly GDP estimation was desired by several policy makers and other (key) users.

It was against this background that the National Accounts Working Group set up the **task force for 'GDP flash estimates at t+30 days'** in May 2013. Its mandate was 'to assess whether it is feasible to estimate a sufficiently reliable flash GDP for the euro area and the European Union, based on the information available at t+30 days, including internal national estimates'.

The task force worked on the following topics:

- sharing information and best practices regarding early quarterly GDP estimates;
- country-by-country overview of available sources for GDP estimates between t+25 and t+60 days;
- developing a guidance document on methods and national estimation techniques;
- developing estimation methodology for the euro area and EU GDP t+30 estimates;
- developing quality acceptance criteria to assess test results;
- preparing test estimates;
- developing a communication plan.

The results of the task force's work and the GDP t+30 flash project in general are documented in two statistical working papers. This document provides an overview of methods and techniques that national accountants can use to estimate GDP soon after the end of the reference period. Compilers of early national GDP estimates will benefit from the information it provides. It will also help users of early GDP estimates to interpret their characteristics and quality.

Acknowledgements

This statistical working paper was produced by the sub-group on methods and estimation techniques within the Eurostat task force for ‘GDP flash estimates at t+30 days’.

Eurostat would like to thank all the members of the sub-group who were involved in researching and compiling this paper. More specifically, thanks are due to Filippo Moauro (ISTAT, Italy) for drafting sections 1, 2, 3.1, 3.2, 3.4, 3.5.2 and the annex to this document. Kristina Kiriliauskaite (Statistics Lithuania) was responsible for sections 2.5 and 3.5.1, Michal Široký (Czech Statistical Office) for section 3.3.1, Samu Hakala (Statistics Finland) for section 3.3.2, Claudia Cicconi (ISTAT, Italy) for section 3.6.2., Stefan Leist and Kornel Mahlstein (both from the State Secretariat for Economic Affairs (SECO), Swiss Confederation) for section 4. The work benefited from discussion with all task force participants and from precise and expert revision by Ben Gardiner (Cambridge Econometrics), Arto Kokkinen (Eurostat) and Johannes Wouters (Eurostat).

Special thanks go to Filippo Moauro for supervising and coordinating the work, and for editing this document.

Contents

Executive summary

European GDP flash estimates, coordinated by Eurostat, are currently released at t+45 days. To investigate whether the deadline for release could be cut to t+30 days, and to share experiences among Member States, Eurostat set up a task force in May 2013. Sixteen EU countries and Switzerland took part in its work in 2013, 2014 and 2015. There was a particular focus on the methods used to estimate GDP, and this document is the main outcome of a dedicated sub-group of the task force. It gives a general overview of estimation methods used in different EU countries, and presents a number of practical suggestions. After the introductory chapter, the document discusses the preliminary analysis of available data, model strategies, and methods for forecasting missing data, including extensions to multivariate setups and the analysis of results.

Two main estimation strategies have been identified:

- the 'direct' approach: the last quarter of a given GDP component is estimated directly, and
- the 'indirect' method: the last quarter is extrapolated using a temporal disaggregation procedure.

The difference is given by the model used for extrapolation. Under the former approach, the set of related data is directly used to extrapolate the last quarter of the GDP component. Under the latter, extrapolation is carried out in two steps: firstly, the indicator in the last quarter is extrapolated, then the temporal disaggregation for the GDP component of interest is run. Following the same logic, when no indicator is available a one- or two-step time series method is applied, for the direct or the indirect approach respectively.

A summary of the answers to a questionnaire on estimation methods is also presented in the annex to this document. The questionnaire helped researchers understand Member States' experiences, their approach in terms of accounting method, detail of compilation, price evaluation, type of adjustment on data, whether a preliminary analysis is effectively carried out, details of the more commonly used indicators and, of course, the more technical aspects of the methods used.

In short, this document is an attempt to merge differing experiences in different EU countries into a unified view. It allows the reader to follow a step–by-step approach to flash GDP estimation, with explanations and analysis to help them conduct the exercise efficiently. The approach is pragmatic, with links to available software, references to the literature, and a particular stress on the elements shared by advanced users and beginners. The aim is to share information about the technical aspects of estimation methods within the community of people working in the area of quarterly accounts. The best practices presented here are thus a way of standardising flash GDP estimation at national level and, indirectly, of improving the quality of European aggregates released after a short period of time.

1. Introduction

In 2013 Eurostat set up a task force involving a large number of Member States to assess the feasibility of estimating quarterly GDP in the euro area and the European Union at t+30 days. The main purpose was to share practice in this area and to equip the group with a set of guidelines to help both experts and beginners working on flash estimates of quarterly national accounts.

Estimating quarterly GDP at t+30 days poses the technical problem that data for the last observation are often missing. Moreover, modelling solutions are needed that could efficiently combine the scarce available information with forecasts. The most common situations are as follows:

— the last month (or the last two months) of the quarter t for which monthly short-term business statistics are available is/are missing, or
— there is no data at all for the last quarter.

This document introduces a technical framework for flash estimation of GDP using a pragmatic approach. The focus is on methods used by Member States taking part in the task force's work. These are summarised, but their relative merits are not assessed. The summary covers not only forecasting methods, but also other important practical aspects, including preliminary analysis of available data and comparative analysis of results relative to alternative methods. For this reason the document also proposes a strategy for producing GDP flash estimates, taking into account the debate on technical aspects of estimation within the task force, and the responses to a specific survey conducted on Member States.

Each paragraph of this overview also provides a list of options for approaching the specific issue from the practical point of view. References and information about available software can be found at the end of the document.

The goal of this document is to give those responsible for making flash estimates a concise description of the methods used in various Member States and the tools used to make a rational choice between the available methods. It provides explanations and analysis to enable the reader to make an efficient flash estimate of GDP by following step-by-step instructions.

These guidelines are by no means exhaustive, and the authors acknowledge that there is an extensive literature on forecasting which is in constant development. This material should therefore be viewed as a first attempt by those working in the data production field to compile some suggestions and reflections on flash estimates. It can certainly be improved on.

The document is structured as follows: section 1 presents background information on the GDP flash estimation at t+30 days (section 1.1) and the main purposes of the document (section 1.2). Section 2 is devoted to preliminary analysis, defining objectives (section 2.1), analysis of available data (section 2.2), detail of compilation (section 2.3), graphical analysis (section 2.4), and outlier detection and correction (section 2.5). Forecasting missing data is the subject of section 3: after an introduction (section 3.1) and a set of model strategies (section 3.2), the most commonly used methods are presented, split into direct (section 3.3) and indirect approaches (section 3.4), pure forecasting methods (section 3.5) and multivariate extensions (section 3.6). Section 4 covers the analysis of results, presenting commonly used measures for a comparative analysis of results (section 4.1) and a short discussion of forecast encompassing (section 4.2) and forecast combination (section 4.3). Finally, the annex presents the questionnaire on estimation methods with a summary view of the responses received from Member States participating in the task force and a short discussion of the main evidence.

1.1 Background

Quarterly GDP data are probably the most relevant short-term economic statistics produced by the European Statistical System (ESS). Their release is coordinated by Eurostat. They are disseminated around 45 (GDP flash), 60 (GDP and main aggregates) and 90 (GDP and other sector accounts data) days after the end of the reference quarter.

In recent national and international discussions involving national statistical institutes, users and practitioners, a clear need emerged to cut the time-lag in the GDP flash estimate from 45 to 30 days. Among European countries, the UK, Belgium, Lithuania and Spain were already publishing a GDP flash estimate at t+30 days ([1]). Other countries are testing or have recently tested the feasibility of reducing this time-lag.

There are several reasons for aligning GDP flash estimates on a deadline of t+30 days: policymakers and stakeholders want timely information, and the recent economic crisis has made the need for an efficient early warning system more acute. Moreover, a flash estimate at t+30 days would align the European calendar on that of the US, where GDP is released according to a 30-60-90-day timetable.

This is why, in 2013, Eurostat set up a task force to assess the feasibility of estimating the quarterly GDP of the euro area and the European Union at t+30 days.

A number of technical conditions were defined from the very outset:

- the **focus** of the estimates would be limited to the seasonally adjusted quarterly growth rates in volumes;
- the **methodological frameworks** for estimating GDP t+30 flash data, following the source and methods set out in a series of handbooks issued by Eurostat, ranging from the ESA 2010 manual to handbooks on price and volume figures and, ultimately, the recent edition of the handbook on quarterly national accounts;
- the **limits of these references**, bearing in mind that GDP flash at t+30 days poses two problems: the difficulty of producing data under time restrictions, and the problem of missing data in more recent periods;
- Eurostat's objective was to compile flash estimates of euro area and EU GDP by applying the '**direct method**', i.e. using Member States' estimates and deriving missing countries implicitly at the end of the process;
- finally, 2016q1 would be proposed as the first quarter for which a t+30 flash estimate for the euro area and EU quarterly GDP should be compiled, assuming that it was decided to introduce the method outlined above.

It is clear to all institutions involved in the project that data at Member State level are required in order to reach the final target – the compilation of European aggregates. However, Eurostat decision to change the deadline for the GDP flash estimate to t+30 days, could encourage Member States to publish their estimates by the same deadline.

Statistical agencies, central banks and other stakeholders around the world forecast economic indicators every day, and real GDP is a very important variable for this purpose. Consequently, there is an ongoing discussion about modelling solutions and tools that efficiently combine the scarce available information with the forecasts produced.

1.2 Purpose

The purpose of this document is to provide guidance to national accountants on the correct use of methods and techniques for compiling flash estimates. The emphasis is on the elements common to advanced and less experienced compilers, and the information in this document may facilitate communication within the community of compilers of quarterly accounts. The best practices presented in this statistical working paper are a way to harmonise the GDP flash estimation process at national level and, indirectly, to help produce robust European aggregates.

([1]) More recently, Austria, Latvia and France have started publishing GDP flash estimates 30 days after the end of the reference quarter.

The document also provides a design for flash estimation, from preliminary analysis of available data to outlier detection and correction, model strategy and analysis of results.

The application of guidelines by statistical agencies promotes transparency for data users with a strong interest in the details of flash estimation. This concerns the production process, the full model specification, the reliability of estimates and elements for a comparative analysis of results.

2. Preliminary analysis

2.1 Objective

Description

Flash estimates of quarterly GDP components means estimating recent quarters, when the most suitable set of data for computing is not yet available.

The main objective of the preliminary analysis is to collect and analyse all data available within a short period of time after the quarter, which could help in estimating GDP for the last quarter and ensure a robust estimate. These indicators should be selected from among data measuring a similar phenomenon of the GDP component. Related data should also show a similar path to the levels of the GDP component or to the series after transformation (e.g. logarithmic transformation, difference operators, moving averages).

The available indicators should be collected in accordance with the breakdown of components on the basis of which quarterly accounts are currently compiled. If seasonal adjusted indicators are not available, seasonal adjustment should be applied to this set of data. Comparison at aggregated level should also be considered when the comparison between GDP components and related series is poor at detailed level.

Options and suggestions

When time and resource constraints make it impossible to run a detailed preliminary analysis, it is advisable to focus the analysis on the available indicators that are currently used and the most relevant components.

2.2 Analysis of available data

Description

Analysis of available data starts from the structure of GDP components provided by quarterly national accounts for volume measures. The purpose is to find a relationship between these data with relevant indicators available in time and able to provide useful information to estimate GDP in the last quarter. Related indicators could be official short-term statistics, other relevant economic data, or business and consumer surveys. If indicators with the same detail as GDP components are not available, the analysis could be performed at aggregated level (e.g. broad or aggregated NACE). The available indicators could be quarterly and/or monthly data. Among monthly data, all the indicators that can provide information on at least one month of the last quarter should be considered. Analysis of available data is limited to volume indicators. For data expression of nominal measures (such as turnover indexes and retail sales) a suitable deflation should be performed first.

Options and suggestions:

A detailed search of available indicators among a reasonably wide spectrum of data sources is the best option. A feasible alternative is to limit the analysis to official statistics and/or to more relevant GDP components.

