Structural Time Series Analyser
and Modeller and Predictor

STAMP™ 8

Siem Jan Koopman
Andrew C. Harvey
Jurgen A. Doornik
Neil Shephard

Published by Timberlake Consultants Ltd
www.timberlake.co.uk
www.timberlake-consultancy.com
www.oxmetrics.net

Contents

Figures

Tables

Preface

The full name of *STAMP* is *S*tructural *T*ime series *A*nalyser, *M*odeller and *P*redictor. Structural time series models are formulated directly in terms of components of interest. Such models find application in many subjects, including economics, finance, sociology, management science, biology, geography, meteorology and engineering. STAMP bridges the gap between the theory and its application – providing the necessary tool to make interactive structural time series modelling available for empirical work. (Another such tool is *SsfPack*, by Koopman, Shephard and Doornik (1999), which provides more general procedures but with a programmatic interface, see `www.ssfpack.com`.)

STAMP uses the Kalman filter and related algorithms to fit unobserved component time series models. We are excited to present the new version 8 of STAMP, which provides another big step forward from the previous version. The new version is updated for the new OxMetrics environment and it therefore provides even higher standards in program functionality: by clicking the mouse a few times, everyone is able to start with a full exploratory, statistical or econometric analysis of the time series at hand using the powerful capabilities of the STAMP 8.

Earlier versions of the program were written by Andrew Harvey and Simon Peters, while the data management side was dealt with by Bahram Pesaran. These projects were supported by the Economic and Social Research Council. Version 5 was entirely rewritten in C by the current authors of STAMP. Much of the data management and the graphical interface of STAMP 5 was shared with the PcGive 7 and 8 programs of Jurgen A. Doornik and David F. Hendry. STAMP 6 was the first version of STAMP in which the front-end program GiveWin has been made separate from the econometric module STAMP. Other modules included PcGive™, TSP™ (by TSP International) and X12Arima for GiveWin (based on X12-ARIMA program of the US Census Bureau). STAMP 7 was a separate module of the OxMetrics 4 program that was developed by Jurgen A. Doornik. The current multivariate program STAMP 8 is developed as a module for the OxMetrics 5 program.

Scientific Word in combination with MikTEX and DVIPS eased developing the documentation for version 7 in LaTeX, further facilitated by the more self-contained nature of STAMP 7 and its in-built help system. We thank Maxine Collett who typed and retyped various sections of the earlier versions of this book. We are also thankful to Tirthankar Chakravarty for his excellent help on editing this book.

We thank the London School of Economics and Political Science for hosting the developments of early versions of STAMP. Furthermore, we are grateful for the comments and encouragements of John Aston, Charles Bos, Jim Durbin, Irma Hindrayanto, Maarten van Kampen, Roy Mendelssohn, Marius Ooms, Jeremy Penzer, Tommaso Proietti, Marco Riani, Thomas Trimbur, Brian Wong and the many old and new users of STAMP who send us suggestions for improvements and bug reports. We hope that you will continue to write us (`s.j.koopman@feweb.vu.nl`) with ideas for improving and extending STAMP (and *report any bugs*!).

The Authors

Siem Jan Koopman is Professor of Econometrics at the VU University Amsterdam and Tinbergen Institute Amsterdam. He is author of the textbooks 'Time Series Analysis by State Space Methods' (with J. Durbin) and 'An Introduction to State Space Time Series Analysis' (with J. Commandeur). He is an Associate Editor of *Journal of Forecasting*.

Andrew C. Harvey is Professor of Econometrics at the University of Cambridge. He is the author of the textbooks 'Time Series Models' and 'Econometric Analysis of Time Series'. He has also published the monograph 'Forecasting, Structural Time Series Models and the Kalman filter'. He is a Fellow of the Econometric Society and the British Academy. He is also an Associate Editor of the *J Time Series Analysis*.

Jurgen A. Doornik is a Research Fellow of Nuffield College, Oxford. He is the main developer of *OxMetrics*, the originator of *Ox*, an object-oriented matrix programming language, and author (with David F. Hendry) of *PcGive* and *PcFiml*.

Neil Shephard is Professor of Economics at the University of Oxford and Research Director of the Oxford-Man Institute of Quantitative Finance. He is a Fellow of the Econometric Society, the British Academy, Nuffield College and is an Associate Editor of *Econometrica*.

Part I

Prologue

Chapter 1

Introduction

1.1 Overview of the STAMP book

(1) *Read this chapter and the next for an introduction to the* STAMP *system.*
This introductory chapter explains how to use the documentation and provides a brief overview of the special features that make STAMP such a powerful tool for time series modelling. Chapter 2 explains how to start STAMP, and gives a brief overview of OxMetrics. The OxMetrics program facilitates data input, text editing, graphics output and much more.

(2) *Read Chapter 3, at the beginning of Part II, for an introduction to how STAMP may be used to model and forecast a time series, and Chapters 4–7 to learn in detail about how structural time series models are applied in practice.*
Chapter 3 provides a simple hands-on introduction to structural time series modelling with STAMP. It is followed by three chapters which teach you how to go about structural time series modelling, and build models using explanatory variables and interventions. The examples provided use a variety of real data sets, from the consumption of spirits in the UK to the water level of the Nile. These data sets are introduced in §1.10. The last chapter in this part of the book deals with applications to economic and financial time series.

(3) Chapters 4–7 require virtually no prior knowledge of structural time series modelling or the theory of time series analysis in general. Hence, the way in which the program may be used is readily accessible to the practitioner. Statistical theory is kept to a minimum with more detailed points being discussed in slanted typescript. Chapter 6 is fully devoted to the analysis and modelling of multivariate models in STAMP.

(4) Chapter 8 systematically describes how univariate and multivariate models are formulated, estimated and evaluated. Users already familiar with structural time series modelling may find it more efficient to consult this chapter first and omit the earlier chapters in Part II.

(5) *Turn to Chapters 9 and 10 in Part III for a description of the statistical output of* STAMP. These chapters define the models, estimation procedures and test statis-

tics available in STAMP. An Appendix lists the commands which are available for the batch use of STAMP.

1.2 General information

STAMP is an interactive menu-driven time series modelling program. Version 7, to which this documentation refers, runs under Microsoft Windows XP. Particular features of the program are its ease of use, edit facilities, flexible data handling, model building, and forecasting capability.

STAMP is designed for modelling time series data using unobserved components. The full name of the program is **S**tructural **T**ime series **A**nalyser, **M**odeller and **P**redictor. The current available module is for the fitting and checking of univariate time series models built out of time varying components of interest such as trends and seasonals. The models can include explanatory variables, interventions, lagged endogenous variables and many other features. The theory of structural time series models is laid out in Harvey (1989), with some additional developments described in Harvey and Shephard (1993) and Harvey (2006). A selection of key papers on unobserved components time series models (statistical theory, estimation and testing, methodology, forecasting) are collected in Harvey and Proietti (2005). An introductory textbook about the structural time series model and the state space approach to time series analysis is Commandeur and Koopman (2007). Basic aspects of structural time series modelling are described from a Bayesian perspective in West and Harrison (1997).

The ease with which structural time series models can be implemented in practice has been demonstrated repeatedly with the previous version of STAMP which was written for the MS-DOS operating system. The new version has been rewritten to enable use of the exciting possibilities offered by the OxMetrics system. The interactive framework of OxMetrics, with extensive use of graphics, offers new scope for modelling. The power of modern machines, coupled with the theoretical and computational advances of state space methods described in Durbin and Koopman (2001) and Koopman, Shephard and Doornik (1999), means that, structural time series modelling is a practical option.

The documentation fully explains the time series methods, the modelling approach and the techniques used, as well as bridging the gap between the time series theory and empirical practice. Detailed tutorials describe the use of the program while teaching the methods. The aim is to provide an operational approach to time series modelling using the most sophisticated yet easy-to-use software available. The context-sensitive on-line help system offers help on the program.

This chapter discusses the special features of STAMP and describes how to use the documentation. The OxMetrics system is described in a separate book, however the next chapter also provides a brief overview of its main features.

1.3 Features introduced in STAMP 7

Many new features were introduced in version 7 of STAMP and they are all kept in the current version 8. Some of these features are

- **Missing values** can be treated in this version of STAMP. In all levels of the analysis, missing observations are accounted for automatically. Model-based estimates of the missing values can be produced.
- **Confidence intervals** can be included in the graphs with estimated components. The **predicted**, **filtered** and **smoothed** estimates of the components can be presented simultaneously.
- The observation **weight functions** that are used for the estimation of the unobserved components are given as output. The associating **spectral gain and phase functions** can also be produced.
- Model-based **forecasting** and **backcasting** can be carried out.
- **Time-varying regression coefficients** can be included as part of the model. Three specifications can be selected: (i) random walk, (ii) spline specification (or smooth trend) and (iii) return-to-normality (deviations from a fixed coefficient follow an autoregressive process of order 1).
- Unobserved stationary **autoregressive processes** of orders 1 and 2 can be included in the model together with three stochastic cycle components.
- Higher order **smooth cycles** with band-pass filter properties can be included in the model. Higher order **smooth trends** can also be included and lead effectively to application of Butterworth filters.
- The **trigonometric seasonal component** consists of separate processes for the different seasonal frequencies. These processes can be selected separately in the model. Also different seasonal variances can be attached to the seasonal processes.
- More flexible options for the handling and estimation of **parameters** are provided. The parameters can be treated without transformation.
- Dates for **outliers and breaks** can be stored and remain part of the model once selected.

1.4 New in STAMP 8

Many new features have been introduced in version 8 of STAMP. The most notable are:

- **The multivariate capabilities of the STAMP program are back**. The multivariate structural time series model where the unobserved components become vectors and the disturbance variances become disturbance variance matrices can be considered for the analysis of a set of multiple time series. The number of multivariate options has increased considerably compared to earlier versions of the program:

- *Select components by equation*: Different components can be selected for different equations. This enables the user to analyse time series with different dynamic characteristics jointly. For example, consider two time series where one series may be subject to seasonal dynamics while the other series does not require a seasonal component. The trends of the two time series may move together. STAMP 8 allows the user to select a seasonal component for the first series but not for the second series. This applies to all components in STAMP: trend, seasonal, cycle, autoregressive, irregular, time-varying regressions, etc.
- *Select regressions and interventions by equation*: An option for selecting different explanatory variables and interventions for different equations has been available in STAMP versions 5 and 6. However, the current facility of distributing explanatory variables over different equations has improved and is more flexible.
- *Design a dependence structure for each component*: Multivariate models in STAMP 5 and 6 were limited in their choice of variance matrices: only full variance matrices of different ranks could be considered. A reduced-rank variance matrix implies common features in multiple time series. This option remains in STAMP but the specification has changed slightly. The disturbance variance matrix imposes a dependence structure within the component vector (between the different equations). This dependence can be designed by the user in a simple way and for each component separately. For example, the cycle component in equation 1 can be forced to depend on the cycles in equations 2 and 3 only.
- *In STAMP 8 different variance matrices for different disturbances can be chosen*: The range of variance matrices includes scalar and diagonal matrices, scaled matrices of ones (when applied to the slope component, this implies balanced growth) and one rank plus diagonal matrices. The latter case implies that a vector component can be decomposed into common and idiosyncratic effects. In many applications, these different specifications can be interpreted easily and can be highly interesting.
- *The multivariate options extend to all models introduced in STAMP 7*: This includes the higher-order smooth trend models, the higher-order (bandpass) cycle components and the (vector) autoregressive components of orders 1 and 2.
- *Missing observations*: They can also be handled within multivariate time series models. This allows the interpolation of missing observations through time but also through different time series.
- *Forecasting of multivariate time series made simple*: In particular, STAMP 8 allows the incorporation of available future observations for the explanatory variables in the database. Furthermore, future observations of dependent variables are considered in graphical presentations of forecasts and for the measurement of forecast accuracy (using standard measures such as the root mean squared forecast error (RMSE) and the mean absolute percentage error (MAPE).
- *Estimation of parameters in multivariate time series models is based on exact procedures*: The diffuse initialisation of the Kalman filter is implemented, the exact likelihood function is computed and the score function with respect to variance parameters is computed analytically and fast. This leads to a robust estimation procedure in STAMP 8 and a relatively fast estimation process.
- *The number of graphical output for multivariate models is increased*: STAMP 8 offers an easy handling of the graphical output. An option for graphics output selection for each equation is provided. The powerful tools in OxMetrics 5 to edit graphical output are fully available to STAMP 8 users.

- **Automatic outlier and break detection**: Another major development in STAMP 8 is the implementation of a new automatic detection procedure for outliers and breaks in univariate and multivariate time series models. The following features are available:
  - STAMP 8 is able to propose a set of potential outliers and trend breaks for univariate and multivariate time series. It is a basic but effective two-step procedure based on the auxiliary residuals. First the selected model is estimated and the diagnostics are investigated. Then a first (larger) set of potential outliers and trend breaks are selected from the auxiliary residuals. After re-estimation of the model, only those interventions survive that are sufficiently significant. In the multivariate case, this selection procedure is carried out jointly for each equation in the model.
  - After the automatic selection, the results are reported. All considered outliers and breaks are kept in the intervention dialog and they can be deleted from the model or added to the model in the usual way and implemented as in STAMP 7. For future use, the interventions can be saved. It prevents the manual input of outliers and breaks altogether.
  - The automatic selection procedure can be repeated with the inclusion of a fixed set of explanatory and intervention variables.
- **Other new features**:
  - Each parameter in the models of STAMP 8 can be edited directly. Parameters can be kept fixed at a particular value. Variances can be kept fixed at values relative to a particular variance of another component (q-ratio). This

facility also applies to multivariate models.

- General forecasting options have been extended and made more flexible. The number of output options for prediction and forecasting have been increased. Future values of explanatory variables available in the database can be used for the forecasting of dependent variables.
- More output diagnostics are presented for predictions (one-step and multi-step), auxiliary residuals and weight and gain functions.

1.5 The special features of STAMP

(1) *Ease of use*

- STAMP is easy to use, being a **fully interactive and menu-driven** approach to time series modelling: pull-down menus offer available options and dialog boxes provide access to the available functions.
- STAMP has a very **high level of error protection** making it suitable for students or practicioners acquiring experience in time series modelling or with computers, for live teaching in the classroom or fraught late-night research.
- STAMP provides an extensive context-sensitive **help system** explaining both the program usage and the time series methodology.
- The **user friendliness and screen presentations are of the high quality** provided by the OxMetrics system, operating in an edit window to allow documentation of results as analysis proceeds, with easy review of previous results and cutting and pasting within or between windows.

(2) *Advanced graphics*

- STAMP supports **text and graphics together on screen**, with easy adjustment of graph types, layout and colours.
- **Graphs can be documented and edited** via direct screen access with reading from the graph.
- **Time series and cross-plots** are supported with flexible adjustment and scaling options. Spectra, correlograms, histograms and data densities can be graphed.
- **Forecasts and components** can be graphed in many combinations.

(3) *Flexible data handling*

- The **data handling system provides convenient storage** of large data sets with easy loading to STAMP either as a unit, or for subsamples or subsets of variables.
- **Excel and Lotus spreadsheet files** can be loaded directly, or using 'cut and paste' facilities.

- **Large data sets** can be analysed, with as many variables and observations as memory allows.
- Database variables can be transformed by a **calculator**, or by entering **mathematical formulae** in an editor with easy storage for reuse: the database is easily viewed, incorrect observations are simple to revise, and variables can be documented on-line.
- **Appending** across data sets is simple, and the data used for estimation can be any subset of the data in the database.
- **Several data sets** can be opened simultaneously, with easy switching between the database.

(4) *Model specification and estimation*

- STAMP is designed for **modelling time series** data using models constructed out of interpretable components which are usually specified after the inspection of some of the many graphical displays available in the program. The models are easily setup by clicking onto menus and dialogs in the program.
- The models are **estimated** using maximum likelihood: powerful numerical optimisation algorithms (which exploit analytic derivatives) are embedded in the program with easy user control. Graphical procedures allow the visual inspection of profile likelihood functions.
- Efficient **signal extraction** of time varying components, such as trends and seasonals, is easy. Seasonal adjustment and model-based detrending are available as specific options.
- A **Batch language** allows automatic estimation and evaluation of models, and can be used to prepare a STAMP session for teaching.

(5) *Diagnostic checking and predictive testing*

- **Equation mis-specification tests are automatically provided**. Checks are included on the innovations from the model for residual autocorrelation, heteroskedasticity and normality. More detailed diagnostic checking, including plots of residual density functions is possible. Auxiliary residuals, which are smoothed estimates of the noise in the components of the model, can be used to detect outliers and structural breaks.
- Post-sample and within-sample **predictive testing** is available with full graphical support, including plots of predictions against outcomes with one-step and multi-step error bars.

(6) *Forecasting*

- It is easy to produce forecasts of the series, together with error bars, and forecasts of trend, seasonal and cyclical **components**.
- Forecasts can be made conditional on future values of the **explanatory variables** which are constructed by assuming a particular growth rate, are fed in

manually or are themselves forecasted by building a structural time series model. This makes it easy to forecast under different **scenarios**.

