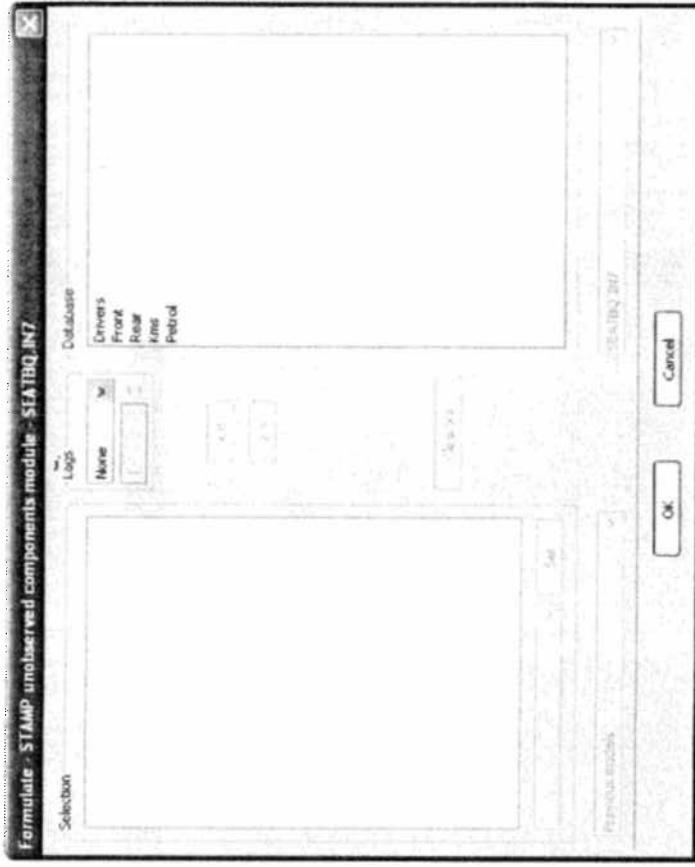
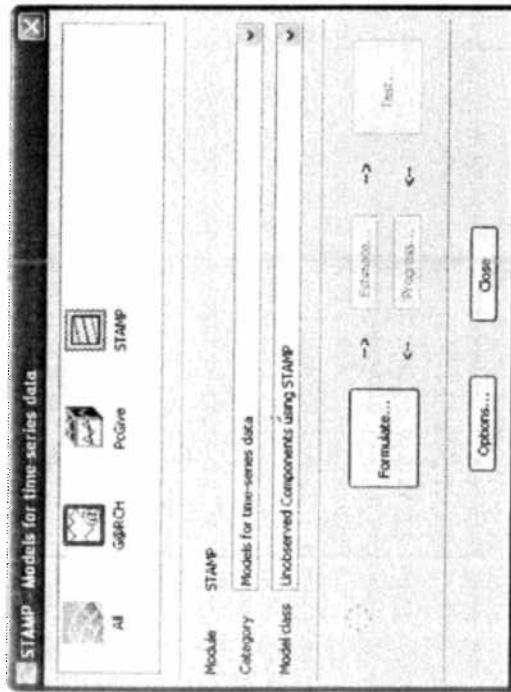


Chapter 8

Tutorial on Model Building and Testing

This chapter explains how to set up models in STAMP and evaluate the results. First you need to load the data SEATBQ into OxMetrics. This tutorial will be centred around the **Model** window of OxMetrics that is typically started by pressing Alt+Y:



8.1 Specification of univariate models

To start the process of specifying a model, press Alt+Y to start the **Model** window. When the STAMP pictogram is not highlighted, click on the 'STAMP' pictogram, select 'Models for time-series data' as 'Category' and 'Unobserved Components using STAMP' as 'Model class'. To start formulate a model, press the 'Formulate' button:

8.1.1 Formulate a model

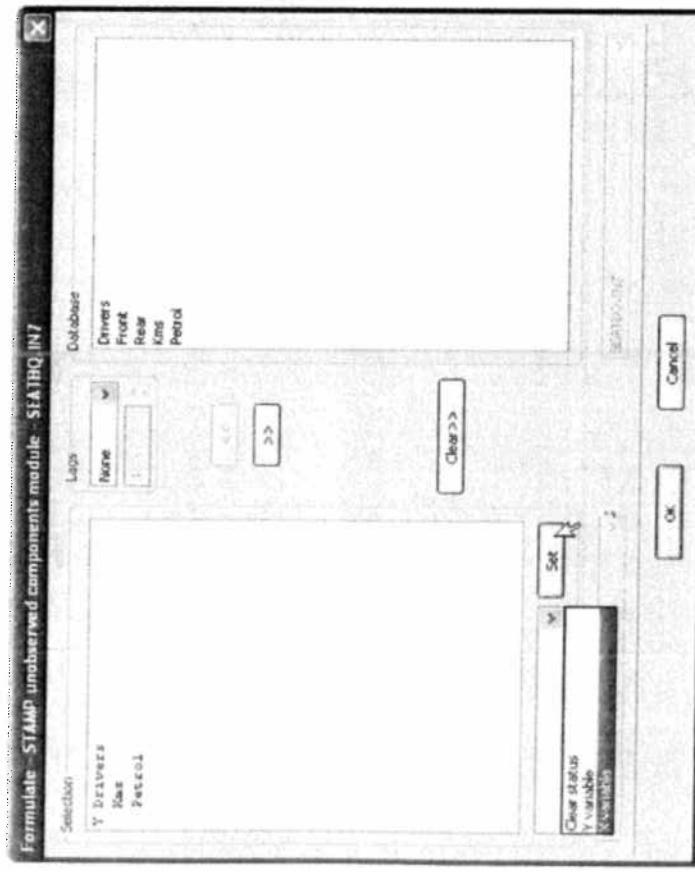
The Formulate dialog specifies the variables to be included in a model, assigns their status as dependent (Y) or explanatory, and creates lagged values. You can move around the options in the usual way, with the mouse or with Tab and Arrow keys.

In this seat belt example, the 'Drivers' series will be used as the dependent variable, with Kms and Petrol being explanatory variables.

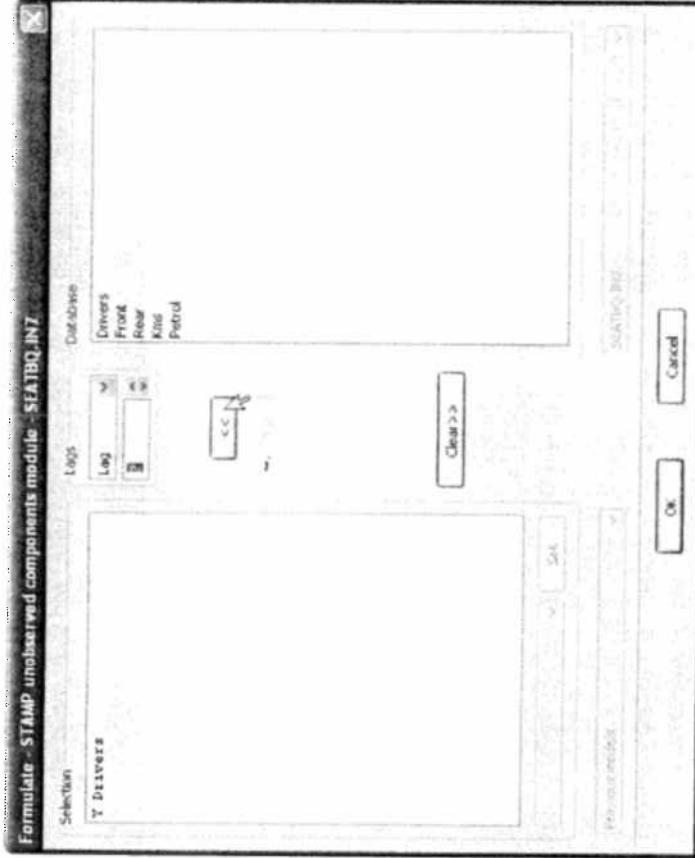
Familiarity with the ideas underlying structural time series modelling is assumed. The aim is to systematically describe the various dialogs rather than to explain the underlying statistical methodology. Those interested in the substance of this application, which concerns the assessment of the effect of the seat belt law in Great Britain, should read Harvey and Durbin (1986).

Select the series from the 'Database' listbox which need to be transferred to the 'Selection' listbox. One way is to click on one database variable in the 'Database' listbox and press the << button. You may also use the multiple selection facility by clicking on database variables and holding the Ctrl key simultaneously. Then, press

the << button. The first series from the multiple selection will be labelled as dependent (Y). The other variables are not labelled which imply that they are explanatory variables (X). To change the status of a variable in the model, select it and use the droplist below the 'Selection' listbox where you can choose from 'Y variable', 'X variable' and 'Clear status'.



To remove variables from the 'Selection' listbox, select the appropriate variables in this listbox and press the >> button. To create lagged variables, mark the relevant variables in the 'Database' listbox, change the option from 'None' to 'Lag' in the droplist of the 'Lags' section (between the 'Selection' and 'Database' listboxes). Select the appropriate lag number in the editbox below the droplist. Then press the << button. Alternatively, you may also want to select 'Lag 0 to' in the droplist of the 'Lags' section. In this case all lags from 0 to the specified lag will be moved to the 'Selection' listbox.



The lagged variables which are not required can be deleted by selecting the variables in the 'Selection' listbox and pressing the >> button.

The model for Drivers does not require lags so if you created them, delete them now. Two further options in the Formulate dialog are offered by the two droplists at the bottom of the dialog:

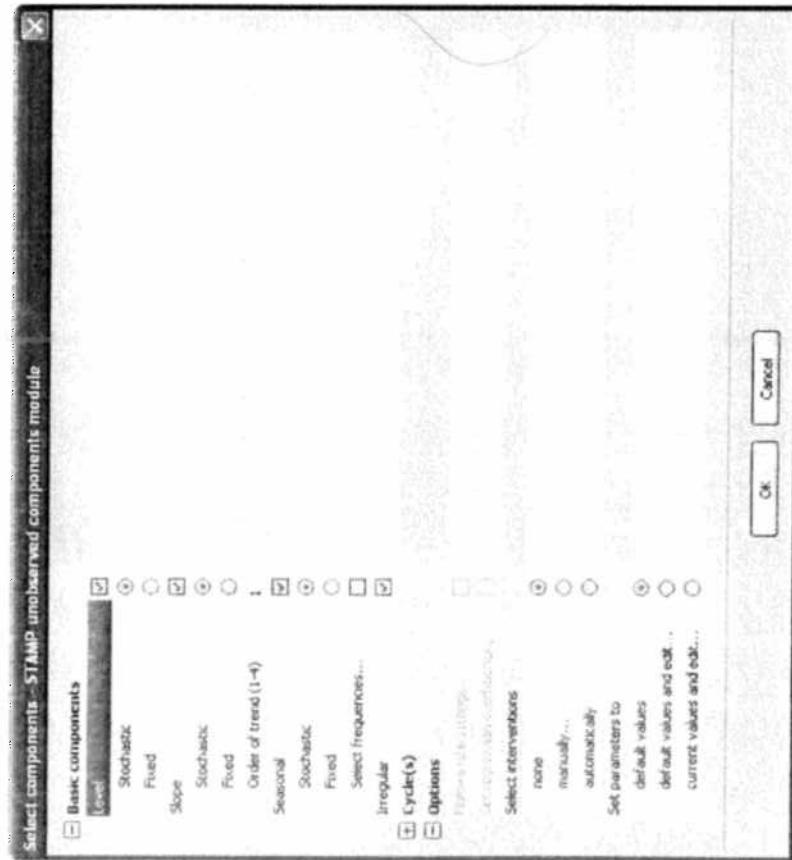
- In each STAMP session, previous models are stored in memory and its settings can be reloaded using the 'Previous models' droplist. Various details of the model are reloaded including component specifications, interventions, parameter estimates and restrictions.
- When multiple databases are active in OxMetrics, one can switch between databases by changing the database using the droplist below the 'Database' listbox.

The button Clear>> removes the current model or the model that is formulated. Once this button is pressed, all settings are lost and model formulation can be restarted in this dialog.

The OK button leads you to the Select components dialog.

8.1.2 Selection of components

The Select components dialog:



This dialog allows you to specify the unobserved components to be included in the model. The meaning of the different choices should be clear to anyone with a knowledge of structural time series models. If they are not, the tutorial on components, Chapter 4, should be consulted. The combinations needed to specify some of the more common time series models of trend are shown below:

- (1) Local level or random walk plus noise – ‘Level, Stochastic’ selected and ‘Slope’ not selected;
- (2) Local level with drift – ‘Level, Stochastic’ and ‘Slope, Fixed’ selected;
- (3) Smooth trend – ‘Level, Fixed’ and ‘Slope, Stochastic’ selected;
- (4) Generalised trend – ‘Level, Fixed’ and ‘Slope, Stochastic’ selected and value ‘Order of trend’ is 2, 3, or 4;

Note that certain combinations are inadmissible. For example, you cannot have a slope with no level. When ‘Level’ is not selected but ‘Slope’ is selected, the program automatically activates the level component as ‘Level, Fixed’.

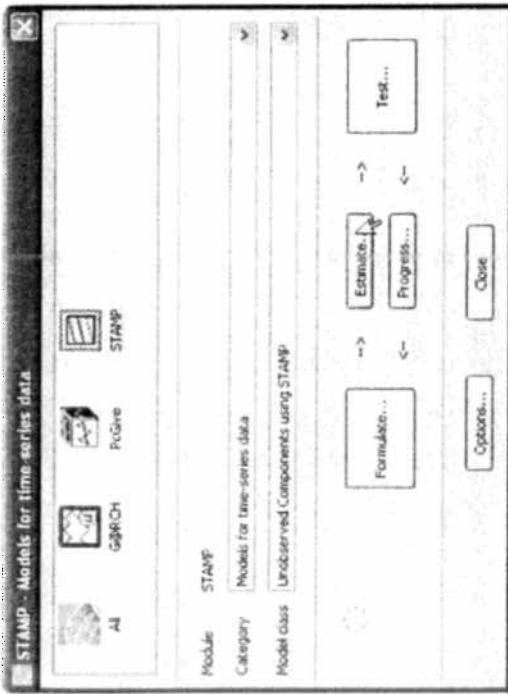
The default model is the *basic structural model*. To change the setting, move the cursor to the appropriate box in the usual way; that is, use the mouse. Alternatively, you can use the Arrow keys to change options within a group. The Tab key can be used to change from one group of options to another. The first three groups of components are selected by check boxes. The options ‘Stochastic’ and ‘Fixed’ can be specified using radio buttons. The slope component is further determined by the ‘Order of trend’; this value must be an integer and can only take the value 1 (default), 2, 3 and 4. The groups associated with the level and slope and seasonal components belong to the section ‘Basic components’.

To select other (stationary) components, the section ‘Cycles’ need to be expanded by clicking on ‘Cycles’. Here, for example, two out of three of the ‘Cycle’ boxes may be marked together with ‘AR(1)’ or ‘AR(2)’ (or even both). The three cycles are distinguished from each other by having different starting values for the period (or frequency) of the cycle. These are starting values for the numerical optimisation procedure. They can be changed in another dialog later.

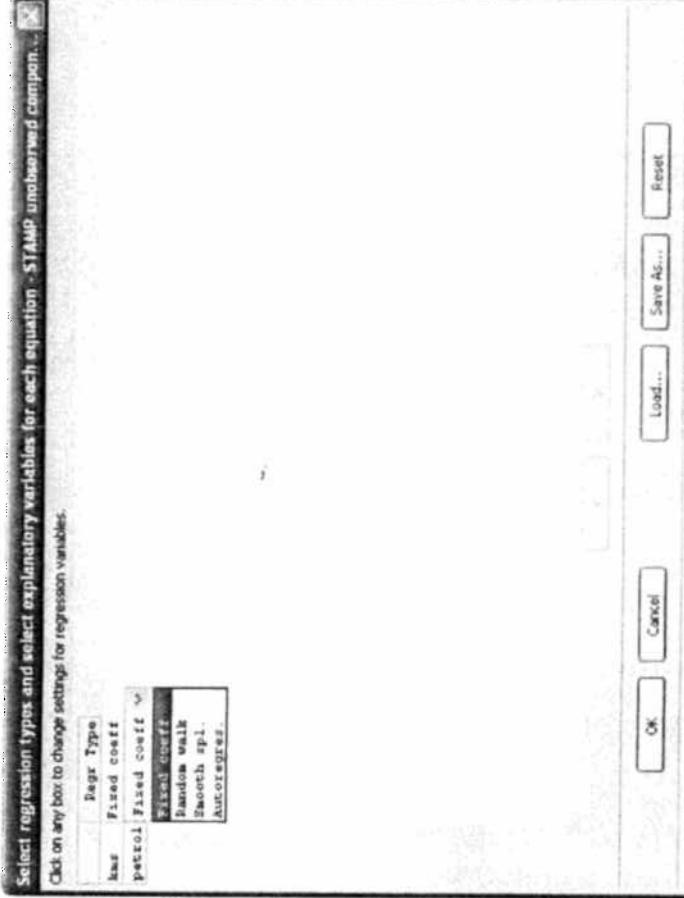
Having specified the components to enter into the model, there are two ways of proceeding for univariate models. The first possibility is to move directly to the Estimate dialog. This is appropriate if the model is fully specified and can be achieved by keeping the default settings in the ‘Options’ section (‘Select interventions none’ and ‘Set parameters to default values’) and by pressing OK. The Estimate dialog is discussed in §8.2. The second possibility is to choose one or more options in the ‘Options’ section. When explanatory variables are selected, the option ‘Set regression coefficients’ is activated by default. This option opens the ‘Regression coefficients’ dialog and allows the user to select time-varying regression parameters with a choice of different specifications.

The option ‘Select interventions manually’ opens the Select interventions dialog while the option ‘Select interventions automatically’ will activate the automatic outlier and break detection procedure during the estimation process. The options ‘Set parameter values to current values and edit’ and ‘Set parameter values to current values and edit’ open the Edit and fix parameters dialog. In the former case, the default values are displayed while in the latter case, the current parameter values are displayed (when available). You may also want to return to the Formulate dialog which requires pressing Cancel button and type Alt+y (within the OxMetrics program).

After leaving the Select components dialog, it can be re-entered at any time from the Model window (press Alt+y within OxMetrics) and press the Estimate button.



The newly re-entered Select components dialog will always display the current specified model which can be amended at any time.



For each explanatory variable, one can select the nature of the regression coefficient. The default option is a fixed parameter for each regression coefficient. However, time-varying regression parameters can also be considered and you can choose three different specifications for a specific time-varying regression coefficient δ_t :

- Random walk: $\delta_t = \delta_{t-1} + u_t$, where $u_t \sim NID(0, \sigma_u^2)$;
- Smoothing spline: $\Delta\delta_t = \Delta\delta_{t-1} + u_t$, where $u_t \sim NID(0, \sigma_u^2)$;
- Return to normality: $\delta_t - \delta = \rho_\delta(\delta_{t-1} - \delta) + u_t$, where $u_t \sim NID(0, \sigma_u^2)$ and δ is the "long-term" fixed coefficient.

The additional parameter σ_u^2 and, possibly, ρ_δ will be simultaneously estimated with the other parameters in STAMP. For the return to normality specification, the additional coefficient δ is part of the fixed regression coefficients and is placed in the state vector. Once the different specifications are selected for each regression parameter, press OK.

For the Drivers model, the regression coefficients are taken as *fixed parameters and the default settings can be accepted*.

8.1.4 Selection of interventions

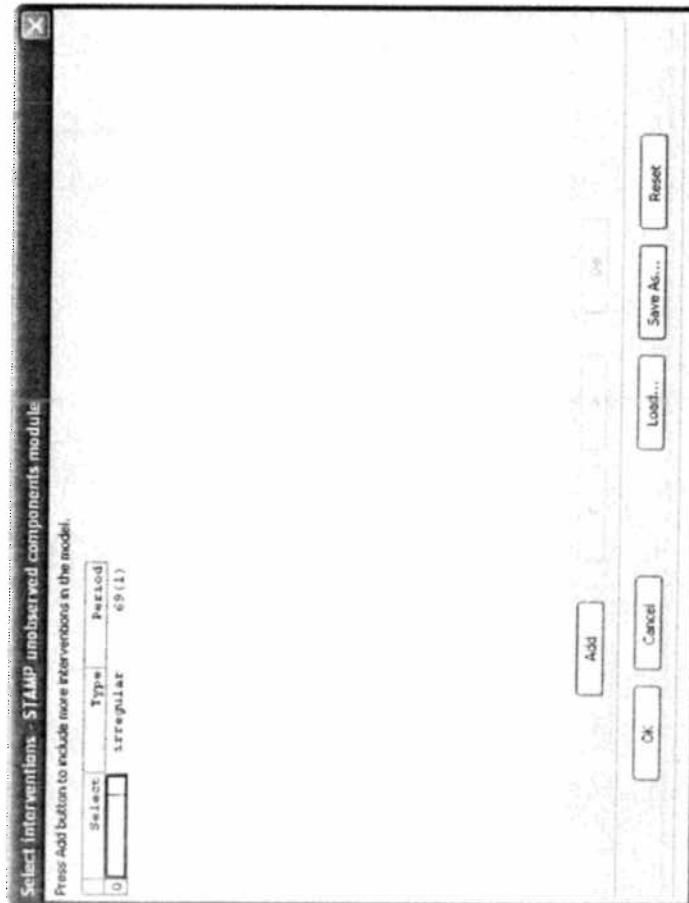
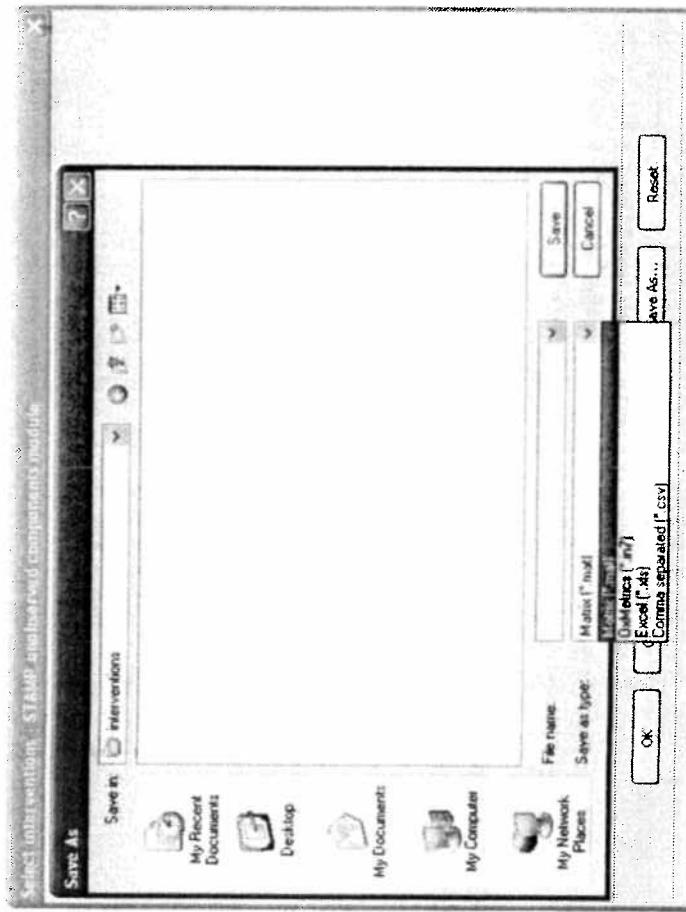
The Select interventions dialog

8.1.3 Specify regression coefficients

The Regression coefficients dialog consists of a series of dropdown boxes:

Add, at the bottom of the dialog, is pressed. In this way a long list of interventions can be created from which interventions can be included or excluded in the model (using the 'Select' column).