2.3 Detail of compilation

Description

Flash estimates of GDP are the result of a multi-step exercise in which elementary estimates relative to sub-components are aggregated following a pre-defined scheme. The elementary components could reflect the maximum detail of quarterly national accounts or a reasonable simplified split, if insufficient data are available. The exercise could be applied to sub-components from the production side or the demand side, or it could follow both approaches. In the latter scenario, the final GDP estimate could be benchmarked to ensure consistency of approaches.

In the presence of timely indicators, the preliminary analysis could start from same breakdown of GDP used in the current production of quarterly national accounts. In the following steps, the analysis could develop towards lower detail of aggregation if there is evidence that indicators fit macro-aggregates better.

Options and suggestions

In the presence of indicators from both production and demand side, the national accountants should do the exercise according to both approaches. However, compilation could initially be limited to a single approach, even if data from the other side were available. In the presence of a detailed set of information, one can foresee a detailed compilation scheme. Such detail should ensure the best use of the information provided by the composition of quarterly national accounts. Complementing the compilation with higher levels of aggregation (e.g. the split of value added into the 3 or 10 categories of the NACE classification) is also suggested, since two or more levels of aggregation allow a wider comparative analysis of results. For first implementation, the compilation could be limited to a reasonable aggregation scheme including most relevant aggregates and taking into account the available split of indicators.

2.4 Graphical analysis

Description

A first graphical comparison in the time domain provides the analyst with useful information on the utility of available indicators in forecasting GDP components. Graphs should look at data in levels or logarithms (better after standardisation), first or seasonal differences or growth rates. The analysis could also be complemented by a very first run of the forecasting software using default methods; a first analysis of the residuals from this fit would enable a preliminary idea to be formed of most suitable model specifications.

The analyst could collect information on:

- the structure of the GDP component in terms of trend-cycle-seasonal components;
- the presence of seasonality in the related indicator;
- the presence of outliers in the indicators;
- the fit of related series to the GDP component.

Although it may be time-consuming, especially in the presence of a detailed breakdown of GDP into components by the demand and/or the supply side, it is useful to conduct a preliminary graphical analysis for at least most relevant time series.

Options and suggestions

The national accountants should run a detailed graphical analysis component by component, complementing the analysis with plots of transformed data so as to add elements in the choice of more suitable model specification. It could also be good practice to use an automatic procedure.

2.5 Outlier detection and correction

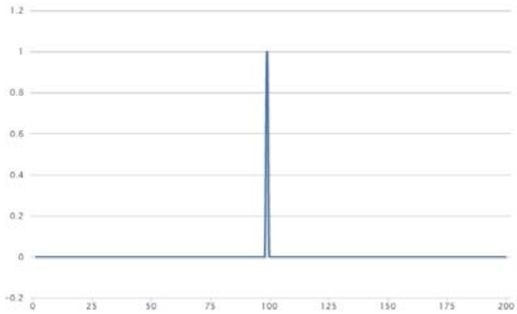
Description

Economic time series are affected by many political, social and environmental factors, both internal and external. Factors which do not fit within the normal distribution of time series are called **outliers**. These differ dramatically from the typical pattern of the time series and its trend. Outliers adversely affect the quality of final forecasts and estimates of components. They should therefore either be removed if they relate to data errors, or treated separately in model specifications if associated with specific events (such as strikes and earthquakes).

There are three main types of outlier:

- An **additive outlier** (AO) occurs on a given date t_0 and affects one observation only. It may be expressed as follows:

$$AO_t^{t_0} = \begin{cases} 1\ for\ t = t_0, \\ 0\ for\ t \neq t_0 \end{cases}$$



The typical example is a strike taking place during a specific period of time ([2]).

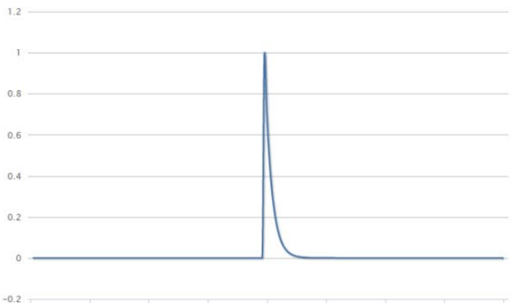
— A **level shift** (**LS**) is an abrupt and permanent change in trend level beginning on a given date t_0. It is modelled by the regression variable

$$LS_t^{t_0} = \begin{cases} -1\ for\ t < t_0, \\ 0\ for\ t \geq t_0 \end{cases}$$

A level shift takes place when there is a structural change in the economy, such as a financial crisis. One example is a raise in the VAT standard rate during a given period.

— A **temporary change** (**TC**) is an abrupt and temporary level change beginning on the given date t_0. It is modelled by the regression variable

$$TC_t^{t_0} = \begin{cases} 0\ for\ t < t_0, \\ \alpha^{t-t_0}\ for\ t \geq t_0 \end{cases}$$



where α is a rate of decay back to the previous level such that $0 < \alpha < 1$. An example of a temporary change is a natural disaster that hits the economy hard at a given point, after which its impact lessens. Another example is a strike that starts during one calendar period and ends after an interval.

([2]) The illustrations on this page are taken from: https://ec.europa.eu/eurostat/cros/content/e-learning-courses-seasonal-adjustment-0_en

Outliers must be identified and treated as regressors in the model adopted for flash estimation. Of course, any that are not significant are excluded from the model specification.

Options and suggestions

The analyst should run a detailed identification procedure for outliers and their corrections. The use of an automatic procedure for detecting and correcting outliers, such as those of TRAMO/SEATS and X-12-ARIMA, is also recommended.

3. Forecasting missing data

3.1 Introduction

Description

To estimate a given component of GDP over the last quarter, one can apply either a direct or an indirect approach. In the latter case, the estimate is extrapolated by applying a temporal disaggregation procedure to annual data, using quarterly indicators as suitable related information. The model used for extrapolation is what makes the difference. Under the **direct** approach, the set of related data is used directly to extrapolate the last quarter of the GDP component. Under the **indirect** approach, the extrapolation is carried out in two steps: firstly the indicator in the last quarter is extrapolated, then the temporal disaggregation for the GDP component of interest is run.

In a flash estimation process, the choice between a direct and an indirect strategy often reflects the same choice in the compilation of the regular quarterly national accounts with complete information set. Since flash estimation is characterised by a partial lack of data, sticking to the same production practices as in the regular quarterly accounts should ensure closeness in final results.

When, at t+30 days, the usual information set is completely missing or only alternative indicators are available (such as business and consumer surveys), it makes no difference which of the two approaches is chosen. Only an empirical analysis of errors produced by a preliminary in-sample rolling forecasting exercise can lead to the most appropriate choice. Other significant factors are knowledge and years of experience with a particular practice, which may determine an NSI's preference for one of the two approaches.

Options and suggestions

Under the direct approach, a GDP component should be extrapolated using a partial set of short-term indicators, complemented by business and consumer surveys or alternative indicators. Under indirect extrapolation, the current process for estimating quarterly accounts follows an indirect temporal disaggregation methodology.

Sub-optimal methods are as follows:

- direct estimation using naïve extrapolation methods (e.g. forecasts obtained from growth rates of related indicators)
- naive extrapolation without related data (e.g. extrapolated growth rate = 0 or equal to the growth rate of a previous quarter).

3.2 Model strategies

Description

There are two aspects to forecasting missing data: the first dimension relates to model strategy (see previous section for the distinction between direct and indirect approaches), the second the availability and use of related short-term indicators for last quarter estimation. Related data could be fully available, partial or totally absent. Where related data is available, the challenge is to find the best way to bridge the gap between available short-term data and the GDP component. If there are no related data available, the problem is reduced to finding the best estimate on the basis of a pure forecasting method.

As regards modelling, in the presence of related indicators, the simplest extrapolation of a GDP component can be based on a regression method. Under a direct strategy, commonly used methods are **autoregressive distributed lag** (ADL) models and **factor models**, whereas under the indirect approach temporal disaggregation regressions adopting either an ARIMA structure on the residuals or ADL forms are used. All these basic methods have multivariate extensions in the literature.

As regards data availability, a frequent situation is when the indicator is monthly and only one or two months of the last quarter are available. In such circumstances, quarterly forecasting requires the additional step of forecasting the missing months of related data,which is usually solved through a pure forecasting method. These methods fall into the class of ‘bridge models’.

In the absence of related data, pure forecasting methods (e.g. ARIMA models) can be adopted. In such a case, it is good practice to follow the same methodology, strategy and detail of compilation as adopted in current production of quarterly accounts.

Options and suggestions

Regression based methods using indicators are the basic and more suitable model strategy. Where partial information on related data is available, the bridge modelling approach is to be recommended. By contrast, in the absence of related data, pure forecasting methods (ARIMA, structural models, etc.) are the best solutions. Multivariate models are possible extensions of the usual univariate models. However, careful consideration should be given to applying them to extensive data production, given their complexity, especially as regards the stability of results in the estimation phases. Basic alternatives are estimating the last quarter by simply imputing the previous available quarter or other naive methods (such as the same growth rate as a related variable). This is a common practice among national accountants, especially for estimating small entities when the usual sources are not available. However this practice should be avoided when considering relatively more important variables, since it corresponds to giving an implicit modelling structure to the data without any empirical evidence of its validity.

3.3 Direct approach

3.3.1 ADL models

3.3.1.1 Model selection

Description

Autoregressive distributed lag (ADL) models are a type of regression in which lagged values of the dependent variable and current and lagged values of one or more explanatory variables are included among regressors. Inclusion of lagged values is justified by the dynamic nature of most economic processes. A substantial period of time may pass between the economic decision-making period and the final impact of a change in a variable to be estimated. Furthermore, the dynamic behaviour of an economy can reveal itself through a dependence of the current value of an economic variable on its own past values. For the sake of brevity, we will present here only a simple ADL(1,1) form, i.e. a model whose regressors include a constant term α_0, the lagged dependent variable Y_{t-1} with regression coefficient α_1 and one explanatory variable X_t taken at lags 0 and 1 and with regression coefficients given respectively by γ_0 and γ_1:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + u_t$$

In the equation above u_t is the error term with values that have a zero mean, constant variance and are serially uncorrelated.

Most macroeconomic time series are trended and therefore non-stationary (their mean and/or variance depend on time). The problem with non-stationary data is that the standard regression procedures can easily lead to incorrect conclusions. In these cases a very significant relationship can be found, while the variables used in the analysis have no interrelationships. This is called a spurious regression. It is therefore necessary to test whether the variables are really related. In this case we say that the variables are cointegrated. It means that there is a genuine long-run equilibrium between variables, which takes the form:

$$Y_t^* = \beta_0 + \beta_1 X_t^*$$

If the variables are cointegrated, it is convenient to express the ADL(1,1) model in an equivalent re-parametrised form called the error-correction model (ECM):

$$\Delta Y_t = \gamma_0 \Delta X_t - \pi\left(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}\right) + u_t \text{ (3)}$$

This form has the advantage of including both long-run and short-run information. This is because the long-run equilibrium $Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$ is included in the model together with the short-run dynamics captured by the differenced term. In this model, γ_0 measures the short-run effect that a change in X_t will have on a change in Y_t. On the other hand, $\pi = 1 - \alpha_1$ is the feedback effect and shows how much of the disequilibrium is corrected, i.e. the extent to which any disequilibrium in the previous period affects any adjustment in Y_t. It provides us with information about the speed of adjustment in cases of disequilibrium. $\beta_1 = (\gamma_0+\gamma_1) / (1- \alpha_1)$ is the long-run multiplier. If the variables are in logs then β_1 is the long-run elasticity of Y with respect to X.

The ECM is a convenient model measuring the correction from disequilibrium of the previous period, and which also has a good economic implication. In the case of cointegration, the disequilibrium error term is a stationary variable. It implies that there is some adjustment process preventing the errors in the long-run relationship from becoming larger and larger. Another important advantage of the ECM is the ease with which it can fit into the general to specific approach to econometric modelling, which is in fact a search for the most parsimonious model that best fits the given data sets.