(7) *Output*

- **Graphs can be saved** in several file formats including for later recall, further editing, and printing, or for importing into many popular word processors, as well as directly by 'cut and paste'
- **Results window** information can be saved as an ASCII (human readable) document for input to most word processors, or directly input by 'cut and paste'.
- Model residuals and recursive output can be **stored in the database** for additional graphs or evaluation.

1.6 Basics of the program

1.6.1 Data storage and input

The primary mode of data storage is the IN7/BN7 format. This gives complete compatibility with the other OxMetrics modules. The IN7/BN7 format is based on a pair of files with extensions **.IN7** and **.BN7**. The latter is a binary file containing the actual data, whereas the former holds the information on the contents of the binary file such as variable names and sample periods. The information file is human-readable, whereas the binary file is not. Several real data sets are supplied with STAMP and used in the tutorials.

OxMetrics can read also read other data file formats, including the following spreadsheet files:

- Excel: **.XLS** files;
- Lotus: **.WKS** and **.WK1** files.

Once the data has been entered, they are best stored in IN7/BN7 format. The program checks for potential overwriting of files, and if this is going to occur it offers options for proceeding to overwrite, appending to a file that already exists or selecting another file name. Files may be stored in different directories or drives to those on which the program resides. On input, a search procedure across directories and drives is easily implemented within the Open file dialog if the desired file is not found initially.

The data options allow easy archiving of data. For data samples, reference is by the absolute date in the form *Year Period* to *Year Period*, for example, 1965 1 to 1985 3. Whenever a sample choice has to be made, STAMP will show the maximum available. STAMP can also handle daily data and intra-daily data, see OxMetrics documentation.

1.6.2 Menus and dialogs

The STAMP program is interactive and menu-driven. That is, at each stage a set of options is available, any one of which may be selected. Choices are mainly made by placing the mouse cursor on the menu option and clicking with the left button. It is of course only possible to choose options that make sense: unavailable options are shown in dimmed format. The heart of the STAMP program lies in the **Formulate** and **Test** menus. The dialogs set out the available settings and the user chooses the appropriate configuration. Many dialogs have a default which can be accepted simply by clicking the OK key (using the mouse) or pressing the Enter key ↵. Changing the settings in the dialogs can be carried out using the mouse. In addition, the `Tab` key can also be used to move the cursor between options. Variables are typically contained in 'List boxes' and the `Spacebar` plays an important role in that it marks a selection of variables for an operation. The OK button tells the program to carry out the specified operation and/or move on to the next dialog in the sequence, while Cancel returns to the Results or Graphics window.

Other important keys are: the `Arrow` keys (←↑↓→) and paging keys (`PgUp`, `PgDn`, `Home` and `End`) which move around windows; the `Esc` key which cancels instructions or escapes from dialogs; `F1` which provides context-sensitive help.

Most operations can be conducted by clicking the left mouse button to select, clicking twice to implement, and clicking and holding down the left button while moving the mouse to drag the cursor (for example, to mark a block of text for cutting and pasting). Any combination of mouse and keyboard is feasible when either would do the job alone.

Several dialogs have list boxes in which a selection can be made, for example to select for a list of variables. Occasionally, only one selection is possible, but often multiple variables can be selected. With the keyboard, you can mark a range of variables by holding the shift key down while using the arrow up or down keys. With the mouse there is more flexibility:

- single click to select one variable;
- hold the left mouse button down to select a range of variables;
- hold the `Ctrl` key down and click to select additional variables;
- hold the `Shift` key down and click to extend the selection range.

1.6.3 Help system

STAMP incorporates an extensive, cross-referenced help system which offers advice before crucial decisions are made and can be accessed at any time, either by pressing the `F1` key (giving context-dependent help) to call the help menu. If in doubt when using STAMP, press `F1` to get help: this facilitates its use as you are much less likely to get stuck. The help system is html-based which means that it works as a webpage. The help system starts the default web-browser on your computer automatically to display the context-dependent help.

Many help, advisory, and warning messages are interspersed throughout the program to pop up as needed. Most of these are self-explanatory queries or comments.

1.6.4 Results storage

All results are shown in the Results window of OxMetrics as calculations proceed, but are not stored on disk unless specifically requested. Results window storage is normally in files with an .OUT extension. Such files can be on a different drive or directory from STAMP. OxMetrics will issue a warning when trying to overwrite an existing file, offering the choice between overwriting the file, appending to the file, or selecting a new destination file name.

1.7 Using STAMP documentation

The documentation comes in four main parts.

I. **Prologue** — introduces the main features provided by the program and sketches how to use it.
II. **Tutorials on Structural Time Series Modelling** — explain how STAMP can be used to model and forecast real data.
III. **STAMP Tutorials** — explain the general working of the program, graphics, data input and modelling.
IV. **Statistical Output** — sets out the statistical details of the models, describes how they are estimated and provides precise definitions of all the statistics and plots in the output.

Assuming the program has been installed, you should proceed as follows after reading the remainder of this chapter:

(a) If you are new to structural time series modelling, follow the example in Chapter 3 to see what the program does at its most basic level.
(b) If you are familiar with the ideas of structural time series modelling, follow the tutorial in Chapter 8.
(c) If you wish to begin by using your own data set, work through the first three tutorials in Part III.
(d) If you would like to read about some of the ways in which STAMP allows you to explore new approaches to modelling economic and financial time series, turn to Chapter 7.
(e) If you want to know about the statistical treatment of the models and the definitions of the diagnostic output, turn to Chapters 9–10.

To use the documentation, either check the index for the subject, topic, menu or dialog that seems relevant, or look up the part relevant to your current activity in the **Contents**, and scan for the most likely keyword. The references point to relevant publications which analyse the methodology and methods embodied in STAMP.

Equations are numbered as (chapter.number); for example, (8.1) refers to equation 8.1, which is the first equation in Chapter 8. References to sections have the form §chapter.section, for example, §8.1 is the first section in Chapter 8. Some sections are starred. These can be left out on first reading for they are concerned with some of the more technical issues.

The typeface conventions adopted in the tutorials are as follows:

(1) Instructions to be typed into the keyboard are shown in `Typewriter` font.
(2) Keyboard commands, including combinations such as `Alt+f`, are also shown in this font.
(3) Menus names are given in **bold**, unless they are just being mentioned in passing.
(4) Dialog names, sometimes appearing as options in menus, are given in Sans serif.
(5) The labels on 'buttons' in dialogs are shown in Sans serif.
(6) The labels used to describe option boxes in dialogs are enclosed in 'raised commas'.
(7) *Italics* are used for technical names.
(8) *Slanted* is used for points to be stressed or comments which are discussing a point not directly related to the main line of argument.

1.8 Citation

To facilitate replication and validation of empirical findings, the STAMP book should be cited in all reports and publications involving its application. The appropriate reference is Koopman, Harvey, Doornik and Shephard (2007).

1.9 World Wide Web

Consult `http://stamp-software.com/` for pointers to additional information relevant to the current and future updates and versions of STAMP.

1.10 Tutorial data sets

STAMP comes with a number of real data sets which are used to illustrate how the program is used for structural time series modelling. All are stored in IN7/BN7 format. This section describes the data sets, in alphabetical order, and discusses some of the questions which they raise. In some cases we indicate how STAMP sets out to answer these questions and take the opportunity to illustrate some of the graphical output of the program.

1.10.1 ENERGY: energy consumption in the UK

This consists of a number of series of quarterly energy consumption (in millions of useful therms). The series are for coal, gas and electricity for the 'Other industries' and 'Other final users' of the economy. The series are given in logged and unlogged form. Thus 'ofuGASI' is the logarithm of gas consumption by Other final users. For further details and a listing of the data, see Harvey (1989).

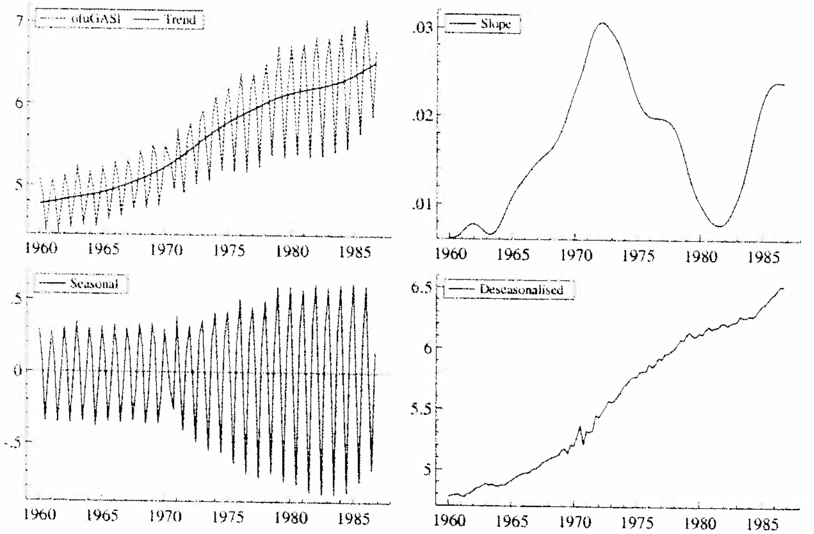


Figure 1.1 Analysis of gas consumption by final users. Original series and estimated: trend, slope of trend, seasonal and seasonally adjusted series.

It is informative to decompose series of this kind into trend, seasonal and irregular components. STAMP does this, and some of its output is shown in Figure 1.1 for 'ofuGASI'. Although the model was formulated in logarithms, the graphs show the implications for the raw data. The first graph shows the trend; there is a dramatic increase in gas usage with the introduction of natural gas from the North sea in the early 1970s. This also shows up in the graph to the right which plots the annualised growth rate of the trend throughout the sample period. The seasonality is shown in the graph in the bottom left hand corner in terms of its multiplicative effect on the trend. Thus at the end of the period it can be seen that, on average, consumption in the winter quarter is running at about 70% above the underlying trend. The greater dispersion in the seasonal pattern over time is due to a higher proportion of gas being used for heating

as usage increased in the 1970s. The final graph shows the seasonally adjusted series produced by fitting the structural time series model.

1.10.2 EXCH: daily exchange rates for the US dollar

The four series here are the first differences of the logarithms of daily exchange rates of the dollar against the pound, deutschmark, yen and Swiss franc. The data were recorded at the end of each weekday from 1/10/81 until 28/6/85. Thus the sample size is 946. The interesting issue here is the volatility in the series. Figure 1.2 shows the absolute value of the first difference of the logs of the series together with an estimate of the corresponding underlying volatility in the Deutschmark based on a stochastic volatility model fitted using STAMP.

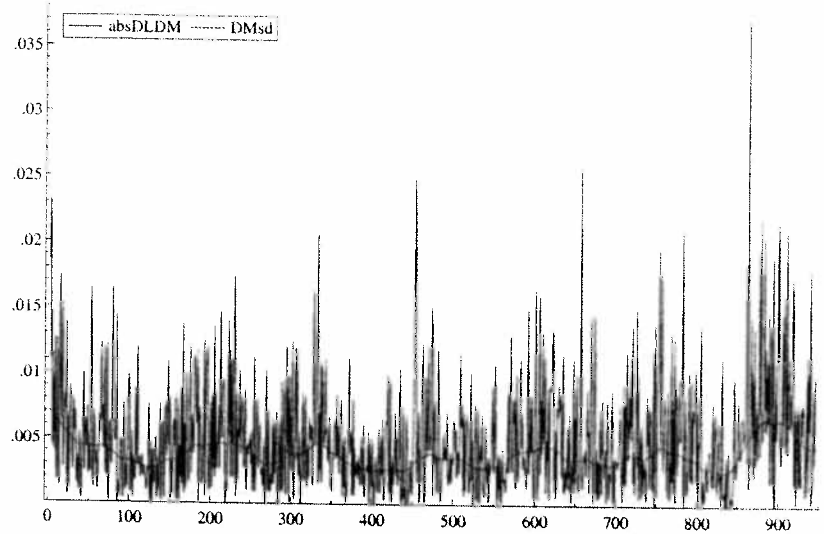


Figure 1.2 The estimate of volatility for the DM.

1.10.3 INTEREST: short- and long-term interest rates

The series 'long' is the monthly yield on 20-year UK gilts. The 'short' series is the 91 day UK Treasury bill rate. Questions to ask are whether the rates contain an underlying component and whether this component emerges if the series are modelled jointly. The data are from Mills (1993, p. 225).

1.10.4 MINKMUSK: minks and muskrats in Canada

The file MINKMUSK contains two series showing the numbers of skins of minks and muskrats traded annually by the Hudson Bay Company in Canada from 1848 to 1909. These data have been studied extensively in the time series literature and as a result provide a useful test-bed for new techniques. There is a known prey-predator relationship between the two species and this gives rise to interlinked cycles. Chan and Wallis (1978) carried out analysis and modelling within a multivariate ARMA framework. They first detrended by quadratic regression on both series. By contrast, STAMP allows the fitting of trend components as part of the overall model. The remaining movements in the series can be captured by a first-order vector autoregression.

1.10.5 NILE: level of the Nile

This is a series of annual observations on the volume of the flow of the Nile in $\times 10^8$ cubic metres; see Balke (1993). Interesting issues which arise include the possibility of a structural break with the construction of the Aswan dam in 1899 and the existence of outliers. A cycle might also be present.

1.10.6 RAINBRAZ: rainfall in north-east Brazil

This is a series of annual observations on rainfall in Fortaleza in North-East Brazil, starting in 1849. The unit of measurement is centimetres, with the data being recorded to the nearest millimetre. The series is one of the longest records of tropical rainfall available, and it has been subject to a great deal of study since the region inland from Fortaleza frequently suffers from severe droughts. The main issue concerns the existence of a cycle, or cycles, in the pattern of rainfall; see, for example, Kane and Trivedi (1986).

1.10.7 SEATBELT and SEATBQ: road casualties in Great Britain and the 1983 seat belt law

The SEATBELT data consists of monthly observations on the numbers of drivers, front seat passengers and rear seat passengers who were killed or seriously injured (KSI) in road accidents in cars in Great Britain. Data on the number of kilometres travelled and the real price of petrol is also included. Harvey and Durbin (1986) used these data to carry out an extensive study of the effect of the seat belt law of January 31st, 1983.

The file SEATBQ contains only the last observation in each quarter, (that is March, June, September and October) in order to make the analysis more manageable.

The data provide a good illustration of the way in which the intervention analysis of the effect of the seat belt law can be extended by using control groups.

1.10.8 SPIRIT: consumption of spirits in the UK

The SPIRIT file contains the per capita consumption of spirits in the UK from 1870 to 1938, together with per capita income and relative price; see Prest (1949). The consumption of spirits can only be partly explained by income and price, presumably because of a trend component representing changes in tastes and social conditions. Such variables cannot be measured, but including a deterministic time trend leaves considerable serial correlation in the residuals. The solution is to include a stochastic trend component in the regression model.

1.10.9 TELEPHON: telephone calls to three different countries

This file contains data on the logarithms of paid minutes of telephone calls from Australia to three different countries.

1.10.10 UKCYP: consumption, income and prices in the UK

The data on quarterly UK real personal disposable income (Y), consumer non-durables (C) and the GDP deflator (P) can be found in the file named UKCYP. Lower case letters denote the variables in logarithms.

1.10.11 USYCIMP: US macroeconomic time series

This file consists of data on quarterly, seasonally adjusted, logarithms of observations on three real, per capita US series. These are 'private' GNP (that is, not including government expenditure), consumption and investment, denoted 'y', 'c' and 'i' respectively. In addition, the file includes the price level as represented by the GNP deflator, 'p', and 'm', the money supply, or, more precisely, M2; see King, Plosser, Stock and Watson (1991).

The first question which often arises with such economic data is what are its main features, particularly with respect to cyclical behaviour. Economists sometimes refer to such features as 'stylised facts'. STAMP offers one possible way of obtaining a set of stylised facts. For example, we can fit a model consisting of a trend and a cycle to GNP. Figure 1.3 shows the result of fitting such a model and its implications for forecasting. The data are in logarithms and the forecasts are for five years from the end of 1985. The final panel shows the forecasts with a band of one RMSE on either side.