The options 'Load' and 'Save As' allow the inclusion of a list of interventions that is created before and saved by the 'Save As' option. The 'Save As' dialog

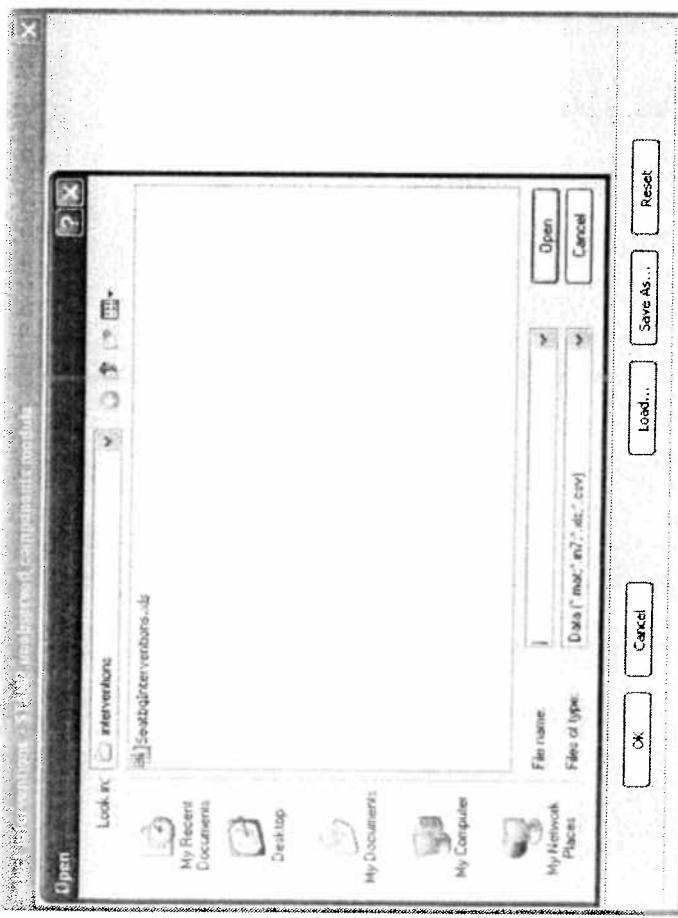


allows you to construct three types of interventions:

- *Impulse/Irregular* - an impulse intervention only affects the observation in question. It can be regarded as picking up an abnormal value of the irregular component, perhaps caused by a measurement error or a strike. In other words the observation in question is regarded as an *outlier*.
- *Level/Step* - a step intervention is a permanent increase in the level of the trend. It represents a structural break, perhaps caused by a change in policy.
- *Slope/Staircase* - a slope intervention leads to a permanent change in the direction of the trend.

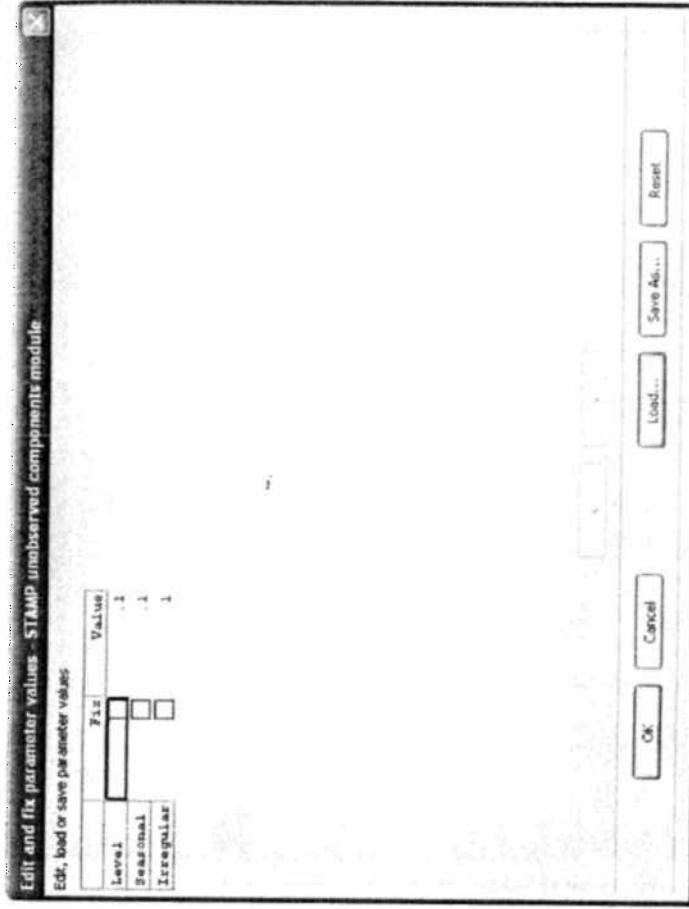
All interventions are incorporated as explanatory variables in the model directly. To construct an intervention variable, enter a date in the column 'Period' either by typing the date or by double-clicking on the date and using the mouse and/or arrow keys. The so-called spin keys can be used to increase or decrease the year number or period number. The default type of an intervention is Irregular. By double-clicking on this entry in the 'Type' column, a dropdown appears from which you can choose the appropriate intervention: irregular, level and slope. Once the appropriate intervention type and date is entered, the intervention becomes part of the model when the checkbox in the first column 'Select' is activated. A new entry for an intervention appears when the button

can save the list of interventions in different formats and the 'Load' dialog



can load a list of interventions in different formats.

The British Government brought in a law at the end of January 1983 which made the wearing of seatbelts compulsory. In our seat belt example we might represent the effect of this law by a level intervention in 83Q1.



In this dialog you can

- set initial values of parameters that are used for starting the numerical maximisation of the likelihood function in the Estimate dialog, see below.
 - set parameter values and fix them at these values so that these will **not** be estimated.
- In the entries of the last column 'Value', parameter values can be edited in the usual way. In the first column 'Fix' it is indicated whether parameters need to be fixed at their current values (checkbox is signed) or whether these values are used as starting values for estimation (checkbox is empty). In this dialog, the options 'Load' and 'Save As' are also available to load and save, respectively, parameter value settings.

For the Seatbelt case, just keep the default of estimating all parameter values and press OK.

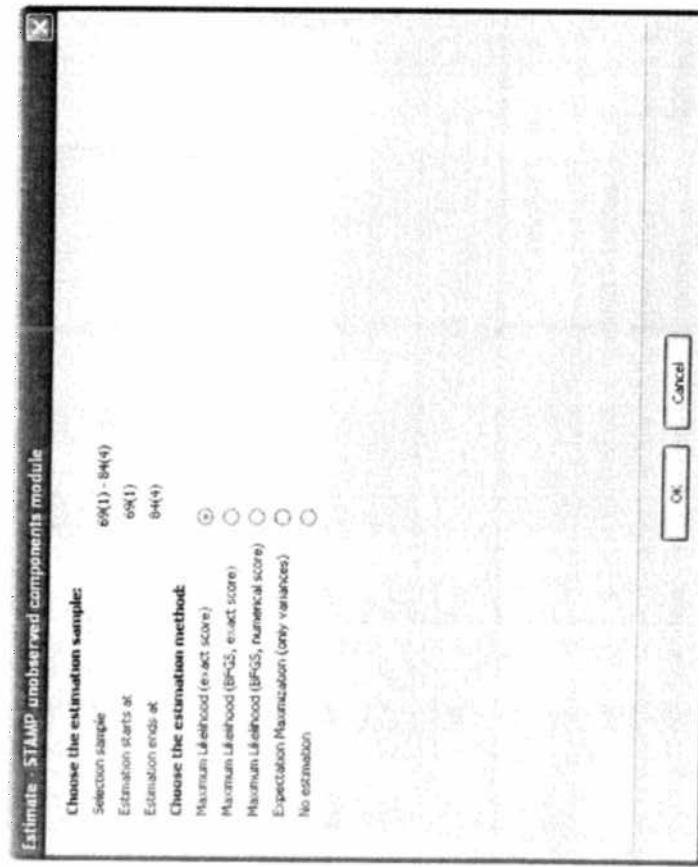
8.2 Estimate a model

The Estimate Model dialog gives the start sign to the estimation procedure of STAMP. When this has finished, some basic estimation results appear in the Results window of

8.1.5 Edit and fix parameter values

The dialog Edit and fix parameter values can be accessed when the option 'Edit and fix parameter values' is activated in the Select components dialog:

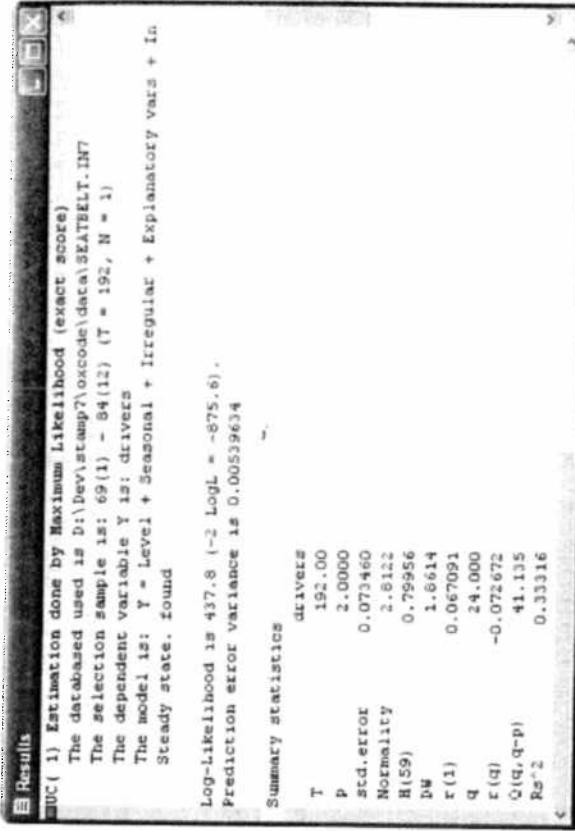
8.2.1 Estimate dialog



The Estimate Model dialog offers the user a number of options. Firstly the 'Estimation sample' may be set to something other than the 'Selection sample' implied by the database in OxMetrics. Changes are made to the sample period by accessing the appropriate edit boxes using the mouse and inputting the new date. The Estimate Model dialog provides different ways of estimating a structural time series model by the method of maximum likelihood (ML). The different procedures are described in the next section.

8.2.2 Maximum likelihood

The first option 'Maximum Likelihood (exact score)' is the fully automatic procedure which is the one to choose when you are a first user of STAMP.



The estimation procedure automatically enters into an initial estimation routines. The details of these routines are given in §9.6. The initial estimation is based primarily on the EM algorithm and a naive nonlinear estimation method. Once the initial estimation routine has obtained a solution close to the ML estimator, the program switches to computing full ML estimates using a numerical optimisation procedure. Again you should refer to §9.6 to understand exactly what is happening here. It should be stressed, however, that STAMP has been designed in such a way that, for most users, the estimation phase can be regarded as a "black box".

Once estimation is in progress, some dots are printed; it confirms that the estimation is in progress. In most situations, the on-line printing of the dots will not be noticed. However, when the time series is very long and/or the model is large, the estimation process may take some time and printing the dots may be useful. When the optimisation routine has converged or it has reached the maximum number of iterations, it prints some output to the Results window about the model and the estimation procedure. Furthermore, the program produces plots of the estimated components in the Model graphics window. More output and, very importantly, more graphical output can be generated from the dialogs of the Test menu; see §8.3.

There are four different estimation methods:

- Maximum likelihood (exact score): the default procedure developed especially for STAMP;
- Maximum likelihood (BFGS, exact score): the standard BFGS method of numerical optimisation where exact scores are computed; it is implemented for the `ox` programming language and known as the function `MaxBFGS()`.

- Maximum Likelihood (BFGS, numerical score): as the previous method but based on numerical scores;
 - Expectation Maximisation (only variances): the standard EM method using the Ox/Ssfpack functions, it is only implemented for the estimation of variances. Other parameters are kept fixed during the estimation method.
- Finally, the estimation part can also be circumvented in STAMP. Either by cancelling the dialog (press the red box in the upper-right corner of the dialog) or by selecting 'No estimation' and proceed as normal by pressing OK. In this case, the current (initial) parameter values are taken and output can be generated as detailed below. To emphasize that the model is not estimated, the default output in the Results window is limited to

UC(1) No estimation done

The databased used is SEATBQ.INT

The selection sample is: 69(1) - 84(4) (T = 64, N = 1)

The dependent variable Y is: Drivers

The model is: Y = Level + Seasonal + Irregular + Expl...

instead of

UC(1) Estimation done by Maximum Likelihood (exact score)

The databased used is SEATBQ.INT

The selection sample is: 69(1) - 84(4) (T = 64, N = 1)

The dependent variable Y is: Drivers

The model is: Y = Level + Seasonal + Irregular + Expl...

Steady state, found

Log-Likelihood is 136.987 (-2 LogL = -273.974).

Prediction error variance is 0.00642936

Summary statistics

	Drivers
T	64.000
p	2.0000
std. error	0.080183
Normality	2.7592
H(19)	0.81746
DW	2.3173
r(1)	-0.17758
q	8.0000
r(q)	-0.027749
Q(q,q-p)	7.9327
Rs^2	0.58262

Finally, the estimation part can also be circumvented in STAMP. Either by cancelling the dialog (press the red box in the upper-right corner of the dialog) or by selecting 'No estimation' and proceed as normal by pressing OK. In this case, the current (initial) parameter values are taken and output can be generated as detailed below. To emphasize that the model is not estimated, the default output in the Results window is limited to

UC(1) No estimation done

The databased used is SEATBQ.INT

The selection sample is: 69(1) - 84(4) (T = 64, N = 1)

The dependent variable Y is: Drivers

The model is: Y = Level + Seasonal + Irregular + Expl...

instead of

UC(1) Estimation done by Maximum Likelihood (exact score)

The databased used is SEATBQ.INT

The selection sample is: 69(1) - 84(4) (T = 64, N = 1)

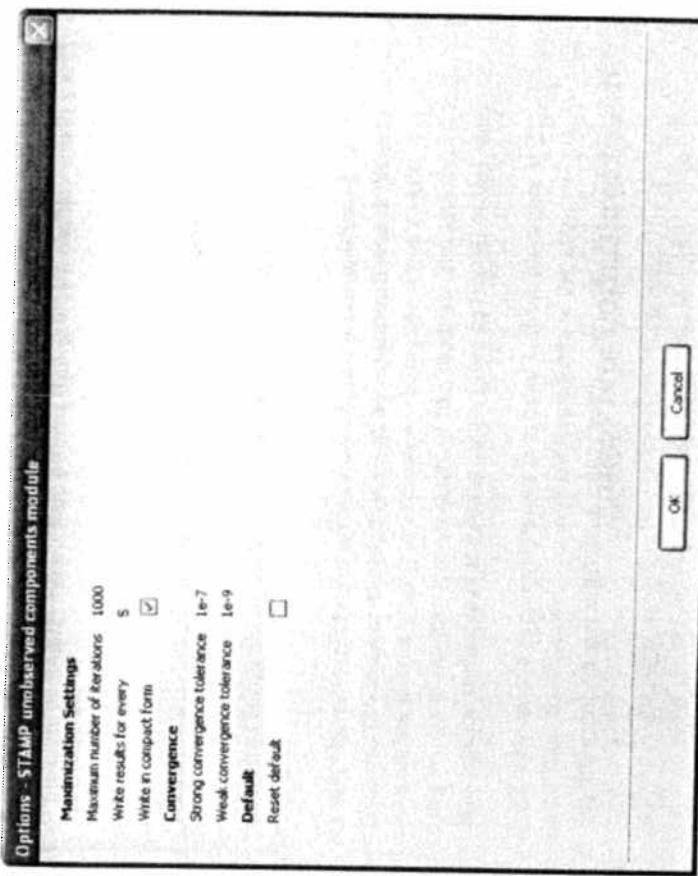
The dependent variable Y is: Drivers

The model is: Y = Level + Seasonal + Irregular + Expl...

Steady state, found

Log-Likelihood is 136.987 (-2 LogL = -273.974).

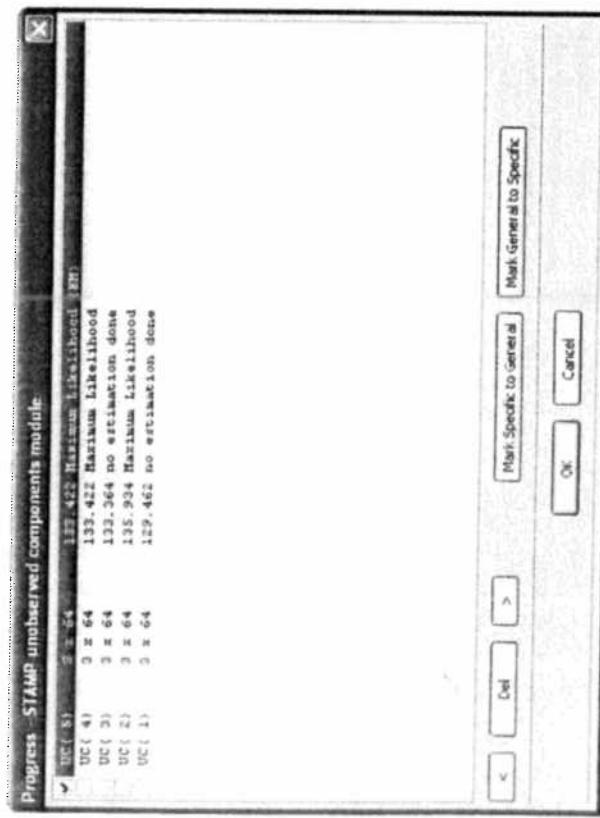
Prediction error variance is 0.00642936



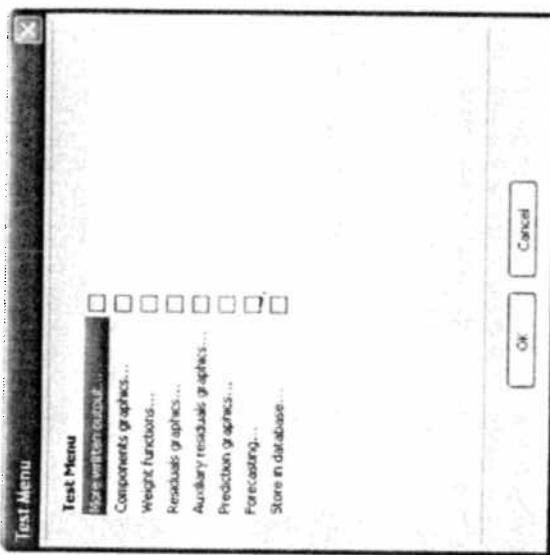
Further you can request more output during estimation by reducing the times that output is written to the Results window and by de-activating the option for 'Write in compact form'.

8.2.4 Progress

In the main Model menu (to be accessed via Alt+y, the button Progress gives access to the the Progress dialog



This dialog allows the user to regain the settings of previous models that are considered in STAMP. It also facilitates comparisons between models.



Each of the eight available options here offers a dialog which enables you to specify exactly what statistics you would like. The first dialog produces written output only, the second dialog is mainly concerned with graphical output of estimated components. The third dialog provides more insights about the signal extraction carried out by STAMP. The fourth to sixth dialogs give graphical diagnostics based on various residual series. They further allow the user to carry out diagnostic checking by written. The seventh dialog offers forecasting facilities. The final eighth dialog gives options to store estimated components in the database of OxMetrics. They are described in the next section.

The output is generated using advanced statistical methods such as the Kalman filter, smoothing algorithms, etc. Most graphics and diagnostic statistics generated from the Test menu are standard and easy to understand. However, if you require some guidance and examples, go to the previous chapters of this manual. The technical details of the algorithms and the formulae of the statistics are given in Part III of this book.

All the written and graphical output is sent to OxMetrics where it may be edited and saved. Thus, for example, you may add comments to the output as you go along or delete any information, perhaps on an unsuccessful run, which you regard as superfluous.

8.3 Model evaluation and testing

Once estimation is complete, STAMP outputs an estimation report, a diagnostic summary report, the estimated variances of the disturbances and a state analysis in the Results window. Further output may be obtained from the Test menu. The Test menu is accessed from the Model menu (press Alt+y) and by pressing the button 'Test'.

The results shown below apply to the Drivers quarterly seat belt model with a stochastic level, no slope, stochastic seasonal, two explanatory variables (Kms and Peto), level intervention at 83.1 and irregular. The restriction described for the Parameter vector dialog is not imposed. The model is estimated by Maximum likelihood.

8.3.1 Estimation report

The most important piece of information in the estimation report is the message 'Very strong/Strong/Weak/No convergence in ... iterations'. The precise definition of these terms can be found in §9.6, but from a practical point of view, the appearance of the word 'strong' is to be welcomed. If convergence is not strong, you have several possibilities:

- Proceed to examine the model output in order to see how well it fits and where, if anywhere, it appears to deficient.
- Return to the Options dialog, and increase the number of iterations. Re-start the estimation process.
- Formulate a new model.
- Go to the Edit and fix parameter values dialog and start estimation from a new set of initial parameter values. Also some parameter values may be altered and restricted before the estimation process is re-started.

Convergence in the 'Drivers' model is 'Very strong'. The estimation part of the output is given below:

```
Estimating...
Strong convergence relative to 1e-006
- likelihood cvg 0
- gradient cvg 0.0000337282
- parameter cvg 0
- number of bad iterations 3
Estimation process completed.

UC( 1) Estimation done by Maximum Likelihood (exact score)
The database used is SEATBQ.INT
The selection sample is: 69(1) - 84(4) (T = 64, N = 1)
The dependent variable Y is: Drivers
The model is: Y = Level + Seasonal + Irregular + Expl...
Steady state. found
```

based on the first p autocorrelations; it should be tested against a χ^2 distribution with q degrees of freedom. The classical Durbin–Watson test is also given.

The heteroskedasticity test statistic, $H(h)$, is the ratio of the squares of the last h residuals to the squares of the first h residuals where h is set to the closest integer of $T/3$. It is centred around unity and should be treated as having an F distribution with (h, h) degrees of freedom. A high (low) value indicates an increase (decrease) in the variance over time.

The normality test statistic is the *Bowman–Shenton* statistic based on third and fourth moments of the residuals and having a χ^2 distribution with 2 degrees of freedom when the model is correctly specified. The 5% critical value is thus 5.99. High values are often caused by outliers. The model should not necessarily be rejected, but further investigation, perhaps by looking at the auxiliary residuals, is required.

The basic goodness of fit measure is the *prediction error variance*. Its square root is shown as the equation standard error, 'Std. Error'. The appropriate 'R²' (R^2) measure is also given; the statistics R_d^2 and R_s^2 are more suitable in time series as they compare the fit with a random walk plus drift and a random walk with fixed seasonal dummies, respectively.

Part of the default output is also concerned with the estimated state vector. This output will be discussed in more detail below.