Options and suggestions

The final model should be selected after a careful comparison of several hypotheses of suitable explanatory variables. It should have an economic interpretation and all explanatory variables should be statistically significant in terms of t-tests or regression coefficients. If the data are not seasonally adjusted, seasonality should be captured by introducing seasonal dummy variables into the model. Different lags of the variables and different data transformations should be evaluated. When the selection of regressors is carried out automatically, the significance of regression coefficients should be checked.

3.3.1.2 Estimation and diagnostics

Description

A convenient way to estimate the ECM was introduced by Engle and Granger (1987), involving four steps:

Step 1: Testing the variables for their order of integration

By definition, cointegration necessitates that the variables are integrated of the same order. (Integration of order d means that the series, after differentiating d times, becomes stationary.) Thus the first step is to test each variable to determine its order of integration. The augmented Dickey-Fuller (ADF) test is the most frequently used. Other suitable tests include the Phillips-Peron test and the KPSS test. If all variables are integrated of the same order then we proceed with step two.

Step 2: Estimation of the long-run relationship

If the results of step 1 indicate that all variables are integrated of the same order, the next step is to estimate the long-run equilibrium relationship and obtain the residuals of this equation. The adequacy of the model should be checked with the help of a test for serial correlation of residuals (e.g. Durbin-Watson test, Breusch-Godfrey test) and a test for heteroscedasticity of residuals (e.g. White test). Moreover, the multicollinearity of the explanatory variables should be assessed.

Step 3: Checking for the order of integration of the residuals

In order to determine if the variables are actually cointegrated, it is necessary to test if the estimated residuals from the long-run relationship are stationary. Then the variables are cointegrated. This can be done by performing an ADF test (or another stationarity test) on the residual series. The critical values differ from the

(3) As Stock and Watson (2007, pp. 541–544, 601) show, one may also consider using the autoregressive distributed lags model without the long-run cointegration part for short-run forecasts. If the differenced (and possibly logarithmised) Y and X are stationary (and differenced X is an exogenous variable) it is possible to explain and forecast the difference of Y with its own lags and with the lags of differenced X. The statistical inference on the model parameters is always possible when explaining a stationary variable with other stationary variables.

standard ADF values, being more negative. If we find that the residuals are stationary then we can reject the hypothesis that the variables are not cointegrated.

Step 4: Estimation of the error-correction model

If the variables are cointegrated, the residuals from the equilibrium regression can be used to estimate the error-correction model and to analyse the long-run and short-run effects of the variables, as well as to see the adjustment coefficient, which is the coefficient of the lagged residual terms of the long-run relationship identified in step 2. At the end we always have to check for the adequacy of the model by performing diagnostic tests for serial correlation and heteroscedasticity of residuals.

A drawback of the Engle-Granger approach is that when there is more than one explanatory variable, there may be more than one cointegrating relationship, and the procedure using residuals from a single relationship cannot accommodate this possibility. It can be resolved by applying the Johansen test, which provides us with the number of cointegrating relationships.

Options and suggestions

The statistical properties of the model should be verified using a variety of diagnostic tools, covering integration and cointegration of variables, serial correlation, heteroscedasticity and multicollinearity. In addition, a specification test (e.g. RESET) and a test for stability of parameters (e.g. CUSUM) should be performed, as should tests to establish the order of integration of the variables and their possible cointegration. There is a risk that estimation without diagnostic checking could result in serious misspecification of the model and spurious regression.

3.3.1.3 Available software

Description

One of the advantages of the Engle-Granger approach described in the previous chapter is that it is easy to understand and implement. When the conditions for the classical regression model are fulfilled, the parameters of both the long-run relationship and the error correction model can be safely estimated using ordinary least squares. This implies that virtually any statistical/econometric software can be applied. Even the commonly used spreadsheet packages contain tools for performing simple regression tasks. However, the correct application of the Engle-Granger procedure calls for many statistical tests. Using a specialised software, which includes a variety of suitable tests (e.g. EViews), is highly recommended. Advanced statistical/econometric software packages also contain their own programming languages that enable the whole procedure to be automated and facilitate comparison of alternative specifications of the model.

Options and suggestions

For work on ADL models, the best solutions are those provided by specialised statistical/econometric software complemented by suitable macros for routine and automated activities. The advantage of using this software is also that it usually covers all the tests described above. Using general software which includes (at least) the ordinary least squares estimation method is a sub-optimal solution. Spreadsheet packages could also be adopted for calculations, but are not recommended, as they do not provide the required tests for model selection and diagnostic checking.

3.3.2 Dynamic factor models ([4])

Dynamic factor models can be useful if available source data is too scarce to calculate or estimate a component of interest at a more detailed level (e.g. gross domestic product using the production approach at NACE 2-digit level). It should be noted that instead of trying to forecast future observations the emphasis is on trying to nowcast an observation at a time period for which there is some, but not very much, data available. The idea is to use all available micro-level data by composing factors that describe movements in latent variables that are themselves unobserved. If possible, it is recommended that factors be interpreted as comprehensible

([4]) Fornaro (2013) discusses the method and shows some promising results of test calculations in his working paper, which describes nowcasting economic activity using dynamic factor models. This chapter is based on this working paper.

phenomena, because gaining a better understanding of model dynamics can only be useful in modelling. However, from a pure production point of view this is not necessary.

3.3.2.1. Model selection

Factor extraction

Description

Extracting common factors from a dataset is the first stage of modelling. One important issue in the estimation process stems from the factor selection, i.e. how many factors should be included in the nowcasting model. The factor selection can be based on information criteria, such as the Bayesian Information Criteria (BIC), or factor-based regression criteria suggested by Groen and Kapetanios (2009). Josse and Husson (2012) provide an algorithm that estimates the optimal number of principal components for a given dataset presenting missing values. At the second stage these factors – whether or not they are interpreted as related elements – are used in nowcasting (estimating) the component(s) of interest. The out-of-sample performance of the various nowcasting models should also be considered when trying to determine the optimal number of factors to be extracted.

By defining $\boldsymbol{X_t}$ as a dataset that contains micro-level data of *N* observations (e.g. firms), we can define a factor model:

$$X_t = \Lambda F_t + e_t$$

with *r* latent factors included in $\boldsymbol{F_t}$. $\boldsymbol{\Lambda}$ is the matrix of factor loadings and $\boldsymbol{e_t}$ is the *Nx1* vector of idiosyncratic components that are allowed to be both serially and cross-sectionally correlated, making this model resemble the approximate factor model by Chamberlain and Rothschild (1983). The factors are estimated by principal components, i.e. the estimate of $\boldsymbol{F_t}$ is given by the eigenvectors corresponding to the largest eigenvalues of the matrix ***XX’***. The *TxN* matrix ***X*** consists of the observed datasets at time periods *t=1,...,T*. The basics of principal component estimation require that all the series be of the same length. This can be achieved either by selecting from the original data a subset that fulfils the criteria or by imputing the missing values. Selecting a subset that does not include missing observations risks losing a substantial amount of information. On the other hand, imputing missing values complicates the factor extraction process.

Options and suggestions

The best solution would be to exploit all potentially useful variables, imputing missing values or selecting a subset that does not include missing values. Then the optimal number of factors is determined using appropriate criteria (like BIC) and the extracted factors need to be interpreted. Sub-optimal solutions are those that do not meet all the requirements mentioned or that simplify the preliminary planning.

Nowcasting model

Description

The estimated factors can be used as predictors in the nowcasting model

$$y_t = \beta \hat{F}_t + \varepsilon_t$$

Where y_t measures economic activity, t is the reference period we are interested in and ε_t is the nowcasting error. Parameters β are estimated using the Ordinary Least Squares (OLS) method.

The nowcasting model above is presented in a very simple form. Before rushing into estimating regression coefficients and producing nowcasts there are some important things to consider. The dependent variable and explanatory variables – in this case extracted factors – should be examined. It is very common for economic time series not to be stationary, that is, mean, variance or autocorrelation structure is not independent of time. One solution is to try to obtain stationary time series by using, for example, first order and seasonal differencing and log-transformations. Stationary time series can be used safely in a simple regression model. A more ambitious approach would be to investigate possible cointegration between variables and to build a more sophisticated model, such as the error-correction model (ECM) presented in chapter 3.3.1. More about checking and obtaining stationarity can be found in chapter 3.5.1.

Options and suggestions

The dependent variable and extracted factors should be investigated to detect non-stationarity and possible cointegration. After that, a reliable nowcasting model is formulated. If a simple regression model is used, the variables are transformed to render them stationary. It is important to avoid using a simple regression model without investigating the stationarity of the variables.

3.3.2.2. Estimation and diagnostics

Description

Estimation techniques and diagnostic tools depend on the nowcasting model chosen. Using a simple regression model allows parameters to be estimated using OLS. In addition, a simple regression model does not necessarily require as many diagnostic checks as a more sophisticated model might require. This makes using simple models appealing. On the other hand, building a meticulously well-thought-out model and performing the required diagnostic checks guarantees that the nowcasting model should work as well as possible, at least in theory.

Options and suggestions

The analyst should perform all the usual diagnostic checks (such as the significance of regression coefficients, their logical sign, stationarity of residuals), accompanied by an appropriate graphical analysis. If a more sophisticated nowcasting model is used, the requisite diagnostic checks should be performed accordingly. Out-of-sample performance should also be evaluated. In the light of diagnostic checks, the performance of a chosen model can be expected to be steady, so frequent re-modelling can be avoided.

3.3.2.3. Available software

R offers an open source solution for dynamic factor modelling. The following R packages are needed:

- Base package for handling data
- Utils package for utility functions
- missMDA for imputing missing values
- Stats package for model estimation

Many other packages may prove beneficial as well. For instance, grDevices and graphics packages include tools for graphical visualisation. See http://www.r-project.org/ for more information about R.

Moreover, at least SAS, and probably many other commercial software packages, can be used in dynamic factor modelling. The factor procedure for principal components analysis is included in SAS/STAT. SAS/ETS includes an SSM procedure for modelling state space models. An example of a dynamic factor model is presented in the user manual (the SSM procedure, example 27.7).

3.4 Indirect approach

3.4.1 Model selection

Description

Common methods of temporal disaggregation are based on simple linear regressions between a dependent variable Y_t and a set of related indicators $X_t = (x_{1t}, \dots, x_{kt})$. Y_t is available in the form of a sum or an average over a given interval of observation (e.g. every year), whereas the X_t is available at a higher frequency (e.g. every quarter). A general setup could be the ADL(1,1) model formulated at the higher frequency of observations, as in Section 3.3.1.1.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + u_t$$

where the dependent variable Y_t could be also a differenced series (see Proietti, 2005).

The ADL(1,1) model nests popular forms of temporal disaggregation developed by Chow and Lin (1971), Fernandez (1981) and Litterman (1983). In particular, the Chow and Lin model occurs when the series is modelled in levels and $\gamma_1 = -\alpha_1\gamma_0$. In this case, the ADL(1,1) model becomes a stationary regression model with AR(1) residuals. The case by Fernandez occurs when the dependent variable $\mathrm{Y_t}$ is an expression of a differenced series, $\alpha_1 = 0$ and $\gamma_1 = 0$; finally, Litterman's model occurs when $\mathrm{Y_t}$ is a differenced series and $\gamma_1 = -\alpha_1\gamma_0$, i.e. a regression with non-stationary ARI(1,1) residuals. Further forms imply an ADL(1,0) model.

The model specification can be adapted to include intervention variables such as impulse or step dummies and ramp effects. Moreover, $\mathrm{Y_t}$ can be subject to logarithmic transformation.

Model selection could be based on a general to specific approach, comparing significance and summary statistics on estimated residuals available at low frequency of observation and the significance of regression coefficients. It is crucial in flash estimates to run an ex-ante comparative analysis of the performances of model specifications in terms of their errors of extrapolation (in an annual-quarterly disaggregation this error is the difference between the sum of the 4 extrapolated quarters and the true annual value).