1.10.12 USmacro07: more recent US macroeconomic time series

This file contains the latest US data on key variables and essentially updates the previous file. It contains quarterly data on real U.S. Gross Domestic Product, Investment and Consumption from 1947(1) to 2007(2), obtained from the Department of Commerce (website: www.bea.gov). The logarithms are LGDP, LINV and LCONS. The Consumer

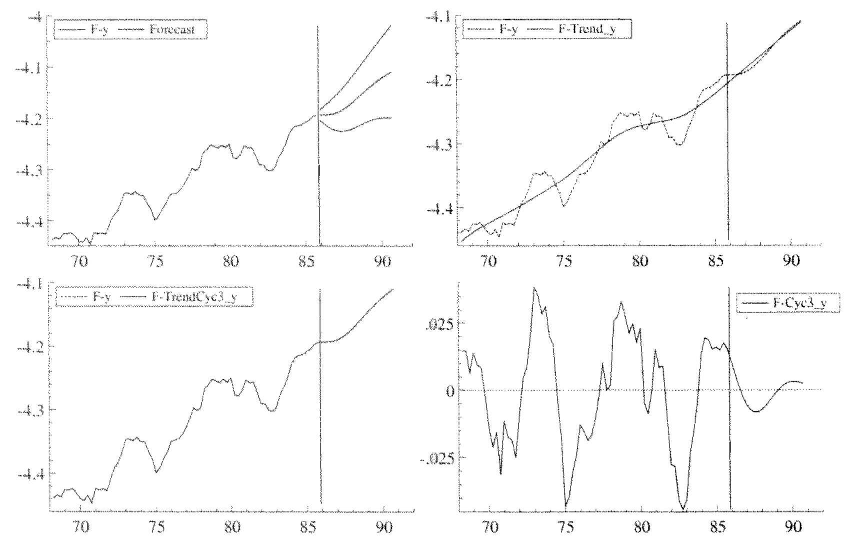


Figure 1.3 Stylised facts of US GNP. Series and estimated trend and cycle. Forecast and forecast confidence intervals.

Price Index CPI is from the U.S. Bureau of Labor Statistics (website: `www.bls.gov`) and the series INFLcpi is the (annualized) rate of inflation, measured as the first differences of the quarterly CPI multiplied by four hundred. INFLdef is constructed in the same way but from the GDP deflator. Unlike the real series, CPI is not seasonally adjusted.

1.11 Data sets used in exercises

The following data sets are used in the exercises which appear at the end of some of the tutorials:

AIRLINE: This is a well-known data set consisting of the number of UK *airline passengers* in thousands from January 1949 to December, 1960; see Box and Jenkins (1970).

EMPL: The two series in this file, 'empl' and 'output', are the logarithms of *employment and output in UK manufacturing*. The data are seasonally adjusted and employment is measured in thousands while output is an index with the value in 1983 set to 100. The dynamic employment-output equation needs a stochastic trend component to allow for changes in productivity; see Harvey, Henry, Peters and Wren-Lewis (1986).

ICEVOL: This data set consists of 220 points, corresponding to intervals of 2,000 years, on oxygen-18 levels measured from deep sea cores. These measurements act as proxies for *global ice volume*; see Newton, North and Crowley (1991). One reason for being interested in forecasting from this data set is the question of whether greenhouse effect warming could be offset by a future ice age. The data are thought to contain at least three cycles, but it is an open question as to whether or not they are strictly deterministic.

LAXQ: This consists of quarterly observations of *US exports to Latin America*; see Bruce and Martin (1989). There is the possibility of outliers due to dock strikes.

PURSE: This file contains the number of *purses snatched* in the Hyde Park area of Chicago each lunar month. Since the data consist of small integers, it is interesting to explore transformation, such as taking square roots, before a model is fitted.

1.12 STAMP and PcGive

STAMP shares the same structure as PcGive with regard to data handling and the output of results. Furthermore the conventions adopted in the modelling dialogs are similar. As a result users of PcGive are not only able to use data files directly with STAMP, but will also find the program easy to use.

Despite the superficial similarities, the analysis and modelling carried out by STAMP is quite different to that in PcGive. The programs are designed for different purposes and are essentially *complementary*. PcGive is primarily an econometrics packages, while STAMP addresses a wide range of time series applications in many disciplines. Indeed the structural time series approach which it is designed to implement is much more of a competitor to the ARIMA approach originally popularised by Box and Jenkins (1970). The case for using structural time series models rather than ARIMA models is set out at some length in Harvey (1989), West and Harrison (1997) and Durbin and Koopman (2001).

PcGive is primarily a program for carrying out single equation regression, but with the option of using techniques such as instrumental variables and non-linear least squares. STAMP will also carry out regression, but this is not its prime purpose and for serious economic work, PcGive is to be preferred for its more comprehensive facilities and output. PcGive enables the user to carry out some univariate time series analysis, but this is within an autoregressive framework, stressing the role of unit root tests. It is not the same as STAMP's modelling and analysis of components.

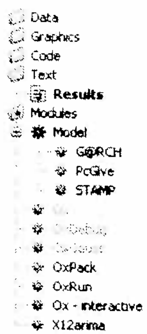
Chapter 2

Getting Started

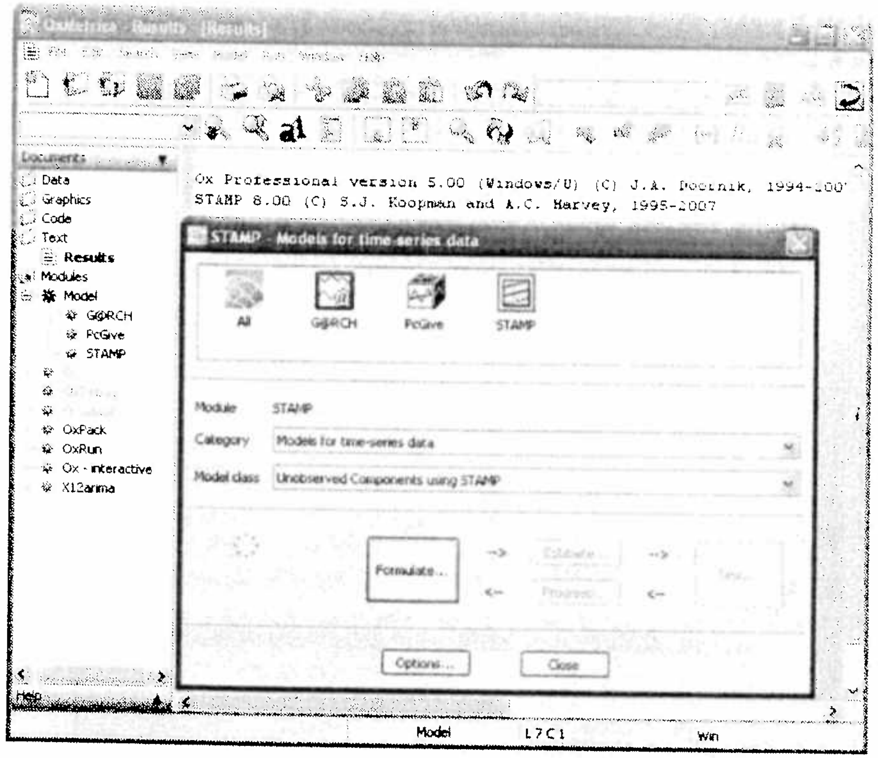
We now discuss some of the basic skills required to get started with STAMP. We will give a short introduction to the services provided by OxMetrics. OxMetrics provides the data which STAMP analyses, and receives all the text output and graphs which you create in STAMP. When you start STAMP, it automatically starts up OxMetrics. Note that STAMP can also be started directly from OxMetrics. The aim of this tutorial is just to load data and create graphs: for instructions on how to transform data using the calculator or algebra, consult the OxMetrics book.

2.1 Starting STAMP

First start OxMetrics, then STAMP can be started from the Modules menu of OxMetrics:



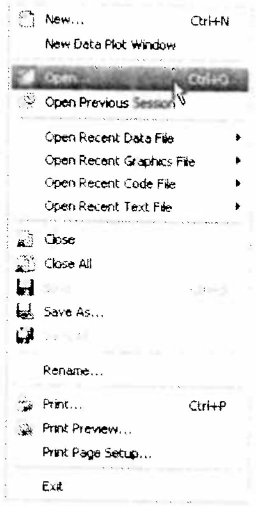
Alternatively, start STAMP from the taskbar or from the OxMetrics group. If this is the first time you have used STAMP, your initial screen could look like the capture shown below.



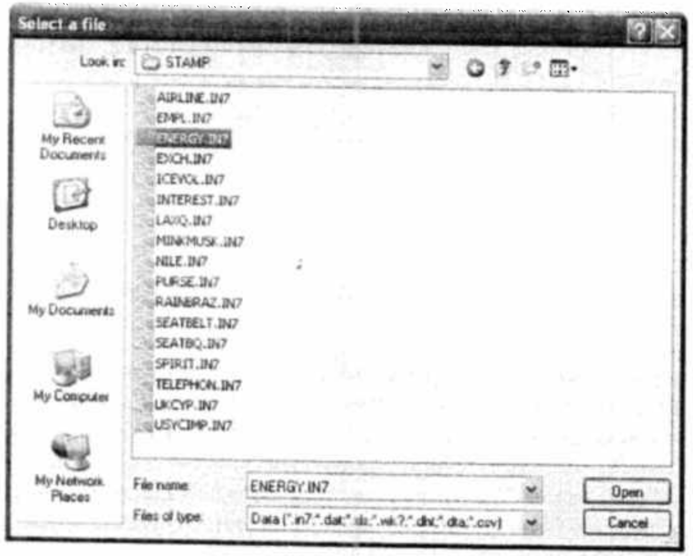
2.2 Loading and viewing the tutorial data set

Without data, STAMP cannot operate, so the first step is to load data into OxMetrics which STAMP can then access. Here we load the energy consumption data set, which was introduced in §1.10.1. The data set is called ENERGY.IN7, and it contains a number of series on quarterly energy consumption in the UK. The IN7 file is a human-readable file, describing the data. There is a companion BN7 file, which holds the actual numbers (in binary format, so this file cannot be edited). OxMetrics can handle a wide range of data files, among them Excel (.XLS) and Lotus files (.WKS and .WK1), and of course plain human-readable (ASCII) files. You can also cut and paste data between Excel and OxMetrics. Details are in the OxMetrics book.

To load the tutorial data set, access the File menu in OxMetrics:

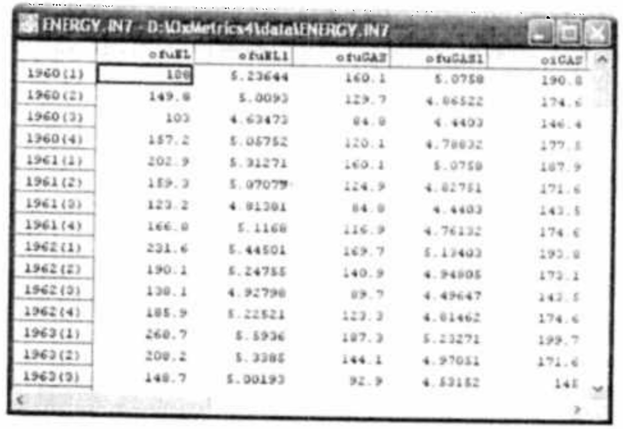


and choose Open. If you installed in the default directory structure the data will be in the `\Program Files\STAMP` directory, so locate that directory and select ENERGY:[1]



[1] You may or may not see the .IN7 file extension of the data files. This depends on the settings in the Explorer options. ENERGY does have a little OxMetrics picture, to indicate that the file type is associated with OxMetrics.

The data file will be loaded, and displayed in OxMetrics:



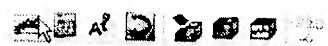
Double clicking on the variable name shows the documentation of the variable if any is available. It would also allow renaming the variable. The data can be manipulated, much like in a spreadsheet program. Here we shall not need these facilities. Do not click on the cross of this window: this closes the database, thus removing it from OxMetrics and so from STAMP.

2.3 OxMetrics graphics

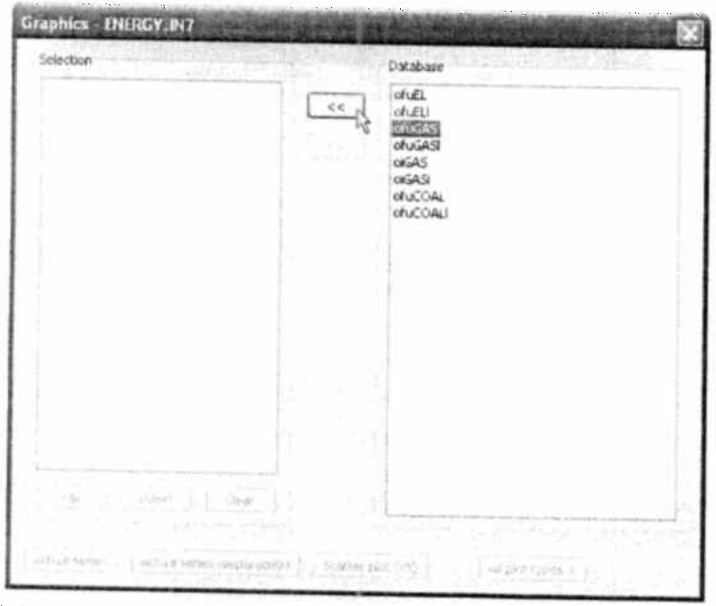
The graphics facilities of OxMetrics are powerful yet easy to use. This section will show you how to make time-plots and cross-plots of variables in the database. OxMetrics offers automatic selections of scaling etc., but you will be able to edit these graphs, and change the default layout such as line colours and line types. Graphs can also be saved in a variety of formats for later use in a word processor or for reloading to OxMetrics.

2.3.1 A first graph

Graphics is the first entry on the Model menu. It can also be activated by clicking on the cross-plot graphics icon on the toolbar:



Activate the command to see the following dialog box:



This is the first example of a dialog with a multiple selection list box. In such a list box you can mark as many items as you want. Here we select the variables we wish to graph. With the keyboard you can only mark a single variable (by using the arrow up and down keys) or range of variables (hold the shift key down while using the arrow up or down keys).
With the mouse there is more flexibility:

- single click to select one variable;
- hold the left mouse button down to select a range of variables;
- hold the `Ctrl` key down and click to select additional variables;
- hold the `Shift` key down and click to extend the selection range;

In this example we select *ofuGAS*, as shown in the capture above, and then press the << button. Next select *ofuGASI*, press << and, finally, the button Actual series (separately). The graph which appears looks very much like Figure 2.1. It confirms the claim made in §1.10.1, that the seasonality is much more constant after taking logarithms.[2] Note that there are many on-screen edit facilities. For example, you can move the position of the legend by picking it up with the mouse. The OxMetrics book describes the edit facilities in more detail.

[2]With logarithm we always mean the natural logarithm, denoted log(·), unless a different base is explicitly given.

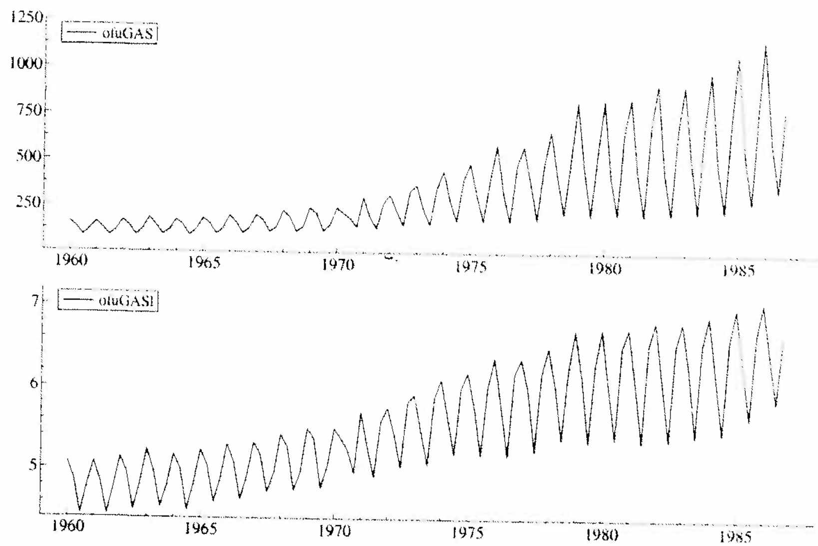


Figure 2.1 Time-series plot of *ofuGAS* and *ofuGASI*.

2.3.2 Graph saving and printing

To print a graph directly to the printer, click on the printer icon in OxMetrics. You can preview the result first using the Print Preview command from the File menu.
Graphs can be saved to disk in various formats:

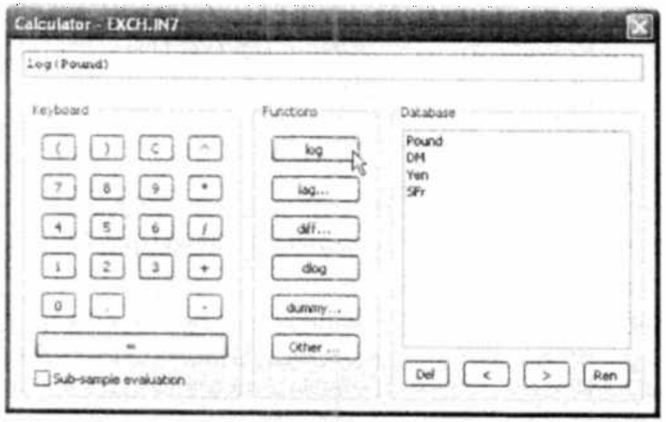
- Enhanced metafile (.EMF);
- The MS Word print format (.PNG);
- Encapsulated PostScript (.EPS), which is the format used to produce all the graphs in this book;
- PostScript (.PS), this is like EPS, but defaulting to a full page print.
- OxMetrics Graphics File (.GWG).