Again the Drivers model is given below as an illustration.

```
UC( 2) Estimation done by Maximum Likelihood (exact score)
The database used is SEATBQ.INT
The selection sample is: 69(1) - 84(4) (T = 64, N = 1)
The dependent variable Y is: Drivers
The model is: Y = Level + Seasonal + Irregular + Explanatory vars + Interventions
Steady state. found

Log-Likelihood is 136.987 (-2 LogL = -273.974).
Prediction error variance is 0.000642936

Summary statistics
```

	Drivers
T	64.000
P	2.0000
std. error	0.080183
Normality	2.7592
H(19)	0.81746
DW	2.3173
r(1)	-0.17758
q	8.0000
r(q)	-0.027749
Q(q,q-p)	7.9327
Rs^2	0.58262

8.3.2 Diagnostic summary report

This output is by default automatically generated after the estimation phase. It presents the basic goodness of fit and diagnostic statistics. Based on this information you can probably decide whether it is worth proceeding to the Test menu or whether the model is so poor that you need to reconsider its specification. When convergence is not strong, it may be that iterating further will eventually produce an estimated model which satisfies the diagnostics.

An excessive amount of residual serial correlation is a strong indication that the model is not adequately capturing the dynamic structure of the series. The statistics denoted $r(j)$ give the autocorrelation at lag j , while $Q(p, q)$ is the Box–Ljung statistic

Level	Variance of disturbances
0.000354007	(q-ratio) 0.07211

It provides a variety of further information for the estimated model. In the 'Print parameters' section it reports on the parameters (usually estimated by maximum likelihood): variances, all parameters organised by each component and the vector of parameters including the transformations that are used during the estimation process and standard errors. These options print estimated standard deviations and estimated parameters associated with the cycle (damping factors and periods) and the autoregressive components. The estimated standard deviations of the disturbances in the model are printed. The figures in parentheses, the *q-ratios*, are the ratios of each standard deviation to the standard deviation associated with the largest variance.

```
Seasonal      1.20909e-005   ( 0.002463)
Irregular    0.00490946   ( 1.0000)

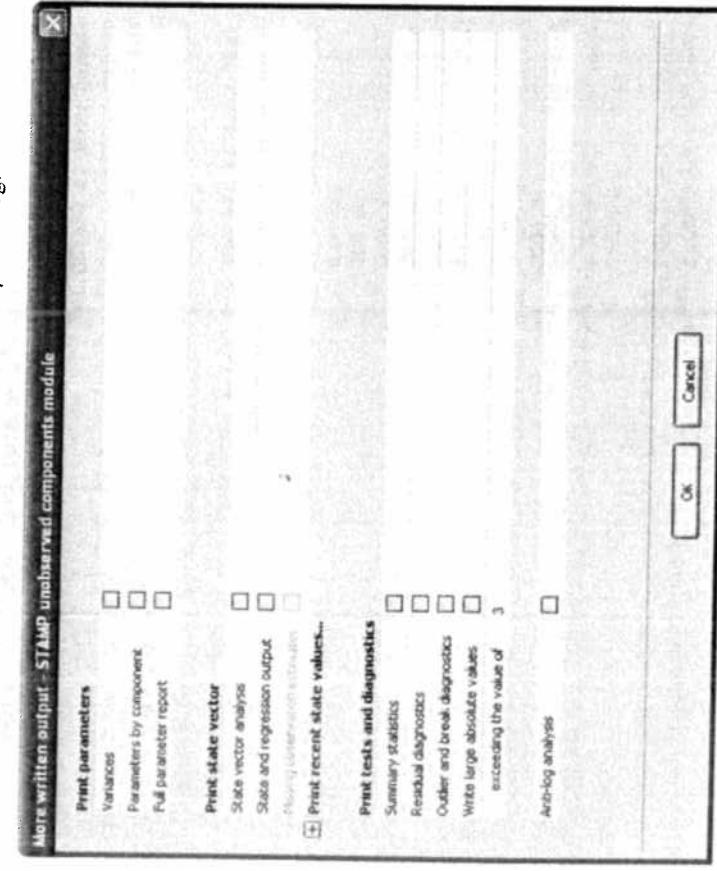
State vector analysis at period 84(4)
Value          Prob
Level          4.65465 [0.00835]
Seasonal chi2 test        73.05202 [0.00000]

Seasonal effects:
Period      Value          Prob
1           -0.07398 [0.00639]
2            -0.14188 [0.00001]
3            -0.01431 [0.60657]
4            0.23017 [0.00000]

Regression effects in final state at time 84(4)
Coefficient   RMSE       t-value     Prob
Level break 83. 1   -0.21893  0.05362  -4.08288 [0.00014]
Kms           0.22452  0.17712  1.26767 [0.21007]
Petrol         -0.26798  0.12235  -2.19021 [0.03262]
```

8.3.3 More written output

The first dialog in the Test menu is the More written output dialog.



The section 'Print state vector' in the Written more output dialog contains options that print values taken by the various components at the end of the sample. This information is stored in what we call the *state* vector. The default output from the 'Static and regression output' option has the following particular feature. The figure in square brackets after the *t-value* is a two-sided *Prob.* value which shows the probability of getting an absolute value of a standard normal variable greater than the *t-value* if the true parameter is zero. Additional output is generated for the seasonal and cycle components when it is relevant to the model. The generated output is discussed in Chapter 10.

In many cases the data are in logs and it is useful to activate the check box 'Anti-log analysis'. This provides additional information. For example, the actual value of the level is given and the seasonals can be interpreted as the factors by which to multiply the trend. Finally, the default option to 'Get steady state' is discussed in Chapter 9 and it is advised to keep this option activated.

The regression output for the explanatory variables and the interventions are given below.

State vector anti-log analysis at period 84 (4)

It is assumed that time series is in logs.

Value Prob

Level (anti-log)	105.07251	[0.00035]
Level (bias corrected)	448.19351	[.NaN]
Seasonal chi2 test	73.05202	[0.00000]
Seasonal effects:		

Period	Value	Prob	%Effect
1	0.92869	[0.00639]	-7.13079
2	0.86772	[0.00001]	-13.22766
3	0.98579	[0.60657]	-1.42060
4	1.25881	[0.00000]	25.88118

State vector at period 84 (4)

Coefficient	RMSE	t-value	Prob
4.65465	1.70328	2.73276	[0.00835]
0.18603	0.02482	7.49486	[0.00000]
Seasonal 2	-0.02984	0.02108	-1.41529
Seasonal 3	0.04414	0.01595	2.76714
			[0.00761]

Regression effects in final state at time 84(4)

	Coefficient	RMSE	t-value	Prob
Level break 83. 1	-0.21893	0.05362	-4.08288	[0.00014]
Kms	0.22452	0.17712	1.26767	[0.21007]
Petrol	-0.26798	0.12235	-2.19021	[0.03262]

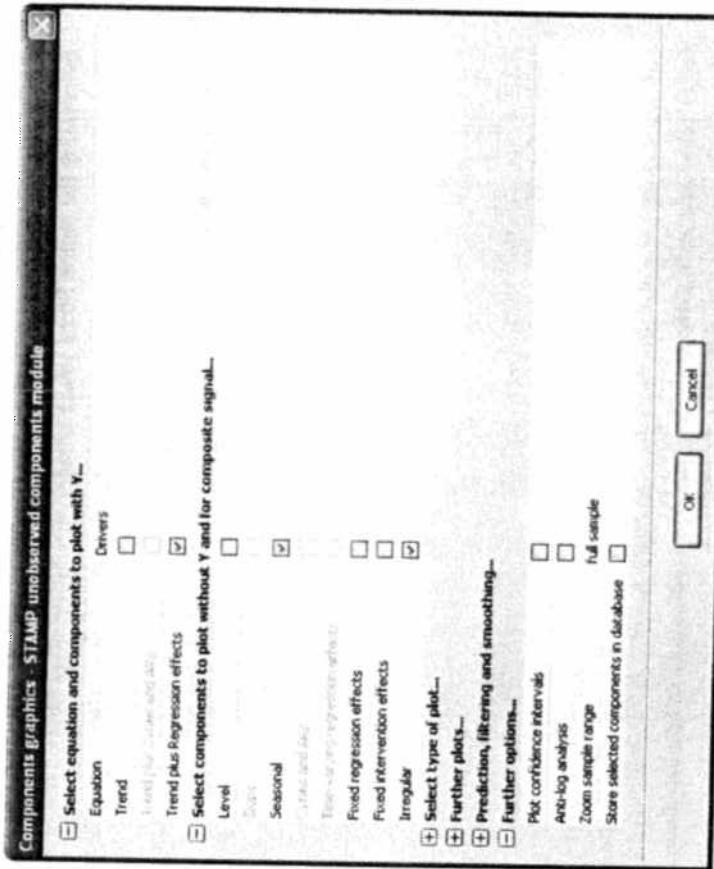
Note that in our example Kms' is not statistically significant but 'Petrol' is significant.

The intervention estimate is -0.22 , implying a fall in the level of front seat passengers killed and seriously injured of $100(1 - \exp(-0.22))$ which is around 20%.

The option 'Print recent state values' allows the user to investigate the different estimates (prediction, filtering and smoothing) of the different components in the recent period. This information can be useful to assess the level of revisions in estimates. For the level and seasonal components, the output is given by

Final state values of Level	Coef(t-1)	Coef(t)	Coef(T)	Rmse(t-1)	Rmse(t)	Rmse(T)	Rmse(T)
83 (4)	4.855	5.002	4.636	1.753	1.746	1.700	
84 (1)	5.002	4.887	4.642	1.746	1.738	1.701	
84 (2)	4.887	4.886	4.647	1.738	1.738	1.701	
84 (3)	4.886	4.839	4.652	1.738	1.738	1.702	
84 (4)	4.839	4.655	4.655	1.738	1.738	1.703	
Final state values of Seasonal							
Coef(t-1)							
83 (4)	0.2374	0.2259	0.2300	0.03459	0.03247	0.03006	0.02426
84 (4)	-0.07659	-0.07073	-0.07379	0.02623	0.02478		

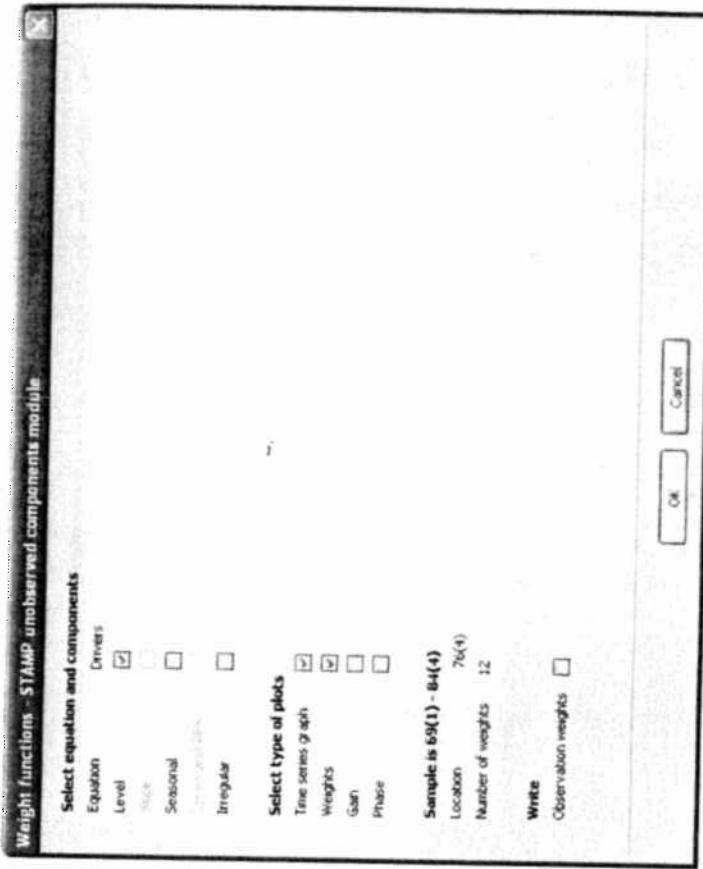
The information provided here is fundamental to the interpretation of the model. The smoothed estimates of the components are obtained using all the information in the sample; that is, they are constructed using observations which come after as well as those which come before. This is sometimes referred to as *signal extraction*. The Components graphics dialog controls the way this information is presented. In particular it provides options to graph particular components and combinations of components. The use of the 'Zoom sample range' option is particularly useful here as it allows an enormous amount of information to be very quickly assimilated.



Some interesting features of this dialog are given below.

- All components included in the model can be plotted together with the sum of regression effects ('Fixed regression effects') and the sum of intervention effects

- (‘Fixed intervention effects’). Confidence intervals can be graphed when the option ‘Plot confidence intervals’ is activated.
- The ‘Detrended Y’ and ‘Seasonally adjusted Y’ options in the ‘Further plots’ section allow subtraction of the estimated trend and seasonal component, respectively, from a particular series. These options often show some interesting features of the data which may not be clear at first sight.
- Each seasonal effect j , for $j = 1, \dots, s$, can be graphed against the years of the data-set. This set of s (the number of seasonals) graphs is obtained by marking the checkbox ‘Individual seasonals’ in the ‘Further plots’ section.
- When the data is recorded in logs, the components may be transformed by taking anti-logs (exp) which can be requested in the ‘Further options’ section using ‘Anti-log analysis’.
- Sometimes it is more useful to focus on the predicted and/or filtered estimates; that is, the one-step ahead predictions and/or the concurrent estimates of the components, respectively. These options are available in the section ‘Prediction, filtering and smoothing’.
- All series generated through this dialog can be saved in the database of OxMetrics by pressing the button **Store select components in database**.
- The signal extraction sample can be changed in this dialog using option ‘Signal extraction sample’. This option is different than ‘Zoom sample range’. The latter option has no effect on the sample selection, it just focuses on a specific sample range in the plots.



It may be interesting to view how the weight and gain functions change when the ‘Location’ in the sample approaches the end of the sample.

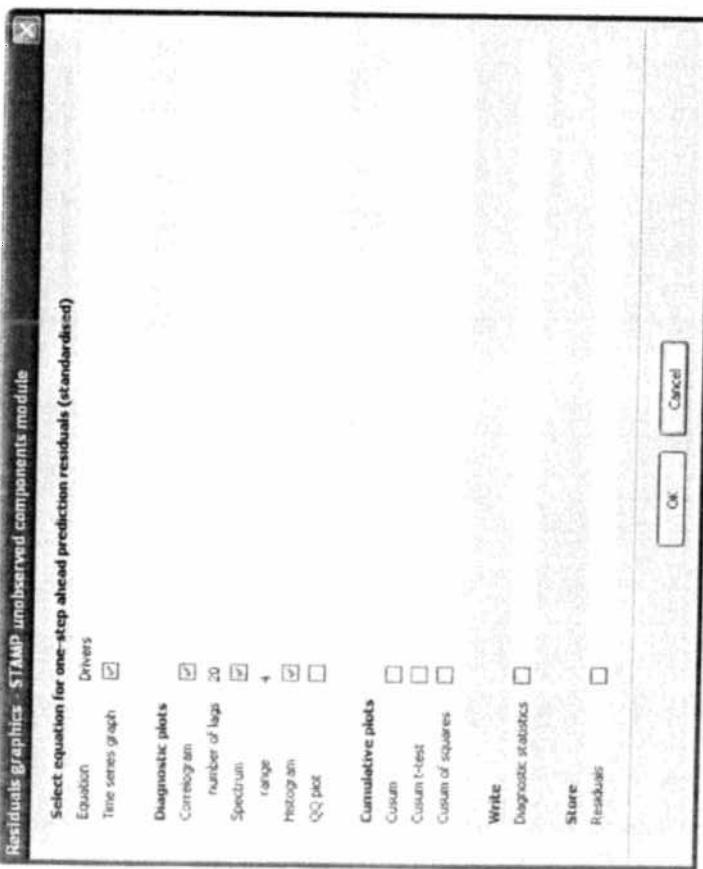
8.3.6 Residuals graphics

The basic way of assessing the suitability of a time series model is by examining its (standardised) residuals. Based on the true values of the parameters, they are independent and normally distributed with zero mean and unit variance. Diagnostic checking of the model may be carried out by plotting the residuals against time and checking the *Correlogram* to see whether the autocorrelations are small. A graph of the distribution of the residuals can also be constructed. An alternative way of examining the residuals is by looking at their frequency domain properties using the *Spectrum*.

The Residuals graphics dialog, with the default options marked, is

To gain insights in how the estimated components are computed, weight functions can be quite useful. Estimation is based on appropriate weighting of observations. The weights are imposed by the estimated model, see Harvey and Koopman (2000). The frequency domain counterparts of the weighting function are the spectral gain function and the phase. All these functions can be requested from the Weight functions dialog for each estimated component and for each time period in the sample.

8.3.5 Weight functions

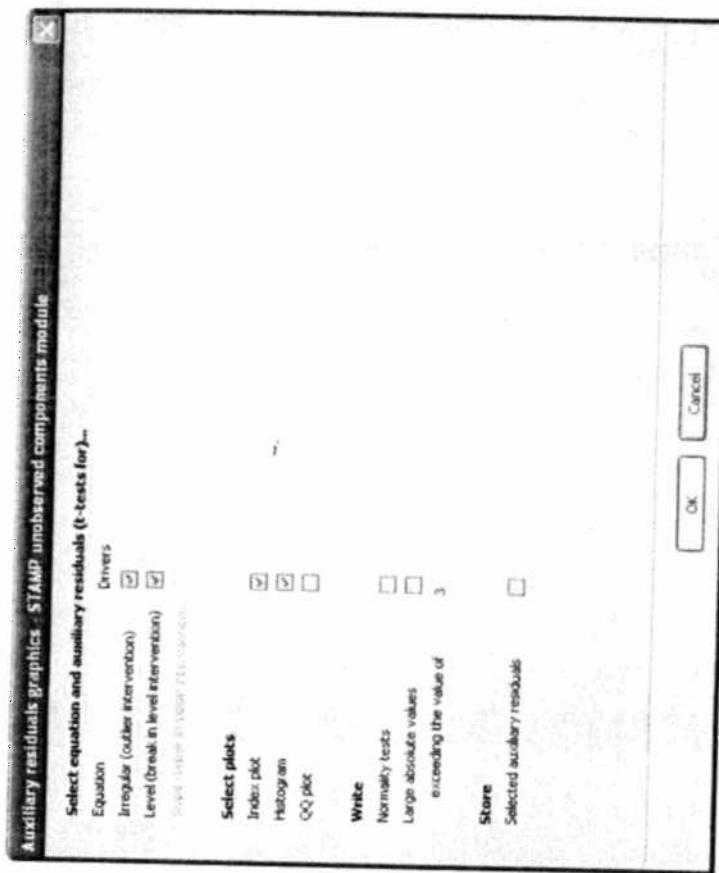


The graph options in the Residuals graphics dialog can be requested in the usual way. The diagnostics checking tools are discussed in § 10.6 which also gives some appropriate definitions and formulae. The option ‘Write diagnostic tests’ sends some serial correlation statistics and normality tests to the OxMetrics Results window. The Store section allows you to save the residuals (which are sometimes called the *innovations*).

8.3.7 Auxiliary residuals graphics

The auxiliary residuals are smoothed estimates of the irregular, level and slope disturbances. Although they are neither serially uncorrelated nor uncorrelated with each other, they play a valuable role in that they go some way towards separating out pieces of information which are mixed up together in the innovation residuals. In particular they are helpful in detecting and distinguishing between outliers and structural change.

In the ‘Drivers’ case there is some indication of a downward shift in the levels of the series after the oil crisis in the mid-1970s. Thus there may be a role for additional level intervention variables. Can you comment on the relation between the level residuals of ‘Drivers’?



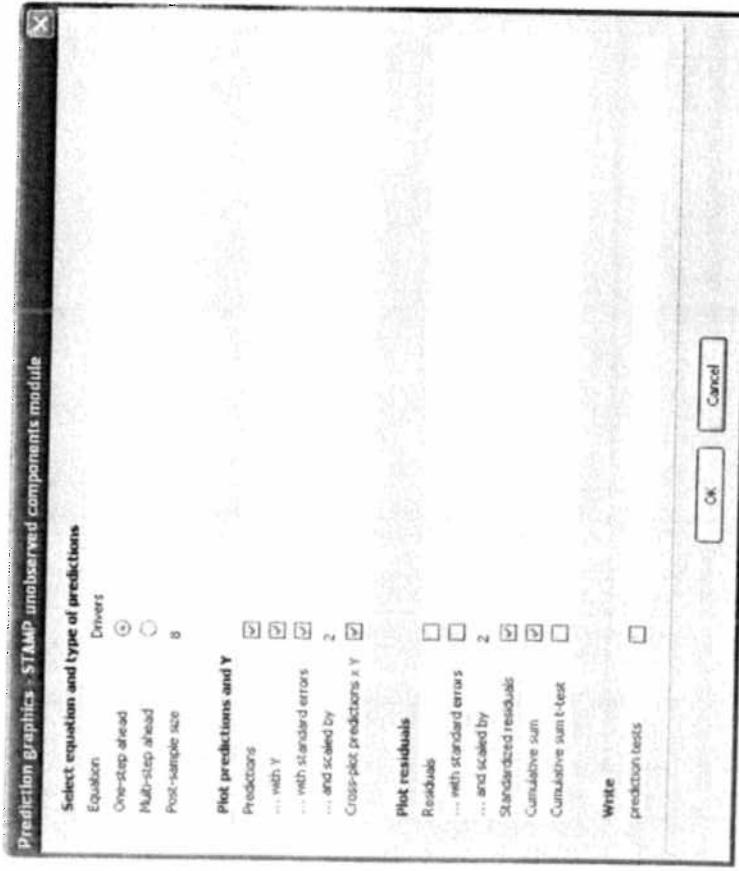
The options of the Auxiliary residuals graphics dialog can be requested in the usual way. The first three check boxes indicate which auxiliary residuals should be subjected to the requested diagnostics. A particularly useful option is ‘Large absolute values .. exceeding the value of’ in the ‘Write’ section which prints information of large values in specific auxiliary residuals. This includes the actual value, which must be larger than 3.0 (default) in absolute values, and the corresponding date. The normality tests, which are adjusted for serial correlation, are printed in the Results window when the checkbox ‘Normality tests’ in the ‘Write’ section is marked.