Options and suggestions

The analyst should operate a statistically sound selection from a class of ADL regressions, including all the cases of models for data in levels, differences, log-levels and log-differences. If the wider class of ADL methods is not available, it could be enough to make a selection from among the most popular temporal disaggregation techniques, adjusting the fit with covariates' trends, including reasonable intervention variables. To avoid doing this is the naive approach consisting of extrapolating the last quarter on the basis of trends in related indicators.

3.4.2 Estimation and diagnostics

Description

The most commonly used methods of temporal disaggregation are a special type of regressions; they require the residuals to be modelled in such a way that fitted values distribute the annual totals within the desired time span (e.g over the quarters). In other words, the standard hypothesis of white noise residuals used in the context of ordinary least squares (OLS) methods cannot be applied. Consequently, popular methods of the Chow and Lin class or ADL(1,1) models find a solution within generalised least squares (GLS) regressions. In particular, these forms have the advantage that all the regression coefficients and the error term variance can be easily concentrated out from the log-likelihood function, which becomes a function of the parameter α_1 only. Then, the maximum of the likelihood function is conveniently carried out via a grid search method, scanning a set of tentative values of α_1 over the interval (-1, 1).

When the dependent variable is subject to logarithmic transformation, estimation requires nonlinear statistical treatment, as temporal disaggregation becomes a nonlinear problem. For details, see Proietti, (2005), where there is also a discussion on revisited ADL models and temporal disaggregation within state space forms.

Estimation and model selection could also interact, defining a general to specific approach within a defined ADL(1,1) class. In this case, estimation should be complemented by information criteria (e.g. AIC and BIC statistics) and diagnostics on innovations.

The adjustment procedure popularised by Denton (1971) is also very commonly used among NSIs. In this context, final distributed estimates are obtained from a two-step constrained optimisation, where an initial set of data or initial estimates are adjusted in such a way to ensure time continuity. The results preserve the growth rates of initial data.

Options and suggestions

The analyst should base the model selection on an ADL framework following a general to specific approach, making use of a set of consistent diagnostics. Intervention analysis is also strongly suggested. Also useful is the analysis of goodness of fit of disaggregated trends (especially in growth rates terms) and sign correlations with indicators. For inexperienced users, the analyst could limit estimation and model selection to classical methods of temporal disaggregation. Denton's adjustment procedure could also be adopted for extrapolation. Naive extrapolation without either a modelling strategy or a consistent statistical treatment is not recommended.

3.4.3 Available software

A temporal disaggregation package, Ecotrim, is in use at Eurostat, which uses the most commonly used methods. R also has some libraries on temporal disaggregation, the results of which should still be tested. ISTAT adopts temporal disaggregation routines of classical regression methods and the Denton approach in a form implemented by the Bank of Italy and incorporated in the Speakeasy program.

The largest set of methods of the ADL(1,1) class referred to in Proietti (2005) is available in Ox. A console version of Ox, available free on the internet, can be used for temporal disaggregation.

3.5 Pure forecasting methods

3.5.1 ARIMA models

Description

ARIMA models are appropriate for modelling time series with trend characteristics, random walk processes, and seasonal and non-seasonal patterns. The ARIMA (p, d, q) model describes the value of a time series as a function of the order of the autoregressive (p), integrated (d), and moving average (q) parts of the model, where p, d, q are non-negative integers. The number of values (lags and differences) involved in AR, I and MA processes is referred to as the order of the respective effects. Additional stationary regressors can be added if they improve the criteria on the basis of which the models are selected – these are known as ARIMAX models.

Autoregressive process

The autoregressive part (AR) is similar to a linear regression model, i.e. it is an autoregressive model that assumes the current time series values depend on past values from the same series. An autoregressive process (of order p) can therefore be presented as:

$$y_t = \alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$$

where α is the constant, $y_{t-1}, y_{t-2}, \cdots, y_{t-p}$ are the past series values (lags), $\varphi_1, \varphi_2, \cdots, \varphi_{t-p}$ are the lags' coefficients and $\varepsilon_t$is a random shock (variable with mean zero and constant variance).

Theoretically, the current time series value may be represented by a model of large numbers of past observations, but in practice one or two lagged observations are enough to obtain a significant estimate of the current observation. The number of periods that affect the time series is indicated by the parameters p (regular autoregressive order).

Moving average process

In a moving average (MA) model, the current value of $\mathrm{y_t}$ is described as a linear function of concurrent shocks (errors) and past shocks (errors). The moving average process (or order q) can be presented as:

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-q}$ are the past shocks (lags), $\theta_1, \theta_2, \cdots, \theta_q$ are the lags' coefficients, ε_t is a random shock (variable with mean zero, constant variances and zero covariances with respect to $\varepsilon_{t-j}, j = 1,2, ..$).

The number of past shocks that affect the time series is indicated by the parameters q.

3.5.1.1 Model selection, estimation, diagnostics and the forecasting stage

Description

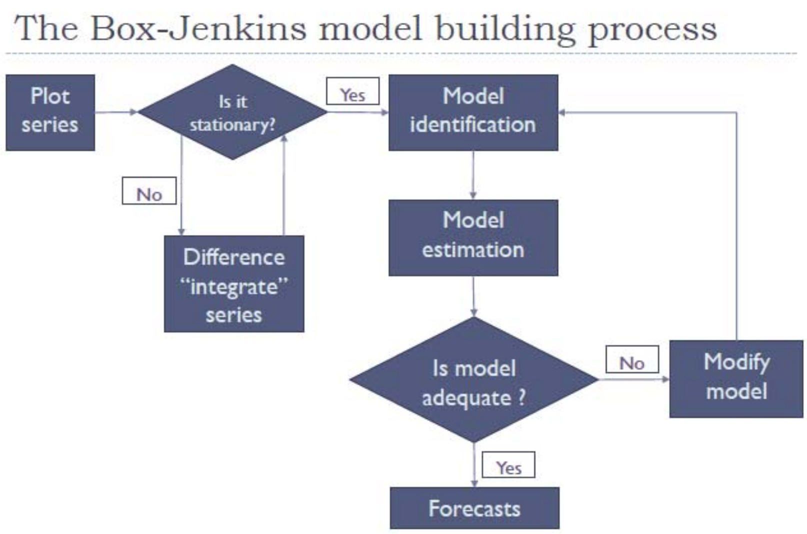
The ARIMA model selection procedure comprises three main iterative steps:

1. Identification (choice of the model parameters)

2. Estimation and diagnostic checking stage
3. Forecasting

The ARIMA model selection process is very well presented by the Box-Jenkins procedure:

Figure 1: The Box-Jenkins model



Source: http://documents.mx/documents/lecture16ts3.html slide number 11

Identification stage

The ARMA models can be applied only to a stationary data series or a series transformed to make it stationary. A time series is stationary if its mean, variance and autocorrelation structure do not change over time:

$$(E(y_t) = \mu_y = const,\ E\left(y_t - \mu_y\right)^2 = \sigma_y^2 = const,\ E\left(y_t - \mu_y\right)\left(y_{t-k} - \mu_y\right) = const.$$

Stationarity should be confirmed by statistical tests like the Augmented Dickey-Fuller Test (ADF) or the autocorrelation function (ACF). If the time series is not stationary, it is often possible to transform it so that it becomes stationary, using one or both of the following techniques:

1. Differencing the data, which means that the series y_t are transformed in the following way:

$$\Delta y_t = y_t - y_{t-1}$$

The new series, Δy_t, is usually stationary. If the new series is again not stationary, it can be differenced a second time.

2. Taking the logarithms or square roots of the series in order to stabilise the variance in the case of non-constant variance time series (in both cases data should be strictly positive).

If time series are stationary, we move on to step 2 to model the identification of parameters. The major tools used at the identification stage are the autocorrelation function (ACF), the partial autocorrelation function (PACF) and the Akaike information criterion (AIC). By looking at the autocorrelation function (ACF) and partial autocorrelation (PACF) plots of the differenced series, we can tentatively identify the significant number of AR(p) and/or MA(q) terms. The Akaike Information Criterion (AIC) is a way of selecting a model from a set of models. The chosen model is the one which has the lowest AIC .

Estimation and diagnostic checking stage

At the estimation and diagnostic checking stage (assessment of the parameters' values) we are checking and diagnosing the model, going back to the model identification stage if the previous assumptions are not satisfied. At this stage, the assumptions of the ARIMA model are checked, e.g. the hypothesis of errors being independently and normally distributed or the statistical significance of coefficients.

We can use maximum likelihood, the method of least squares, or Yule-Walker equations for the parameters estimation:

— The least squares estimator of the parameters $(\varphi, \theta, \alpha)$ is obtained by minimising the sum of squares of model errors.

$$S = \min_{r_i} \sum_{i=1}^{n} {r_i}^2 , r_i = y_i - f(x, \varphi, \theta, \alpha)$$

— The maximum likelihood method is about how to find such estimates of parameters. It maximises the system components and the probability of matching y_i.

$$f(y_1, y_2, \cdots, y_n | \hat{y}_1, \hat{y}_2, \cdots, \hat{y}_1, \sigma^2) \rightarrow max$$

— The Yule-Walker equations are the following set of equations

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0}$$

where $m = 0, ..., p$, yielding p + 1 equations. Here γ_m is the autocovariance function of y_t, σ_ε is the standard deviation of the input noise process, and $\delta_{m,0}$ is the Kronecker delta function.

Forecasting stage

The estimated model is used to generate forecasts and their confidence limits. The forecasting process expression can be found recursively the prediction after step 1 given by:

$$\hat{y}_{t+1} = \alpha + \varphi_1 y_t + \varphi_2 y_{t-1} + \cdots + \varphi_p y_{t-p} + \theta_1 \hat{\varepsilon}_t + \theta_2 \hat{\varepsilon}_{t-1} + \cdots + \theta_q \hat{\varepsilon}_{t-q}$$

Options and suggestions

The ARIMA model selection process should be based on the Box-Jenkins procedure, following all the model selection stages. Beyond statistical significance, evaluation of variables and results should have a sound economic interpretability. Moreover, an automatic model selection procedure (see, for instance, that used by TRAMO-SEATS) could be adopted, if it were accompanied by a reasonable analysis of results and by routine revisions of the models selected.

3.5.1.2 Available software

Description

The main possible software packages for ARIMA modelling and estimation are R, SAS, EViews, Demetra+ and JDemetra+.

R:

R is a free, open-source software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS (more information about R: http://www.r-project.org/). For time series analysis and estimation, we need R package forecast (> install.packages('forecasts')). R has the main functions required for ARIMA modelling, analysis and estimation stages. For the identification stage, R has useful graphical visualisation, estimates of the autocovariance or autocorrelation, and partial autocorrelations functions. R has the possibility to return the best ARIMA model according to either AIC value automatically or to select a model according to choices and criteria made at the estimation and diagnostic checking stage. R is a good tool for forecasting time series. It provides forecasts from models fitted by ARIMA and provides prediction intervals. R has good forecast graphical visualisation possibilities.

SAS:

SAS software provides analogous possibilities for ARIMA modelling and estimation as R (more information: https://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_arima_sect017.htm).

EViews:

EViews, the software for estimation, forecasting, statistical analysis, graphics, simulation, data management, is an all-in-one powerful, graphical object-oriented interface (more information: http://www.eviews.com/Learning/index.html).

However, SAS and EViews are not free and open-source software packages. Their possibilities are very similar and provide qualitative results. The differences between these programs are their graphical interface and their function syntaxes.

Demetra+ and JDemetra+:

Demetra+ and JDemetra+ were developed by the National Bank of Belgium and Eurostat for seasonally adjustment implementation and time series analysis. These programs are user-friendly tools for checking seasonality and seasonally adjusted results, working with time series and forecasting. Demetra+ and JDemetra+ include two seasonal adjustment methods: X-12-ARIMA and TRAMO/SEATS (more information: http://www.cros-portal.eu/content/demetra).

Options and suggestions

The analyst should build up macro-programs using statistical software for ARIMA modelling and estimation and for optimising the entire process of estimation until the final results are obtained. Alternatively, modelling and estimation could be approached manually in some parts of the process.