When you save a graph in any format, the GWG file is automatically saved alongside it. Then, when loading a previously saved EPS file (say), OxMetrics can use the GWG file to reload the actual graph.

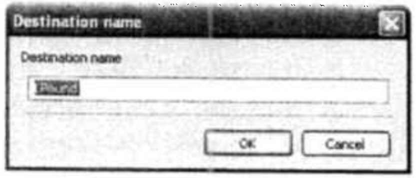
2.4 Data transformations

To complete this short introduction to OxMetrics, we do some data transformations using the OxMetrics Calculator. Section 1.10.2 introduced the EXCH data set. Load this data set into OxMetrics. The status line at the bottom of OxMetrics now indicates that this is the active database (to change between databases, change the selection in the toolbar).

The objective is to create the logs of the first differences of the \$/£ exchange rate. Activate the Calculator, select Pounds from the list of variables, and click on the log button. The expression `log(Pound)` will be written in the expression window:



Press `Enter` (or click on the = button). You will be prompted for a name:



Accept the default. Now use the diff button to take first differences of *LPound* in a similar way. This creates *DLPound*. Finally, to take the absolute values, select *DLPound*, then press the button Other... locate the function in the Functions dailog, and double click on `fabs`:

The expression reads `fabs(DLPound)`. Press =, and give the variable a name, *absDLPound* for example. The transformations were logged in the Results window:

```
Algebra code for EXCH.IN7:
LPound = log(Pound);
DLPound = diff(LPound,1);
absDLPound = fabs(DLPound);
```

This code can be saved, and rerun later. More information on Algebra is in the OxMetrics book. Figure 2.2 shows the created absolute values of the \$/£ exchange rates in the upper part, together with a spline line (obtained using automatic bandwith selection, see the Cross Plot page in the OxMetrics Graphics dialog). Figure 2.2 plots in the bottom part the sample autocorrelation function of the absolute values.

This completes the getting started chapter. We hope that you're equipped now for more substantial structural time series modelling exercises.

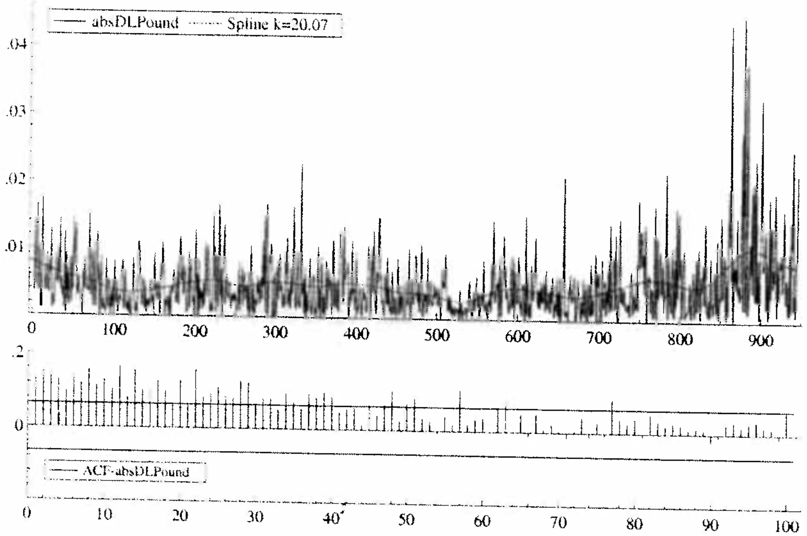


Figure 2.2 Time series plot of *absDLPound* and sample ACF of *absDLPound*.

Part II

Tutorials on Structural Time Series Modelling

Chapter 3

Introduction to Univariate Modelling

This chapter provides a simple introduction to structural time series modelling using STAMP. A recent introductory textbook treatment for this class of time series models is provided by Commandeur and Koopman (2007). Most attention in this chapter is focused on the **Formulate** and **Test** menus of STAMP. The previous chapter showed how to start STAMP, and load the ENERGY data set in OxMetrics.

Prior to modelling we graph the 'ofuCOAL' series in OxMetrics, see Figure 3.1. The graph of this series, which is quarterly consumption of coal by 'Other final users' shows a clear downward trend and a seasonal pattern which seems to decrease over time. The downward trend is still apparent when the logarithm of the series, 'ofuCOALl', is graphed. However, the seasonal pattern appears more stable and so, as in many cases, the log transform is to be preferred.

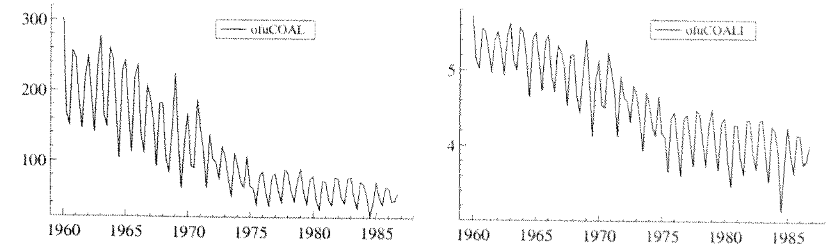


Figure 3.1 Coal consumption by final users. Original series and logarithms.

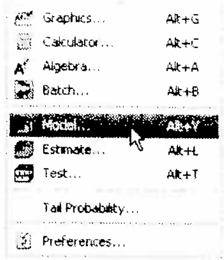
If you wish, you can explore more transformations and examine the correlograms of the observations after various types of differencing. This can be done using the tools provided by Calculator and Algebra and using OxMetrics Graphics in the Model menu.

3.1 Model formulation

Structural time series modelling in STAMP starts with the Model dialog. Here, the variable(s) to be modelled are selected, and additional model components specified.

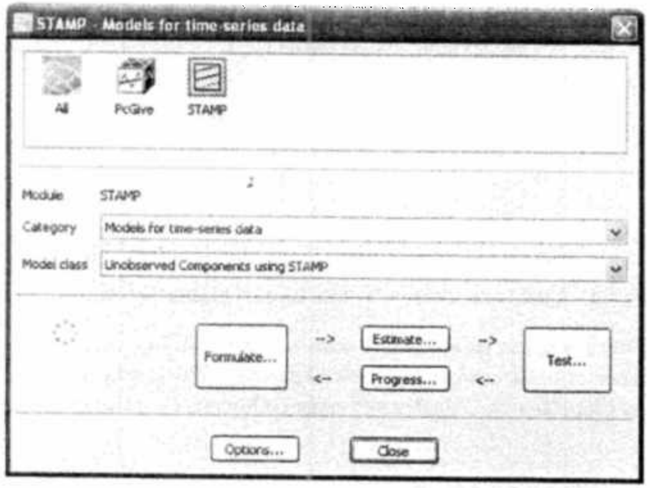
After successful estimation, model evaluation proceeds from the **Test** dialog.

The Model dialog can be activated in several ways. For example, it can be selected from the **Model** menu in OxMetrics.

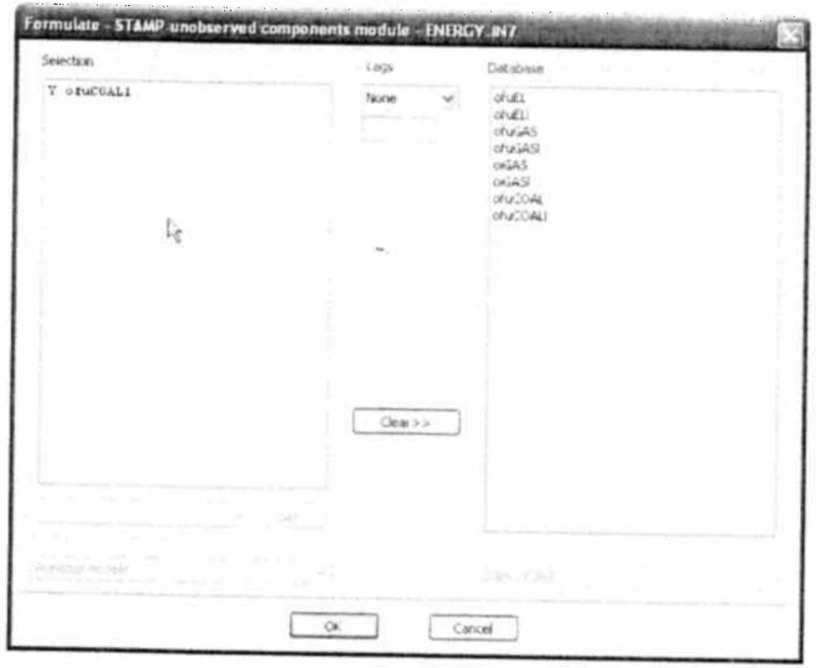


You can also click on the Model icon in the OxMetrics toolbar:

Finally you can select it from Modules in the Documents window or just by pressing `Alt+y`:



Inside the Model dialog, choose Formulate by clicking on its button. The Formulate dialog allows you to mark 'ofuCOALI' by using the mouse and press the << button (or double-click on the variable). You will see:



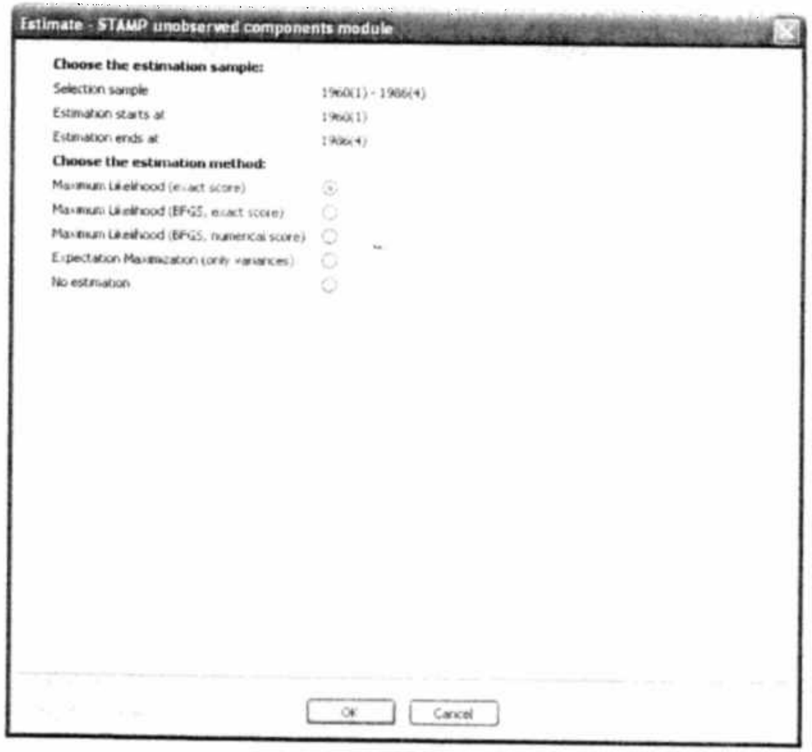
The 'ofuCOALI' series has now appeared under 'Selection' as a Y, or dependent, variable. As we will be using no explanatory variables in this chapter, we are ready to proceed. Press OK, or **Enter** since OK is highlighted.

The next stage is to select a suitable set of components, on the basis of your knowledge of the salient features of the series. Most quarterly economic series display trend and seasonal movements and in the case of ofuCOALI such movements were apparent from the graph.

The Select components dialog lists the available options:

The default setting in the Select components dialog is the *Basic Structural Model* (BSM). This specifies a 'Stochastic Level' and 'Stochastic Slope', making up the trend, a 'Stochastic Seasonal' and an 'Irregular'. These settings may be changed by moving around the dialog using the `Tab` or `Arrow` keys (up and down only in this case) and marking the desired box with the `Spacebar`. Alternatively, click on the desired box with the mouse. For the moment accept the BSM by choosing OK; you can simply hit ↵ since the OK button is highlighted; this leads to the Estimate Model dialog. It allows you to change the sample period if you wish. The other options in the dialog are for different estimation methods and may be ignored at this stage. Press Enter or click OK.

Once estimation is complete, some basic information appears in the Results window. In order to read everything you may need to move up or down the screen using the `Arrow` keys or the mouse. Using the default options, the output will read:




```
UC( 1) Estimation done by Maximum Likelihood (exact score)
    The selection sample is: 1960(1) - 1986(4) (T = 108, N = 1)
    The dependent variable Y is: ofuCOAL1
    The model is:  Y = Trend + Seasonal + Irregular
    Steady state.......... found without full convergence

Log-Likelihood is 188.015 (-2 LogL = -376.031).
Prediction error variance is 0.0205058
Summary statistics
                  ofuCOAL1
 T                  108.00
 p                  3.0000
 std.error         0.14320
 Normality          6.6991
 H(34)              1.5112
 DW                 1.8476
 r(1)             0.063574
 q                  12.000
 r(q)            -0.063351
 Q(q,q-p)           6.0221
 Rs^2              0.35124
```

```
Variances of disturbances:
                   Value      (q-ratio)
Level          0.000749318   (  0.04554)
Slope          2.75126e-006  (0.0001672)
Seasonal          0.000000   (   0.0000)
Irregular        0.0164540   (    1.000)

State vector analysis at period 1986(4)
                                Value      Prob
Level                         3.90276 [0.00000]
Slope                        -0.00685 [0.36652]
Seasonal chi2 test          485.58515 [0.00000]
Seasonal effects:
                Period          Value      Prob
                     1        0.27043 [0.00000]
                     2       -0.09705 [0.00002]
                     3       -0.40679 [0.00000]
                     4        0.23342 [0.00000]
```

The 'Estimation report' tells us that convergence was Very Strong. This is good news. Maximum likelihood estimation has just been carried out by numerical optimisation and STAMP is telling us that this was successful. A failure to satisfy the various convergence criteria may be an indication of a poorly specified model

The 'Diagnostic summary report' provides some basic diagnostics and goodness-of-fit statistics. In particular, we have the Box–Ljung Q-statistic, $Q(12, 9)$. This is a test for residual serial correlation, which is based on the first 12 residual autocorrelations and should be tested against a chi-square distribution with 9 degrees of freedom. It is not significant here. If you wish to find the 'p-value', go to the OxMetrics Model/Tail Probability menu. Select 'Chi^2(n1)' in the dialog and enter '9' for 'n1' and the value of the test statistic in 'value'. The result is `Chi^2(9) = 6.0221 [0.7377]`.

Since the estimation procedure converged and the diagnostics appear satisfactory, we can be reasonably confident that we have estimated a sensible model (though it may not be the best).

The variances govern the movements in the components. They are part of the standard model output, and were already printed in the Results window:

```
Variances of disturbances:
                   Value      (q-ratio)
Level          0.000749318   (  0.04554)
Slope          2.75126e-006  (0.0001672)
Seasonal          0.000000   (   0.0000)
Irregular        0.0164540   (    1.000)
```

Interpreting the actual numbers is not easy for the inexperienced user and, in fact, the information they contain is effectively displayed in the plots of the estimated components. However, a *zero* parameter estimate does convey information since it tells us that the corresponding component is fixed. Thus in the present example the seasonal pattern is fixed. As part of the default output, the components graphics can be viewed from the Documents window, in Graphics/Model, see Figure 3.2.

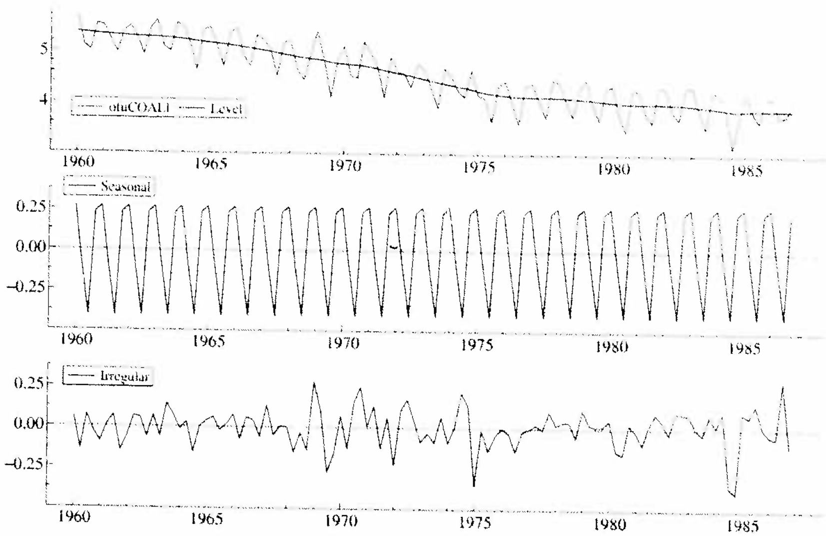


Figure 3.2 Components graphics output of model for coal consumption.

To investigate the parameter estimates more extensively, more detailed diagnostics and forecasts can be accessed from the **Model** dialog by clicking on the **Test** button. The **Test** dialog appears.