8.3.8 Prediction testing

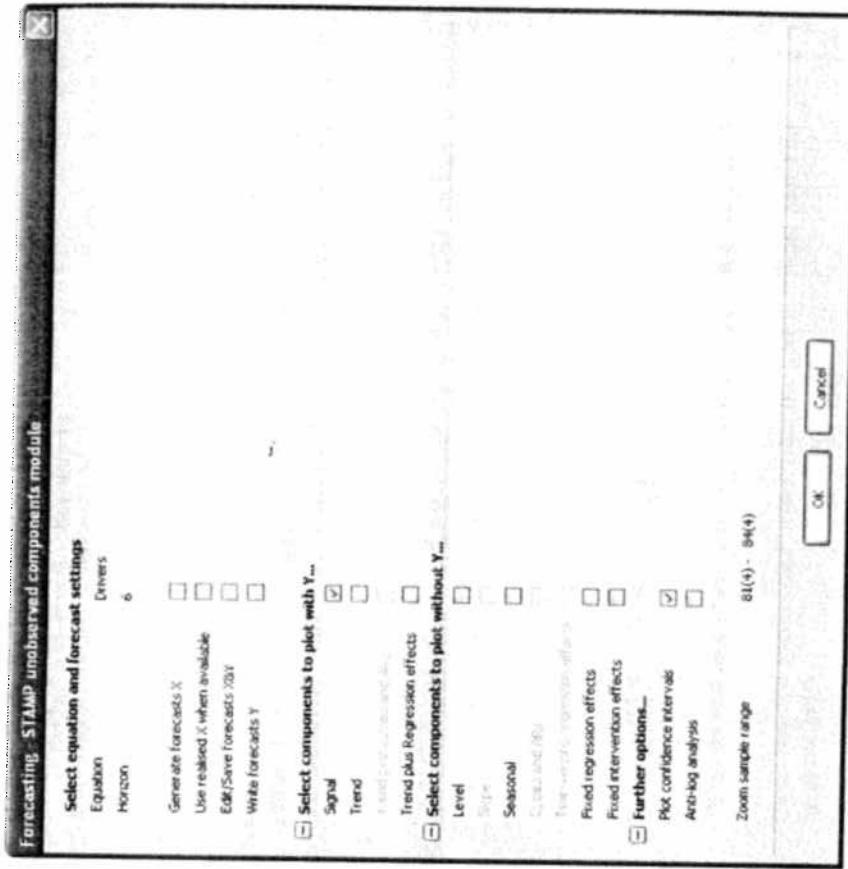
This dialog is for making predictions over periods at the end of the sample for which we have taken out the most recent observations from the estimation sample. Note that the predictions are therefore made within the parameter estimation sample. The formulae of the related statistics are discussed in § 10.8.

One-step ahead predictions can be made and their consistency with the predictions in the sample period can be assessed by looking at graphs of their predictive performance and examining statistics such as the *predictive failure test* which are produced when ‘Prediction tests’ option is marked in the ‘Write’ section. The statistics are reported in the Results window. The default settings of the Prediction graphics dialog is

given below.



When 'Multi-step ahead' is chosen, the predictions are made using the information at the end of the sample period minus the post-sample size but the predictions are not updated in the post-sample with the arrival of new observations. In other words they are extrapolations or multi-step predictions.



8.4.1 Without explanatory variables

Forecasting without explanatory variables is straightforward. As well as the series itself, the individual components may be forecasted by marking the appropriate boxes. Graphs of forecasts together with a prediction interval of 68% (that is, one RMSE on either side) can be produced. The 'Anti-log analysis' option takes the anti-log (adjusted as described in §10.5). The forecasts are printed in the Results window when 'Write forecasts Y' is marked. The length of the forecast horizon can be altered using the first edit field in this dialog.

To use the forecasts later for another model, the forecasts may be edited and saved to a file by marking 'Edit/Save forecasts X'. This gives entrance to a matrix editor in which the forecasts (implied by the model) can be edited and can be saved. Note that when the forecasts are changed in the Matrix Editor, STAMP will keep its forecasts unaltered.

8.4 Forecasting

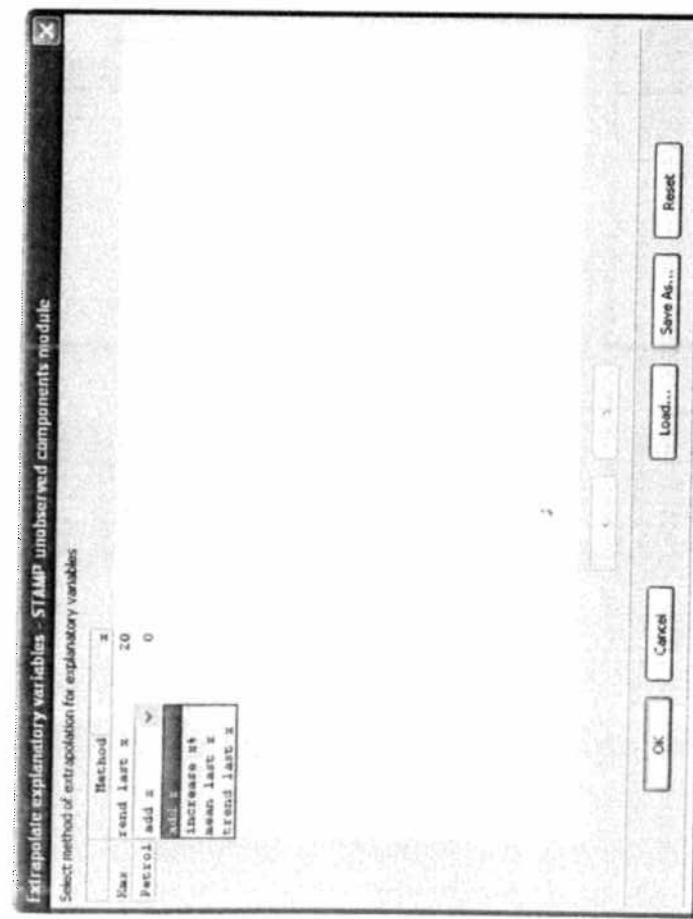
The Forecasting command in the Test menu leads to the Forecasting dialog.

8.4.2 Interventions

If the interventions have been created using the Intervention dialog, appropriate future values are created automatically.

8.4.3 Explanatory variables

When forecasts are to be made conditional on future values of an explanatory variable, the default is to set these values equal to the last value in the sample. However, you may want to alter the default values in four different ways using the option 'Generate forecasts X' in the 'Forecast settings' section:



- All explanatory variables can be extrapolated by a pre-specified method for each explanatory variable and altering the associated edit field appropriately. If the data are in logarithms the increment corresponds to a growth rate. The following methods are available:
 - add x : for every post-sample period, value of x is added, consecutively.
 - increase $x\%$: for every post-sample period, value is increased by percentage x , consecutively.

- mean last x : values for post-sample period is extrapolated by the mean of the last x in-sample observations.
- trend last x : values for post-sample period is extrapolated by the trend of the last x in-sample observations.

- The option 'Edit/Save forecasts X&Y' can be used to edit the extrapolated values for each X variable.

- When the end date of the estimation sample is before the end date of the database, the available observations of the explanatory variables can be used by choosing the option 'Use realised X when available'. The graph will also present the realised Y .

- Finally, the option 'Edit/Save forecasts X&Y' activates the Matrix Editor and allows editing the forecasts of explanatory variables. Such forecasts can be generated by separate models for the explanatory models. Their forecasts can be written in the Results window by the option 'Write forecasts Y'.

Forecasts with 68% confidence interval from period 84(4) forwards:

Forecast	start err	leftbound	rightbound
1	7.14066	0.08518	7.05547
2	7.07239	0.08904	6.98335
3	7.20078	0.09065	7.11013
4	7.44433	0.09052	7.35382
5	7.14066	0.09359	7.04706
6	7.07239	0.09711	6.97527

When the end date of the estimation sample is before the end date of the database, the Results window will also output forecast accuracy measures such as the root mean squared (percentage) error (RMSE) and the mean absolute (percentage) error (MAPE).

Part III

Statistical Treatment

Chapter 9

Statistical Treatment of Models

To understand fully the output of STAMP it is useful to understand the way in which the models are handled statistically. The algorithms used to carry out the computations are based on the state space form. The interested reader may find explanations of the underlying ideas and proofs of the algorithms in Anderson and Moore (1979), Harvey (1989, Chs. 3 and 4), de Jong (1991) and Koopman (1993). This chapter simply sets out the algorithms and gives the exact technical details of implementation. We first formally define the models which were used in Chapters 4–6.

9.1 Model definitions

9.1.1 Univariate time series models

A univariate model may be written as

$$y_t = \mu_t + \gamma_t + \psi_t + \nu_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2), \quad t = 1, \dots, T, \quad (9.1)$$

where μ_t is the trend, γ_t is the seasonal, ψ_t is the cycle, ν_t is a first-order autoregressive component and ϵ_t is the irregular. The model can be extended with two similar cycle components; see below.

The stochastic trend component is specified as

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

where β_t is the slope or gradient of the trend μ_t . The irregular ϵ_t , the level disturbance η_t and the slope disturbance ζ_t are mutually uncorrelated. The slope component β_t can be excluded from the trend specification when this is appropriate. Some special trend specifications are listed in Table 9.1.

The seasonal component has the trigonometric seasonal form and is given by

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}$$

Table 9.1 Some special level and trend specifications.

Level	σ_ϵ	σ_η
constant term	*	0
local level (LL)	*	*
random walk (RW)	0	*
Trend	σ_ϵ	σ_ζ
deterministic	*	0
LL with fixed slope	*	0
RW with fixed drift	0	*
local linear (LLT)	*	*
smooth trend	*	*
second differencing	0	*
Hodrick-Prescott	*	.025 σ_ϵ

Asterix * indicates any positive value.

where each $\gamma_{j,t}$ is generated by

$$\begin{bmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{bmatrix}, \quad j = 1, \dots, [s/2], \\ t = 1, \dots, T,$$

where $\lambda_j = 2\pi j/s$ is the frequency, in radians, and the seasonal disturbances $\omega_{j,t}$ and $\omega_{j,t}^*$ are two mutually uncorrelated NID disturbances with zero mean and common variance σ_ω^2 . For s even, the component at $j = s/2$ collapses to

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \omega_{j,t}.$$

The statistical specification of a cycle, ψ_t , is given by

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho_\psi \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad t = 1, \dots, T,$$

where ρ_ψ , in the range $0 < \rho_\psi \leq 1$, is a *damping factor*; λ_c is the frequency, in radians, in the range $0 \leq \lambda_c \leq \pi$; κ_t and κ_t^* are two mutually uncorrelated NID disturbances with zero mean and common variance σ_κ^2 . Note that the *period* of the cycle is equal to $2\pi/\lambda_c$. There may be two additional cycles of this form incorporated in the model.

A first-order autoregressive, AR(1), process is given by

$$\nu_t = \rho_\nu \nu_{t-1} + \xi_t, \quad \xi_t \sim \text{NID}(0, \sigma_\xi^2),$$

with ρ_ν in the range $0 < \rho_\nu < 1$. The AR(1) component is actually a limiting case of the stochastic cycle when λ_c is equal to 0 or π , though it is specified separately

in STAMP, partly to avoid confounding and partly because it is not a limiting case in multivariate models.

Finally, the disturbances driving each of the components in the model are mutually uncorrelated.

9.1.2 Explanatory variables and interventions

A single equation model may include exogenous explanatory variables, lagged values of the dependent variable and intervention variables, as well as unobserved components such as trend, seasonal and cycle. Thus (9.1) can be extended as

$$y_t = \mu_t + \gamma_t + \phi_t + \psi_t + r_t + \sum_{\tau=1}^p \phi_\tau y_{t-\tau} + \sum_{i=1}^k \sum_{\tau=0}^q \Delta_{it} x_{i,t-\tau} + \sum_{j=1}^h \lambda_j w_{j,t} + \varepsilon_t \quad (9.2)$$

where x_{it} is an exogenous variable, w_{jt} is an intervention (dummy) variable and ϕ_τ, Δ_{it} and λ_j are unknown parameters.

9.1.3 Multivariate time series models

Multivariate models have a similar form to univariate models, except that y_t is now an $N \times 1$ vector of observations which depends on unobserved components which are also vectors. Thus in the special case of a multivariate local level model

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \Sigma_\eta), \end{aligned} \quad (9.3)$$

where Σ_ϵ and Σ_η are both $N \times N$ variance matrices, and η_t and ϵ_t are mutually uncorrelated in all time periods. The other disturbances in the general model (9.1) similarly become vectors which have $N \times N$ variance matrices. Models of this kind are called *seemingly unrelated time series equations* (SUTSE).

The other components are incorporated in a multivariate model in a similar manner by letting the components and disturbances become $N \times 1$ vectors and letting the variances become $N \times N$ variance matrices. In the case of the cycle, the parameters ρ and λ are the same for all series. As regards the disturbances,

$$E(\kappa_t \kappa_t^*) = E(\kappa_t^* \kappa_t^*) = \Sigma_\kappa \text{ and } E(\kappa_t \kappa_t^*) = \mathbf{0}. \quad (9.4)$$

The specification of the cycle model can also be written as

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \left\{ \rho_\psi \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \otimes \mathbf{I}_N \right\} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \quad (9.5)$$

where

$$Var \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} = \mathbf{I}_2 \otimes \Sigma_\kappa, \quad \psi_t \text{ and } \psi_t^* \text{ are } N \times 1 \text{ vectors.}$$

A stationary first-order vector autoregressive may be included in a multivariate model as an alternative to, or even as well as, a cycle. Thus

$$\nu_t = \Phi \nu_{t-1} + \xi_t, \quad \text{and} \quad \text{Var}(\xi_t) = \Sigma_\xi, \quad (9.6)$$

where Φ is a $N \times N$ matrix of coefficients. The condition for ν_t to be stationary is that the roots of the matrix polynomial $I - \Phi L$ should all lie outside the unit circle.

9.1.4 Common factors

In a common factor model, some or all of the components are driven by disturbance vectors with less than N elements. In terms of a SUTSE model, the presence of common factors means that the variance matrices of the relevant disturbances are less than full rank.

Common factors may appear in all components, including the irregular. Common trends arise through common levels, common slopes or a combination of the two.

9.1.4.1 Common levels'

Consider the local level model, (9.3), but suppose that the rank of Σ_η is $K < N$. The model then contains K common levels or common trends and may be written as

$$\begin{aligned} y_t &= \Theta \mu_t^\dagger + \mu_\theta + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t^\dagger &= \mu_{t-1}^\dagger + \eta_t^\dagger, & \eta_t^\dagger &\sim \text{NID}(0, D_\eta), \end{aligned} \quad (9.7)$$

where η_t^\dagger is a $K \times 1$ vector, Θ is a $(N \times K)$ matrix of *standardised factor loadings*, D_η is a diagonal matrix and μ_θ is a $(N \times 1)$ vector in which the first $N - K$ elements are zeros and the last K elements are contained in a vector $\bar{\mu}$. The standardised factor loading matrix contains ones in the ‘diagonal’ positions, i.e. $\theta_{ii} = 1, i = 1, \dots, K$, while $\theta_{ij} = 0$ for $j > i$.

The model may be recast in the original SUTSE form (9.3) by writing $\mu_t = \Theta \mu_t^\dagger + \mu_\theta$ and noting that $\Sigma_\eta = \Theta D_\eta \Theta'$ is a singular matrix of rank K . When there are no common trends so $N = K$, the factor loading matrix is the Cholesky decomposition of Σ_η . As regards $\bar{\mu}$, partition $\Theta = (\Theta'_1, \Theta'_2)'$ where Θ'_1 consists of the first K rows, and partition μ_t similarly. Then

$$\bar{\mu} = -\Theta_2 \Theta_1^{-1} \mu_{1t} + \mu_{2t}, \quad t = 1, \dots, T. \quad (9.8)$$

The vector $\bar{\mu}$ is estimated from the above set of equations using the estimated states at $t = T$. The computations are carried out by back-substitution since the load matrix is lower triangular.

9.1.4.2 Smooth trends with common slopes

Common slopes may be formulated along similar lines. We first consider the case where the variance matrix of the slope disturbances is of rank K_β but the variance matrix of levels is null so that the estimated trends are relatively smooth. The model is therefore

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t &= \mu_{t-1} + \beta_{t-1}^\dagger, \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim \text{NID}(0, \Sigma_\zeta), \end{aligned}$$

but, following (9.7) above, it may be re-formulated as

$$\begin{aligned} y_t &= \Theta \mu_t^\dagger + \mu_{\theta t} + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t^\dagger &= \mu_{t-1}^\dagger + \beta_{t-1}^\dagger, \\ \beta_t^\dagger &= \beta_{t-1}^\dagger + \zeta_t^\dagger, & \zeta_t^\dagger &\sim \text{NID}(0, D_\zeta), \end{aligned}$$

where $\Sigma_\zeta = \Theta D \Theta'$ and the first K_β elements in $\mu_{\theta t}$ are zeros and the remainder are contained in a vector $\bar{\mu} + \bar{\beta}_t$. The $(N - K) \times 1$ vectors, $\bar{\mu}$ and $\bar{\beta}_t$, may both be calculated by expressions like (9.8).

9.1.4.3 Common trends: level and slopes

A general multivariate local linear trend model in which the level variance matrix is of rank K_η while the slope variance matrix is of rank K_β may be written

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t &= \mu_{t-1} + \Theta_\beta \beta_{t-1}^\dagger + \beta_\theta + \eta_t, & \eta_t &\sim \text{NID}(0, \Sigma_\eta), \\ \beta_t^\dagger &= \beta_{t-1}^\dagger + \zeta_t^\dagger, & \zeta_t^\dagger &\sim \text{NID}(0, D_\zeta), \end{aligned}$$

where the $N \times K_\beta$ matrix Θ_β is such that $\Sigma_\zeta = \Theta_\beta D_\zeta \Theta_\beta'$, and $\beta_\theta = (0', \bar{\beta}')'$ with $\bar{\beta}$ a vector of length $N - K_\beta$.

If $K_\beta = 1$, setting Θ_β equal to a vector of ones and letting $\bar{\beta} = 0$ would imply that all series had the same underlying growth rate (when modelling in logs). This might be plausible even if there are no common levels. The implication is that the trends in the forecast function remain parallel, in other words the long-run growth paths are the same. However, unless there are similar restrictions on the levels, the growth paths within the sample will not necessarily have kept together.

The following interpretation can be made if $K_\beta \leq K_\mu$ and the common stochastic slopes only affect the observations *via* the common stochastic trends:

$$\begin{aligned} y_t &= \Theta_\mu \mu_t^\dagger + \mu_{\theta t} + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \Sigma_\epsilon), \\ \mu_t^\dagger &= \mu_{t-1}^\dagger + \Theta_{1\mu}^\dagger \Theta_{\mu\beta} \beta_{t-1}^\dagger + \eta_t^\dagger, & \eta_t^\dagger &\sim \text{NID}(0, D_\eta), \\ \beta_t^\dagger &= \beta_{t-1}^\dagger + \zeta_t^\dagger, & \zeta_t^\dagger &\sim \text{NID}(0, D_\zeta), \end{aligned}$$

where $\Theta_{1\mu}$ is the first K_μ rows of Θ_μ , $\Theta_{\mu\beta}$ is the first K_μ rows of Θ_β , $\mu_{\theta t} = (\mathbf{0}', \beta'_1 t, \bar{\mu}' + \beta'_2 t')'$ with β'_1 is a vector of length $K_\mu - K_\beta$, $\bar{\beta}_2$ is a vector of length $N - K_\mu$ and $\bar{\beta} = (\beta'_1, \bar{\beta}'_2)'$. The fixed components satisfy, for any $1 \leq t \leq T$,

$$\bar{\mu} = -\Theta_{2\mu} \Theta_{1\mu}^{-1} \mu_{1t} + \mu_{2t} - \bar{\beta}_2 t \quad \text{and} \quad \bar{\beta} = -\Theta_{2\beta} \Theta_{1\beta}^{-1} \beta_{1t} + \beta_{2t}$$

where $\Theta_{1\beta}$ is the first K_β rows of Θ_β and $\Theta_{2\beta}$ is the last $N - K_\mu$ rows.

At present STAMP does not offer the option of making the restrictions implied by the above formulation. However, if $K_\mu = K_\beta = 1$, it must be the case that the above restriction implies $\Theta_{2\beta} = \Theta_{2\mu}$. More generally for $K_\mu = K_\beta$ we have

$$\mu_t^\dagger = \mu_{t-1}^\dagger + \Theta_{1\mu}^{-1} \Theta_{1\beta} \beta_{t-1}^\dagger + n_t^\dagger, \quad n_t^\dagger \sim \text{NID}(\mathbf{0}, D_\eta)$$

and setting $\Theta_{1\beta} = \Theta_{1\mu}$ implies the further restriction that the slopes in the common trends are mutually independent. Thus the common trends are fully independent in both level and slope disturbances.

9.1.4.4 Common seasonals

Common factors in seasonality implies a reduction in the number of disturbances driving changes in the seasonal patterns. It does not imply any similarity in seasonal patterns unless the deterministic seasonal components outside the common seasonals are set to zero. Thus suppose, for simplicity, that the series contain only seasonals and irregular. Then

$$y_t = \Theta \gamma_t^\dagger + \gamma_{\theta t} + \epsilon_t,$$

where the last $N - K$ elements of $\gamma_{\theta t}$ contain fixed seasonal effects.

9.1.4.5 Common cycles

A model with common trends and common cycles could be written

$$y_t = \Theta_\mu \mu_t + \mu_\theta + \Theta_\psi \psi_t + \epsilon_t,$$

where Θ_ψ is $N \times K_\psi$. Since the expectation of a cycle is zero, it is not necessary to include a vector of constant terms corresponding to the vector μ_θ which is needed for common levels.

As with other components, the common cycle can be re-formulated as a SUTSE model in which the disturbance variance matrices are $\Sigma_\kappa = \Theta_\psi D_\kappa \Theta_\psi'$. This is only possible because ρ_ψ and λ are the same in all the common cycles.

9.1.5 Explanatory variables

Explanatory variables and interventions may be included in multivariate models. Thus (9.2) generalises to

$$y_t = \mu_t + \gamma_t + \psi_t + r_t + \sum_{\tau=1}^r \Phi_\tau y_{t-\tau} + \sum_{\tau=0}^s \delta_\tau x_{t-\tau} + \Lambda w_t + \epsilon_t, \quad t = 1, \dots, T,$$

where x_t is a $K \times 1$ vector of explanatory variables and w_t is a $K^* \times 1$ vector of interventions. Elements in the parameters matrices, Φ , δ_τ and Λ may be specified to be zero, thereby excluding certain variables from particular equations. In addition, the unobserved components may be subject to common factor restrictions as described in the previous subsection.

9.2 State space form

The statistical treatment of structural time series models is based on the state space form (SSF). We use a particular SSF which is close to the 'state space model' of de Jong (1991). It joins a *measurement* equation

$$y_t = Z_t \alpha_t + X_t b + G_t u_t, \quad t = 1, \dots, T, \quad (9.9)$$

with a *transition equation*, which allows the *states* α_t to evolve according to a first-order vector autoregressive process. We write this as

$$\alpha_{t+1} = T_t \alpha_t + W_t b + H_t u_t. \quad (9.10)$$

The transition equation is initialised at the first time point by setting

$$\alpha_1 = W_0 b + H_0 u_0. \quad (9.11)$$

The model is completed by making assumptions about the error process u_t and the vector of regressors b . We assume they are independently and normally distributed with

$$u_t \sim \text{NID}(\mathbf{0}, \sigma^2 I), \quad b = c + B\delta \quad \text{and} \quad \delta \sim N(\mu, \sigma^2 \Lambda). \quad (9.12)$$

This formulation of the SSF is slightly different from that exploited in books such as Harvey (1989) and West and Harrison (1989). It is used to produce rather more elegant treatments of diffuse initial conditions and fixed effects; see de Jong (1991).

The state system matrices Z_t and T_t are fixed matrices which merely contain known values. The regression system matrices X_t and H_t are always known. The error system matrices G_t and W_t are also sparse but most non-zero values are unknown and are regarded as "hyper" parameters. The disturbance u_t is transformed into

noise for the SSF by the error system matrices \mathbf{G}_t and \mathbf{H}_t . All system matrices can be regarded as selection type matrices.