3.5.2 Structural time series models

3.5.2.1 Model selection, estimation, diagnostics and the forecasting stage

Description

A popular alternative to ARIMA models is provided by structural time series (STS) models. Here the series is expressed in terms of components of interest like trend, cycle, seasonal and the irregular term. A first simple STS model for non-seasonal time series y_t is given by the local linear trend (LLT) model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim (0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + \beta_t + \eta_t, & \eta_t &\sim (0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \zeta_t & \zeta_t &\sim (0, \sigma_\zeta^2) \end{aligned}$$

where μ_t is a stochastic trend with slope β_t and ε_t is a white noise disturbance. It can easily be shown that the trend μ_t is a generalisation of a linear trend $\mu + \beta t$ (where μ is the constant or 'level' and β the deterministic slope) in cases where both variance disturbances of level μ_t and slope β_t, respectively σ_η^2 and σ_ζ^2, are null. Moreover, the LLT model adapts to time series with an adaptive slope: when σ_ζ^2 is positive: the slope is smooth if σ_η^2 is either zero or relatively low with respect to σ_ζ^2; alternatively, the slope is more erratic when the ratio $\sigma_\eta^2/\sigma_\zeta^2$ increases. A positive value for σ_η^2 implies a time series requiring a double difference to become stationary or integrated of order two denoted as I(2). By contrast, when σ_η^2 is positive and σ_ζ^2 is null the trend becomes a random walk with 'drift' or constant slope. This latter occurrence is for difference stationary I(1) time series. To sum up, despite its simplicity a LLT model is flexible enough to fit a variety of time series, depending on the relative size of disturbance variances. Moreover, this framework modernises other modelling approaches and popular forecasting methods such as 'exponential smoothing', popularised by Holt (1957 and 2004) and known as the Holt-Winters method.

The LLT model is generalised to incorporate cycles and/or seasonal components. The basic structural model (BSM) is the form most frequently used for seasonal time series, showing several similarities with the Airline ARIMA(0, 1, 1)(0, 1, 1) model.

STS models have a standard statistical treatment within the Kalman filter which allows quite simple forecasting, estimation of missing observations, regression effects, outliers, seasonal adjustment and temporal disaggregation. The main differences between STS and ARIMA models are that the latter have the advantage of parsimony in model representation, whereas the former are designed according to the analyst's desire to see given components in the form describing the data evolution according to a given model.

Options and suggestions

The best way for the analysis is to fit a general form (BSM or LLT) to the time series under analysis, including, when required, a sound strategy for log transformation, outlier detection, the treatment of deterministic effects and seasonal adjustment. In a second step, the analysis is addressed towards model selection and diagnostic checking. Finally, towards the computation of the desired forecasts, some elements of automation could also be considered in the procedure for model selection, estimation and forecasting. Another way of simplifying the process is to limit model selection and estimation to the more representative forms. What should be avoided is fitting a model of deterministic structural components (like a linear trend or seasonal dummies) to observed time series by means of simple OLS regressions.

3.5.2.2 Available software

The STAMP program, Structural Time Series Analyser and Modeller and Predictor, is the best choice for the adoption of STS models. Other routines are available upon request in Ox (console version available free for research purposes on the internet at http://www.oxmetrics.net/), as referred to in the work of Moauro and Savio (2005). Even if developed in a temporal disaggregation domain, these routines allow estimation of standard univariate and multivariate STS models. Further routines allow quite easy forecasting, testing and diagnostic checking SsfPack must be installed for them to be usable. An Ox package for general state space models and the Kalman filter is also available free on the internet at http://www.ssfpack.com/.

3.6 Multivariate extensions

3.6.1 Introduction

Description

Both ARIMA and STS models are models for univariate time series whose use could be extended to multivariate settings. In a multivariate model, both the target and the explanatory variables are treated as a cross–section of time series. Following Harvey (1989, p.429), under a multivariate setup '… it is assumed that the different series are not subject to any cause-and-effect relationship between them. However, they are subject to the same overall environment, such as the prevailing business climate, and so a multivariate model will seek to link them together'. The main advantage of multivariate models is thus that they overcome the assumption of exogeneity of explanatory variables implicit in regression methods. Furthermore, such models often provide more useful information on the dynamic properties of the series and produce more accurate forecasts. By contrast, their statistical treatment is more complicated, as the number of unknown parameters to be estimated increases rapidly with the number of series to be treated together, and their identifiability might become complicated.

The possibilities include those listed below:

— VAR models (i.e. vector autoregressive models) are the multivariate extension of ARIMA models without the MA component. This class of models is popular for simplicity of statistical treatment.

— Multivariate extension of STS models, i.e. seemingly unrelated time series equation SUTSE models. This class includes dynamic factor analysis for which, given a representation of a cross-section of time series into components (trend, cycle, etc), there might exist a specification that shares certain components. In other words, a reduced number of these components is informative for the entire set, simplifying the model specification. When common factors are found in the trend, the model is co-integrated.

— A very interesting way of approaching the problem of flash estimate is the extension of multivariate settings to models handling mixed frequency data. As discussed in previous sections, in the real world data are often available for different time spans (in the example above, data were monthly and quarterly) and the possibility to forecast together in a unique setting could simplify their statistical treatment. In particular, adopting multivariate models with mixed-frequency data overcomes the multi-step procedure of bridge models, and the flash estimate finds a solution in one step. Here, missing data at the end of the sample – the ragged edge problem - are estimated together in one step once unknown model parameters have been obtained. For recent contributions see, among others, the works by Kuzin et al. (2009), Clements and Galvao (2008) and Banbura at al. (2013).

— Temporal disaggregation has also been treated by the literature in a context of mixed-frequency multivariate time series models. See the contributions by Frale et al. (2010 and 2011) on constructing a euro area monthly indicator of economic activity based on factor models and by Moauro and Savio (2005) and Moauro (2014) on an application to employment based on SUTSE models. In this case estimates of quarterly data are a by-product, as they are obtained through temporal aggregation of monthly estimates.

— A list of other methods is, of course, not exhaustive and there are several other possible modelling strategies. They include the nonlinear class of state-dependent models (SDM) or switching regime models (SRM) introduced by Brian Priestley in 1980. Here, the dynamic paths are governed by a set of autoregressive parameters, a set of moving average parameters and a local intercept, each of them dependent on past information. For a general outline of nonlinear models, the classic references are Priestley (1988), Tong (1990) and Granger and Teräsvirta (1993).

Options and suggestions

The analyst should select an appropriate multivariate model strategy, according to the data available, the relevance of the exercise and the resources available, then conduct a comparative analysis of results with more common univariate frameworks, to understand the real usefulness or value-added of a complex multivariate framework. When a well-experimented multivariate set-up is used, comparative analysis could be avoided. In any case, a proper preliminary analysis of available indicators and diagnostic checking are strongly recommended. The use of complex multivariate frameworks without experience and without comparative analysis using univariate methods should be avoided.

3.6.2 VAR modelling approach

3.6.2.1 Modelling and statistical treatment

Description ([5])

Vector autoregressive (VAR) models have a long tradition as tools for multivariate time series analysis. They became popular when Sims (1980) proposed them as an alternative to simultaneous equation models, which had represented the mainstream for time series modelling since the 1950s. VAR modelling received increasing attention for reasons including the greater availability of longer time series, the need to consider dynamics, the fact that it overcame the problem of the exogeneity property of simultaneous equation modelling, and the availability of a more powerful computing system. VAR models offer a useful framework for structural analysis and variance decomposition, but can be also used for forecasting. They can be used to treat both stationary and integrated time series. More recent developments of VAR relevant for forecasting purposes concern both factor-augmented VARs (FAVAR), mixed-frequency VAR and Bayesian VAR (BVAR).

Vector autoregressive process

A vector autoregressive process of order p – VAR(p) - can be represented as

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t$$

([5]) This chapter draws on Luetkepohl (2007). See it and the references it contains for a more formal and detailed description of the methodology.

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})'$ is a $K \times 1$ vector of time series variables, the A_i ($i = 1,2,..., p$) are ($K \times K$) matrices of parameters and the error term $u_t = (u_{1t}, u_{2t}, \ldots, u_{Kt})'$ is a K–dimensional zero mean white noise process with covariance matrix Σ_u. If $\det(I_K - A_1 z - \ldots - A_p z^p) \neq 0$ for $|z| \leq 1$, y_t is integrated of order zero $I(0)$ and the VAR is stable. However, since most of the economic variables and short-term indicators are non-stationary, it is worth considering the case of non-stable VARs. If the determinant above has a root for z=1 and all other roots outside the unit circle, then some or all of the time series in y_t are integrated (6) and two cases can be distinguished, depending on whether or not cointegrating relationships exist among the time series components (7). In the latter case, y_t can simply be modelled as a VAR in differences:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t.$$

If one or more cointegrating relationships exist, the model should preferably include an additional term in the VAR in differences, to account for the dynamic adjustment of the variables towards their long-run equilibrium. Such models are generally referred to as **vector error correction models** (VECM) and can be represented as

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

where $\Pi = -(I_K - A_1 - \ldots - A_p)$ is a matrix of rank r with $0 < r < K$ (8) and $\Gamma_j = -(A_{j+1} + \ldots + A_p)$ for $j = 1, \ldots, p-1$.

Then, r is the *cointegrating rank* of the process, Πy_{t-1} is the *equilibrium correction* or *error correction term* and captures the long-run relationship among the time series, while the short-run movements are determined by the *short-run parameters* Γ_j s.

VAR models in levels, differences or VECM form are in general well suited for forecasting. Assuming the model and the parameters are known (9), the optimal h-step ahead forecast of y_t in period T (10), $y_{T+h/T}$, is the conditional expectation of y_{T+h} given the information up to T :

$$y_{T+h/T} = E(y_{T+h/T} \mid y_T, y_{T-1}, \ldots) = A_1 y_{T+h-1/T} + \ldots + A_p y_{T+h-p/T}$$

where the forecasts for $h = 1,2,\ldots$ can be computed recursively given the initial condition $y_{T+h/T} = y_T$ for $h = 0$. The forecast error is then given by

$$y_{T+h} - y_{T+h/T} = u_{T+h} + \Phi_1 u_{T+h-1} + \ldots + \Phi_{h-1} u_{T+1}$$

where the $\Phi_i = \sum_{j=1}^{i} \Phi_{j-1} A_j$, $i = 1,2,\ldots$ can be computed recursively given $\Phi_0 = I_K$ and $A_j = 0$ for $j > p$. It is unbiased and its covariance matrix is given by $\sum_y(h) = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi'_j$. The results are valid for both $I(0)$ and

(6) In this case it is said that y_t itself is $I(1)$.

(7) For a survey of the literature on cointegration see, for instance, Johansen (2006) and the references it contains.

(8) If $r=K$, y_t is $I(0)$.

(9) If unknown parameters are replaced by their estimates, as is the case in practice, this has an impact on the precision of the forecasts. The forecast error, and hence the mean square forecast error, is the sum of two components: the difference between the true realisation of the process and the forecast obtained with the true parameters, and that between the latter and the forecast obtained by replacing the parameters with their estimates.

(10) Notice that what follows is also valid for VECMs, as they can be rewritten in level form.

$I(1)$ processes, although the properties of the process of the forecast errors are different in the two cases.

The main drawback of using VARs for forecasting purposes is that the number of parameters increases rapidly with the number of time series in the VAR and the lag order, giving rise to inefficiency in the estimates of the parameters and hence very large confidence intervals of the point forecasts. The loss of efficiency in medium to large VAR models is sometimes solved by using Bayesian VAR models ([11]).

A development that has received some attention in recent literature is the Factor-Augmented Vector Autoregression (FAVAR) model ([12]), in which a small-scale VAR is augmented with dynamic factors that incorporate information from a potentially very large set of indicators.
It is also worth noticing that quarterly VARs with temporally aggregated indicators when one or more of the indicators are available at monthly frequency provide poorer nowcasting performances with respect to mixed frequency VARs (MF-VAR), recently proposed in the literature[13].