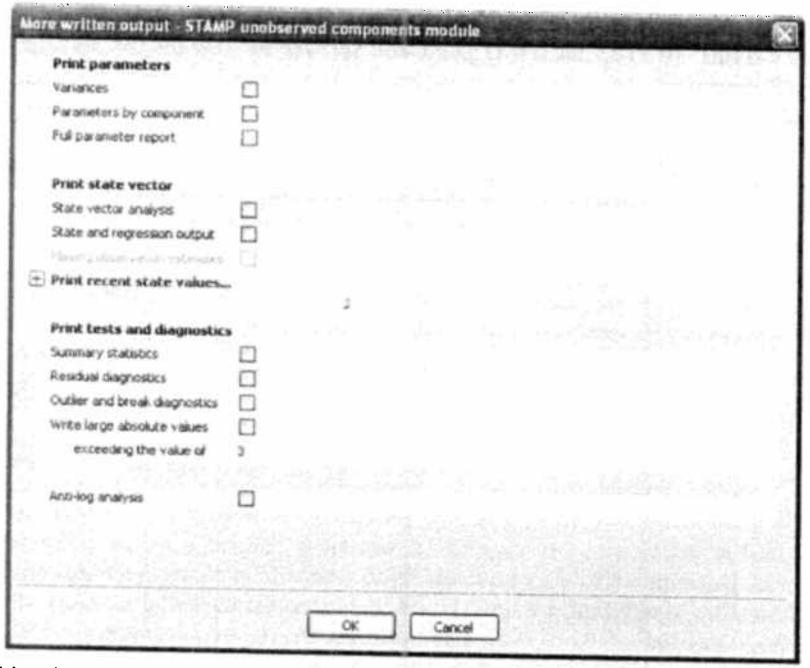
3.2 Evaluating and testing the model

The **Test** dialog contains eight options. The first three are concerned with estimates of parameters and components, the fourth and sixth give additional diagnostics, while the fifth allows the user to carry out diagnostic checking by auxiliary residuals. The penultimate option is for forecasting, and is described in the next section. The last option is for the storage of output in the OxMetrics database. The listed order is quite logical, though the easiest way to determine how well the model is working is often to move straight to Components graphics and examine the various plots.

To access the different **Test** dialogs, one or more options can be selected and click on OK.

3.2.1 More written output

The final state vector contains information on the values taken by the various components at the end of the sample. To see the final state values, access More written output on the **Test** menu, and make sure that the options 'State vector analysis' and 'State and regression output' are checked:



This prints results for the final state vector:


```
State vector analysis at period 1986(4)
                              Value       Prob
Level                       3.90276  [0.00000]
Slope                      -0.00685  [0.36652]
Seasonal chi2 test        485.58515  [0.00000]
Seasonal effects:
                 Period       Value       Prob
                      1     0.27043  [0.00000]
                      2    -0.09705  [0.00002]
                      3    -0.40679  [0.00000]
                      4     0.23342  [0.00000]
Equation ofuCOAL1: coefficients of components in final state at period 1986(4)
               Coefficient       RMSE      t-value       Prob
Level              3.90276    0.06218     62.76158  [0.00000]
Slope             -0.00685    0.00755     -0.90701  [0.36652]
Seasonal           0.16524    0.01768      9.34847  [0.00000]
Seasonal 2         0.33861    0.01768     19.15706  [0.00000]
Seasonal 3         0.06818    0.01242      5.48945  [0.00000]
```

Thus, for example, the slope is -0.00685 and the figure in square brackets after the 't-value' is a two-sided Prob. value; that is, it shows the probability of getting an absolute value of a standard normal variable greater than this value if the true parameter is zero.

In our particular example, as in many other cases, the data are in logarithms. Returning to the More written output menu, we now check the 'Anti-log analysis' option, together with 'State vector analysis':

```
State vector anti-log analysis at period 1986(4)
It is assumed that time series is in logs.
                              Value       Prob
Level (anti-log)           49.53886  [0.00000]
Level (bias corrected)     49.63473  [       -]
Slope (yearly %growth)     -2.73858  [0.36652]
Seasonal chi2 test        485.58515  [0.00000]
Seasonal effects:
                 Period       Value       Prob     %Effect
                      1     1.31052  [0.00000]    31.05223
                      2     0.90751  [0.00002]    -9.24935
                      3     0.66578  [0.00000]   -33.42185
                      4     1.26291  [0.00000]    26.29149
```

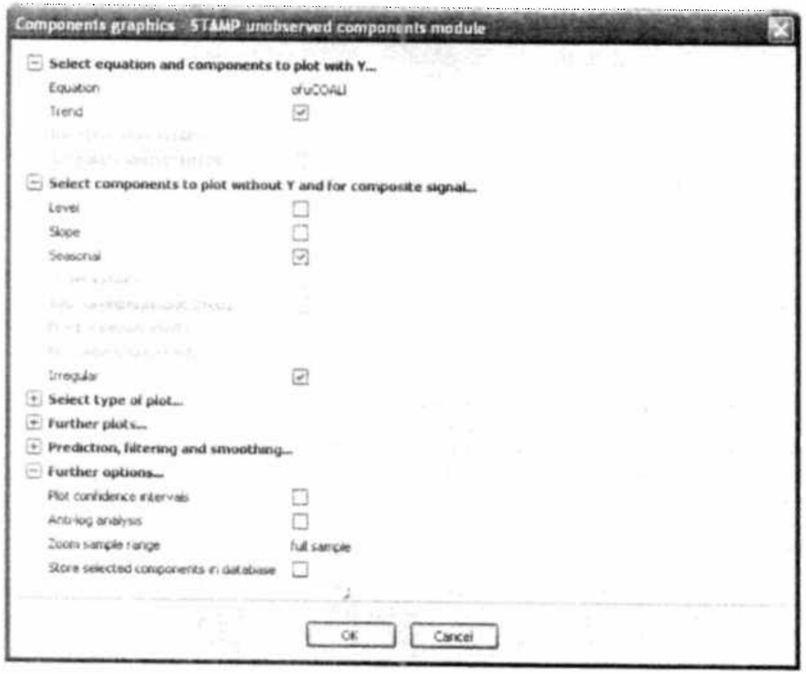
The additional information shows, for example, that the slope can be interpreted as a growth rate and we see that here the estimated current growth rate is -2.74% per year. The seasonals can be interpreted as the factors by which we multiply the trend, so that, for example, consumption of coal is, on average, 31% higher in the winter, Seas 1.

3.2.2 Components graphics

The information provided in Components graphics is fundamental to the interpretation of the model. The default, shown on the next page, provides what we consider to be the most useful plots, but others may be obtained by marking the required boxes in the usual way.

The *smoothed* estimates of the components are obtained using all the information in the sample; that is, they are constructed using observations which come after as well as those which come before. This is sometimes referred to as *signal extraction*.

When the data are in logarithms it is often helpful to look at the trend and seasonal components after exponentiating (anti-logs). To do this mark 'Anti-log analysis'. The interpretation of the seasonals and growth rate graphs, see Figure 3.3 is as in 'State vector analysis' except that now we see any changes which may have taken place over the whole sample period. However, in the present example both are fixed.



3.2.3 Weight functions

Structural time series models belong to the class of linear time series models. Therefore, the smoothed estimates of the components are linear functions of all observations in the selected sample. It may be of interest to investigate the actual obervation weights that are used for the computation of the smoothed estimate at a particular time point within the sample. To facilitate such an analysis, the dialog Weight functions allows, along more specialised output, the graphical representation of the weights. To illustrate this, access the dialog Weight functions:

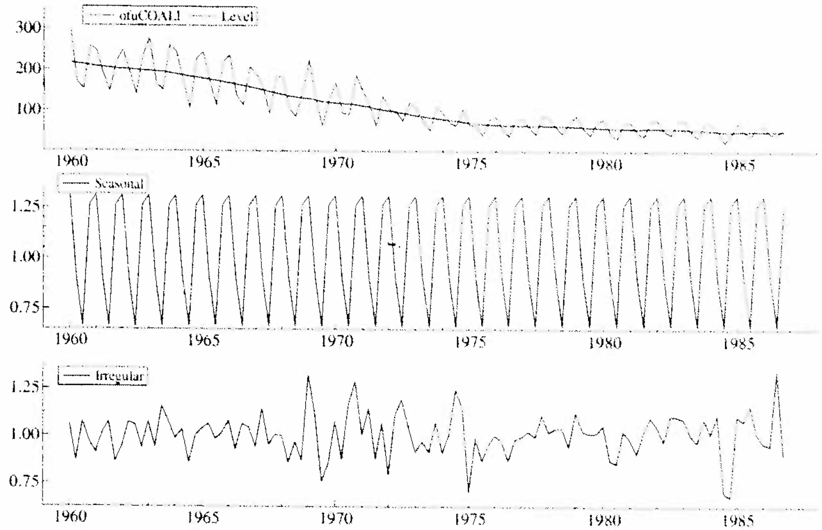
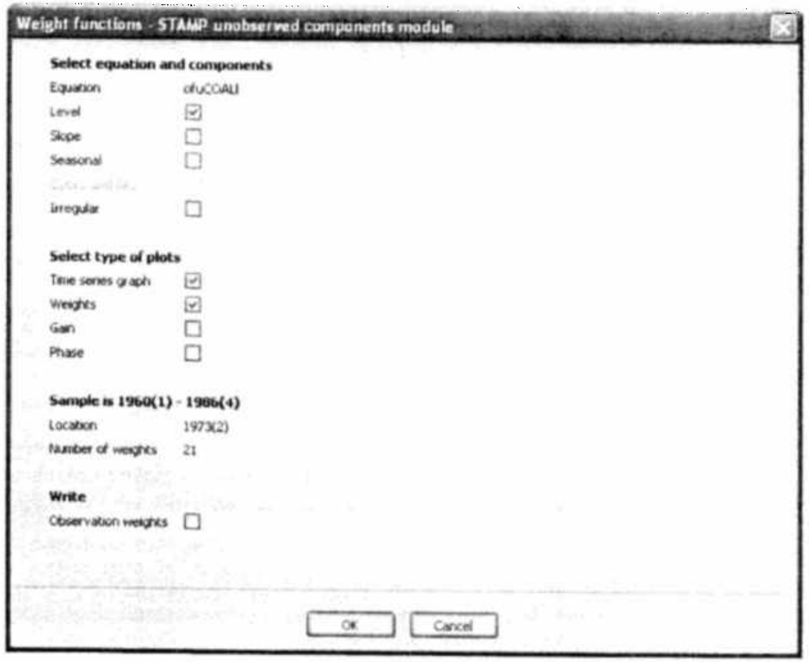


Figure 3.3 Components graphics analysis of model for coal consumption, anti-log analysis.



The default setting produces the graphical output as presented in Figure 3.4 for the estimated level component at period 1973 Q2. It is clear that the weights decay exponentially with the distance of the observations from the period 1973 Q2. Further the symmetric distribution of the weights around period 1973 Q2 is a feature that is common to linear and time-invariant structural time series models.

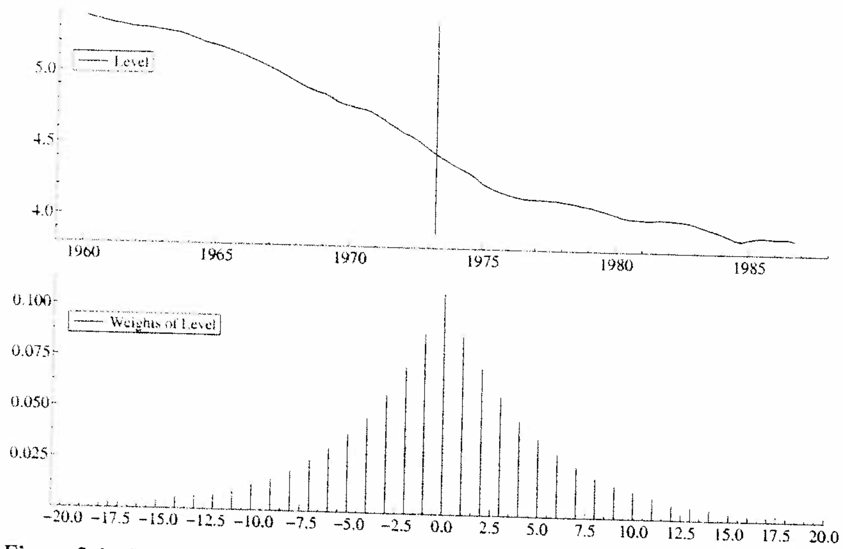
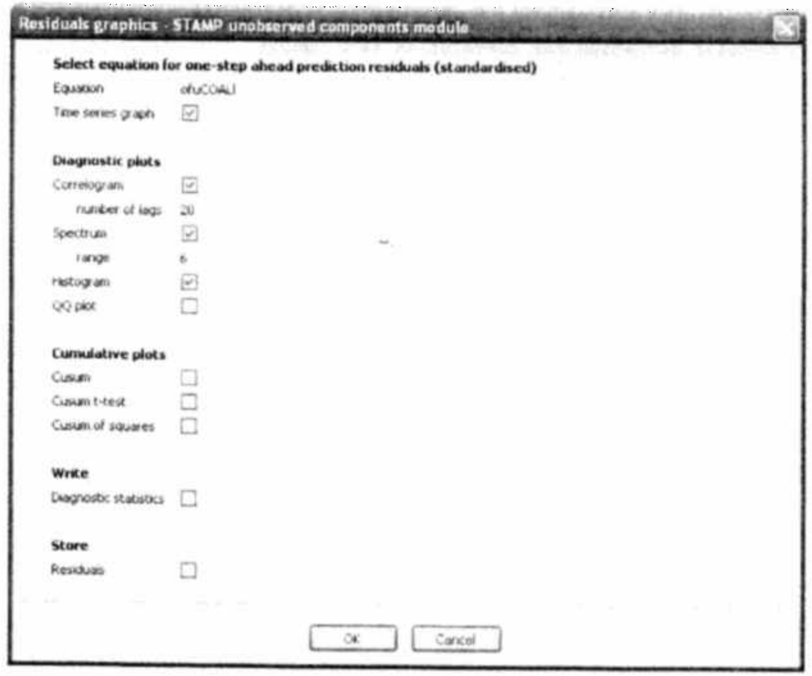


Figure 3.4 Weight graphics analysis for trend of model for coal consumption at period 1973 Q2.

3.2.4 Residuals graphics

Further diagnostic checking of the model may be carried out by plotting the residuals and looking at their full correlogram. A graph of the distribution of the residuals can also be constructed. These options are provided through the Residuals graphics dialog.

The default produces Figure 3.5. Additional output can be requested through the option 'Write diagnostic statistics' and it outputs an extensive list of residual, goodness-of-fit and serial correlation statistics:




```
Normality test for Residuals ofuCOAL1
Sample size             103.00
Mean                  0.024851
St.Dev                 0.99969
Skewness              -0.59764
Excess kurtosis        0.99695
Minimum                -3.0098
Maximum                 2.4635
                         Chi^2        Prob
Skewness                6.1315   [ 0.0133]
Kurtosis                4.2656   [ 0.0389]
Bowman-Shenton          10.397   [ 0.0055]

Goodness-of-fit based on Residuals ofuCOAL1
                                                 Value
Prediction error variance (p.e.v)             0.020506
Prediction error mean deviation (m.d)         0.015444
Ratio p.e.v. / m.d in squares                   1.1223
Coefficient of determination R^2               0.94281
... based on differences Rd^2                  0.89242
... based on diff around seas mean Rs^2        0.31974
Information criterion Akaike (AIC)             -3.7759
... Bayesian Schwartz (BIC)                    -3.6269
Serial correlation statistics for Residuals ofuCOAL1
```

```
Durbin-Watson test is 1.86205
Asymptotic deviation for correlation is 0.096225
  Lag      df     Ser.Corr    BoxLjung      prob
    4       1   -0.0089215      1.4639  [ 0.2263]
    8       5   -0.0052752      2.6772  [ 0.7496]
   12       9    -0.063513      6.3039  [ 0.7092]
```

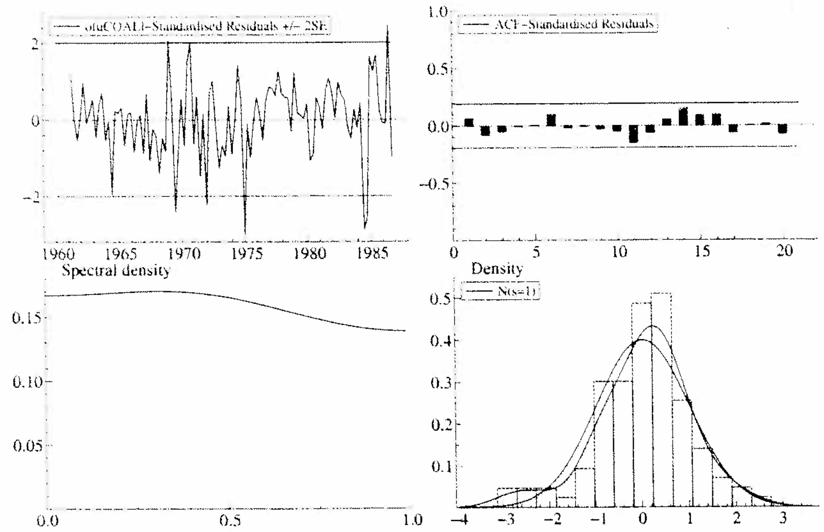
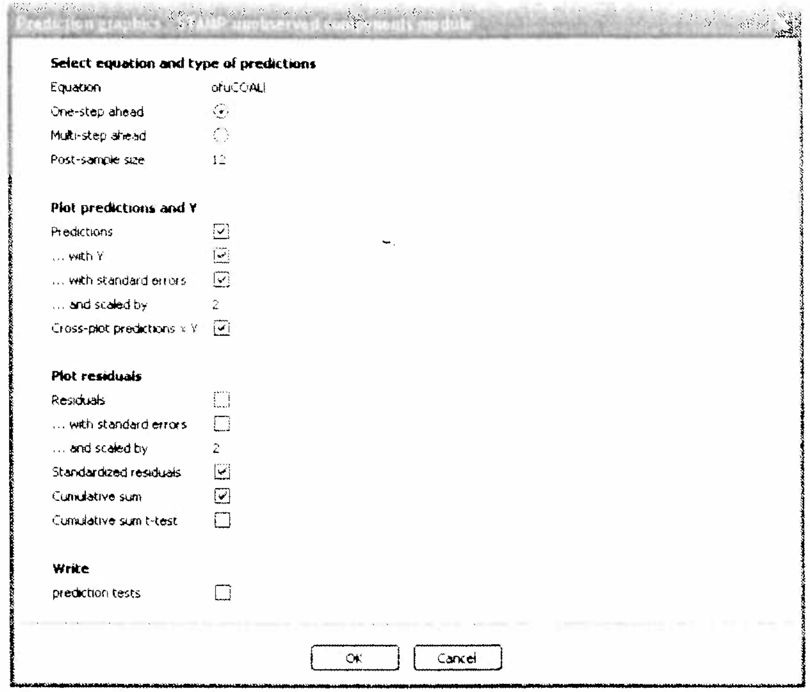


Figure 3.5 Residuals graphics analysis of model for coal consumption.