At first sight, the role for \mathbf{b} in the SSF looks quite complicated. However, the SSF is written in this way to allow a unified treatment of a variety of features for time series models. There are three particularly important ones on which to focus. First, the measurement equation allows \mathbf{b} to influence the observations directly through the regressors \mathbf{X}_t . In general \mathbf{X}_t , like \mathbf{G}_t and \mathbf{H}_t , may be a sparse selection type matrix. Second, the states are directly affected by \mathbf{b} through \mathbf{W}_t . This feature is particularly useful for handling interventions like structural changes (changes in trend) and growth changes (changes in slope). Third, the prior distribution of the initial state vector is (partly) defined via \mathbf{b} .

The random vector \mathbf{b} allows the analysis of models where diffuse distributions are placed on the regressors and on the non-stationary components of the initial state. The vector \mathbf{b} is a linear combination of the known constant vector \mathbf{c} and the random vector δ . The distribution of δ is regarded as *diffuse*, i.e. the variance matrix $\mathbf{\Lambda}$ converges to infinity. Basically, this means that a proper prior distribution for the state cannot be given when components are non-stationary. For a technical discussion on these issues, we refer to de Jong and Chu-Chun-Lin (1994). The connection between diffuse initial conditions and marginal or restricted likelihoods is discussed in Tunnicliffe-Wilson (1989) and Shephard (1993). The unconditional distributions for stationary components, such as cycles, are handled via the second part of (9.11); that is, $\mathbf{H}_0 \mathbf{u}_0$.

The generality of the SSF can be simplified for structural time series models in three different ways.

- (1) The constant vector \mathbf{c} of the specification for \mathbf{b} can be set to a zero vector such that $\mathbf{b} = \mathbf{B}\delta$. The matrix \mathbf{B} is a square selection matrix of zeros and ones. The vector δ can be partitioned such that the regression effects and the initial effects are separated and that δ enter the equations of the SSF directly. This is achieved by

$$\delta = \begin{pmatrix} \delta_x \\ \delta_i \end{pmatrix}, \quad \text{and} \quad \mathbf{B} = (\mathbf{B}_x, \mathbf{B}_i),$$

where the $(k \times 1)$ vector δ_x contains the regression effects, the $(d \times 1)$ vector δ_i contains the initial effects and the dimensions of the matrices \mathbf{B} , \mathbf{B}_x and \mathbf{B}_i are $(k+d) \times (k+d)$, $(k+d) \times k$ and $(k+d) \times d$, respectively. This brings the specification for \mathbf{b} to

$$\mathbf{b} = \mathbf{B}_x \delta_x + \mathbf{B}_i \delta_i$$

and

$$\begin{aligned} \mathbf{X}_t \mathbf{b} &= \mathbf{X}_t^* \delta_x & \text{where } \mathbf{X}_t^* &= \mathbf{X}_t \mathbf{B}_x & \text{and } \mathbf{X}_t \mathbf{B}_i &= \mathbf{0}, \\ \mathbf{W}_t \mathbf{b} &= \mathbf{W}_t^* \delta_x & \text{where } \mathbf{W}_t^* &= \mathbf{W}_t \mathbf{B}_x & \text{and } \mathbf{W}_t \mathbf{B}_i &= \mathbf{0}, \\ \mathbf{W}_0 \mathbf{b} &= \mathbf{W}_0^* \delta_i & \text{where } \mathbf{W}_0^* &= \mathbf{W}_0 \mathbf{B}_i & \text{and } \mathbf{W}_0 \mathbf{B}_x &= \mathbf{0}. \end{aligned}$$

- (2) For all models in STAMP, \mathbf{G}_t and \mathbf{H}_t will be orthogonal; that is, $\mathbf{G}_t' \mathbf{H}_t = \mathbf{0}$, so the noise terms in the two equations, $\mathbf{G}_t \mathbf{u}_t$ and $\mathbf{H}_t \mathbf{u}_t$, are independently distributed. This feature of a model simplifies the statistical analysis of the SSF.
- (3) All models in STAMP can be handled by a time-invariant SSF. This implies that the t subscripts of the state and error system matrices can be dropped; i.e. $\mathbf{Z}_t = \mathbf{Z}$, $\mathbf{T}_t = \mathbf{T}$, $\mathbf{G}_t = \mathbf{G}$ and $\mathbf{H}_t = \mathbf{H}$. A time-invariant SSF also simplifies the statistical analysis of the SSF.

9.2.1 Structural time series models in SSF

Structural time series models fit nicely in the state space form. However, some care should be taken with the initial state vector specification.

9.2.1.1 Univariate models in SSF

Some examples are given to get a flavour of it.

- *Seasonal model.* The first example of a structural time series model in SSF is a model with smooth trend and trigonometric stochastic seasonal with $s = 4$ (quarterly observations). The SSF representation is

$$y_t = (1 \ 0 \ 1 \ 0 \ 0) \boldsymbol{\alpha}_t + (\sigma_\epsilon \ 0 \ 0) \mathbf{u}_t$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \mu_t \\ \beta_t \\ \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\zeta & 0 & 0 & 0 \\ 0 & 0 & \sigma_\omega & 0 & 0 \\ 0 & 0 & 0 & \sigma_\omega & 0 \\ 0 & 0 & 0 & 0 & \sigma_\omega \end{pmatrix} \mathbf{u}_t$$

- *Cycle model.* Another example is a univariate model which consists of a level and cycle component. The SSF representation of this model is

$$y_t = (1 \ 1 \ 0) \boldsymbol{\alpha}_t + (\sigma_\epsilon \ 0 \ 0) \mathbf{u}_t, \text{ where } \boldsymbol{\alpha}_t = (\boldsymbol{\mu}_t, \boldsymbol{\phi}_t, \psi_t)',$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 0 & \sigma_\eta & 0 & 0 \\ 0 & 0 & \sigma_\kappa & 0 \\ 0 & 0 & 0 & \sigma_\omega \end{pmatrix} \mathbf{u}_t$$

- *Explanatory variables.* A univariate structural time series model may include explanatory variables. Consider a local linear trend model with two explanatory variables and a level intervention at some time point $t = j$. The SSF representa-

tion of this model is given by

$$\begin{aligned} y_t &= (\begin{pmatrix} 1 & 0 \end{pmatrix}) \alpha_t + (\begin{pmatrix} x_{1,t} & x_{2,t} & 0 \end{pmatrix}) \delta_x + (\begin{pmatrix} \sigma_\epsilon & 0 & 0 \end{pmatrix}) u_t \\ \alpha_t &= \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 0 & 0 & w_t \\ 0 & 0 & 0 \end{pmatrix} \delta_x + \begin{pmatrix} 0 & \sigma_\eta & 0 \\ 0 & 0 & \sigma_\zeta \end{pmatrix} u_t \end{aligned}$$

with vector $u_t = (\epsilon_t, \eta_t, \zeta_t)'$. The initial state vector is given by $\alpha_1 = \delta_1$ such that $B = I$ and $H_0 = 0$. The series $x_{1,t}$ and $x_{2,t}$ are explanatory variables and w_t is a series of zeros but at some time point $t = j$ its value is unity.

9.2.1.2 Multivariate model in SSF

Any multivariate model in STAMP with common factors can be written in the SUTSE form with lower rank variance matrices; see §9.1.4. The SUTSE form retains the sparse structure of the system matrices Z and T which is important for computational issues related to the estimation of these models. Therefore, this general formulation is adopted to specify a multivariate model in the SSF. Since the SUTSE form is a minor generalisation to the univariate model, only one example will be given. The multivariate model is a smooth trend plus cycle model which can be represented by the SSF as

$$y_t = (\begin{pmatrix} I & 0 & I & 0 \end{pmatrix}) \alpha_t + (\begin{pmatrix} \Gamma_\epsilon & 0 & 0 & 0 \end{pmatrix}) u_t,$$

$$\alpha_t = \begin{pmatrix} I & I & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho \cos \lambda_c I & \rho \sin \lambda_c I \\ 0 & 0 & -\rho \sin \lambda_c I & \rho \cos \lambda_c I \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \Gamma_\zeta & 0 & 0 \\ 0 & 0 & \Gamma_\kappa & 0 \\ 0 & 0 & 0 & \Gamma_\kappa \end{pmatrix} u_t$$

where state vector $\alpha_t = (\mu'_t, \beta'_t, \psi'_t, \psi^*_t)'$, disturbance vector $u_t = (\epsilon_t, \zeta_t, \kappa_t, \kappa_t^*)'$ and each variance matrix is decomposed as $\Sigma = \Gamma \Gamma'$ with the appropriate subscript. The unknown values in transition matrix T rely on the *damping factor* ρ and the *frequency* λ_c . The initial state is given by

$$\alpha_1 = \begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \delta_1 + \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \Gamma_\kappa & 0 \\ 0 & \Gamma_\kappa \end{pmatrix} u_0$$

with vector $u_0 = (\kappa_0, \kappa_0^*)'$ and δ_1 contains the initial effects for μ_t and β_t .

9.2.1.3 Deterministic trend and seasonal components

The deterministic seasonal component is the special case of the stochastic seasonal model with $\sigma_\omega = 0$. Alternatively, the fixed seasonal can be incorporated within X_t

as it is usually done in regression models. The latter one is adopted in STAMP. In a similar way (partly) deterministic trends can be incorporated in X_t or W_t . However, for practical reasons such as the implementation of trend interventions within W_t , the trend components μ_t and β_t are always part of the state vector. The inclusion of a fixed trend in the state vector has some consequences for calculation and interpretation of residuals and the prediction error variance.

9.3 Kalman filter

The Kalman filter (KF) plays the same role for time series models in SSF as least squares computations for a regression model. The KF is primarily a set of vector and matrix recursions. The importance of the KF is based on

- (1) computation of one-step ahead predictions of observation and state vectors, and the corresponding mean square errors;
- (2) diagnostic checking by means of one-step ahead prediction errors;
- (3) computation of the likelihood function via the one-step ahead prediction error decomposition;
- (4) smoothing which uses the output of the KF.

The KF has a variety of forms, but the one exploited in STAMP computes

$$\begin{aligned} a_{t|t-1} &= E(\alpha_t | Y_{t-1}, \delta = 0), \\ \sigma^2 P_{t|t-1} &= E\{(\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})' | Y_{t-1}, \delta = 0\}, \end{aligned}$$

which are referred to as the mean and mean square error (MSE) of the state, respectively, given the past information and setting δ to zero. Generally, we need to perform estimation with δ not zero, so these terms will be corrected by the use of an augmented Kalman filter. This will be discussed in the next subsection.

The recursive equations of the Kalman filter are given by

$$\begin{aligned} v_t &= y_t - X_t b - Z a_{t|t-1}, & F_t &= Z P_{t|t-1} Z' + G G', \\ q_t &= q_{t-1} + v_t' F_{t|t-1}^{-1} v_t, & K_t &= T P_{t|t-1} Z' F_{t|t-1}^{-1}, \\ a_{t+1|t} &= T a_{t|t-1} + W_t b + K_t v_t, \\ P_{t+1|t} &= T P_{t|t-1} T' - K_t F_t K_t' + H H', \end{aligned} \quad (9.13)$$

starting with $a_{1|0} = W_0 b$, $P_{1|0} = H_0 H_0'$ and $q_0 = 0$. Here K_t is called the Kalman gain, while v_t and $\sigma^2 F_t$ are the one-step ahead prediction error (or innovation) and its mean square error, respectively. The scaled innovations $F_t^{-\frac{1}{2}} v_t$ (or generalised least squares residuals) are approximately NID with zero mean and scale identity matrix as its variance matrix in a correctly specified model. Note that the estimate of σ^2 is given by $\hat{\sigma}^2 = q_T / NT$. The proof of the KF is simple and it relies on standard results of linear estimation. Although the KF is presented in terms of time-invariant system and

error matrices (\mathbf{Z} , \mathbf{T} , \mathbf{G} and \mathbf{H}), the time indices can be added to these matrices without any problem.

In the special case of a univariate model, $N = 1$, the innovation vector and the corresponding mean square error matrix reduce to the scalars v_t and $\sigma^2 f_t$, respectively. Also, the Kalman gain matrix reduces to the vector \mathbf{k}_t . This notation is only used for univariate models.

The KF requires a positive (for univariate models) or a non-singular positive definite matrix (for multivariate models) \mathbf{F}_t which may not occur at the start of the KF when, for example, the column rank of \mathbf{G}_1 is less than N and $\mathbf{P}_{1|0} = \mathbf{H}_0 \mathbf{H}'_0 = \mathbf{0}$. An example is when no irregular is present in a structural time series model. In these circumstances, the KF is warmed up by a number of prediction updates; that is, $\mathbf{a}_{j+1|0} = \mathbf{T}\mathbf{a}_{j|0}$ and $\mathbf{P}_{j+1|0} = \mathbf{T}\mathbf{P}_{j|0}\mathbf{T}' + \mathbf{H}\mathbf{H}'$. The initialisation process continues until $\mathbf{F}_j = \mathbf{Z}\mathbf{P}_{j|0}\mathbf{Z}' + \mathbf{G}\mathbf{G}'$ is positive definite. The KF is then re-initialised by $\mathbf{a}_{1|0} = \mathbf{a}_{1|0}$, $\mathbf{P}_{1|0} = \mathbf{P}_{1|0}$ and $q_0 = 0$ where j is the number of required prediction updates.

When \mathbf{F}_t becomes non-positive definite during the KF updating, it is generally assumed that a numerical problem has occurred. Available remedies like a *square-root* version of the KF are not considered. However, the implementation of the KF in STAMP ensures for each time-point t that the MSE matrices $\mathbf{P}_{t+1|t}$ and \mathbf{F}_t are symmetric, which is crucial in avoiding numerical problems. Also, avoiding unnecessary computations, such as adding zeros and multiplying zeros or unity values, helps in this respect.

In case of a time-invariant SSF, the KF may reach a steady state solution at some time s ; that is, when $\mathbf{P}_{s+1|s} = \mathbf{P}_{s|s-1}$ for $p < s \leq T$ where p is the dimension of the state vector. This also implies that \mathbf{F}_j and \mathbf{K}_j are time-invariant for $j = s, \dots, T$. The KF reduces in these circumstances to the equations for $\mathbf{a}_{t+1|t}$, \mathbf{v}_t and q_t which are computationally very cheap. Finally, the steady state can be interpreted as the solution of the *Riccati equation*; see Anderson and Moore (1979).

9.3.1 The augmented Kalman filter

The KF evaluates the one-step ahead prediction of the state vector conditional on $\delta = \mathbf{0}$. When non-stationary components and/or fixed regression effects are present in the model, the unknown random vector δ is treated as diffuse; see [14.2]. In these circumstances an augmented Kalman filter (AKF) is applied:

$$\begin{aligned} \mathbf{V}_t &= -\mathbf{Z}\mathbf{A}_{t|t-1} - \mathbf{X}_t\mathbf{B} \\ \mathbf{A}_{t+1|t} &= \mathbf{T}\mathbf{A}_{t|t-1} + \mathbf{W}_t\mathbf{B} + \mathbf{K}_t\mathbf{V}_t \\ (\mathbf{s}_t, \mathbf{S}_t) &= (\mathbf{s}_{t-1}, \mathbf{S}_{t-1}) + \mathbf{V}_t'\mathbf{F}_t^{-1}(\mathbf{v}_t, \mathbf{V}_t) \end{aligned} \quad (9.14)$$

for $t = 1, \dots, T$ and with $\mathbf{A}_{1|0} = \mathbf{W}_0\mathbf{B}$. The AKF must be regarded as a supplement to the KF equations; the expressions for \mathbf{v}_t , \mathbf{F}_t , \mathbf{q}_t , \mathbf{K}_t , $\mathbf{a}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$ remain as they are. The number of columns for \mathbf{V}_t and $\mathbf{A}_{t+1|t}$ is $k + d$, the same number of

columns as in matrix \mathbf{B} of the SSF. The one-step ahead prediction of the state vector and the corresponding MSE matrix are given by

$$\begin{aligned} E(\alpha_t | \mathbf{Y}_{t-1}) &= \hat{\mathbf{a}}_{t|t-1} &= \mathbf{a}_{t|t-1} + \mathbf{A}_{t|t-1}\mathbf{S}_{t-1}^{-1}\mathbf{s}_{t-1}, \\ MSE(\alpha_t | \mathbf{Y}_{t-1}) &= \sigma^2 \hat{\mathbf{P}}_{t|t-1} &= \sigma^2 (\mathbf{P}_{t|t-1} + \mathbf{A}_{t|t-1}\mathbf{S}_{t-1}^{-1}\mathbf{A}_{t|t-1}'), \end{aligned} \quad (9.15)$$

respectively. The one-step ahead prediction error and the corresponding MSE matrix (or variance matrix) are given by

$$\begin{aligned} \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1}) &= \hat{\mathbf{v}}_t &= \mathbf{v}_t + \mathbf{V}_t \mathbf{S}_{t-1}^{-1} \mathbf{s}_{t-1} \\ MSE(\mathbf{y}_t | \mathbf{Y}_{t-1}) &= \sigma^2 \hat{\mathbf{F}}_t &= \sigma^2 (\mathbf{F}_t + \mathbf{V}_t \mathbf{S}_{t-1}^{-1} \mathbf{V}_t'), \end{aligned} \quad (9.16)$$

respectively. The matrix inversions for \mathbf{S}_t , $t = 1, \dots, T$, can be evaluated recursively using similar methods as recursive regressions; see Appendix 9.7 of this chapter. The estimate of the scalar variance σ^2 is

$$\hat{\sigma}^2 = \frac{1}{NT - d - k} \hat{q}_T \quad \text{where} \quad \hat{q}_T = q_T - \mathbf{s}_T' \mathbf{S}_T^{-1} \mathbf{s}_T. \quad (9.17)$$

The full sample generalised least squares estimate of δ and its MSE matrix are given by

$$\hat{\delta} = \mathbf{S}_T^{-1} \mathbf{s}_T \quad \text{and} \quad \text{MSE}(\hat{\delta}) = \sigma^2 \mathbf{S}_T^{-1} \quad (9.18)$$

respectively. The MSE matrix is used to obtain standard errors and t-statistics for $\hat{\delta}$. In a similar way as vector δ , $\hat{\delta}$ is partitioned as

$$\hat{\delta} = \begin{pmatrix} \hat{\delta}_x \\ \hat{\delta}_i \end{pmatrix}.$$

For a proof and a full exposition of the augmented Kalman filter, see de Jong (1991). When the SSF does not contain values for \mathbf{X}_t and \mathbf{W}_t , the KF can evaluate $\hat{\mathbf{a}}_{t+1|t}$ from $t = d + 1$ onwards. This implies that the AKF is only used for the first d updates. Then, the AKF is collapsed to the KF, mainly using equation (9.15) for $t = d$. On the other hand, when \mathbf{X}_t and \mathbf{W}_t are present in the SSF, a partial collapse with respect to δ is possible. The exact details of the full and the partial collapse are given in Appendix 9.7 of this chapter. Formal proofs on collapsing are given by de Jong and Chu-Chun-Lin (1994).

9.3.2 The likelihood function

The likelihood function can be obtained from the KF by using the innovations and their mean square errors. This is known as the *prediction error decomposition*. For example, the likelihood function, conditional on $\delta = 0$, is defined as

$$\log L(y|\theta^*, \delta = 0) = -\frac{NT}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{\mathbf{q}_T}{2\sigma^2} \quad (9.19)$$

apart from some constant. To limit the number of logarithmic operations in computing the likelihood, it is important to check whether a steady-state has been reached such that $\log |\mathbf{F}_t|$ is constant for all t .

The unknown parameters in the system matrices of the SSF are stacked in vector θ . For a given vector $\theta = \theta^*$, the KF calculates the likelihood and, therefore, it provides means to obtain maximum likelihood estimates of the parameters. These matters are discussed in §9.6.

9.3.2.1 Univariate models: the concentrated likelihood

The diffuse likelihood function (DL) is defined as the likelihood when δ is supposed to be diffuse; see §9.2. For a given vector $\theta = \theta^*$, the DL is given by

$$\log L(y|\theta^*) = -\frac{T^*}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{1}{2} \log |\mathbf{S}_T| - \frac{1}{2\sigma^2} \hat{\mathbf{q}}_T \quad (9.20)$$

with $T^* = T - k - d$. When a (partial) collapse of the AKF has taken place at some time $t = m$, the DL is calculated by

$$\log L(y|\theta^*) = -\frac{T^*}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{1}{2} \log |\mathbf{S}_*| - \frac{1}{2\sigma^2} \hat{\mathbf{q}}_T. \quad (9.21)$$

Appendix 9.7 of this chapter defines \mathbf{S}_* and it gives more details of likelihood calculation when a full or a partial collapse has taken place. The scalar variance σ^2 can be concentrated out of the likelihood DL. This concentrated diffuse likelihood function (CDL) is given by

$$\log L_c(y|\theta^*) = -\frac{T^*}{2} \log \hat{\sigma}^2 - \frac{1}{2} \sum_{t=1}^T \log f_t - \frac{1}{2} \log |\mathbf{S}_*.| \quad (9.22)$$

To calculate the CDL, one variance corresponding to a particular unobserved component is set to a unity value such that this variance is set equal to σ^2 of (9.12). The maximum likelihood estimate of this ‘concentrated out’ variance is $\hat{\sigma}^2$. The other variances of the components in the model are now measured as a ratio to the concentrated variance. For example, in a local level model, where σ^2_ε is concentrated out, we have $\sigma^2_\eta = \sigma^2 q_\eta$. These variance ratios are referred to as *q-ratios*. Parameter estimation is carried out with respect to the *q-ratios*; see §9.6. The CDL is used for parameter estimation of univariate models in STAMP. For a full discussion on diffuse likelihoods, we refer to de Jong (1988).

9.3.2.2 Multivariate models

The diffuse likelihood is used for parameter estimation of multivariate models. Thus, no parameter is concentrated out from the likelihood. The AKF is used for calculation but

it can be (partially) collapsed at some time $t = m$. The multivariate diffuse likelihood is defined as

$$\log L(y|\theta^*) = -\frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2} \log |\mathbf{S}_*| - \frac{\hat{\mathbf{q}}_T}{2}, \quad (9.23)$$

where \mathbf{S}_* is defined in Appendix 9.7 of this chapter.

9.3.3 Prediction error variance

The prediction error variance (*PEV*) of a univariate structural time series model is the variance of the one-step ahead prediction errors in the steady state. As discussed at the start of this section, the steady state may require more updates of the KF than the number of available observations. The formal definition of the estimated *PEV* is

$$PEV = \hat{\sigma}^2 f,$$

where f is the steady state value of f_t ; that is, $f = \lim_{t \rightarrow \infty} f_t$. For a multivariate model, the prediction error variance matrix can be evaluated. When a steady state is not obtained, it may be useful to report a ‘finite’ *PEV*; that is,

$$PEV_n = \hat{\sigma}^2 \hat{f}^2,$$

where n is some large finite number. The ‘finite’ *PEV* at the end of the sample; that is, at $n = T$, might also be of interest.