Options and suggestions

The analyst should investigate the potential non-stationarity of the time series at hand and possible cointegration between them. Depending on the response, choosing among VAR in levels, VAR in differences or VAR in VECM form, the selection of the set of indicators to include in the VAR could proceed together with the 'target' series. Also crucial is the choice of the lag order p, to be carried out on the basis of suitable information criteria. Finally, the analyst should perform a careful in sample model diagnostic analysis prior to forecasting.

If monthly indicators are available, the possibility of using BVARs or FAVARs or MF-VAR could be considered. The use of an automatic procedure for model selection, estimation and forecasting should be avoided.

3.6.2.2 Available software

Packages for standard and Bayesian VAR analysis are available in all the most commonly used econometric software packages.

R:

In R it is possible to perform VAR analysis with a package called 'vars'; the package 'MSBVAR' provides the possibility of running frequentist, Bayesian and Markov switching VARs ([14]);

SAS:

The IML package VARMAX provides the possibility of running VAR analysis in SAS ([15]);

Stata: Stata has a suite of commands for fitting, forecasting, interpreting, and performing inference on VAR models called 'var' ([16]);

EViews and MATLAB:

VAR analysis can also be performed with EViews ([17]) and MATLAB ([18]);

([11]) See Karlsson (2013) and the references it contains.

([12]) See Bernanke, Boivin and Eliasz (2005); Bai, Li and Lu (2014) and the references it contains.

([13]) See, for instance, Zadrozny (1988); Mariano and Murasawa (2010); Foroni and Marcellino (2013) for a survey of models for mixed-frequency data in general.

([14]) See, for instance, http://cran.r-project.org/web/packages/MSBVAR/MSBVAR.pdf.

([15]) See http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/viewer.htm#etsug_varmax_sect001.htm.

([16]) See, for instance, http://www.stata.com/manuals14/tsvarintro.pdf.

([17]) For example: http://www.eco.uc3m.es/~jgonzalo/teaching/timeseriesma/eviewsvar.pdf.

([18]) See http://it.mathworks.com/help/econ/var-models.html.

3.6.3 Multivariate STS models

3.6.3.1 Modelling and statistical treatment

Description

A multivariate generalisation of STS models of section 3.5.2 is given by SUTSE models, which find a wide treatment in the literature. Given a cross-section of N time series $\boldsymbol{y_t} = (y_{1t}, \dots, y_{Nt})$ it is assumed that each y_{it} for i=1, …, N and time t = 1, …,n is not directly related to the others, although the series are subject to similar influences. $\boldsymbol{y_t}$ is expressed in terms of additive N-dimensional unobserved components, e.g. level $\boldsymbol{\mu_t}$, slope $\boldsymbol{\beta_t}$, and irregular $\boldsymbol{\varepsilon_t}$, which can be contemporaneously correlated, considering non-diagonal covariance matrices, respectively $\boldsymbol{\Sigma_\eta}$, $\boldsymbol{\Sigma_\zeta}$ and $\boldsymbol{\Sigma_\varepsilon}$, for their disturbance terms $\boldsymbol{\eta_t}$, $\boldsymbol{\zeta_t}$ and $\boldsymbol{\varepsilon_t}$. A multivariate LLT is represented as follows:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim (\boldsymbol{0}, \boldsymbol{\Sigma}_\varepsilon) \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim (\boldsymbol{0}, \boldsymbol{\Sigma}_\eta) \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\zeta}_t & \boldsymbol{\zeta}_t &\sim (\boldsymbol{0}, \boldsymbol{\Sigma}_\zeta) \end{aligned}$$

A multivariate LLT model and its generalisation including cycles and/or seasonality has a standard statistical treatment within the Kalman filter when $\boldsymbol{y_t}$ is fully observed at the desired frequency. Otherwise, if the components y_{1t}, …, y_{Nt} are available at different frequencies and there is the need for temporal disaggregation, the Kalman filter is also a viable tool, adopting Harvey and Chung's approach (2000) when data are in levels and Proietti and Moauro's approach (2006) when data are in logs. Complexity in estimation concerns only the number of hyper-parameters pertaining to disturbance covariance matrices, which increases with the scale of the system. If N is large, the Kalman filter is still an efficient instrument for log-likelihood evaluation, even if computationally demanding, but maximisation is no longer feasible through gradient-based methods. An alternative solution is provided by the EM algorithm, efficiently implemented in Moauro (2014).

The SUTSE approach is flexible enough to allow for almost any kind of disaggregation problem and to handle outlier detection, missing observations of both raw and seasonally adjusted time series. Another advantage is that common factor restrictions on level, slope, cycle and irregular components can be introduced very easily by imposing rank restrictions on the covariance matrices of the disturbances driving the components of interest.

The use of tests for the choice of the final form of a SUTSE model is another relevant point. We refer to the locally best invariant (LBI) tests for the form of the trend component and for the existence of common factors in a multivariate structural context. Feasible solutions adopted in the literature consist in using both parametric and non-parametric versions of LBI tests, the latter based on the innovations obtained from the unrestricted form, i.e. the model estimated under the alternative hypothesis.

Options and suggestions

The analyst should build up a multivariate setup of GDP components to be forecast together with related indicators and run a suitable model estimation program for SUTSE models. Establish a model selection strategy from general to specific forms in terms of both time series and components to be included. Check results with those obtained by simple univariate methods to evaluate the possible advantage of adopting a complicated strategy. A quicker option is to forecast the desired GDP components by means of a general SUTSE model limiting the analysis to accuracy of results. The analyst should avoid adopting a SUTSE modelling strategy without either a preventive experimental phase, or an accurate comparative analysis of results with simpler strategies.

3.6.3.2 Available software

As for univariate STS models, the STAMP program also represents the best choice for SUTSE models. Other routines are available upon request in Ox (a console version is available free for research purposes on the internet at http://www.oxmetrics.net/), as referred to by Moauro and Savio (2005) and Moauro (2014). In general, Ox is also a nice environment for developing useful software, using the ready-to-use codes and examples of SsfPack for general state space models. The Kalman filter is also available free on the internet at http://www.ssfpack.com/.

4. Analysis of results

4.1 Comparative analysis of results

Description

The activity of forecasting is often linked to building statistical models and estimating parameters using observed historical data. Choosing a forecasting model from a set of viable candidates should be based on in-sample and out-of-sample performance. Some models may perform fairly well in-sample, but not necessarily out-of-sample. As a result, model selection and evaluating forecast accuracy are two complementary tasks. The quality of a forecasting model should thus be judged by how it compares to other models in terms of out-of-sample forecasting accuracy.

The objective of this chapter is to list different criteria that should be considered when assessing the accuracy of different forecasting models.

Terminology

Throughout this chapter the following terminology will be used:

$$y(t) = \{y_1, y_2, \dots, y_{t-1}, y_t, y_{t+1,} y_{t+2}, \dots, y_T\}$$

A time series starts at time 1 and ends at time T. In this context, the past is given by observations going from $\{y_1, \dots, y_{t-1}\}$, the present is denoted by y_t and the future is denoted by $\{y_{t+1}, \dots, y_T\}$. $\hat{y}_{t+h,t}$ is a forecast for the time period t+h with information available at time t.

Theory

An optimal forecast has a certain number of properties. Hence evaluating a forecast means checking whether the properties of an optimal forecast are met.

We assume that the series $y(t)$ which has to be forecast needs to be a covariance stationary time series. According to the Wold representation, any covariance stationary time series can be presented in the following way.

$$y_t = \mu + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \cdots$$
$$\varepsilon_t \sim WN(0, \sigma^2)$$

Hence an observation y_t at time t is a function of its past errors. If a forecast is optimal, the forecasting error is only a function of future shocks (iid). If this is the case, we have used all the information available at time t in order to predict $y_{t+h,t}$. In the context of the Wold representation, a forecast is a function of past shocks.

$$y_{t+h,t} = \mu + b_h\varepsilon_t + b_{h+1}\varepsilon_{t-1} + \cdots$$

Forecasting errors will then be related to future shocks

$$e_{t+h,t} = y_{t+h} - y_{t+h,t} = \varepsilon_{t+h} + b_1\varepsilon_{t+h-1} + \cdots + b_{h-1}\varepsilon_{t+1},$$

With variance

$$\sigma_h^2 = \sigma^2\left(1 + \sum_{i=1}^{h-1} b_i^2\right)$$

Four key properties of **optimal forecasts**, which we can easily check, are as follows:

- Optimal forecasts are unbiased.
- They have 1-step-ahead errors that are white noise.
- They have h-step-ahead errors that are at most MA(h-1).
- They have h-step-ahead errors with variances that are non-decreasing in h and that converge to the unconditional variance of the process.

4.1.1 Measures of dispersion

1. *Optimal forecasts are unbiased*

   If the forecast is unbiased, then the forecast error has a zero mean. In a 1-step-ahead forecast, simply regress the forecast error series on a constant and use the t-statistic to test the hypothesis that the population mean is zero.

   For h-step-ahead forecasts, it follows from the Wold decomposition that errors are serially correlated even if the forecast is optimal. We can regress the errors on a constant, allowing for MA(q) disturbances with q > (h-1). If available information at t is used optimally, moving-average parameters should not differ significantly from zero beyond h-1.

2. *Recursive Estimation Procedures for Diagnosing Accuracy and Selecting Forecasting Models*

   Recursive estimation describes the process of estimating a model with a small sample of data and adding more and more observations while re-estimating the model each time. The idea is very simple; instead of using all the available data to estimate the forecasting model, take a small subset with k observations and estimate the parameters. Then use k+1 observations and re-estimate the same model. If the relationship is stable, parameter values should stabilise as the sample size grows. Based on this idea, there exist a number of methods to test for parameter stability, such as recursive residuals, the Cusum test or recursive coefficients.

   Another popular approach is a rolling window approach. Economic relations may vary or even break down over time. If a forecasting model relies on data that exhibit such characteristics, it is likely to produce suboptimal forecasts if they are not taken into account. A simple way of detecting such instabilities is to perform a rolling window approach. To implement this, you estimate the model based on the sample $\{y_1, \ldots, y_k\}$ and the estimate is again based on the sample $\{y_{1+j}, \ldots, y_{k+j}\}$. The choice of k and j should be based on the whole sample size. A potential choice is k = 40 and j = 10 for quarterly data.

   A number of statistical software packages include tests for parameter stability. This is the case, for instance, in Eviews.

3. *Assessing optimality with respect to an information set*

   If a forecast is optimal, all future errors are not forecastable. Hence all available information today contains no additional information about the future. We can assess optimality with respect to various sets of information by broadening the information set (including alternative variables/indicators) and estimating regressions of the form

   $$e_{t+h,t} = \alpha_0 + \sum_{i=1}^{k-1} \alpha_i x_{it} + u_t$$

   The hypothesis of interest is that all the α's are 0, which is a necessary condition for forecast optimality as regards the information contained in the x's. Instead of finding many new variables, an alternative and much easier way to implement this idea is the Mincer-Zarnowitz regression. The idea is to regress the true value on a constant and on the forecast values.

   $$y_{t+h} = \beta_0 + \beta_1\, y_{t+h,t} + u_t$$

   If the forecast is optimal, there is no bias and $\beta_0 = 0$. And if the information is optimally used $\beta_1 = 1$. In practice, it is very difficult to find a forecast that meets these conditions.

Options and suggestions

The analyst should proceed by testing for unbiasedness, recursive estimates and assessing optimality with respect to a given information set. Graphical inspection by plotting the one-step head forecast residuals is also to be recommended.

4.1.2 Summary error statistics

Suppose that a time series $y(t)$ has to be forecast and there are two possible models ('model 1' and 'model 2') which have been selected on the basis of their in-sample characteristics. They produce forecasts $\hat{y}^{M1}_{t+h,t}$ and $\hat{y}^{M2}_{t+h,t}$ of the time series $y(t)$ at time *t* based on the series itself up to y_t and possibly with outside information available to the forecasters (exogenous variables). The output of these models for the one-step-ahead forecast is:

Forecast horizon	Model 1	Model 2
h = 1	$\hat{y}^{M1}_{t+1,t}$	$\hat{y}^{M2}_{t+1,t}$

The one-step-ahead forecast errors are $e^{M1}_{t+1,t} = \hat{y}^{M1}_{t+1,t} - y_{t+1}$ and $e^{M2}_{t+1,t} = \hat{y}^{M2}_{t+1,t} - y_{t+1}$ for the two models respectively.