The Auxiliary residuals option is concerned with the detection of outliers and structural change. It will be described when interventions are dealt with in Chapter 5.

3.2.5 Prediction graphics

Prediction graphics is for making predictions as if data has been held back at the end of the sample. The number of observations to withhold can be specified in the dialog. In the illustration below, we change the default from 8 to 12. The Prediction graphics dialog can be activated via the **Test** menu:



Apart from changing the post-sample size to 12, accepting the default produces graphs as shown in Figure 3.6. In the first graph it can be seen that the values of 1984 Q3 and Q4 are outside the prediction intervals, set at two root mean square errors (RMSEs). The three subsequent predictions are also poor. The forecast Chow test can also be given (in addition to the 12 post-sample predictions):

```
Equation ofuCOAL1: post-sample predictive tests.
Failure Chi2( 12) test is   29.4566   [0.0034]
Cusum t( 12)      test is    0.2668   [0.7941]
```

The figure shown in square brackets indicates that the probability of a value greater than this magnitude is 0.0034. This provides further evidence that the prediction errors are not consistent with the model.

The above exercise may be repeated with 'Multi-step prediction' marked and with post-sample size set to 12. In this case the predictions are made using the information at the end of 1983 and are not updated with the arrival of each new observation. The result is that although Q3 and Q4 of 1984 are again poorly predicted, the next three observations are predicted very well. We therefore conclude that Q3 and Q4 of 1984 are unusual observations, and in fact this turns out to be the case since the coal miners were on strike then. Hence the model is a good one and what needs to be done is to remove the two strike observations; see Chapter 5 on intervention analysis.

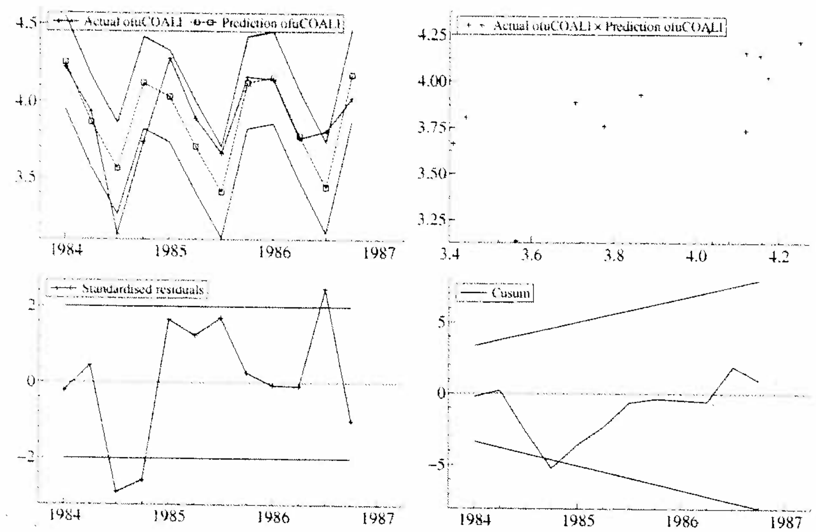
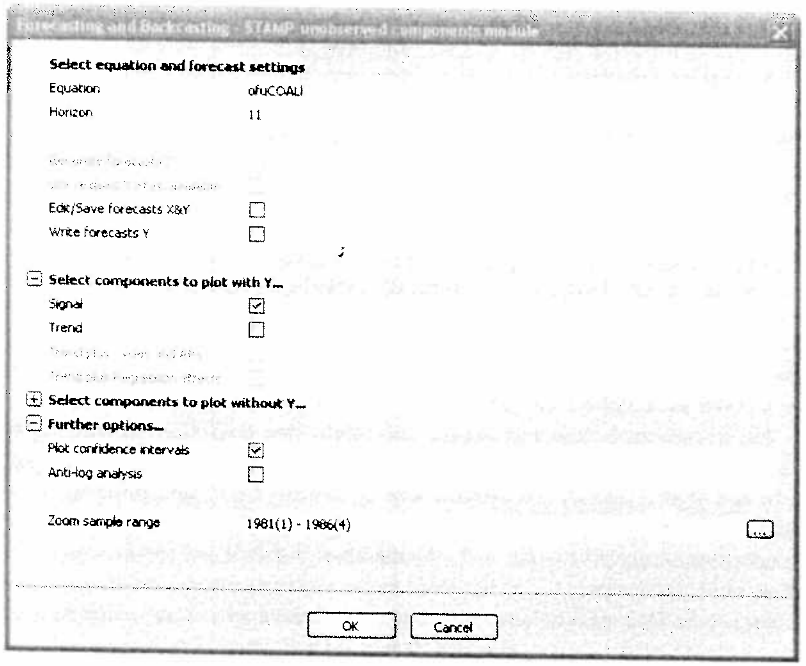


Figure 3.6 Prediction analysis of model for coal consumption.



3.2.6 Forecasting

The Forecasting dialog is for making predictions over a pre- or post-sample period: that is, for a period where no observations on the series are available. The dialog allows us to look at forecasts of components as well as forecasts of the series itself. Just mark the appropriate boxes.

The lines in Figure 3.7 indicate *one* RMSE on either side. Note that the prediction interval increases the further forward we move into the future.

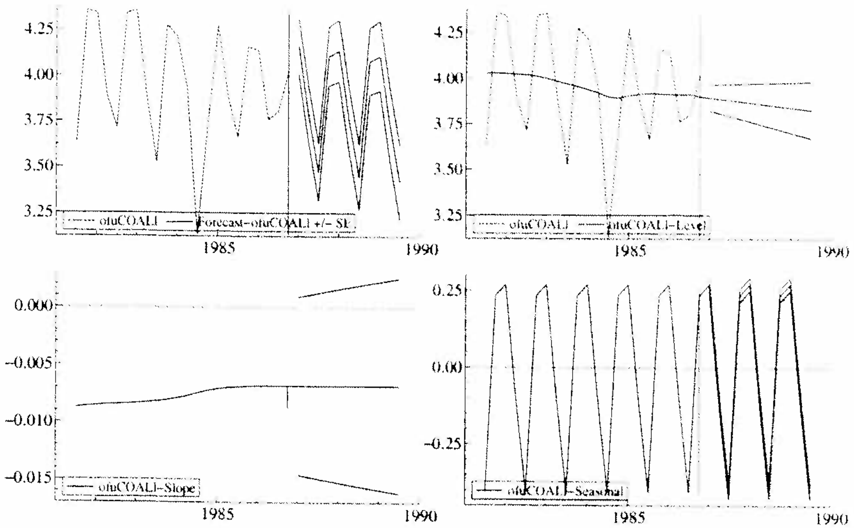


Figure 3.7 Forecast analysis of model for coal consumption.

3.3 Exercises

(1) The remaining series in the ENERGY file illustrate other interesting features. In interpreting the GAS series, it needs to be borne in mind that in 1970 cheaper natural gas became available from the North Sea. The jump in consumption is quite dramatic in 'Other industries' — oiGAS — but STAMP copes with it very well. Show this by fitting a BSM to the logged series. Make forecasts up to five years ahead.

(2) Return to ofuCOAL, but now mark 'Fixed level' in the Components dialog. Show that the result is a much smoother trend and a growth rate which gradually changes over time. Compare the overall fit to that of the unrestricted model. Do you think that this 'smooth trend' specification has more attractive properties?

(3) Fit a model to the electricity series, ofuELECI. Comment on the movements in the trend and seasonal. Compare with the results from ofuGASI in §1.10.1. Carry out post-sample predictive testing over the last two years. What is the value of the Chow statistic?

Chapter 4

Tutorial on Components

4.1 Selection of components

The specification of components is at the heart of structural time series modelling. This section sets out the methodology underlying the meaning and selection of components. The next four sections give details of the various components and provide examples showing how a particular model is specified and evaluated. The statistical specifications of the various components are in separate subsections and a degree of understanding of what the models can achieve is possible without fully understanding them.

Initial specification — An initial judgement about components can often be made on the basis of prior knowledge of the series. For example, a seasonal component will typically be included if the observations are quarterly. A graph of the data often provides confirmation or further information. The various options in OxMetrics 4 allow the exploration of various transformations, such as logarithms. The correlograms of series which have been made stationary by differencing can also be examined, although this type of analysis plays a much less prominent role than it does in the ARIMA 'Box-Jenkins' model selection methodology.

Estimation — Once a model has been specified, it is estimated. If convergence problems arise, this may be an indication of a poor specification or too many parameters.

Parameters — The variance ratios govern the extent to which the components move stochastically. The actual numbers are not easy to interpret. However, a value of zero for a variance indicates that the corresponding component is deterministic . If this is the case, a standard regression type significance test can be carried out on the corresponding component in the state. If it is not significantly different from zero, it may be possible to simplify the model by eliminating that particular component.

Components — In general, the components are stochastic and so can only be assessed by looking at their behaviour throughout the whole sample, rather than at the end. The behaviour of the smoothed components provides a guide as to whether the decomposition implied by the fitted model is useful.

Diagnostics — Once the model has been estimated, various diagnostic and goodness-of-fit statistics are printed out. Further diagnostic checking of the model as a whole can be carried out by accessing **Test**/Residuals graphics dialog.

Predictive testing — Predictions near the end of the series give a good idea of the properties of model and how well it fits. STAMP offers the option of computing and displaying predictions and carrying out predictive tests when data is held back at the end of the sample. These tests are based on one-step ahead predictions. However, it is often informative to examine the behaviour of extrapolative (multi-step) predictions as well.

If we try out many different specifications, there is the danger of data mining. By retaining some observations at the end of the series for post-sample predictive testing, we guard against a spuriously good fit. These post-sample observations are not used to select the model or to estimate the parameters.

4.2 Trend

The trend is the long-run component in the series. It indicates the general direction in which the series is moving.

There are two parts to the trend:

- *Level* — the actual value of the trend;
- *Slope* — this component of the trend may or may not be present.

In order to illustrate the various aspects of trend specification, we will use the data set of US macroeconomic time series — USYCIMP.IN7. These logged observations are all quarterly, but seasonally adjusted. If you are not in STAMP, you may start the program within OxMetrics from the menu **Model** or by pressing `Alt+y`. The database can be loaded into OxMetrics by using the option **File**/Open.

4.2.1 Local level model

The local level model consists of a random disturbance term around an underlying level which moves up and down, but without any particular direction.

We start with the rate of inflation, which is denoted as *Dp*. It is the first difference of the logarithm of the price level, *p*. From the **Model**/Graphics dialog in OxMetrics, you can produce a graph of the series, see Figure 4.1. As you will see, it has the characteristics of a local level model.

To estimate a local level model for Dp, go to the **Model**/Formulate dialog of STAMP and select Dp from the Database listbox bu using your mouse and press << (or double-click on Dp). If you are continuing from the previous section, change the database by selecting the USYCIMP database below the Database listbox (of course, this database must be loaded into OxMetrics first). If Dp appears as an Y variable in the Selection listbox, press OK to go to the Select components dialog. Here we remove the slope component from the model by clicking on the checkbox of 'Slope' (the cross sign

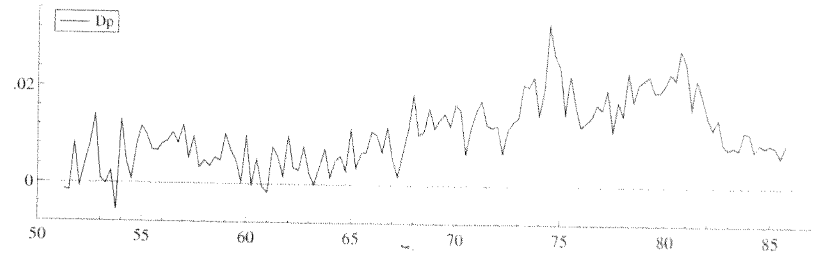


Figure 4.1 Time series of US inflation, from data set USYCIMP.

disappears) and we repeat this for the seasonal component. We keep the other default options. We have specified the local level model. By clicking on OK we enter the Estimate dialog in which we can keep all the defaults and only require to click on OK. The model is estimated and output is reported in the Results window. The goodness-of-fit statistics and diagnostics indicate the fit is fine:

```
    Estimation done by Maximum Likelihood (exact score)
    The databased used is USYCIMP.IN7
    The selection sample is: 51(2) - 85(4) (T = 139, N = 1)
    The dependent variable Y is: Dp
    The model is:  Y = Level + Irregular
    Steady state. found

Log-Likelihood is 746.842 (-2 LogL = -1493.68).
Prediction error variance is 1.97066e-005
Summary statistics
                        Dp
 T                  139.00
 p                  1.0000
 std.error       0.0044392
 Normality          4.1757
 H(46)             0.86535
 DW                 1.9946
 r(1)           0.00097185
 q                  11.000
 r(q)             -0.18240
 Q(q,q-p)           7.1812
 R^2               0.62510

 Variances of disturbances:
                    Value    (q-ratio)
Level         2.84629e-006  (  0.2308)
Irregular     1.23330e-005  (   1.000)
```

To see the implications of the local level specification most clearly, go to Graphs/Model in the Documents window. These graphs can be reproduced by going to **Test**/Components graphics. Selecting 'Trend' in the box 'Plots with Y' and selecting 'Irregular', which are the defaults, and clicking on OK brings up a graph which shows the underlying rate of inflation. You can also move to the Forecast dialog. By accepting the default settings, you will see the underlying rate of inflation and its forecast; see Figure 4.2.

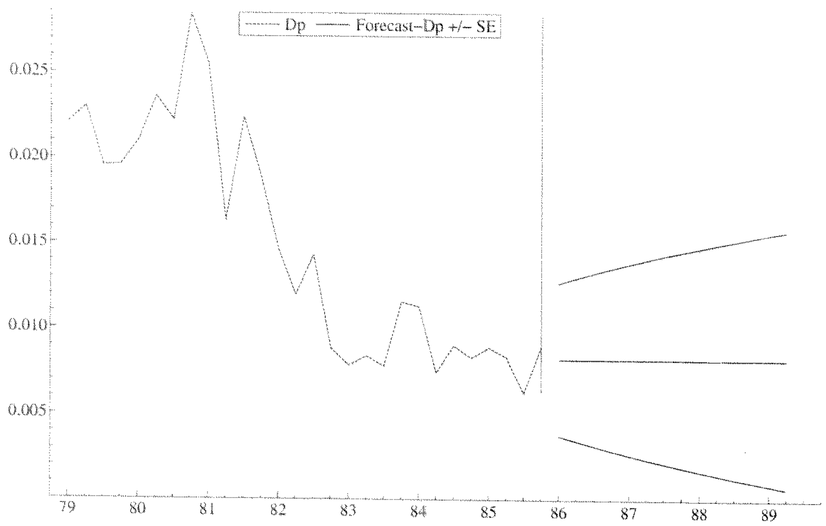


Figure 4.2 The local level model with forecast of US inflation.