9.3.4 The final state and regression estimates

The final state vector $\hat{\mathbf{a}}_{T|T}$ contains, together with $\hat{\mathbf{d}}_x$, all the information from the estimated SSF required to forecast future observations. Therefore, the final state is of special interest. The estimate of the state vector and its MSE matrix at the end of the sample are given by

$$\begin{aligned} \hat{\mathbf{a}}_{T|T} &= \hat{\mathbf{a}}_{T|T-1} + \hat{\mathbf{P}}_{T|T-1} \mathbf{Z}'_T \hat{\mathbf{F}}_T^{-1} \hat{\mathbf{v}}_T, \\ \hat{\sigma}^2 \mathbf{P}_{T|T} &= \hat{\sigma}^2 (\mathbf{P}_{T|T-1} - \hat{\mathbf{P}}_{T|T-1} \mathbf{Z}'_T \hat{\mathbf{F}}_T^{-1} \mathbf{Z}_T \mathbf{P}_{T|T-1}), \end{aligned}$$

respectively. The diagonal part of the MSE matrix is used to form standard errors and *t*-statistics of the final state $\hat{\mathbf{a}}_{T|T}$. Under certain circumstances, some elements of this diagonal part may be equal to zero and calculation of the *t*-statistic is not possible then. Also, for stationary parts of the state vector, it is not useful to calculate the *t*-statistic since stationary components are not persistent throughout the series.

The regression estimates are obtained from

$$\hat{\mathbf{d}}_x = \bar{\mathbf{S}}_T^{-1} \bar{s}_T \quad \text{and} \quad MSE(\hat{\mathbf{d}}_x) = \hat{\sigma}^2 \bar{\mathbf{S}}_T^{-1},$$

where $\bar{\mathbf{S}}_T$ and \bar{s}_T are defined in Appendix 9.7 of this chapter. The diagonal part of the MSE matrix is used to form standard errors and *t*-statistics.

9.3.5 Filtered components

The filtered components are based on the one-step ahead predictions of the state vector; see equation (9.15). When $k = 0$, the AKF is applied until $t = d$. Then, the AKF collapses to the KF which evaluates the filtered states directly. However, when $k > 0$, a partial collapse takes place and filtering is based on (9.15) with matrix $\widehat{\mathbf{S}}_T^{-1}$ calculated recursively; see Appendix 9.7 of this chapter. Note that the filtered states are only defined for $t = d + k + 1, \dots, T$.

9.3.6 Residuals

The residuals in STAMP for univariate models are defined as the standardised innovations

$$v_t = \widehat{v}_t / \widehat{\sigma} \widehat{f}_t^{\frac{1}{2}}, \quad t = d + 1, \dots, T. \quad (9.24)$$

where \widehat{v}_t and $\widehat{\sigma}^2 \widehat{f}_t$ are defined as in (9.16). When $k = 0$, the residuals are obtained from the KF directly after the collapse of the AKF at $t = d$. In a multivariate model the residuals of the j -th equation are given by

$$v_{j,t} = \widehat{v}_{j,t} / \widehat{\sigma} \widehat{F}_{jj,t}^{\frac{1}{2}}$$

for $t = d + 1, \dots, T$, where $a_{i,t}$ is the i -th element of vector \mathbf{a}_t and $A_{ij,t}$ is the (i,j) element of matrix \mathbf{A}_t .

When a SSF contains fixed effects; that is, when $k > 0$, two sets of residuals are distinguished inside STAMP. These residuals are defined in the next two subsections. Note that the two sets are the same when $k = 0$.

9.3.6.1 Generalised least squares residuals

The generalised least squares (GLS) residuals are standardised innovations computed using (9.24), $t = d + 1, \dots, T$, but where \widehat{v}_t and \widehat{f}_t are obtained from the AKF (9.15) applied to a SSF with b_x being replaced by $\mathbf{B}_x \widehat{\delta}_x$. The calculation of these residuals requires two steps. Firstly, the estimates $\widehat{\delta}_x$ and $\widehat{\sigma}$ are obtained from the AKF with $d+k$ columns for $\mathbf{A}_{t+1|t}$. Then, a second AKF is applied, with d columns for $\mathbf{A}_{t+1|t}$, from which \widehat{v}_t and \widehat{f}_t are obtained for $t = d + 1, \dots, T$. Of course, the first AKF may be partially collapsed after $t = d$ and the second AKF may be fully collapsed to a KF after $t = d$. The GLS residuals are not independently distributed with constant variance in small samples. Tests for normality, heteroskedasticity and serial correlation are applied to the GLS residuals.

9.3.6.2 Generalised recursive residuals

The generalised recursive residuals are the standardised residuals (9.24), but now obtained from an AKF, with $d+k$ columns for $\mathbf{A}_{t+1|t}$, accompanied with a set of recur-

sive regressions for the inversion calculations of \mathbf{S}_t ; see Appendix 9.7 of this chapter. A partial collapse of this AKF may take place at $t = d$. These residuals are denoted by w_t , for $t = d + k + 1, \dots, T$, and they are commonly used in predictive tests and diagnostics.

9.4 Disturbance smoother

Smoothing refers to the estimation of the state vector α_t and the disturbance vector \mathbf{u}_t using information in the whole sample rather than just past data. Smoothing is an important feature because it is the basis for

- (1) signal extraction, detrending and seasonal adjustment;
- (2) diagnostic checking for detecting and distinguishing between outliers and structural changes using auxiliary residuals;
- (3) EM algorithm for initial estimation of parameters;
- (4) calculation of the score, defined as the first derivative of the likelihood with respect to the vector of parameters.

We consider the disturbance smoother (DS) of Koopman (1993) which directly estimates the disturbance vector of the SSF, \mathbf{u}_t , and the MSE matrix; that is,

$$E(\mathbf{u}_t | \mathbf{Y}_T, \delta = \mathbf{0}) = \tilde{\mathbf{u}}_t \quad \text{and} \quad \text{MSE}(\tilde{\mathbf{u}}_t) = \sigma^2 \mathbf{C}_t,$$

respectively, with $t = 1, \dots, T$. Note that in the context of structural time series models, the interest in \mathbf{u}_t is limited to elements of \mathbf{G}_t and \mathbf{H}_t and their corresponding MSE matrices.

The DS is based on the same backwards recursions as found by Bryson and Ho (1969), de Jong (1989) and Kohn and Ansley (1989) which are given by

$$\begin{aligned} \mathbf{e}_t &= \mathbf{F}_t^{-1} \mathbf{v}_t - \mathbf{K}_t' \mathbf{r}_t, & \mathbf{r}_{t-1} &= \mathbf{Z}' \mathbf{e}_t + \mathbf{T}' \mathbf{r}_t, \\ \mathbf{D}_t &= \mathbf{F}_t^{-1} + \mathbf{K}_t' \mathbf{N}_t \mathbf{K}_t, & \mathbf{N}_{t-1} &= \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t' \mathbf{N}_t \mathbf{L}_t, \end{aligned}$$

where $\mathbf{L}_t = \mathbf{T} - \mathbf{K}_t \mathbf{Z}$ and $t = T, \dots, 1$. These backwards recursions are initialised with $\mathbf{r}_T = \mathbf{0}$ and $\mathbf{N}_T = \mathbf{0}$. The storage requirements for the KF are limited to the quantities \mathbf{v}_t , \mathbf{F}_t and \mathbf{K}_t for $t = 1, \dots, T$. The smooth estimates for the disturbances of the measurement equation and the MSE matrices are

$$\tilde{\mathbf{G}}_t = \mathbf{G} \mathbf{G}' \mathbf{e}_t \quad \text{and} \quad \sigma^2 \mathbf{G} \mathbf{C}_t \mathbf{G}' = \sigma^2 (\mathbf{G} \mathbf{G}' - \mathbf{G} \mathbf{G}' \mathbf{D}_t \mathbf{G}'), \quad (9.25)$$

respectively. The smooth estimates for the disturbances of the transition equation and the mean square errors are

$$\tilde{\mathbf{H}}_t = \mathbf{H} \mathbf{H}' \mathbf{r}_t \quad \text{and} \quad \sigma^2 \mathbf{H} \mathbf{C}_t \mathbf{H}' = \sigma^2 (\mathbf{H} \mathbf{H}' - \mathbf{H} \mathbf{H}' \mathbf{N}_t \mathbf{H} \mathbf{H}'), \quad (9.26)$$

respectively. The proof of the disturbance smoother is given by Koopman (1993). The DS is the most computationally efficient smoother available yet.

9.4.1 The augmented disturbance smoother

In a similar way as the KF, the DS must be augmented to allow for initial effects with diffuse distributions and fixed components. The supplement recursions for the DS are

$$\mathbf{E}_t = \mathbf{F}_t^{-1} \mathbf{V}_t - \mathbf{K}'_t \mathbf{R}_t \quad \text{and} \quad \mathbf{R}_{t-1} = \mathbf{Z}' \mathbf{E}_t + \mathbf{T}' \mathbf{R}_t,$$

and they will be referred to as the augmented disturbance smoother (ADS). The number of columns for \mathbf{E}_t and \mathbf{R}_t is equal to $d+k$. When the AKF is not collapsed at all, the full sample estimates of the disturbance vectors $E(\mathbf{G}\mathbf{u}_t | \mathbf{Y}_T)$, and the corresponding MSE matrices $MSE(\mathbf{G}\mathbf{u}_t | \mathbf{Y}_T)$, are given by (9.25) with \mathbf{e}_t and \mathbf{D}_t replaced by

$$\hat{\mathbf{e}}_t = \mathbf{e}_t + \mathbf{E}_t \tilde{\boldsymbol{\delta}} \quad \text{and} \quad \hat{\mathbf{D}}_t = \mathbf{D}_t - \mathbf{E}_t \mathbf{S}^{-1} \mathbf{E}'_t, \quad (9.27)$$

respectively, for $t = 1, \dots, T$. In a similar way, $E(\mathbf{H}\mathbf{u}_t | \mathbf{Y}_T)$ and $MSE(\mathbf{H}\mathbf{u}_t | \mathbf{Y}_T)$ are given by (9.26) with \mathbf{r}_t and \mathbf{N}_t replaced by

$$\hat{\mathbf{r}}_t = \mathbf{r}_t + \mathbf{R}_t \tilde{\boldsymbol{\delta}} \quad \text{and} \quad \hat{\mathbf{N}}_t = \mathbf{N}_t - \mathbf{R}_t \mathbf{S}^{-1} \mathbf{R}'_t, \quad (9.28)$$

respectively, for the transition disturbances with Note that $\tilde{\boldsymbol{\delta}} = \mathbf{S}_T^{-1} \mathbf{s}_T$ and $\mathbf{S} = \mathbf{S}_T$. For a proof of the ADS, the reader is referred to Koopman (1993). When the AKF is fully collapsed to the KF at $t = m$, the DS can be used for $t = T, \dots, m+1$. The ADS should be applied for $t = m, \dots, 1$ but when a collapse has taken place, the matrices \mathbf{S}_T and \mathbf{R}_m are not available. Therefore, smoothing is not straightforward when a collapse has occurred during the AKF. However, a suggestion of Chu-Chun-Lin and de Jong (1993) can be used to recover these missing matrices.

9.4.2 The EM algorithm and exact score

The EM algorithm is a recursive method to obtain maximum likelihood estimates for unknown elements in the state and error system matrices of the SSF; see Shumway and Stoffer (1982) and Watson and Engle (1983). This method requires a huge amount of computations and is not easy to implement. Koopman (1993) proposes a simple and computationally efficient EM algorithm for unknown values inside the variance matrices $\Omega_G = \mathbf{G}\mathbf{G}'$ and $\Omega_H = \mathbf{H}\mathbf{H}'$ where the error system matrices are supposed to be time-invariant. Let θ be the stack of these unknown values. Given a set of values for θ , say θ^* , the EM step calculates new variance matrices via

$$\begin{aligned} \Omega_G(\theta^\dagger) &= \Omega_G(\theta^*) + \Omega_G(\theta^*) \Lambda_e \Omega_G(\theta^*), \\ \Omega_H(\theta^\dagger) &= \Omega_H(\theta^*) + \Omega_H(\theta^*) \Lambda_r \Omega_H(\theta^*), \end{aligned}$$

where

$$\Lambda_e = \sum_{t=1}^T \hat{\mathbf{e}}_t \hat{\mathbf{e}}'_t - \hat{\mathbf{D}}_t \quad \text{and} \quad \Lambda_r = \sum_{t=1}^T \hat{\mathbf{r}}_t \hat{\mathbf{r}}'_t - \hat{\mathbf{N}}_t.$$

The quantities $\hat{\mathbf{e}}_t$ and $\hat{\mathbf{D}}_t$ are defined in (9.27) and the quantities $\hat{\mathbf{r}}_t$ and $\hat{\mathbf{N}}_t$ are defined in (9.28). The log-likelihood value of the SSF with these new variance matrices is always larger compared to the log-likelihood value of the SSF with the old variance matrices. This EM algorithm only requires one pass through the KF and one pass back through the disturbance smoother. Note that the EM maximises the diffuse log-likelihood (9.21). For a proof and a discussion of this EM algorithm, the reader is referred to Koopman (1993).

The score vector is defined as the first derivative of the log-likelihood with respect to the vector of parameters θ . The vector of first derivatives is usually evaluated numerically. Koopman and Shephard (1992) present a method to calculate the exact score for any parameter in the state and error system matrices. However, the computational gain is only achieved for exact score calculations with respect to parameters in the error system matrices $\mathbf{G}_t = \mathbf{G}$ and $\mathbf{H}_t = \mathbf{H}$. Again, let θ be the stack of unknown values in the variance matrices of the SSF and let vector θ^* contain specific values for θ , then the exact score is given by

$$\mathbf{q}(\theta^*) = \frac{\partial \log L(\mathbf{y} | \theta^*)}{\partial \theta} = \frac{1}{2} \text{tr} \left\{ \Lambda_e \frac{\partial \Omega_G(\theta^*)}{\partial \theta} \right\} + \frac{1}{2} \text{tr} \left\{ \Lambda_r \frac{\partial \Omega_H(\theta^*)}{\partial \theta} \right\}, \quad (9.29)$$

and the definitions of the EM algorithm apply here as well.

9.4.3 Smoothed components

The smoothed state vector is defined as the estimated state vector using the full data-set; that is,

$$E(\alpha_t | \mathbf{Y}_T) = \tilde{\alpha}_t. \quad (9.30)$$

The smoothed state vector forms the basis for detrending and seasonal adjustment procedures for the SSF. Graphical inspection of a sub-set or a linear combination of the smoothed state vector is an important feature in STAMP. The actual values for (9.30) are obtained by

$$\tilde{\alpha}_{t+1} = \mathbf{T} \tilde{\alpha}_t + \mathbf{W}_t^* \tilde{\delta}_x + \tilde{\mathbf{H}} \tilde{\mathbf{u}}_t, \quad t = 1, \dots, T, \quad (9.31)$$

with $\tilde{\alpha}_1 = \hat{\mathbf{a}}_{1|0} + \hat{\mathbf{P}}_{1|0}(\mathbf{r}_0 + \mathbf{R}_0 \tilde{\boldsymbol{\delta}})$. Hence, the smoothed state vector is calculated using the forward recursion (9.31) after the DS is applied and the vector $\mathbf{H} \tilde{\mathbf{u}}_t = \mathbf{H} \mathbf{H}' \tilde{\mathbf{r}}_t$ is stored for $t = 1, \dots, T$. The MSE matrix of the smoothed state vector is not considered; see de Jong (1989) and Kohn and Ansley (1989). It is shown by Koopman (1993) that this method of state smoothing is computationally very fast. Also note that no extra storage space is required since the storage space of the Kalman gain matrix \mathbf{K}_t can be used to store the vector $\mathbf{H} \tilde{\mathbf{u}}_t$ for $t = 1, \dots, T$.

9.4.4 Auxiliary residuals

The standardised smoothed estimates of the disturbances are referred to as *auxiliary residuals*. Graphs of these residuals, in conjunction with normality tests, are a valuable tool for detecting data irregularities such as outliers, level changes and slope changes. Note that the auxiliary residuals are serially correlated even when the model is correctly specified. Therefore, the normality test should be adjusted for the implied serial correlation. The technical details and some interesting applications can be found in Harvey and Koopman (1992).

The auxiliary residuals are obtained directly from (9.25) and (9.26). The residuals are standardised using the diagonal elements of the appropriate MSE matrix. The variance σ^2 must be replaced by its estimate $\hat{\sigma}^2$.

9.5 Forecasting

An important feature of STAMP is its extensive forecast facility. The model-based extrapolations can be calculated for the actual series but also for any element of the state vector.

9.5.1 Forecasts of series and components

The forecast of the state vector l multi-steps ahead, for $l = 1, \dots, L$, and its MSE matrix are given by

$$\hat{a}_{T+l|T} = E(\alpha_{T+l}|\mathbf{Y}_T), \quad \sigma^2 \hat{P}_{T+l|T} = \text{MSE}(\alpha_{T+l}|\mathbf{Y}_T),$$

respectively. The forecasts and MSE matrices are generated recursively by

$$\begin{aligned} \hat{a}_{T+l|T} &= \hat{T}\hat{a}_{T+1|T} + \mathbf{W}_T^* \tilde{\delta}_{\mathbf{x}}, \\ \hat{\sigma}^2 \hat{P}_{T+l|T} &= \hat{\sigma}^2 (\hat{T}\hat{P}_{T+1|T} \hat{T}' + \mathbf{H}\mathbf{H}'), \end{aligned} \quad (9.32)$$

respectively. The initialisation is directly obtained from the AKF at time T . The forecast for \mathbf{y}_{T+1} and the corresponding MSE matrices are given by

$$\begin{aligned} \hat{y}_{T+l|T} &= \hat{Z}\hat{a}_{T+1|T} + \mathbf{X}_T^* \tilde{\delta}_{\mathbf{x}}, \\ \hat{\sigma}^2 \hat{F}_{T+l|T} &= \hat{\sigma}^2 (\hat{Z}\hat{P}_{T+1|T} \hat{Z}' + \mathbf{G}\mathbf{G}'), \end{aligned} \quad (9.33)$$

respectively. Graphical inspection of the forecasted values can be valuable for checking model validity.

9.5.2 Extrapolative residuals

The extrapolative residuals can be calculated when future values of \mathbf{y}_t become available. They are defined as

$$\hat{v}_{t+l} = \mathbf{y}_{T+l} - \hat{y}_{T+l|T}, \quad (9.34)$$

where $\hat{y}_{T+l|T}$ is defined in (9.33). The variance matrices of the extrapolative residuals are equal to the MSE matrices of $\hat{y}_{T+l|T}$; that is, $\hat{\sigma}^2 \hat{F}_{T+l|T}$ as in (9.33). The extrapolative residuals can be standardised by using the corresponding diagonal elements of the variance matrix.

9.6 Parameter estimation

The model definitions are given in §9.1. It is shown in §9.2 how the models of STAMP can be put in SSF. The filtering and smoothing algorithms of §9.3 and §9.4, respectively, are associated with the SSF and can be applied conditional on known state and error system matrices. The unknown values inside these matrices are treated as parameters which need to be estimated. The main task of STAMP is to estimate these parameters using maximum likelihood estimation (MLE) methods. This subsection sets out the strategy for solving this problem.

9.6.1 The parameters of STAMP

The state and error system matrices of the SSF are time-invariant and contain many zero and unity values. Therefore, they can be regarded as selection matrices. For a univariate model, the unknown values in the error system matrices are the standard deviations of the unobserved components. The time-invariant transition matrix \mathbf{T} of the SSF contains the unknown first-order autoregressive parameter of stationary components such as cycle and (optional) slope. In the case of a cycle, the state system matrix \mathbf{T} also contains cosine and sine terms which rely on the unknown frequency λ . Table 14.2 overviews the set of possible parameters for a univariate model.

The variances are always positive, the cycle *damping factors* must be in the range $0 < \rho < 1$, and the *frequencies* can only take values in the range $0 < \lambda < \pi$. Since the parameters are to be estimated via a numerical method, some special transformations are used to enforce these restrictions on the parameters; see Table 14.2. The set of transformed parameters, relevant for a specific model, are stacked in the vector θ .

As observed from Table 14.2, the variance of the cycle, σ_ψ^2 , is taken as the parameter instead of the variance of its disturbance σ_κ^2 . In this case, $\sigma_\kappa^2 = \sigma_\psi^2(1 - \rho_\psi^2)^{\frac{1}{2}}$ and $\sigma_\kappa^2 \rightarrow 0$ as $\rho_\psi \rightarrow 1$. This allows the estimation of deterministic, but stationary, cycles which is the extreme case of $\rho_\psi = 1$.

9.6.1.1 Variance matrices in a multivariate model

The same parameters exist for a multivariate model as for a univariate model, except that the variances are replaced by variance matrices; see §9.1. A special case is the vector autoregressive component which is discussed in §9.5.2. Since all variance matrices are treated the same, we consider a $N \times N$ matrix Σ , i.e. a non-negative symmetric

Table 9.2 Parameters and transformations.

Variance	Parameter	Transformation
Irregular	σ_ϵ^2	$\exp(2\theta_\epsilon)$
Level	σ_η^2	$\exp(2\theta_\eta)$
Slope	σ_ζ^2	$\exp(2\theta_\zeta)$
Seasonal	σ_ω^2	$\exp(2\theta_\omega)$
Cycle	σ_ψ^2	$\exp(2\theta_\psi)$
AR(1)	σ_ξ^2	$\exp(2\theta_\xi)$
Autoregressive	Parameter	Transformation
Slope	ρ_β	$ \theta_\beta (1 + \theta_\beta^2)^{-\frac{1}{2}}$
Cycle	ρ_ψ	$ \theta_\psi (1 + \theta_\psi^2)^{-\frac{1}{2}}$
AR(1)	ρ_ν	$\theta_\nu(1 + \theta_\nu^2)^{-\frac{1}{2}}$
Frequency	Parameter	Transformation
Cycle	λ	$2\pi/2 + \exp(\theta_\lambda)$

matrix but not necessarily full rank. The *Cholesky decomposition* of Σ enforces these restrictions. The variance matrix is decomposed as $\Sigma = \Theta D \Theta'$ where $N \times p$ matrix Θ is a lower-triangular matrix with unity values on the leading diagonal, $1 \leq p \leq N$, and $p \times p$ matrix D is a non-negative diagonal matrix. The lower-triangular elements of Θ ; that is, θ_{ij} with $i = 2, \dots, N$ and $j = 1, \dots, \min(i-1, p)$, may take any value, including zero values. The diagonal elements of D are transformed as

$$d_i = \exp(2\theta_{ii})$$

where d_i is the i -th diagonal element of D and $i = 1, \dots, p$. Indeed, matrix Θ is the same *factor loadings* matrix as in the model specifications for common components; see §9.1.4. Since matrix Θ has p columns, this would suggest that variance matrix Σ has rank p . In most cases this will be true. However, as any diagonal element of D can be zero, the rank of Σ can be less than p . Note that when $d_i = 0$, the i -th column of Θ does not affect variance matrix Σ and can be set to zero.

The variance matrices are placed into the SSF via the error system matrices. For example, the variance matrix of the irregular enters the SSF via $\mathbf{G} = (\Theta D^{\frac{1}{2}}, \mathbf{0}, \dots, \mathbf{0})'$ where Θ and D correspond to the decomposition of the irregular variance matrix Σ_ϵ and $\mathbf{0}$ denotes a zero matrix.