Can we have any objective statement on which model is doing a better job? In this section, we consider the issue of whether various models can be compared fairly. There are a number of competing criteria. They could include dispersion, mean absolute errors or mean squared error, ,to mention but a few. This chapter focuses on the most common ones.

The penalty received for an error ($e_{t+h,t} = \hat{y}_{t+h,t} - y_{t+h}$) depends on the selected loss function L($\hat{y}_{t+h,t}$, y_{t+h}).

Options

— Mean error (ME)

The mean error measures bias. For a good forecast, the expected error should be zero.

— Error variance (EV)

The error variance measures the dispersion of the forecast errors.

$$EV = \frac{1}{T}\sum_{t=1}^{T}(e_{t+h,t} - ME)^2$$

— Mean squared error (MSE)

Mean squared errors have the disadvantage of not preserving units. For instance, if a time series is in euros, the unit of the mean square errors become euros squared, which is difficult to interpret.

Often the square roots of these measures are used to preserve units, yielding the

— Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e^2_{t+h,t}}$$

— Root mean square percent error

$$RMSPE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[100(e_{t+h,t}/y_{t+h})]^2}$$

To conclude this chapter, a standard methodology would be to take log differences, as most time series are not stationary, and use the root mean square error (RMSE) to assess the accuracy of the forecasts.

— Theil's inequality coefficient

Theil's inequality coefficient is a popular and useful statistic. The idea is to compare forecasts with a naive forecast.

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_{t+1,t} - y_{t+1}\right)}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_{t+1,t}\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_{t+1}\right)^2}}$$

The numerator of Theil's inequality is equal to the RMSE. The scaling, however, is such that U will always lie between 0 and 1. The closer U is to 0, the better the model's forecasting power.

4.2 Forecast Encompassing

Furthermore, there exists the possibility of combining several forecasts to yield a single forecast which may perform better than any of the original ones. In many cases, there is no forecast that is significantly better than the others. In some cases, different forecasts can be combined to yield a composite forecast which does better than all the other single forecasts. The question is whether one forecast incorporates all the relevant information in competing forecasts.

Let us focus on the case of two forecasts, $\hat{y}_{t+h,t}^{M1}$ and $\hat{y}_{t+h,t}^{M2}$, coming from two different models. Consider the regression of the realised values y_{t+h} on the forecasts for the respective point in time:

$$y_{t+h} = \beta_{M1}\hat{y}_{t+h,t}^{M1} + \beta_{M2}\hat{y}_{t+h,t}^{M2} + \epsilon_{t+h}$$

If $(\beta_{M1}, \beta_{M2}) = (1,0)$, then model M1 forecast-encompasses model M2 and if $(\beta_{M1}, \beta_{M2}) = (0,1)$, then model M2 forecast-encompasses model M1. For other (β_{M1}, β_{M2}) values, neither model forecast-encompasses the other, and both forecasts contain useful information about y_{t+h}.

4.3 Forecast Combination

Failure of each model's forecasts to encompass the other model's forecasts indicates misspecification of both models. In such cases there may be gains from forecast combination. The regression-based forecast combination method is a way of usefully incorporating forecasts based on different models into a composite forecast that outperforms the individual forecasts. Typically, we simply estimate the regression

$$y_{t+h} = \beta_{M1}\hat{y}_{t+h,t}^{M1} + \beta_{M2}\hat{y}_{t+h,t}^{M2} + \epsilon_{t+h}$$

We thus build a composite forecast on the basis of past forecasts of the models and observed realisations of the forecast variable. To evaluate the composite forecasts, we can use the same methods as for the evaluation of the forecasts coming from an individual model. If the composite forecast outperforms the individual forecasts, the estimated vector (β_{M1}, β_{M2}) enables the forecaster to combine any future forecasts based on the M1 and M2 models into a composite forecast.

Options and suggestions

The analyst should use several measures and tests for forecast encompassing and evaluating forecast combination. A forecaster wants to know which error measure should be used. The answer is, however, dependent on the situation. If the objective is to forecast a single series, then it is perfectly fine to use the mean squared error as an accuracy measure. If there are time and/or resource restrictions, it may be enough to use just one measure. Graphical inspection is also a recommended practice, especially when it is focused on plots of different forecasts against the true realisation.

References

Section 1 - Introduction

Bloem at al. (2001), *Quarterly National Accounts Manual Concepts,Data Sources and Compilation*, International Monetary Fund.

Elliott G., Granger C.W.J., Timmermann A. (2006), *Handbook of Economic Forecasting*, North-Holland.

Eurostat (2001), *Handbook on price and volume measures in national accounts*, Office for Official Publication of the European Communities, Luxembourg.

Eurostat (2010), *Handbook on quarterly national accounts*, Office for Official Publication of the European Communities, Luxembourg.

Eurostat (2013), *European system of accounts ESA 2010*, Office for Official Publication of the European Communities, Luxembourg.

Section 2.5 – Outlier detection and their removal

Eurostat, 'Seasonal Adjustment eLearning course material'.

Available at: https://ec.europa.eu/eurostat/cros/content/e-learning-courses-seasonal-adjustment-0_en

Gomez V., Maravall, A. (2001), 'Automatic modelling methods for univariate series', *Documento de Trabajo,* Nº 9808, Banco de España - Servicio de Estudios.

Tsay R. (1986), 'Time Series Model Specification in the Presence of Outliers', *Journal of the American Statistical Association*, Vol. 81, No. 393 (Mar.1986), pp.132-141.

Section 3.3.1 – ADL models

Asteriou, D., Hall, S. G. (2011), *Applied Econometrics* (Second edition). Palgrave Macmillan.

Chatfield, C. (1996), *The Analysis of Time Series,* 5th ed. Chapman & Hall, New York.

Engle, R. F., Granger, C. W. J. (1987), 'Co-integration and Error Correction: Representation, Estimation and Testing', *Econometrica*, 55(2), pp. 251-276.

Hamilton, J. D. (1994,. *Time Series Analysis*, Princeton University Press.

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, pp. 549-563.

Pankratz, A. (1991), *Forecasting with Dynamic Regression Models,* John Wiley and Sons, New York.

Stock J. H., Watson M. W. (2007), *Introduction to Econometrics*, 2nd edition, Pearson Education Inc.

Section 3.3.2 – Dynamic factor models

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset market', *Econometrica*, 51(5): 1281–1304, September 1983.

Fornaro, P. (2013), 'Predicting Finnish economic activity using firm-level data'*, Working Papers* 2/2013, Statistics Finland. Available at: http://www.stat.fi/tup/julkaisut/tiedostot/julkaisuluettelo/ywrp2_201300_2013_12316_net.pdf

Groen, J.J.J & Kapetanios, G. (2009), 'Model selection criteria for factor-augmented regressions', *Staff Reports*, 363, Federal Reserve Bank of New York.

Josse, J. & Husson, F. (2012), 'Selecting the number of components in principal component analysis using cross-validation approximations', *Computational Statistics & Data Analysis*, 56(6), pp. 1869–1879, 2012.

Section 3.4 – Indirect approach

Chow G., A.L. Lin (1971), 'Best linear unbiased interpolation, distribution and extrapolation of time series by related series', *The Review of Economics and Statistics*, 53, 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, pp. 99-102.
Fernàndez R.B. (1981), 'A methodological note on the estimation of time series', *The Review of Economics and Statistics*, 63, pp. 471-478.

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

Proietti T. (2005), 'Temporal Disaggregation by State space Methods: Dynamic Regression Methods Revisited', ISTAT, Final report of Commissione di studio sul trattamento dei dati ai fini dell'analisi congiunturale Incaricata di formulare proposte relative alle strategie da utilizzare per la disaggregazione temporale nei conti economici trimestrali, pp.85-116.

Section 3.5.1 – ARIMA models

Box and Jenkins (1976), *Time Series Analysis: Forecasting and control*, Revised edition, San Francisco, Holden Day.

Brockwell, P.J. and Davis, R.A. (2003), *Introduction to Time Series and Forecasting*, Second edition, Springer.

Grandell J. (2015), 'Time series analysis'. Available at: http://www.math.kth.se/matstat/gru/sf2943/ts.pdf.

Judge, G.G., Carter Hill R., Griffiths W.E., Lutkepohl H. (1988), *Introduction to the Theory and Practice of Econometrics*. 2nd. edition, New York, John Wiley & Sons.

Nielsen H.B. (2015), 'Univariate Time Series Analysis, ARIMA models'. Available at: http://www.econ.ku.dk/metrics/Econometrics2_04_II/Slides/univariatetimeseries_2PerPage.pdf

Section 3.5.2 – Structural time series models

Doornik, J. A. (2001), *Ox 7.0 – An Object-Oriented Matrix Programming Language*, London, Timberlake Consultants Ltd.

Grassi S. and Proietti T. (2014), 'Characterizing Economic Trends by Bayesian Stochastic Model Specification Search', *Computational Statistics and Data Analysis*, 71, pp. 359-374.

Harvey A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Holt C. (1957), 'Forecasting seasonal and trends by exponentially weighted moving averages', *ONR Research Memorandum*, 52, Carnegie Institute of Technology, Pittsburg, Pennsylvania.

Holt C. (2004), 'Forecasting seasonal and trends by exponentially weighted moving averages', *International Journal of Forecasting,* 20 (1), pp. 5-10.

Koopman S. J., Harvey A. C., Doornik J. A., Shephard N. (2006), *Structural Time Series Analiser and Modeller and Predictor STAMP 8*, Timberlake Consultants Ltd.

Koopman, S. J., N. Shephard and J. A. Doornik (1999), 'Statistical algorithms for models in state space using SsfPack 2.2', *Econometrics Journal*, 2, pp. 107–160.

Moauro F. Savio G. (2005), 'Temporal disaggregation using multivariate structural time series models', *Econometrics Journal*, Volume 8, pp.214-234.

Proietti T. and Grassi S. (2014), 'Stochastic trends and seasonality in economic time series: new evidence from Bayesian stochastic model specification search', *Empirical Economics*, forthcoming.

Winters P. R. (1960), 'Forecasting sales by exponentially weighted moving averages', *Management Science*, 6, pp. 324-342.

Section 3.6.1 and 3.6.3– Multivariate extensions

Banbura M., Giannone D., Modugno D., Reichlin L. (2013), 'Now-casting and the real-time data flow', W*orking paper series,* No. 1564, July 2013, European Central Bank.

Clements, Michael P., Galvao Ana Beatriz (2008), 'Macroeconomic Forecasting With Mixed-Frequency Data', *Journal of Business & Economic Statistics*, American Statistical Association, Vol. 26, pp.546-554.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2010), 'Survey Data as Coincident or Leading Indicators', *Journal of Forecasting*, 29, pp. 109-131.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2011), 'EUROMIND: A Monthly Indicator of the Euro Area Economic Conditions', *Journal of the Royal Statistical Society - Series A*, no. doi: 10.1111/j.1467-985X.2010.00675.x.

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

Harvey A. C., Chung C. (2000), 'Estimating the underlying change in unemployment in the UK', *Journal of the Royal Statistical Society*, A 163, pp. 303–339.

Kuzin V., Marcellino M. and Shumacher C. (2009), 'MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area', *Discussion paper series 1 economic studies*, N.07/2009, Deutsche Bundesbank.

Moauro F. e G. Savio (2005), 'Temporal disaggregation using multivariate structural time series models', *Econometrics Journal*, 8, pp. 214-234.

Moauro (2014), 'Monthly employment indicators of the euro area and larger Member States: real time analysis of indirect estimates', *Journal of Forecasting*, Volume 33, Issue 5, pp.339-349.

Priestley M. B. (1980), 'State dependent models: a general approach to nonlinear time series analysis', *Journal of Time Series Analysis*, 1, pp. 47–71.

Priestley M.B. (1988), *Nonlinear and nonstationary time series analysis*, Academic Press.