The forecast function is a horizontal straight line. It projects the estimate of the current level into the future and this estimate is effectively constructed by putting exponentially declining weights on the past observations; see Harvey (1989). The RMSE shows how the uncertainty of the forecasts increases with the lead time. Again this is shown in Figure 4.2.

4.2.2 Statistical analysis of the local level model

The statistical specification consists of a random walk component to capture the underlying level, μ_t, plus a random, 'white noise', disturbance term, ϵ_t,

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \mathsf{NID}(0, \sigma_\epsilon^2), \quad t = 1, \ldots, T, \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim \mathsf{NID}(0, \sigma_\eta^2), \end{aligned} \tag{4.1}$$

where η_t is the white noise disturbance driving the level. Both disturbances are normally distributed and independent of each other.

- When STAMP carries out estimation it is computing ML estimates of the variances, σ_η^2 and σ_ϵ^2. In the inflation example, these estimates are 2.85×10^{-6} and 1.23×10^{-5}. The relative variance, known here as the *signal to noise ratio*, is $\sigma_\eta^2/\sigma_\epsilon^2$ or, if we are working in terms of standard deviations, $\sigma_\eta/\sigma_\epsilon$. It is given in parentheses after the variance estimate as the *q-ratio* and is given by 0.23.
- After estimation, STAMP runs a *Kalman filter* through the observations to estimate the *state*, μ_t. The estimate of the final state, μ_T, is available from the Test/More written output dialog by activating the option 'State and regression output' (in the section 'Print state vector'):
  Coefficients of components in final state at period 85(4)

	Coefficient	RMSE	t-value	Prob
Level	0.00823	0.00216	3.80859	[0.00021]

  For Dp it is 0.008, which translates to 3.2% per year ($4 \times 100 \times 0.008$), with a root mean square error (RMSE) of 0.8% ($4 \times 100 \times 0.002$). Thus the current annual rate of inflation, which is to be projected into the future, is about 3.2%.
- The **Test**/Components graphics dialog allows you to estimate the trend at all points in the sample using all the observations. This is known as *signal extraction* or *smoothing*. The *filtered* estimate of the trend can also be computed at all points in the sample if you wish. Unlike the smoothed estimate, the filtered estimate is based only on previous observations and the current observation. The *predicted* estimate is based only on previous observations and can be presented as well. The different estimates are presented in Figure 4.3. The graphs can be obtained from the **Test**/Components graphics by expanding the section 'Prediction, filtering and smoothing' (click on this title) and activating the options 'Predictive filtering' and 'Filtering'. Then press OK.

The inclusion of a random walk component in the model means that it is nonstationary. However, it is stationary in first differences, since

$$\Delta y_t = \eta_t + \epsilon_t - \epsilon_{t-1}, \quad t = 2, \ldots, T. \tag{4.2}$$

The OxMetrics Calculator and Graphics tools allow the user to create the differenced time series of Dp, that is DDp, and to produce graphs for the actual time series Dp and DDp and their associating correlograms. These graphs are presented in Figure 4.4 and they indicate quite clearly that differencing leads to a stationary series. (It can be shown to be equivalent to a first-order MA process.) However, while the transformation to stationarity is insightful, it should be stressed that this kind of analysis plays a much less prominent role in structural time series modelling than it does in the ARIMA methodology of Box and Jenkins (1970).

4.2.3 Local linear trend and smooth trend

Specifying both the level and the slope to be stochastic, which is the default, results in the local linear trend model. The forecast function is a straight line with an upward or downward slope, and both the level and the slope are constructed by putting more weight on the most recent observations.

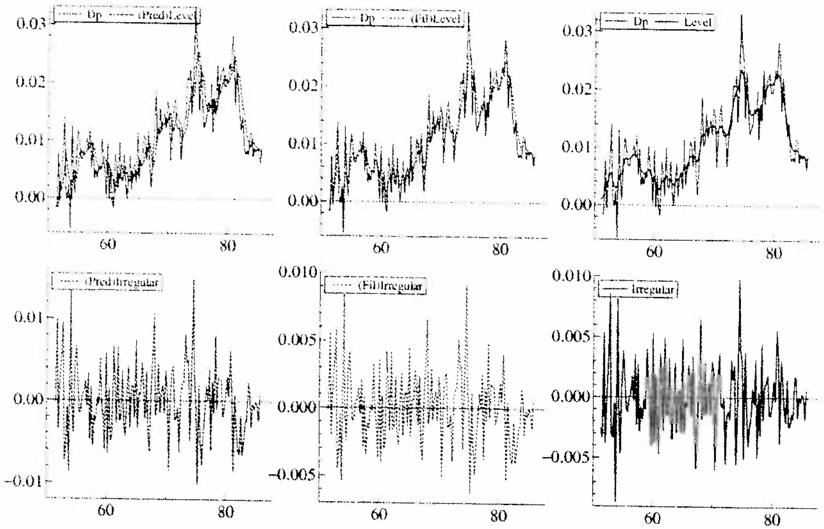


Figure 4.3 The local level model with different estimates of level and irregular components: predicted, filtered and smoothed estimates.

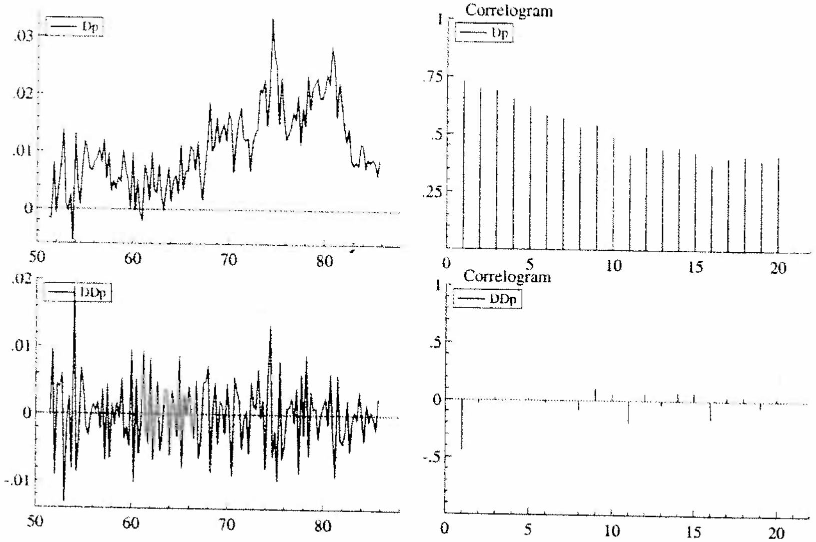


Figure 4.4 Graph and correlogram of US inflation in levels and first differences.

A variant of the local linear trend model induces a somewhat smoother trend by

specifying 'Fixed Level' in the Select Components dialog. This is usually combined with an autoregressive component or cycle: see §4.4.4.

4.2.4 Statistical specification of the local linear trend model

A stochastic trend (local linear trend) model may be written

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim \mathsf{NID}(0, \sigma^2_\epsilon), \tag{4.3}$$

where the trend component is specified as

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim \mathsf{NID}(0, \sigma^2_\eta), \\ \beta_t &= \beta_{t-1} + \zeta_t; & \zeta_t &\sim \mathsf{NID}(0, \sigma^2_\zeta), \end{aligned} \tag{4.4}$$

where μ_t is the level and β_t is the slope. An alternative way of representing the model for the trend μ_t is

$$\Delta\mu_t = \beta_{t-1} + \eta_t, \qquad \Delta\beta_t = \zeta_t, \tag{4.5}$$

where Δ is the difference operator ($\Delta = 1 - L$ with $Ly_t = y_{t-1}$). A more general version of the trend is given by

$$\Delta\mu_t = \beta_{t-1} + \eta_t, \qquad \Delta^p\beta_t = \zeta_t, \tag{4.6}$$

for $p = 1, 2, \ldots$ where $\Delta^p = (1 - L)^p$ with $L^p y_t = y_{t-p}$. When value $p > 1$ is taken, the slope component is more smooth. The estimated trend component from a trend model with $\sigma^2_\eta = 0$ and $p > 1$ can be compared with a so-called Butterworth filter, see Gomez (2001).

The various statistical specifications of the trend in the Components dialog:

can now be defined precisely as follows.

- Level — determines whether level is present in model; if 'Level' is not selected, it is not present in the model and there can be no slope.
  - Stochastic — level specification contains η_t;
  - Fixed — level specification does not contain η_t; that is, σ^2_η is set to zero;
- Slope — determines whether slope is present in the level specification; if 'Slope' is not selected, it is not present in the model.
  - Stochastic — slope specification contains ζ_t;
  - Fixed — slope specification does not contain ζ_t; that is, σ^2_ζ is set to zero.
  - Order of trend — the smoothness value p for the slope component in (4.6).

4.2.5 Specification of the trend

The default specification has $p = 1$ and both level and slope stochastic. The following are important special cases:

- *Local level* or *random walk plus noise* — the trend is a random walk and so we specify 'Level, stochastic' and 'Slope, not selected';
- *Local level with drift* — 'Level, stochastic' and 'Slope, fixed';
- *Smooth trend* — 'Level, fixed' and 'Slope, stochastic';
- *Butterworth trend* — 'Level, fixed', 'Slope, stochastic' and 'Order of trend p' with $p = 2, 3, 4$.

The specification of the trend can sometimes be based on prior knowledge or a plot of the data. If there is doubt, then estimate a general model. Tests of variances being zero have been developed by Nyblom and Makelainen (1983) and Harvey (2001). It is not unusual to find a variance going to zero. We then test if the corresponding element in the state is zero. In particular, if the slope variance (σ_ζ^2) is estimated to be zero, we may wish to test if the slope itself (β), which is now fixed, is also zero. This can be done in the Final state dialog by examining the '$t - value$'. Once restrictions have been imposed we need to check the serial correlation diagnostics.

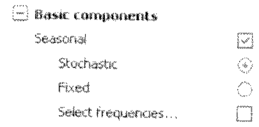
Quite often more than one trend specification may be adequate in terms of dia gnostics and goodness of fit. It is useful to examine the implications of different trends by looking at the smoothed components. It is also useful to check the forecast function. For example, a stochastic slope can sometimes be too sensitive to changes in the series, resulting in very unstable forecasts.

4.3 Seasonal

When appropriate, a seasonal component may be included in a structural time series model. If the data are inputted as quarterly or monthly, the number of seasons in a year, s, is known to STAMP, and a stochastic seasonal option is based on a trigonometric formulation. For daily observations, an intra-weekly pattern is assumed; that is, s is taken to be 7. The definition of the trigonometric form of stochastic seasonality is given in Chapter 9 and discussed at length in Harvey (1989). The important point to understand is that the seasonal pattern becomes deterministic if the seasonal variance parameter is set to zero.

4.3.1 Specifying and testing the seasonal component

There are three options for the seasonal component in the Components dialog:



When a seasonal component is thought to be present, 'Seasonal' is selected. It may sometimes be appropriate to select 'Fixed' for the seasonal componen, perhaps because a previous attempt to fit the model indicated that this was the case. If the number of years is small, it may be reasonable to fix the seasonal as there is not enough data to allow a changing pattern to be estimated.

If it is believed that there is no seasonal component or if the series has been seasonally adjusted, the 'Seasonal' option may be de-activated. If such an assumption is inappropriate, the residuals will tend to show serial correlation, particularly at the seasonal lag, s.

Now select a typical seasonal series, for example, the logarithm of income, y, in the data file UKCYP.IN7. Select the default in Components. This consists of a stochastic level and slope, a stochastic trigonometric seasonal, and an irregular component. It is known as the *Basic Structural Model* (BSM). Estimate the model.

The diagnostic summary and estimated variances of disturbances are printed by default. The dialog Test/More written output provides more output. In this dialog, activate 'Anti-log analysis' and 'State and regression output'. The complete output is:

```
Log-Likelihood is 583.909 (-2 LogL = -1167.82).
Prediction error variance is 0.000351157
Summary statistics
                      y
T                154.00
P                3.0000
std.error      0.018739
Normality        5.6194
H(49)           0.54011
DW               1.9924
r(1)         -0.0051813
q                14.000
r(q)           -0.21686
Q(q,q-p)         31.238
Rs^2            0.16946


Variances of disturbances:
                  Value     (q-ratio)
```

```
Level        0.000198515  (    1.000)
Slope           0.000000  (   0.0000)
Seasonal     1.85395e-006 ( 0.009339)
Irregular    3.71636e-005 (   0.1872)

State vector anti-log analysis at period 93(2)
It is assumed that time series is in logs.
                              Value      Prob
Level (anti-log)        96250.09220 [0.00000]
Level (bias corrected)  96252.77834 [      -]
Slope (yearly %growth)      2.62979 [0.00000]
Seasonal chi2 test          9.05111 [0.02862]
Seasonal effects:
              Period          Value      Prob    %Effect
                   1        0.98435 [0.00941]   -1.56494
                   2        1.00009 [0.98824]    0.00908
                   3        1.00370 [0.57452]    0.37028
                   4        1.01206 [0.05865]    1.20585
```

The features of the output directly relevant to seasonal effects are as follows:

- The seasonal variance parameter is non-zero, indicating that there are changes in the seasonal pattern.
- The state has $s - 1$ elements to capture seasonality. These are not directly interpretable in the seasonal case, but STAMP transforms them into monthly or quarterly effects as appropriate. *Note that the seasonals sum to zero.*
- If the 'Anti-log analysis' option is marked, the seasonal effects are given as factors of proportionality by which to multiply the other components to get the systematic part of the series. Thus, in the case of UK income, the seasonal factors are 0.984, 1.000, 1.004 and 1.012. This indicates that in the fourth season, winter, the level of the trend needs to be multiplied by 1.012; in other words, incomes are, on average, 1.2% higher in quarter 4.
- The importance of the seasonal may be assessed by means of the 'Seasonal test' which is also printed. This tests the statistical significance of the seasonal pattern at the end of the sample period. It is asymptotically chi-square, here $\chi^2(3)$. The test may be misleading if the seasonal pattern is stochastic and it has changed a lot over the sample. Thus if the seasonals have become smaller, they may be insignificant at the end, but not at the beginning. In this case it is important to assess the test in conjunction with a full plot of the seasonal. This point is relevant for UK income, where the seasonal effect becomes smaller over time as the economy becomes less dependent on seasonal factors.

The way in which the seasonal pattern has evolved over time is seen in the smoothed seasonal given by the Components graphics dialog. Selecting only 'Seasonal' (in the 'Select components' section) and 'individual seasonals' (in the 'Further plots' section) produces Figure 4.5. The bottom graph shows the yearly seasonal plots with the evolution of each of the seasonals over time.

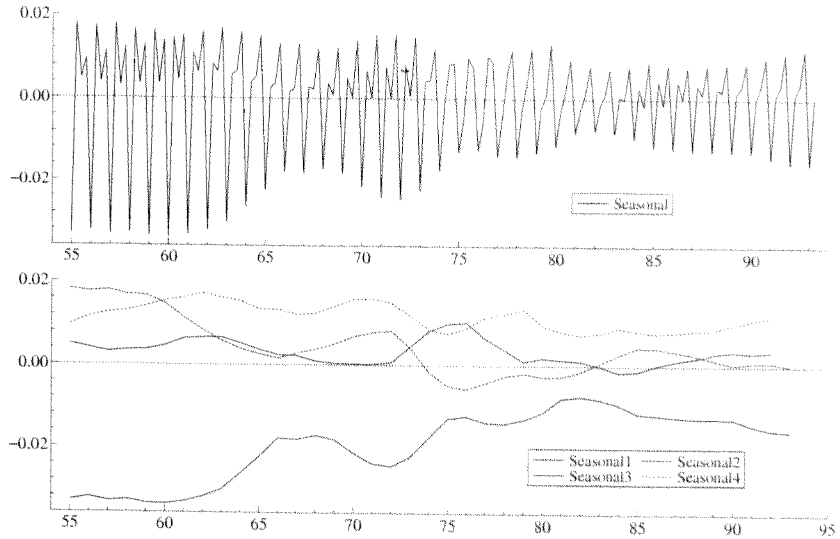


Figure 4.5 The estimated seasonal terms of income in UKCYP.

When forecasting is carried out, the latest seasonal pattern, as can be reported by State and regression output in the More written output dialog, is projected into the future.

```
Coefficients of components in final state at period 93(2)
              Coefficient      RMSE     t-value      Prob
Level            11.47471   0.00747  1535.91044 [0.00000]
Slope             0.00657   0.00114     5.76118 [0.00000]
Seasonal         -0.00595   0.00472    -1.25957 [0.20979]
Seasonal 2        0.00973   0.00481     2.02343 [0.04482]
Seasonal 3        0.00604   0.00356     1.69492 [0.09218]
```

4.3.2 Seasonal adjustment

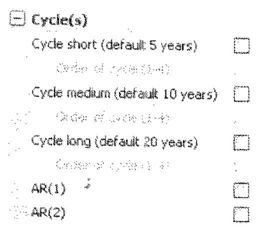
If a model has been fitted with a seasonal component, optimal model-based seasonal adjustment may be carried out within the Components graphics dialog by marking 'Seasonally adjusted Y' in the 'Further plots' section. This option subtracts the smoothed seasonal component from the full series. It can be stored for future use. A more detailed discussion of seasonal adjustment can be found in Chapter 7.