9.6.1.2 Vector autoregressive components

The multivariate model may include a vector autoregression of order 1. The parameters inside the matrix Φ appear in the SSF transition matrix \mathbf{T} . All roots of the matrix polynomial $\mathbf{I} - \Phi \mathbf{L}$ must lie outside the unit circle. Within a numerical optimisation procedure, the constraints on Φ needed to keep the process stationary are imposed by means of an algorithm given by Ansley and Kohn (1986). In the special case of a VAR(1) model, the transformations are:

- The parameters of the VAR component are put into the $N \times N$ unrestricted matrix \mathbf{A} and the $N \times N$ positive definite symmetric matrix Γ .
 - Cholesky transformations are applied on the matrices $\mathbf{I} + \mathbf{A}\mathbf{A}' = \mathbf{B}\mathbf{B}'$ and $\Gamma = \mathbf{L}\mathbf{L}'$.
 - The VAR matrices are defined by $\Phi = \mathbf{L}\mathbf{A}'\mathbf{B}'^{-1}\mathbf{L}^{-1}$ and $\Sigma_v = \Gamma - \Phi\Gamma\Phi'$.

This method ensure a stationary VAR component. However, imposing restrictions on the VAR matrices may be difficult.

9.6.2 BFGS Estimation procedure*

The optimisation method is based on the BFGS *quasi-Newton* method and it aims to maximise the likelihood criterion $\ell(\theta)$. At every call to this updating scheme, it delivers a new estimate for θ , denoted by θ_{i+1} , given a current estimate, θ_i , which brings the object function $\ell(\theta)$ closer to its maximum. The updating scheme can be written as

$$\theta_{i+1} = \theta_i + s_i \delta_i \quad (9.35)$$

where s_i is a step size scalar and δ_i is the search direction vector. Appendix 9.8 of this chapter gives some basic principles on numerical optimisation and in particular the BFGS updating scheme. Some practical details on initialisation, updating, linear search and termination, are given below.

- At the start of the estimation process an initial parameter vector θ_0 and an initial search direction δ_0 must be given. The procedure of setting the initial parameter vector is different for univariate models and for multivariate models; see §9.6.3 and §9.6.4, respectively. The initial direction vector is calculated as

$$\delta_0 = \mathbf{H}^d(\theta_0)\mathbf{q}(\theta_0) \quad (9.36)$$

where $\mathbf{q}(\theta_0)$ is the score vector and $\mathbf{H}^d(\theta_0)$ is the Hessian matrix which only contains values for the diagonal entries. These values are evaluated numerically and are based on parameter vector θ_0 . The off-diagonal entries are zero. The initial score vector is evaluated numerically. For multivariate models, the score vector is evaluated analytically for any next update; see §9.4.2.

- The maximisation process is terminated when all three convergence criteria hold depending on a small value ε or when the number of BFGS updates is equal to the integer M or when irregularities occur. In the latter case, the estimation procedure is stopped with an error code attached to it. The three convergence criteria are

$$\begin{aligned} 1. \text{ Likelihood: } crit_1 &= |\ell(\boldsymbol{\theta}_i) - \ell(\boldsymbol{\theta}_{i+1})| / |\ell(\boldsymbol{\theta}_i)| &< \varepsilon \\ 2. \text{ Gradient: } crit_2 &= \frac{1}{m} \sum_{j=1}^m |q_j(\boldsymbol{\theta}_i)| &< 10\varepsilon \\ 3. \text{ Parameter: } crit_3 &= \frac{1}{m} \sum_{j=1}^m |\theta_{i+1,j} - \theta_{i,j}| / |\theta_{i,j}| &< 100\varepsilon \end{aligned}$$

where $\theta_{i,j}$ denotes the j -th element of the parameter vector $\boldsymbol{\theta}_i$ and $q_j(\boldsymbol{\theta}_i)$ denotes the j -th element of the score vector $\mathbf{q}(\boldsymbol{\theta}_i)$. When all three convergence criteria hold, the estimation procedure terminates and returns one of the following messages

Very strong	$crit_1 < \varepsilon$	$crit_2 < \varepsilon$	$crit_3 < \varepsilon$
Strong	$crit_1 < \varepsilon$	$crit_2 < \varepsilon$	$crit_3 < 10^N \varepsilon$
Weak	$crit_1 < \varepsilon$	$crit_2 < 10^N \varepsilon$	$crit_3 < 10^N \varepsilon$
Very weak	$crit_1 < 10^N \varepsilon$	$crit_2 < 10^N \varepsilon$	$crit_3 < 10^N \varepsilon$

where N is the number of elements in the vector \mathbf{y}_t of the state space form; see §9.2. For univariate models, N is equal to unity. The default values are $\varepsilon = 10^{-7}$ and $M = 100$.

- The direction vector $\boldsymbol{\delta}_i$ is obtained via the BFGS scheme; see Appendix 9.8 of this chapter. To ensure stability in the optimisation routine: the BFGS updating of (9.45) does not take place when $\mathbf{d}'\mathbf{g}$ in (9.45) is smaller than ε .
- When the likelihood function, the score or the diagonal elements of the Hessian matrix cannot be evaluated for some reason, the estimation procedure terminates with an error message.

More specific details of implementation for univariate models and multivariate models are given below.

9.6.3 Estimation of univariate models*

For univariate models, STAMP maximises the concentrated diffuse log-likelihood $\log L_c(y|\boldsymbol{\theta})$ which is treated as an unconstrained non-linear function of $\boldsymbol{\theta}$, denoted by $\ell(\boldsymbol{\theta})$. The likelihood kernel $\ell(\boldsymbol{\theta})$ is appropriately scaled by a function of T , the number of observations.

9.6.3.1 Initial values

The elements of the vector of transformed parameters $\boldsymbol{\theta}$ are given in Table 14.2. The elements related to standard deviations, θ^σ , are set equal to -0.5 , except for θ_η (level) which initial value is -1 whereas θ_ζ (slope) and θ_ω (seasonal) are started off with -1.5 and -2 , respectively.

One element of $\boldsymbol{\theta}^\sigma$ needs to be concentrated out, which effectively means it must be restricted to zero (that is, a unity standard deviation). The choice of the concentrated parameter must be chosen appropriately. At the start of the optimisation routine, the element corresponding to the irregular standard deviation is concentrated out. When no irregular is present in the model we choose from level, slope, cycles, seasonal and autoregressive component, consecutively.

The initial values for ρ_ψ and λ , corresponding to a particular cycle, are provided by the user and STAMP transforms them automatically to $\boldsymbol{\theta}_0$. The initial values for θ_β and θ_ν are set to 2 (which corresponds to ρ values of approximately 0.9).

Before the quasi-Newton optimisation is started, a simple method is constructed to get $\boldsymbol{\theta}^\sigma$ to the neighborhood of the optimal solution in a stable but quickly way. The initialisation method is similar to the quasi-Newton updating scheme (9.35). The j -th element of the direction vector $\boldsymbol{\delta}_i$ is given by

$$\delta_{i,j} = H_{j,j}(\boldsymbol{\theta}_i)\mathbf{q}_j(\boldsymbol{\theta}_i)$$

where $q_j(\boldsymbol{\theta}_i)$ denotes the j -th element of the score vector $\mathbf{q}(\boldsymbol{\theta}_i)$ and $H_{j,j}(\boldsymbol{\theta}_i)$ denotes the (j,j) element of the Hessian matrix $\mathbf{H}(\boldsymbol{\theta}_i)$. The score and the diagonal elements of the Hessian matrix are numerically evaluated at $\boldsymbol{\theta}_i$. The initialisation updating steps are repeated five times but it is stopped earlier if the stopping criterion

$$\sum_j \delta_{i,j} \nabla_j(\boldsymbol{\theta}_i) < 0.01$$

holds.

During this simple estimation procedure, a choice is made of which parameter should be concentrated out of the log-likelihood. At every step i , the parameter related to the largest value in vector $\boldsymbol{\theta}_i$ is concentrated out when its corresponding value in $\boldsymbol{\delta}_i$ is larger than 4. This rule reflects the desire to concentrate out the largest standard deviation of the likelihood. When the choice of concentrated parameter is changed, the other standard deviations should be adjusted, so they have the correct ratios with respect to the new concentrated parameter. This might create instabilities in the search process. Therefore, the difference between the old and the new adjusted value is enforced to lie between -1 and 2 . This strategy is a result of ‘fine tuning’ after it has been applied to a wide range of different models and data-sets.

9.6.3.2 Strategy for setting parameters fixed

During the estimation procedure, certain parameters might take values close to their boundaries. When a parameter moves closer and closer to its boundary, it may slow down the optimisation procedure as a whole. In these circumstances, it is advised to fix parameters at their boundary values but to allow estimation to be continued for the remaining parameters. The strategy for setting parameters fixed is as follows. In the case of standard deviations, when the transformed parameter value is smaller than -5 and its gradient value is smaller than 0.0001 (in absolute terms), it is set to a constant implying a zero standard deviation.

Similarly, the *damping factor* ρ of a cycle is set to unity, when the element corresponding to θ_ρ in vector θ_i is larger than 25 . If the value related to θ_λ (corresponding to the period of a cycle) is larger than 7 or less than -7 , the program reduces the cycle component to an autoregressive component. In both cases, it is assumed that the corresponding gradient value is smaller than 0.0001 (in absolute terms). The transformed parameters related to ρ_β (slope) and ρ_v (autoregressive) don't apply to this strategy.

9.6.3.3 Strategy for determining the concentrated parameter

The initial estimation procedure has determined a concentrated standard deviation from θ^σ . However, during the main estimation process, it might be more appropriate to concentrate out another parameter. When the largest transformed parameter value (corresponding to θ^σ) exceeds the value 1.5 , this particular parameter will be the new concentrated parameter. Then, the other parameters corresponding to θ^σ obtain values as ratios to the new concentrated parameter, the gradient vector $\mathbf{q}(\theta_i)$ is re-calculated and the Hessian matrix is replaced by the diagonal matrix $\mathbf{H}^d(\theta_i)$.

9.6.4 Estimation of multivariate models*

The estimation procedure for multivariate models follows mainly the same route as for univariate models but there are some differences. The likelihood kernel $\ell(\theta)$ is the diffuse log-likelihood $\log L(y|\theta)$ but properly scaled by a function of T . So there is no need to find an appropriate concentrated parameter. The EM algorithm plays a prominent role in the procedure to obtain initial variance matrices for the disturbances; see §9.4.2. The strategy of setting elements to zero in the Cholesky decompositions of the variance matrices is set out below. The parameters for the VAR component cannot be set fixed during the estimation procedure.

9.6.4.1 Initial values

The initial settings for the multivariate model are discussed in terms of the variance matrix, the Cholesky decomposition and the VAR coefficient matrix, rather than directly in

terms of the parameter vector θ . The initialisation procedure automatically transforms the elements of these matrices into the vector θ_0 .

The initial parameter settings of the variance matrices is similar to the univariate settings of the *q-ratios*, except that now a specific *q-ratio* applies to all N diagonal entries of the \mathbf{D} matrix. The factor loading matrix Θ is set to the identity matrix. At this stage, the rank conditions of the variance matrices are ignored. The VAR matrix is set equal to a diagonal matrix with all entries equal to 0.9 . The extra initial parameters for the cycle component are set manually by the user.

The Kalman filter is applied to the multivariate model and the sample variance matrix of the innovations is calculated. The variance matrices of the disturbances are set equal to this matrix but properly scaled by their appropriate *q-ratios*.

Then, the EM algorithm is applied; see §9.4.2. The number of EM steps is equal to 5 which has proved to be appropriate in most cases. The VAR matrix and the other stationary parameters (ρ and λ) remain fixed during the EM. After the EM steps the parameters are transformed as indicated in §9.6.1. For example, the Cholesky decomposition is used to get the diagonal matrix \mathbf{D} and the factor loading matrix Θ from the variance matrix. At this stage, the rank conditions are taken into account. Finally, the parameters are stacked into θ_0 .

9.6.4.2 Strategy for fixing parameters

The need to restrict parameters to their boundary values is important to speed up the process of estimation. This is even more relevant for multivariate models since the estimation process is likely to take much more time than univariate models because filtering and smoothing involve matrices with larger dimensions and more parameters need to be estimated.

The diagonal entries of \mathbf{D} are treated in the same way as the standard deviations in a univariate model. The restrictions for the *damping factor* and the *frequency* parameter are also done in a similar fashion as for univariate models. No restrictions can be made with respect to the parameters related to the VAR component. The *factor loading* matrix elements enter the parameter vector without transformation.

When the j -th diagonal element of \mathbf{D} is restricted to zero, the rank of the variance matrix Σ is reduced by one. All elements of the corresponding j -th column of the factor loading matrix can also be restricted to zero since these elements don't have any effect on the variance matrix Σ . Any specific element of the factor loading matrix Θ is set to zero when its value is smaller than 0.0001 and its corresponding gradient is also smaller than 0.0001 (both in absolute terms).

9.7 Appendix 1: Diffuse distributions*

This appendix discusses technical details about the way we treat initial and regression effects for the SSF. Most components in STAMP are initialised using diffuse priors and so we must treat these priors carefully in our calculations to ensure stable and exact results. This is carried out using the idea of augmented filtering and smoothing. To ensure computational efficiency the Kalman filter (KF) needs to be collapsed at some point.

9.7.1 Collapse of the augmented KF

For a SSF with $k = 0$; that is, $\mathbf{X}_t = \mathbf{0}$ and $\mathbf{W}_t = \mathbf{0}$, $t = 1, \dots, T$, is treated by the AKF for $t = 1, \dots, m$ and the AKF is collapsed to the KF using (9.15) at $t = m$. Then, the KF evaluates $\hat{\mathbf{a}}_{t+1|t}$ and $\hat{\mathbf{P}}_{t+1|t}$ directly, $t = 1, \dots, T$. The collapse also contains the operation

$$\hat{q}_m = q_m - \mathbf{s}'_m \mathbf{S}_m^{-1} \mathbf{s}_m.$$

Indeed, the KF evaluates the residual vector $\hat{\mathbf{v}}_t$ and its (unscaled) variance matrix $\hat{\mathbf{F}}_t$ directly as well. The value for m is restricted by $T \geq m \geq d$ and $|\mathbf{S}_m| > 0$ but usually it is set to $m = d$.

For a SSF with $k > 0$, a full collapse cannot be made since \mathbf{X}_t and \mathbf{W}_t are time-varying. However, a partial collapse can take place with respect to δ_i . This means that the AKF is still applied but with a reduced number of columns. Therefore, we partition the AKF matrices

$$\begin{aligned} \mathbf{A}_{t|t-1} &= \begin{pmatrix} \mathbf{A}_{\mathbf{x}, t|t-1}, & \mathbf{A}_{\mathbf{i}, t|t-1} \end{pmatrix}, & \mathbf{V}_t &= \begin{pmatrix} \mathbf{V}_{\mathbf{x}, t}, & \mathbf{V}_{\mathbf{i}, t} \\ \mathbf{S}_{\mathbf{x}, t} & \mathbf{S}_{\mathbf{i}, t} \end{pmatrix} \\ \mathbf{s}_t &= \begin{pmatrix} \mathbf{s}_{\mathbf{x}, t} \\ \mathbf{s}_{\mathbf{i}, t} \end{pmatrix}, & \mathbf{S}_t &= \begin{pmatrix} \mathbf{V}_{\mathbf{x}, t} \\ \mathbf{S}_{\mathbf{x}, t} \end{pmatrix} \end{aligned} \quad (9.37)$$

When $\mathbf{S}_{\mathbf{i}, t}$ is invertible at $t = m$, the AKF can be partially collapsed to

$$\begin{aligned} \bar{\mathbf{A}}_{t+1|t} &= \frac{\mathbf{A}_{\mathbf{x}, t|t-1}}{\bar{\mathbf{P}}_{t+1|t}}, & + \frac{\mathbf{A}_{\mathbf{i}, t+1|t} \mathbf{S}_{\mathbf{i}, t}^{-1} \mathbf{S}_{\mathbf{i}, t}}{\mathbf{P}_{t+1|t}} \\ \bar{\mathbf{s}}_t &= \frac{\mathbf{s}_{\mathbf{x}, t}}{\bar{\mathbf{P}}_{t+1|t}}, & + \frac{\mathbf{A}_{\mathbf{i}, t+1|t} \mathbf{S}_{\mathbf{i}, t}^{-1} \mathbf{A}_{\mathbf{i}, t+1|t}}{\mathbf{P}_{t+1|t}} \end{aligned}$$

and

$$\begin{aligned} \bar{q}_t &= q_t - \mathbf{s}'_t \mathbf{S}_{\mathbf{i}, t}^{-1} \mathbf{s}_t, \\ \bar{\mathbf{s}}_t &= \mathbf{s}_{\mathbf{x}, t} - \mathbf{S}'_{\mathbf{i}, t} \mathbf{S}_{\mathbf{i}, t}^{-1} \mathbf{s}_t, \\ \bar{\mathbf{S}}_t &= \mathbf{S}_{\mathbf{x}, t} - \mathbf{S}'_{\mathbf{i}, t} \mathbf{S}_{\mathbf{i}, t}^{-1} \mathbf{S}_{\mathbf{x}, t}. \end{aligned}$$

Note that matrix \mathbf{B} of the (partially collapsed) AKF is replaced by $\bar{\mathbf{B}} = \mathbf{B}_{\mathbf{x}}$. The matrices $\bar{\mathbf{A}}_{t+1|t}$, $\bar{\mathbf{V}}_t = -\mathbf{Z}_t \bar{\mathbf{A}}_{t+1|t} - \mathbf{X}_t \bar{\mathbf{B}}$ and $\bar{\mathbf{B}}$ have k columns.

9.7.2 Likelihood calculation

At the end of a fully collapsed AKF, the estimate of σ^2 is given by $\hat{\sigma}^2 = \hat{q}_T / (NT - d - k)$ where $k = 0$ and \hat{q}_T is obtained from the collapsed KF. When $k > 0$, \hat{q}_T is given by

$$\hat{q}_T = \bar{q}_T - \bar{\mathbf{s}}'_T \bar{\mathbf{S}}_T^{-1} \bar{\mathbf{s}}_T.$$

Likelihood calculation also requires the log value of the term $|\mathbf{S}_*|$ where $\mathbf{S}_* = \mathbf{S}_T$. If $k = 0$ and a full collapse has taken place at $t = m$, then $|\mathbf{S}_*| = |\mathbf{S}_m|$. If $k > 0$ and a partial collapse has taken place at $t = m$, then $|\mathbf{S}_*| = |\mathbf{S}_{\mathbf{i}, m}| |\bar{\mathbf{S}}_T|$. Note that the determinants are obtained without any extra computational costs since the inverses of these matrices are calculated earlier.

9.7.3 Recursive regressions

The recursive evaluation of least squares estimates is common practice in standard regression analysis since it is a way to get the recursive residuals. The same technique can be applied to the recursive evaluation of \mathbf{S}_t which consists of the equations

$$\begin{aligned} \hat{\mathbf{a}}_{t|t-1} &= \mathbf{a}_{t|t-1} + \mathbf{A}_{t|t-1} \mathbf{c}_{t-1}, & \hat{\mathbf{P}}_{t|t-1} &= \mathbf{P}_{t|t-1} + \mathbf{A}_{t|t-1} \mathbf{A}'_{t|t-1}, \\ \hat{\mathbf{v}}_t &= \mathbf{v}_t + \mathbf{V}_t \mathbf{c}_{t-1}, & \hat{\mathbf{F}}_t &= \hat{\mathbf{F}}_t + \mathbf{V}_t \mathbf{C}_{t-1} \mathbf{V}_t, \\ \mathbf{c}_t &= \mathbf{c}_{t-1} + \mathbf{K}_t^* \hat{\mathbf{v}}_t, & \mathbf{K}_t^* &= \hat{\mathbf{F}}_t^{-1} \mathbf{V}_t \mathbf{C}_{t-1}^{-1}, \\ & & \mathbf{C}_t &= \mathbf{C}_{t-1} - \mathbf{K}_t' \hat{\mathbf{F}}_t \mathbf{K}_t^*. \end{aligned} \quad (9.38)$$

where $\mathbf{c}_t = \mathbf{S}_t^{-1} \mathbf{s}_t$ and $\mathbf{C}_t = \mathbf{S}_t^{-1}$.

9.8 Appendix 2: Numerical optimisation*

Numerical optimisation methods are used to obtain the maximum likelihood estimates for the parameter vector θ . Here we outline some of the basic principles behind the estimation procedure of STAMP.

There is a vast literature on non-linear optimisation techniques (see, among many others, Gill, Murray and Wright (1981), Cramer (1986) and Thisted (1988)). Note that many texts on optimisation focus on minimisation, rather than maximisation, but of course: $\max \ell(\theta) = -\min -\ell(\theta)$.

A first approach to obtaining the MLE $\boldsymbol{\theta}$ from $\ell(\theta)$ is to consider solving the score equations, assuming the relevant partial derivatives exist:

$$\nabla \ell(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \mathbf{q}(\theta). \quad (9.39)$$

Then $\mathbf{q}(\theta) = \mathbf{0}$ defines the necessary conditions for a local maximum of $\ell(\theta)$ at $\boldsymbol{\theta}$. A

sufficient condition is that

$$\nabla^2 \ell(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} = \frac{\partial \mathbf{q}(\theta)'}{\partial \theta} = -\mathbf{Q}(\theta) \quad (9.40)$$

also exists and is negative definite at θ . If $\mathbf{Q}(\cdot)$ is positive definite for all parameter values, the likelihood is concave, and hence has a unique maximum; if not, there could be local optima or singularities.

9.8.1 Newton type methods

When $\mathbf{q}(\theta)$ is a set of equations that are linear in θ , $\mathbf{q}(\theta) = \mathbf{0}$ can be solved explicitly for θ . More generally, $\mathbf{q}(\theta)$ is non-linear, yielding a problem of locating θ which is no more tractable than maximising $\ell(\theta)$. Thus, we consider iterative approaches in which a sequence of values of θ (denoted θ_i at the i^{th} iteration) is obtained approximating $\mathbf{q}(\theta_i) = \mathbf{0}$, and corresponding to non-decreasing values of the criterion function $\ell(\cdot)$: $\ell(\theta_{i+1}) \geq \ell(\theta_i)$:

$$\theta_{i+1} = \mathbf{h}(\theta_i) \text{ for } i = 1, 2, \dots, I \leq M \quad (9.41)$$

where M is a terminal maximum number of steps, from an initial value θ_0 . A convergence criterion is used to terminate the iteration, such as $|\mathbf{q}(\theta_{i+1})| \simeq 0$ or $|\ell(\theta_{i+1}) - \ell(\theta_i)| \leq \epsilon$. These implementation-specific aspects are discussed in §9.6.

Expand $\mathbf{q}(\theta) = \mathbf{0}$ in a first-order Taylor's series:

$$\mathbf{q}(\theta) \simeq \mathbf{q}(\theta_1) - \mathbf{Q}(\theta_1)(\theta - \theta_1) = \mathbf{0}, \quad (9.42)$$

written as the iterative rule:

$$\theta_{i+1} = \theta_i + \mathbf{H}(\theta_i) \mathbf{q}(\theta_i), \quad (9.43)$$

where $\mathbf{H}(\cdot) = \mathbf{Q}^{-1}(\cdot)$ is the *Hessian* matrix. The gradient, $\mathbf{q}(\theta_i)$, determines the direction of the step to be taken, and $\mathbf{H}(\theta_i)$ modifies the size of the step which determines the metric: this algorithm is the Newton-Raphson technique, or *Newton's method*. Even when the direction is uphill, it is possible to overshoot the maximum in that direction. In that case it is essential to add a line search to determine a step length $s_i \in [0, 1]$:

$$\theta_{i+1} = \theta_i + s_i \delta_i, \quad (9.44)$$

where $\delta_i = \mathbf{H}(\theta_i) \mathbf{q}(\theta_i)$ and s_i is chosen to ensure that $\ell(\theta_{i+1}) \geq \ell(\theta_i)$.