Tong H. (1990), *Nonlinear time series, a dynamical system approach,* Oxford University Press.

Section 3.6.2– VAR modelling approach

Bai J., K. Li and L. Lu (2014), 'Estimation and inference of FAVAR', *MPRA Paper,* No. 60960.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005), 'Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly Journal of Economics*, Vol.120, Issue 1, pp. 387–422.

Foroni C. and Marcellino M. (2013), 'A survey of econometric methods for mixed-frequency data', *Working Paper* 2013/6, Norges Bank.

Johansen S. (2006), 'Cointegration: a survey', in Mills T.C. and Patterson K. (eds) *Palgrave Handbook of econometrics*, Volume 1, Econometric theory, Palgrave Macmillan, Basingstoke.

Karlsson, S. (2013), 'Forecasting with Bayesian Vector Autoregression', *Handbook of Economic Forecasting*, Volume 2, Part B, pp. 791–897.

Luetkepohl H. (2007), 'Econometric analysis with Vector Autoregressive models', *EUI Working Papers*, ECO 2007/11.

Mariano, R., and Y. Murasawa (2010), 'A coincident index, common factors and monthly real GDP', *Oxford Bulletin of Economics and Statistics*, 72(1), pp. 27-46.

Zadrozny, P. A. (1988), 'Gaussian-Likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies', *Econometric Theory*, 4(1), pp. 108-124.

Section 4 - Analysis of results

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica*, 51(5), pp.1281–1304, September 1983.

Diebold, X. Francis (2006), *Elements of Forecasting*, *Thomson Higher Education*, Fourth edition, Mason, USA.

Fornaro, P. (2013), 'Predicting Finnish economic activity using firm-level data', *Working Papers* 2/2013, Statistics Finland. Available at: http://www.stat.fi/tup/julkaisut/tiedostot/julkaisuluettelo/ywrp2_201300_2013_12316_net.pdf

Groen, J.J.J & Kapetanios, G.(2009), 'Model selection criteria for factor-augmented regressions', Staff Reports 363, Federal Reserve Bank of New York.Josse, J. & Husson, F. (2012), 'Selecting the number of components in principal component analysis using cross-validation approximations', *Computational Statistics & Data Analysis*, 56(6), pp.1869–1879, 2012.

Annex - Questionnaire on estimation methods

Introduction

The National Accounts Working Group decided in its meeting on 22-23 May 2013 to establish a Task Force to assess the feasibility to produce a flash estimate of the euro area and European Union quarterly GDP at 30 days after the end of the reference period. For this purpose the input of national t+30 quarterly GDP estimates of Member States will be necessary. (The publication of national t+30 estimates will solely depend on the Member States' decision.)

In order to support Member States in producing t+30 GDP estimates, the Task Force aims to produce an overview of estimation methods and procedures that are currently used by countries. Such an overview will enable countries to learn from each other's methods and estimation practices.

For this purpose we would like to ask you kindly to complete the questionnaire below. The first part of most questions has a multiple choice character: you can simply tick the box that applies. The second part of most questions is an open question: you are requested to provide some additional information, if needed/applicable.

Part A - General information

1. Is your country engaged in GDP t+30 flash estimation?
   ☐Yes officially with published results
   ☐Yes experimentally
   ☐No but we produced some estimations in the past
   ☐No but some studies has been carried out
   ☐Not at all
   Please provide further information on your experience:……………………………………………….
   …….

2. Is your country involved in GDP t+45 flash estimation?
   ☐Yes officially with published results
   ☐Yes experimentally
   ☐No but we produced some estimations in the past
   ☐No but some studies has been carried out
   ☐Not at all
   Please provide further information on your experience:……………………………………………….
   …….

3. How long is the experience of your country in GDP t+30 flash estimation?
   ☐More than 10 years
   ☐Between 5 and 10 years
   ☐Less than 5 years
   ☐Started in occasion of the Task Force
   Please provide further information on your experience:……………………………………………….

4. How long is the experience of your country in GDP t+45 flash estimation?
   ☐More than 10 years
   ☐Between 5 and 10 years
   ☐Less than 5 years
   Please provide further information on your experience:……………………………………………….
   …….

Part B - Main characteristics of the experiment

5. How detailed is your t+30 days estimate experiment?
    a. Accounting approach
        ☐ From the production side
        ☐ From the expenditure side
        ☐ Both approaches
    b. Price evaluation
        ☐ Limited to volume measures
        ☐ Including current price measures
        ☐ current price measures are inflated from volumes
        ☐ volumes are deflated from current prices
        ☐ mixed
    c. Seasonal adjustment
        ☐ Limited to seasonal adjusted data
        ☐ Extended to seasonally and calendar adjusted data
        ☐ Extended to raw estimates
    d. Frequency
        ☐ Quarterly
        ☐ Monthly
        ☐ Other

    Please provide further information on your experience:……………………………………………….
    ……

6. Is your estimates relative to data after transformation?
    a. Data transformation ([19])
        ☐ No transformation, i.e. estimation in levels
        ☐ Estimation in logarithms
        ☐ other
    b. order of differencing
        ☐ quarterly difference
        ☐ annual difference
        ☐ other

    Please provide further information on your experience:………………………………………………

Part C - Preliminary analysis

7. In the current experiment of GDP t+30 flash estimate, which is the role given to preliminary analysis?
    ☐ Detailed analysis of indicators
    ☐ Analysis limited to indicators currently used for QNA or t+45 GDP flash estimate
    ☐ Any preliminary analysis

    Please provide further information on your experience:……………………………………………….
    ……

.

([19]) For the approximation of growth rates with log-differences, if estimation refers to growth rates please tick 'estimation in logarithms' and choose the appropriate order of differentiation.

8. In the current experiment of GDP t+30 flash estimate, which type of indicators do you use?
   - ☐ Official short term statistics
   - ☐ Business and consumer surveys
   - ☐ Administrative data
   - ☐ Big data
   - ☐ other

   Please provide further information on your experience:……………………………………………….
   ……

9. At which detail is carried out preliminary analysis?
   a. From the production side
      - ☐ Following the wider split of components (please specify # of industries ….…)
      - ☐ Aggregated split of components (please specify # of industries …….)
      - ☐ Only the total
      - ☐ Production side not estimated
   b. From the expenditure side
      - ☐ Following the wider split of components (please specify # of functions/products ….)
      - ☐ Aggregated split of components (please specify # of functions/products …..)
      - ☐ Only the total
      - ☐ Expenditure side not estimated
   c. From both sides
      - ☐ Following the wider split of components (please specify # of functions/products …..)
      - ☐ Aggregated split of components (please specify # of functions/products …..)
      - ☐ Only the total

   Please provide further information on your experience:……………………………………………….
   ……………………………………………………………………………………………………….

10. Is preliminary analysis of source data/indicators accompanied by a graphical or other analysis?
    a. Simple comparison of data/indicators with the correspondent target GDP component
       - ☐ In levels (please specify type of transformation/standardization ……………………..)
       - ☐ In differences or growth rates (please specify the order …….)
    b. Advanced tools
       - ☐ Time series components (trend, seasonal, cycle and irregular components)
       - ☐ Correlograms
       - ☐ Analysis in the frequency domain
       - ☐ Analysis to identify the presence of outliers
       - ☐ Fit of data/indicators to related GDP component by means of a first simple model

    Please provide further information on your experience:……………………………………………….
    ………………………………………………………………………………………………………

Part D – Forecasting exercise

11. Which is your model strategy?
    a. Direct approach
       - ☐ ADL models
       - ☐ Dynamic factor models
       - ☐ Multivariate models (please specify)
       - ☐ Bridge models
       - ☐ Other (please specify ………………………)

b. Indirect approach
☐ Denton adjustment methods
☐ Chow-Lin/Fernandez/Litterman regression approach
☐ State space methodology
☐ Multivariate approach
☐ Other (please specify ………………………)

c. Pure forecasting methods
☐ ARIMA models
☐ Structural time series models
☐ Other (please specify ………………………)

Please provide further information on your experience:……………………………………………….
………

12. Do you use forecast combination methods?
☐ Simple combination
☐ Advanced methods
☐ Other

Please provide further information on your experience:……………………………………………….
………

13. Do you adopt methods of analysis of results?
☐ Measures of dispersions
☐ Summary error statistics
☐ Other

Please provide further information on your experience:……………………………………………….

Concise presentation of the questionnaire and brief overview of responses received

The questionnaire contains 13 questions, split into four sections:

- 4 questions on general information
- 2 questions on main characteristics of the experiment
- 4 questions on preliminary data analysis
- 3 questions on the forecasting exercise.

The form is structured in such a way as to avoid overburdening Task Force members. Most questions are closed. There is space to provide more details where appropriate.

The general part of the questionnaire is designed to investigate Member States' experience of producing GDP flash estimates at t+30 and t+45 days. It also provides useful material that links up to the rest of the questionnaire.

Part B tracks the general approach of the experiment in terms of accounting method, price evaluation, seasonal and calendar adjustment and type of data transformation. **Part C**, which deals with preliminary analysis, allows to understand several elements of the work done by Member States before estimation: in particular type of collected indicators (e.g. hard and/or soft data), sectoral detail, accounting approach, and if these data are object of graphical analysis or other type of statistical analysis. **Part D** is for understanding model strategy: in particular the adoption of a direct or indirect approach, or a pure forecasting method. The final questions deal with the use of forecast combination methods and adopting measures for the analysis of results.

The responses received from Member States taking part in the Task Force may be summarised as follows:

- As regards **general information**, most Member States (10) view the t+30 GDP flash estimate as an experimental exercise. A few cases (3) lead to official estimates, while only 2 cases represent sporadic research. In seven cases, the method was used for the first time under the task force's auspices. In four cases its use started recently, while another four Member States have more experience. In all cases but two, the majority of Member States are involved in flash estimation at t+45 days.
- As regards the main characteristic, the experiment is conducted using one accounting approach only most of the time (11 occurrences): 10 times from the production side and only one from the demand side. By contrast, four Member States have developed both demand and production approaches. The exercise is limited to volume measures only in three cases, also including current price measures in all other occurrences. As regards seasonal adjustment, too, it emerges quite clearly that the experiment is in general rather complex: most Member States (12) run a process of flash estimation which involves both calendar and seasonal adjustment, and only three of them limit the forecasting exercise to data which have already been seasonally adjusted. Frequency of estimation is generally quarterly (12 Member States), and rarely monthly (three Member States). Estimation is performed on data transformed into logarithms in six cases, whereas in seven cases data are treated directly in their levels. Data are often modelled in differences (10 cases), mostly quarterly difference (five cases), annual differences (three cases), with another three Member States for which both orders of differentiation are relevant;
- As regards preliminary analysis, only one country says it does not perform any type of analysis. The objective of the analysis is always official short-term statistics, also accompanied by business and consumer surveys (8 cases), administrative data (10 cases), whereas big data are never quoted. The detail concerns the wider split of GDP components in the majority of cases, as well as an aggregated level by nine Member States. Simple graphs are used by 10 Member States, whereas very often the study involves more advanced elements such as time series components (4 cases), correlograms (5 cases), the presence of outliers (5 cases), or even the fit of indicators to related GDP components by means of a first simple/default model estimation (6 cases);
- The answers to the questions on the forecasting exercise show that Member States use both the direct and the indirect approaches (six of each), but that it is even more common for them to adopt a pure forecasting method of estimation (10 times). Within the direct approach, two Member States use ADL models, three use dynamic factor models, and one uses bridge models as well as a naïve method which extrapolates the target GDP component by means of the two-month average of a related indicator. Of the indirect approaches, Denton’s adjustment method is applied once, while three Member States use the Chow and Lin/Fernandez/Litterman regressions and, interestingly, another two use a multivariate method. Among users of pure methods of forecasting, ARIMA models prevail (nine times) over both structural time series models and Holt-Winter methods (once each). Five Member States adopt methods of forecast combinations (advanced in two cases) and 10 Member States perform an analysis of results based either on summary error statistics (nine times) or measures of dispersion (twice) at the final stage.

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

KS-TC-16-006-EN-N

Publications Office

doi:10.2785/51658