If the data has already been seasonally adjusted by another method, the 'No seasonal' option will normally be chosen. However, not all seasonal adjustment procedures succeed in removing seasonality. If this is the case, the residuals will tend to show serial correlation, particularly at the seasonal lag, s. A more thorough check can be carried out by including a seasonal component and seeing if it appears to be significant.

4.4 Cycle

A deterministic cycle is a sine-cosine wave with a given period. A stochastic cycle is constructed by shocking it with disturbances and introducing a damping factor. Such stochastic cycles are capable of modelling the kind of pseudo-cyclical behaviour which is characteristic of many time series, particularly economic and social ones. A deterministic cycle emerges as a limiting case. In many areas, such as meteorology, the question of whether cycles should be deterministic is an open one and so having such cycles as a special case of a more general model is very useful.

STAMP allows the inclusion of up to three stochastic cycles in a model. These are specified in the Select components dialog by marking boxes. If you have less than three cycles, which boxes you mark is a matter of convenience, determined by the default settings.



4.4.1 A simple cycle plus noise model

The RAINBRAZ.IN7 file consists of a single series, RainFort, which is annual data on the number of centimetres of rainfall in Fortaleza, a town in north-east Brazil. The series goes up to 1992 but we will conduct the initial analysis and modelling on the data up to 1984 only.

It is instructive to carry out a preliminary analysis on the data using the OxMetrics options. Select the Graphics dialog from the **Model** menu in OxMetrics. Select the only variable in the 'Database' listbox and press the 'All plot types >' button. In the section 'Actual series', select 'Sample' (double-click on 'full sample' or press button '...') and change the end of the sample to 1984. Then click on the graphical button

'Actual series' in the left-side panel and press 'Plot'. To view the correlogram, move to the time series section by clicking on 'Time-series properties' at the top-panel of the Graphics dialog and click on the graphical button 'Autocorrelation function (ACF)' (in left-side panel) and press 'Plot'. To view the spectrum, click on 'Spectrum' (more below) and press 'Plot'. The histogram of the series can be selected from the section 'Distribution'. Cancel the dialog to view the graphs as shown in Figure 4.6. We would expect a rainfall series to have a constant mean and this seems to be borne out by a graph of the series. Although the individual autocorrelations are quite small, the correlogram shows evidence of a cycle buried within the noise. The same message appears in the estimated spectrum, but more clearly. On this graph, the period is given by dividing 2 by the scaled frequency, on the horizontal axis. Thus there appears to be a cycle with a period of around 12 or 13 years. Re-estimating the spectrum with the window implied by 50 lags, indicates the possibility of a second cycle of around 25 years. There is also a smaller peak in the spectrum at around four years.

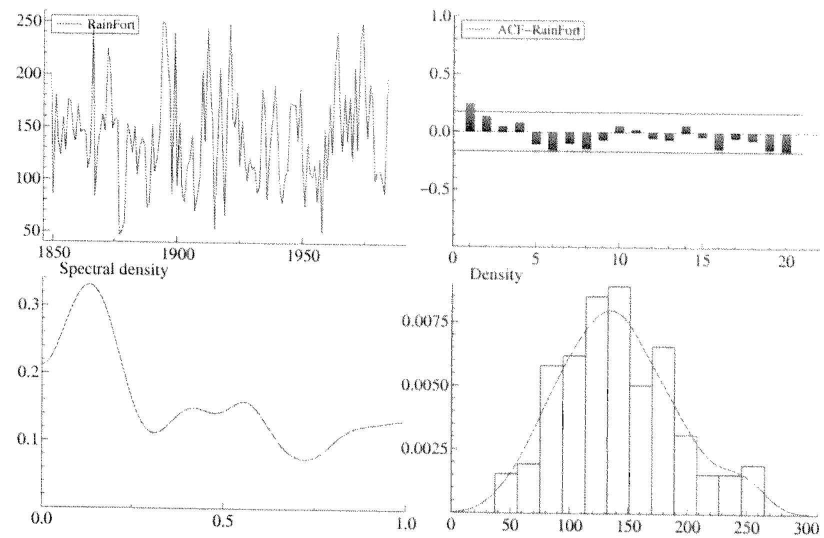


Figure 4.6 Summaries of the 'RainFort' series.

For the moment we will assume a single cycle. The possibility of such a cycle leads us to formulate a model consisting of a 'Level, Fixed', an 'Irregular' and a 'Cycle'. The 'Slope' component is not selected. To select the cycle, open the section 'Cycle(s)' in the Select components dialog and select 'Cycle medium'. This choice is convenient since the default is for the estimation procedure to start off with a period of 10. Set the end of the sample to 1984 in the Estimation dialog.

The most easily interpretable part of the output is the graph of the 'Cycles and ARs' and the 'Trend plus Cycles plus ARs' in the Test/Component graphics dialog. It can be seen in Figure 4.7 that the cycle is somewhat irregular in period and amplitude , and is dominated by the irregular component. The forecasts show that the cycle damps down towards zero as the lead time increases.

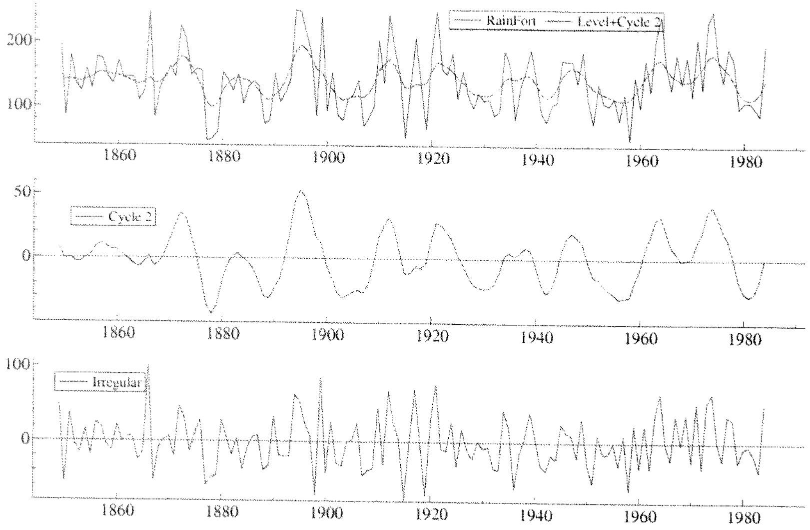


Figure 4.7 Estimated cycle for 'RainFort' series.

More precise information about the fitted cycle can be found in the More written output dialog. Selecting 'Parameters by component' prints for the cycle and irregular components:

```
Parameters in Cycle

Variance                    726.30
Period                      14.748
Frequency                  0.42603
Damping factor             0.85031

Parameters in Irregular

Variance                    1582.1
```

The parameters, which are defined formally in the next subsection, are as follows:

- A variance parameter which is responsible for making the cycle stochastic, σ_κ^2;
- A period (in years), $2\pi/\lambda_c$
- A frequency (in radians), λ_c.
- A damping factor, ρ;

The information in the frequency is more usefully presented in terms of the period (this is 2π divided by the frequency λ_c). The output shows this to be just under 15 years.

The variance of the cycle itself, as opposed to the variance of the disturbance term generating it, is also shown. The sum of this variance and the variance of the irregular should, in theory, be equal to the variance of the observed variable.

4.4.2 Statistical specification

The statistical specification of a cycle, ψ_t, is as follows

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1,\ldots,T \tag{4.7}$$

where λ_c is the frequency, in radians, in the range $0 < \lambda_c < \pi$, κ_t and κ_t^* are two mutually uncorrelated white noise disturbances with zero means and common variance σ_κ^2, and ρ is a *damping factor*. Note that the *period* is $2\pi/\lambda_c$. The stochastic cycle becomes a first-order autoregressive process if λ_c is 0 or π.

In estimating the model, the variance of the cycle itself, σ_ψ^2, rather than σ_κ^2, is taken to be the fixed parameter. Since $\sigma_\kappa^2 = (1-\rho^2)\,\sigma_\psi^2$, it follows that $\sigma_\kappa^2 \to 0$ as $\rho \to 1$ and (4.7) reduces to the deterministic, but stationary, cycle (4.8)

$$\psi_t = \psi_0 \cos\lambda_c t + \psi_0^* \sin\lambda_c t, \quad t = 1,\ldots,T, \tag{4.8}$$

where ψ_0 and ψ_0^* are uncorrelated random variables with zero mean and common variance σ_ψ^2.

The cycle plus noise model of the previous subsection may be written as

$$y_t = \mu + \psi_t + \epsilon_t, \tag{4.9}$$

where μ is the mean of the series and ϵ_t is a white noise term which is uncorrelated with ψ_t and has variance σ_ϵ^2. The decomposition of the variance of y_t may be written as:

$$\sigma_y^2 = \sigma_\psi^2 + \sigma_\epsilon^2. \tag{4.10}$$

4.4.3 Higher order cycles

The cycle component specification can be generalised as proposed by Harvey and Trimbur (2003). Smoother cycle processes can be specified as $\psi_t = \psi_t^{(k)}$ where

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix}, \tag{4.11}$$

for $j = 1,\ldots,k$, and where $\kappa_t = \psi_t^{(0)}$ and $\kappa_t^* = \psi_t^{*(0)}$ are two mutually uncorrelated white noise disturbances with zero means and common variance σ_κ^2 and for

$t = 1, \ldots, n$. A higher value for k leads to more pronounced cut-offs of the band-pass gain function at both ends of the range of business cycle frequencies centered at λ. For example, the model-based filter with $k = 6$ leads to a similar gain function as the one of Baxter and King (1999). Different variations within this class of generalised cycles are discussed in Harvey and Trimbur (2003) who refer to specification (4.11) as the balanced cycle model. They prefer this specification since the time-domain properties of cycle ψ_t are more straightforward to derive and it tends to give a slightly better fit to a selection of U.S. economic time series.

The value of k determines the smoothness of the cycle. In the case $k = 1$, the generalised cycle reduces to (4.7). The value of k, the 'Order of cycle', can be set in the 'Cycle(s)' section of the Select components dialog and has possible values 1, 2, 3, 4. Illustrations of the higher order cycle are given in §7.1.

4.4.4 Trend plus cycle

The series on US GNP is the y series in USYCIMP. It yields an important example of a trend plus cycle model.

In the Select components de-select 'Seasonal' component in the 'Basic components' section and 'Cycle short' in the 'Cycle(s)' section. Also select 'Level, Fixed' so as to give a smooth trend component. *The use of a smooth trend with a cycle often leads to a more attractive decomposition.* Estimate the model. The period of the cycle is about five years, corresponding to a plausible business cycle. The forecast plot, which was displayed earlier in Figure 1.3, is even more informative. It turns out that the peaks and troughs follow the movements charted by the National Bureau of Economic Research (NBER).

Other features of the results include the fact that the irregular variance has been estimated to be zero. Thus it can be — and in fact effectively has been — dropped from the model.

4.4.5 Multiple cycles

The model in (4.9) can be extended so as to include several cycles at different frequencies. Returning to RAINBRAZ and marking cycles 'Cycle medium' and 'Cycle long' (in addition to fixed level and no slope) gives the following results with data up to 1984:

```
UC( 1) Modelling RainFort by Maximum Likelihood (using RAINBRAZ.IN7)
    The selection sample is: 1849 - 1984
    The model is:   Y = Level + Irregular + Cycle 2 + Cycle 3

Log-Likelihood is -511.383 (-2 LogL = 1022.77).
Prediction error variance is 1740.01

Summary statistics
 std.error          41.713
 Normality          0.69437
```



```
H(45)               0.91860
r(1)               0.035794
r(10)               0.15912
DW                   1.8722
Q(10,7)              7.9241
R^2                 0.24116

Variances of disturbances.

Component                         Value    (q-ratio)
Level                           0.00000  ( 0.0000)
Cycle                        0.00067795  ( 0.0000)
Cycle 2                      0.00066725  ( 0.0000)
Irregular                        1748.8  ( 1.0000)

Parameters in Cycle

Variance                         338.97
Period                           12.958
Frequency                       0.48489
Damping factor                   1.0000

Parameters in Cycle 2

Variance                         224.96
Period                           24.587
Frequency                       0.25554
Damping factor                   1.0000

State vector analysis at period 1984
 - level is 143.356 with stand.err 3.59988.
 - amplitude of Cycle 2 is 25.0662
 - amplitude of Cycle 3 is 20.0908
```

Using more mathematical notation, this can be summarized as:

$$\begin{array}{llllll}
\tilde\sigma^2_{\psi 1} = 339, & \tilde\rho = 1, & \tilde\lambda_c = .485(\text{period} = 12.96), & \tilde\sigma = 41.7, & & \\
\tilde\sigma^2_{\psi 2} = 225, & \tilde\rho = 1, & \tilde\lambda_c = .256(\text{period} = 24.59), & Q(10,7) = 7.9, & N = .69, \\
\tilde\sigma^2_{\epsilon} = 1749, & & & R^2 = .24, & H = .92.
\end{array}$$

Thus estimation of the two-cycle model gives two deterministic cycles. The sum of the variance components is 2313; this is close to the series variance of 2293. The diagnostics are satisfactory and if a third cycle is included in the model, it is either very small or disappears completely, depending on the starting valuesused. The results are similar to those reported in Kane and Trivedi (1986) and Morrettin, Mesquita and Rocha (1985) where the authors used a regression model with the two frequencies determined from prior ideas or an inspection of the periodogram.

The two-cycle model appears to be stable and its predictive performance is rather good. Re-estimating the model using observations up to 1992 changes the parameter estimates very little and extrapolating from 1972 would have clearly predicted the droughts of the early 80s and 90s. In the stochastic cycle model the cycle in the forecast function tends to die away after a few years.

4.5 Autoregression

First-order and second-order autoregressive — AR(1) and AR(2) — components can be included in a structural time series model by marking the 'AR(1)' and/or 'AR(2)' box in the Select components dialog. The process is constrained to be stationary; that is, the AR coefficients are restricted to represent a stationary process. If this were not the case there would be a risk of them being confounded with the random walk component in the trend. As noted in the previous section, the stochastic cycle becomes an AR(1) if λ_c is 0 or π, but this possibility is avoided by constraining λ_c to be strictly between 0 and π. Obviously if it ends up being close, reformulating the cyclical component as an AR is probably appropriate.

If the series itself is a pure AR, of any order, rather than an AR plus other stochastic components, it may be estimated by OLS by using the 'Lag' option in the Formulate dialog to create lagged values of the dependent variable. As an example, consider the inflation series in USYCIMP, Dp, used in §4.2, and formulate a model for the first differences with three lags. To include the lags, you select in the 'Lags' section the option 'Lag 0 to' (instead of the default 'None') and type 3 in the next box. Then double-click on 'Dp' in the 'Database' selection or press on << after selecting 'Dp'. As a result, the actual series 'Dp' appears as a Y variable and the three lags are included in the model as explanatory variables.

After estimation, the AR coefficients are printed using the 'Parameters by component' option in Test/More written output. *A useful theoretical exercise is to expand the reduced form MA(1) process obtained from equation (4.2) as an infinite AR and compare the implied coefficients with those of the fitted AR(3); see Harvey (1993, p. 133).*

4.6 Exercises

(1) Fit a local linear trend model to the price level, p, in USYCIMP. Compare the results with the local level model.

(2) The PURSE file contains observations on the number of purses (handbags) snatched in the Hyde Park area of Chicago. Fit a local level model, trying various transformations, such as logarithms, square roots and other powers, for example 2/3. Which is best? What happens if you include a slope in the model?

(3) The series in USYCIMP have been seasonally adjusted by the US Census Bureau's X-11 program. Fit a local level model to Dp. What do you conclude about the effectiveness of X-11 in this case?

(4) Estimate the RAINBRAZ series with a stochastic level. What do you conclude?

- In the article by Newton, North and Crowley (1991), it is suggested that the global ice volume series, contained in ICEVOL, contains cycles of around 100,000, 41,000 and 22,000 years, but that these may not be strictly periodic. Bearing in mind that the points are 2,000 years apart, estimate a model with these three cycles and make predictions for the next 100,000 years.

(5) Fit a basic structural model to ofuGASI in ENERGY. Compare the fit when the seasonal is constrained to be fixed. Do the diagnostics indicate a misspecification?

(6) Seasonally adjust the series in the previous question using the BSM. Are there any seasonal effects remaining?

(7) Estimate the rainfall series in RAINBRAZ with three cycles. How does the predictive performance compare with the two-cycle model?