STAMP uses the so-called variable metric or Broyden-Fletcher-Goldfarb-Shanno (BFGS) update *quasi-Newton* methods. These approximate the *Hessian* matrix $\mathbf{H}(\theta_i)$ by a symmetric positive definite matrix \mathbf{K}_i which is updated at every iteration but converges on $\mathbf{H}(\cdot)$.

Let $\mathbf{d} = \mathbf{s}_i \delta_i = \theta_i - \theta_{i-1}$ and $\mathbf{g} = \mathbf{q}(\theta_i) - \mathbf{q}(\theta_{i-1})$; then the BFGS update is:

$$\mathbf{K}_{i+1} = \mathbf{K}_i + \left(1 + \frac{\mathbf{g}' \mathbf{K}_i \mathbf{g}}{\mathbf{d}' \mathbf{g}} \right) \frac{\mathbf{d} \mathbf{d}'}{\mathbf{d}' \mathbf{g}} - \frac{\mathbf{d} \mathbf{g}' \mathbf{K}_i \mathbf{g}'}{\mathbf{d}' \mathbf{g}}. \quad (9.45)$$

This satisfies the quasi-Newton condition $\mathbf{K}_{i+1} \mathbf{g} = \mathbf{d}$, and possesses the properties of hereditary symmetry (\mathbf{K}_{i+1} is symmetric if \mathbf{K}_i is), hereditary positive definiteness, and super-linear convergence.

9.8.2 Numerical score vector and diagonal Hessian matrix

When the analytic formula for $\mathbf{q}(\theta)$ cannot be obtained easily, STAMP uses BFGS with a numerical approximation to $\mathbf{q}(\cdot)$ based on finite difference approximations. The numerical derivatives are calculated using:

$$q_j(\theta) = \frac{\partial \ell(\theta)}{\partial (\iota'_j \theta)} \simeq \frac{\ell(\theta + \varepsilon \iota_j) - \ell(\theta)}{\varepsilon} \quad (9.46)$$

where ι_j is a unit vector (for example, $(0 \ 1 \ 0 \ \dots \ 0)'$ for $j = 2$), ε is a suitably chosen step length. Thus, ε represents a compromise between round-off error (cancellation of leading digits when subtracting nearly equal numbers) and truncation error (ignoring terms of higher order than ε in the approximation).

The diagonal entries of the Hessian matrix are numerically evaluated as

$$H_{j,j}^d(\theta) = -\frac{1}{Q_{j,j}(\theta)} \quad (9.47)$$

where

$$Q_{j,j}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial (\iota'_j \theta) \partial (\iota'_j \theta)} = \frac{\partial q_j(\theta)}{\partial (\iota'_j \theta)} = \frac{\partial q_j(\theta + \varepsilon \iota_j) - q_j(\theta)}{\varepsilon} \quad (9.48)$$

with the same step-length ε .

Chapter 10

Statistical Model Output

This chapter defines the output of STAMP. The first section explains when and where the various parts of the output appear. The remaining sections explain the output is presented for every option of the Test menu.

10.1 Output from STAMP

This section details the output which occurs when a model is fitted to data. There are three different forms of output. Firstly, the iterations of the numerical optimisation algorithm are displayed in the Message window. Secondly, once the estimation algorithm has stopped (hopefully having converged!), the Results window reports details of the model specification and the results of the optimisation algorithm, and gives a summary of diagnostic statistics. Finally, the Test menu can be used to prompt STAMP to output specific details of the parameter estimates, state variables, model fit and diagnostics.

10.1.1 Model estimation

In order to compute the maximum likelihood estimates, STAMP first computes initial estimates and then it moves on to maximise the likelihood function directly by numerical optimisation; see Chapter 9. During this second phase various statistics are reported in the Message window:

- *Parameters* - the current values of the transformed parameters;
- *Gradients* - the derivatives of the likelihood function with respect to the transformed parameters;
- *Function value* - the kernel of the log-likelihood function;
- *Initial value* - the value of the function after the initial phase;
- *% change* - the percentage increase of the current likelihood kernel over the initial value.

10.1.2 Selected model and estimation output

The initial output, written to the Results window, gives details of the selected model and various results from the numerical optimisation procedure; an example is reproduced in §8.3.1. In particular, the selected components are listed and the number of parameters is given. The sample period used for parameter estimation is also given.

A variety of statistics emerge from the numerical optimisation. The likelihood kernel evaluated at $\hat{\theta}$ is reported together with the number of iterations required for convergence. Three different convergence measures are reported and these are categorised by some term like STRONG CONVERGENCE; see §9.6.

10.1.3 Summary statistics

An example of the diagnostic summary report, which is printed to the Results window after fitting, is given in §8.3.2. Broadly this summary gives the basic diagnostics and goodness-of-fit statistics. The precise definitions are given in §10.4 and §10.7.

- *Log-likelihood* - this is the actual log-likelihood function at its maximum;
- *Prediction error variance (PEV)* - the variance, or variance matrix, of the one-step ahead prediction errors in the steady state, or the finite PEV if accompanied by a message of 'No steady state';
- *Std. Error* - the equation standard error is the square root of the PEV;
- *Normality* - the Doornik–Hansen statistic, which is the Bowman–Shenton statistic with the correction of Doornik and Hansen (1994), distributed approximately as χ^2_2 under the null hypothesis;
- *H(h)* - a test for heteroskedasticity, distributed approximately as $F(h, h)$;
- $r(\tau)$ - the residual autocorrelation at lag τ , distributed approximately as $N(0, 1/T)$;
- DW - Durbin–Watson statistic, distributed approximately as $N(2, 4/T)$;
- $Q(P, d)$ - Box–Ljung Q-statistic based on the first P residual autocorrelations and distributed approximately as χ^2_d ;
- R_D^2, R_S^2 or R_d^2 - the most suitable coefficient of determination.

Notice that the loss in the degrees of freedom d of the Box–Ljung Q-statistic takes account of the number of relative parameters in θ . It does not take account of any lagged dependent variables and their presence necessitates further adjustment.

10.1.4 The sample period

The parameters are estimated using the *model sample* as recorded from the Estimation dialog in the **Model** menu. The calculation of diagnostics and goodness-of-fit measures are calculated using the *state sample*. The default is that the model sample and the

state sample are the same. However, the state sample might be changed to check subsamples isolated from the rest of the sample. This can be done in the option Final state. Of course, both samples must be within the boundaries of the *data sample*.

10.1.5 Test

The nine items of the Test menu are self-explanatory. The headings of the following sections correspond to them. Note that the output for Goodness of Fit is requested by the Final state option and that many of the conventional time series diagnostics (built out of the innovations, or one-step ahead prediction errors) will come up in the Residual option.

10.2 Parameters

The variance parameters together with frequencies and damping factors for cycles and coefficients for autoregressive components are referred to as parameters. The maximum likelihood estimates of the parameters can be requested from the More written output dialog. This dialog also possibly outputs the transformed parameters with their scores and approximate standard errors.

10.2.1 Variances and standard deviations

This gives the variances of the disturbances driving the various components. Thus, the level variance is σ_η^2 , the slope variance is σ_ζ^2 and so on. The standard deviations are the square roots of the variances. The *q-ratios* are the ratios of each variance or standard deviation to the largest.

10.2.2 Cycle and AR(1)

For a stochastic cycle component, this option outputs the damping factor ρ , the frequency λ , the period $2\pi/\lambda$, and the period in terms of 'years'; that is, $2\pi/s\lambda$, where s is the number of 'seasons'. In addition it gives the variance of the cycle, $\sigma_\psi^2 = \sigma_\kappa^2/(1 - \rho^2)$. For the AR(1) component in a univariate model, the autoregressive coefficient is reported.

10.2.3 Transformed parameters and standard errors

The vector of estimated parameters, which are transformed as in §9.6, can be written to the Results window using this option. It also reports the score vector, associated with the estimated vector of parameters, and the numerical standard errors of the transformed parameters.

The scores are calculated as described in §9.6. The standard errors are obtained from the diagonal entries of the numerically evaluated Hessian matrix Q. Formally, the standard error of the i -th estimated parameter is defined as

$$se_i = -1/\sqrt{Q_{ii}}$$

where Q_{ii} is the i -th diagonal entry of matrix Q.

10.3 Final state

The estimate of the final state vector contains all the latest information on the components in the model. Together with the regression estimates of the explanatory variables and the interventions, the final state contains all the information needed to make forecasts.

The Final state dialog also provides useful information about the estimated seasonal pattern and the amplitude of the cycle. Some goodness-of-fit statistics can be generated, but these are discussed in §10.4.

10.3.1 Analysis of state

Following the notation as in Koopman, Shephard and Doornik (1999), the final state is the filtered estimate at time T ; that is, $\hat{\alpha}_{T|T}$. The square roots of the diagonal elements of the corresponding MSE matrix, $\hat{\sigma}^2 \hat{\mathbf{P}}_{T|T}$, are the root mean square errors (RMSEs) of the corresponding state elements. The *t-value* is the estimate of the state divided by the RMSE. Applied to the i -th element of the state, this is

$$t\text{-value}_i = a_i / RMSE_i,$$

where a_i is the i -th element of vector $\hat{\alpha}_{T|T}$.

The Prob values shows the probability of the absolute value of a standard normal variable exceeding the absolute value of the *t-value*. In a hypothesis testing situation, a Prob. value of less than 0.05 corresponds to rejection of the hypothesis of a zero value on a two-sided test at the 5% level of significance. A Prob. value of less than 0.05 is indicated by *, while a value less than 0.01 is denoted **.

The various unobserved components which make up the structural time series model all appear in the final state.

- *Trend* - 'Lvl' is the estimate of the level of the trend μ_T while 'Slp' is the estimate of the slope β_T . The *t-value* might not appear for the level when no irregular is specified in the model or when the irregular variance is estimated as zero.
- *Trigonometric seasonal* - the final state contains the estimates for $\gamma_{1,T}$, $\gamma_{1,T}^*$, $\gamma_{2,T}$, $\gamma_{2,T}^*$, etc. and are denoted in the output by 'Sea.1', 'Sea.2' and so on. Note

that the meaning of 'Sea_1', 'Sea_2', etc., is different to what it is in dummy variable seasonality, and so the final estimated seasonal pattern cannot be identified directly. It must be requested in 'Seasonal tests'.

- *Cycle* - the terms 'Cy_k_1' and 'Cy_k_2' give the estimates of ψ_T and ψ_T^* for cycle k , where k is 1, 2 and/or 3. Since a cycle component is not persistent throughout the series, a *t-value* is not appropriate here.
- *Autoregressive* - the estimate of ψ_T is denoted as 'Ar1'.

10.3.2 Regression analysis

The estimates of the regression parameters, the parameters of the intervention variables and the fixed seasonal dummies are computed as described in Koopman, Shephard and Doornik (1999). These estimates are reported upon request (if part of the specified model) and they may be interpreted in much the same way as in a standard regression model. Because they are deterministic (not time-varying), their RMSEs are standard deviations. The *t-values* would have t-distributions if the (relative) parameters were known. Although they are asymptotically normal, a t-distribution may provide a better approximation to small sample properties. As was noted earlier, the Prob. values are based on the normal distribution.

10.3.3 Seasonal tests

If the seasonal component is deterministic, either because it is specified to be 'fixed' at the outset or its disturbance variance is estimated to be zero, a joint test of significance can be carried out on the $s - 1$ seasonal effects. The test is essentially the same as a test for the joint significance of a set of explanatory variables in regression. Under the null hypothesis of no seasonality, the large sample distribution of the test statistic, denoted by 'Seasonal Chi²($s - 1$)' in the output, is χ_{s-1}^2 . The Prob. value is the probability of a χ_{s-1}^2 variable exceeding the value of the test statistic. In the case of a stochastic seasonal, the joint seasonal test is also produced although a formal joint test of significance of the seasonal effect is inappropriate. However, the seasonal pattern is persistent throughout the series and when the seasonal pattern changes relatively slowly, which is usually the case, the test statistic can provide a useful guide to the relative importance of the seasonal.

The formal definition of the test statistic is

$$\mathbf{a}' \mathbf{P}^{-1} \mathbf{a}$$

where \mathbf{a} contains the estimates of the $s - 1$ seasonal effects at time T and \mathbf{P} is the corresponding MSE matrix. In the case of a fixed seasonal, \mathbf{a} and \mathbf{P} are obtained from the matrix ($\mathbf{s}_T, \hat{\sigma}^2 \mathbf{s}_T$). For a stochastic seasonal, \mathbf{a} and \mathbf{P} come from the matrix ($\hat{\mathbf{a}}_{T|T}, \hat{\sigma}^2 \hat{\mathbf{P}}_{T|T}$); see Koopman, Shephard and Doornik (1999).

10.4 Goodness of fit

The goodness-of-fit statistics in STAMP all relate to the residuals, which are the standardised innovations; see Koopman, Shephard and Doornik (1999). When explanatory variables are specified in the model, the generalised least squares residuals are used. The diagnostic checking facility for these residuals are provided in the Residuals option; see §10.7.

10.4.1 Prediction error variance

The basic measure of goodness-of-fit is the PEV as defined in Harvey (1989). The PEV is the variance of the residuals in the steady state and so it corresponds to the variance of the disturbance of the reduced form ARIMA model. When the Kalman filter does

The final estimate of the seasonal effect for each season; that is, the estimated seasonal pattern at time T , is given here as well.

10.3.4 Cycle tests

The amplitude of the cycle is

$$\sqrt{a^2 + a^{*2}}$$

where a is the element of $\hat{\mathbf{a}}_{T|T}$ corresponding to ψ_T and a^* corresponds to ψ_T^* . The reported ratio is the comparison of the *amplitude* of the cycle with the level of the trend, and is also reported. This gives an indication of its relative importance. When the cycle is deterministic, but stationary, a joint significance χ^2 test (the same as the seasonal test) is produced as well.

10.3.5 Data in logs

When it is indicated that a model has been estimated in logarithms, some additional output for specific components appears in the Results window.

- *Trend* - The value $\exp(a)$ is reported where a is the element of $\hat{\mathbf{a}}_{T|T}$ corresponding to μ_T . When a slope component is specified, the estimated annual growth rate is reported as well. The annual growth rate is the estimate of β_T multiplied by 100s.
- *Seasonal* - The multiplicative effect of the seasonal pattern on the level of the series is given by $\exp(a_j)$, where a_j is the estimate of the seasonal effect in season j . Also the percentage effect is calculated; that is, $100(\exp(a_j) - 1)$. This output is added to the output generated for the option 'Seasonal tests'.
- *Cycle* - Additional to Cycle tests, the amplitude as a percentage of the level is computed. This is simply multiplying the estimated amplitude by 100.

10.4 Goodness of fit

The goodness-of-fit statistics in STAMP all relate to the residuals, which are the standardised innovations; see Koopman, Shephard and Doornik (1999). When explanatory variables are specified in the model, the generalised least squares residuals are used. The diagnostic checking facility for these residuals are provided in the Residuals option; see §10.7.

not converge to the steady state, the finite PEV is used. For a univariate model, the PEV is denoted by $\tilde{\sigma}^2$, irrespective of whether or not it is the finite PEV.

10.4.2 Prediction error mean deviation

Another measure of fit is the mean deviation of the residuals, $v_t, t = d + 1, \dots, T$, defined by

$$md = \frac{\tilde{\sigma}}{T-d} \sum_{t=d+1}^T |v_t|$$

where the PEV $\tilde{\sigma}^2$ is defined in Harvey (1989). In a correctly specified model the reported ratio

$$\frac{2\tilde{\sigma}^2}{\pi md^2}$$

should be approximately equal to unity.

When it is indicated that the series are modelled in logs, the relative measure of error is also given. This measure is easily computed as $100md$.

10.4.3 Coefficients of determination

In econometric modelling the traditional measure of goodness-of-fit is the coefficient of determination R^2 . For a structural time series model this measure is defined by

$$R^2 = 1 - \frac{(T-d)\tilde{\sigma}^2}{\sum_{t=1}^T (y_t - \bar{y})^2},$$

where y_t is the original series and \bar{y} is its sample mean. This coefficient of determination is most useful when the series appears to be stationary with no trend or seasonal.

When the series y_t shows trend movements, it is better to compare the PEV with the variance of first differences, $\Delta y_t = y_t - y_{t-1}$. This leads to

$$R_D^2 = 1 - \frac{(T-d)\tilde{\sigma}^2}{\sum_{t=2}^T (\Delta y_t - \overline{\Delta y})^2},$$

where $\overline{\Delta y}$ is the sample mean of Δy_t . It should be noted that R_D^2 may be negative, indicating a worse fit than a simple random walk plus drift model.

For seasonal data with a trend, it is more appropriate to measure the fit against a random walk plus drift and fixed seasons. This requires the sum of squares, $SSDSM$, obtained by subtracting the seasonal mean from Δy_t . The coefficient of determination is then

$$R_S^2 = 1 - \frac{(T-d)\tilde{\sigma}^2}{SSDSM},$$

which can again be negative.

10.4.4 Information criteria : AIC and BIC

The fit of different models can be compared on the basis of the PEV. However when models have different number of parameters, it is more appropriate to compare them using the Akaike information criterion (AIC) or the Bayes information criterion (BIC).

These criteria are based on the formula

$$\log PEV + cn/T$$

where c is a constant and m is the number of parameters in θ plus the number of non-stationary components in the state vector. The AIC takes c as 2 and the BIC takes c as $\log T$.

10.5 Components

The ‘Components’ option facilitates the computation of filtered or smoothed estimates of the specified components in the model. The filtered estimates are based on the past observation and can be calculated as described in Koopman, Shephard and Doornik (1999). The smoothed estimates are based on all the observations. In what follows, the estimates are denoted by \tilde{x}_t , where x is some component which can be estimated either way.

10.5.1 Series with components

This option plots the original series, y_t , with one of the following (as requested):

- the estimated trend, $\tilde{\mu}_t$ (which includes interventions on the level and slope);
- the estimated trend plus cycle(s) plus AR(1), $\tilde{\mu}_t + \tilde{\psi}_t + \tilde{v}_t$;
- the estimated trend plus cycle(s) plus AR(1) plus explanatory variables, $\tilde{\mu}_t + \tilde{\psi}_t + \tilde{v}_t + \tilde{x}_t' \delta_x$, where the last part includes all the explanatory variables in the equation.

10.5.2 Detrended

The detrended series is $y_t - \tilde{\mu}_t$ for $t = 1, \dots, T$. The smoothed estimates are usually most appropriate for detrending. Note that level and slope interventions are included in the trend.

10.5.3 Seasonally adjusted

The seasonally adjusted series is $y_t - \tilde{\gamma}_t$ for $t = 1, \dots, T$. The smoothed estimate of the seasonal component is normally used for seasonal adjustment.

10.5.4 Individual seasonals

The estimated seasonal component $\tilde{\gamma}_t$ is re-arranged into s different series, each corresponding to one of the seasons. The graph can be interpreted as a set of annual plots for the individual seasonal effects.

10.5.5 Data in logs

If the observations are in logarithms, the output of the various components may be amended in two ways: (1) 'Anti-logs' can be calculated; or (2) 'Modified anti-logs' can be calculated. These amendments do not apply in a similar way for all components. The following list gives the formal amendments:

- *Trend* - (1) $\exp(\tilde{\mu}_t)$ and (2) $\exp(\tilde{\mu}_t)$;
- *Growth rate (Slope)* - (1) $100s\tilde{\beta}_t$ and (2) $100s\tilde{\beta}_t$;
- *Seasonal, Cycle, AR(1) and Irregular* - (1) $\exp(\tilde{a})$ and (2) $\exp(\tilde{\mu}_t + \tilde{a}) - \exp(\tilde{\mu}_t)$, where \tilde{a} is a specific component;
- *Detrended* - (1) $\exp(y_t - \tilde{\mu}_t)$ and (2) $\exp(y_t) - \exp(\tilde{\mu}_t)$;
- *Seasonally adjusted* - (1) $\exp(y_t - \tilde{\gamma}_t)$ and (2) $\exp(y_t - \tilde{\gamma}_t)$.

10.6 Residuals

The residuals are the standardised one-step-ahead prediction errors or innovations and they are defined in Koopman, Shephard and Doornik (1999). The residual series are simply denoted by v_t for $t = d+1, \dots, T$. For a correctly specified model with known variance parameters, the residuals are assumed to be $\text{NID}(0, 1)$. Diagnostic statistics and graphs are the tools to validate this proposition.

10.6.1 Correlogram

The residual autocorrelation at lag τ is defined by

$$r_\tau = \frac{\sum_{t=\tau+d+1}^T (v_t - \bar{v})(v_{t-\tau} - \bar{v})}{\sum_{t=d+1}^T (v_t - \bar{v})^2}, \quad \text{where } \bar{v} = \frac{1}{T-d} \sum_{t=d+1}^T v_t. \quad (10.1)$$

The approximate standard error for r_τ is $1/\sqrt{T-d}$. The Durbin and Watson (1950) statistic, is given by

$$DW = \frac{\sum_{t=2+d}^T (v_t - v_{t-1})^2}{\sum_{t=1+d}^T v_t^2} \simeq 2\{1 - r(1)\}.$$

In a correctly specified model, DW is approximately $\text{N}(2, 4/T)$.

A general test for serial dependence is the portmanteau Box-Ljung Q -statistic, which is based on the first s residual autocorrelations; see Ljung and Box (1978). It is given by

$$Q(P, d) = T(T+2) \sum_{j=1}^P \frac{r_j^2}{T-j}.$$

and distributed approximately as χ_d^2 under the null, where d is equal to $P-n+1$ and n is the number of parameters. Notice that the loss in the degrees of freedom d of $Q(P, d)$ takes account of the number of relative variance parameters. It does not take account of any lagged dependent variables and their presence necessitates further adjustment.

10.6.2 Periodogram and spectrum

These diagnostic graphs are defined in Harvey (1989). They are applied to v_t in exactly the same way. Note that the theoretical spectrum is a horizontal straight line for a white-noise series.

10.6.3 Cumulative statistics and graphs

The 'Cusum' and 'Cusum of squares' can be calculated using the residuals v_t for $t = d+1, \dots, T$.

The 'Cumulative periodogram' is another cumulative graph which is useful since it smooths out the periodogram. The cumulative periodogram cp_j is defined as

$$cp_j = \frac{\sum_{i=1}^j p_i^2}{\sum_{i=1}^n p_i^2}$$

where p_i is the periodogram ordinate at i for $i = 1, \dots, n$ and $n = [(T-d)/2]$. The cumulative periodogram cp_j is plotted against $j = 0, \dots, n$. A useful diagnostic statistic is to measure the maximum deviation of the cp_j from the diagonal line connecting $cp_0 = 0$ and $cp_n = 1$; that is, $\max|cp_j - j/n|$ over $j = 1, \dots, n-1$; see Durbin (1969). This measure is outputted on request in the Results window. Some typical critical values for this two-sided test are .44 for $T \simeq 80$, .40 for $T \simeq 100$ and .30 for $T \simeq 200$.

10.6.4 Distribution statistics

The Bowman and Shenton (1975) statistic for normality is calculated using the residuals v_t for $t = d+1, \dots, T$.

10.6.5 Heteroskedasticity

A basic non-parametric test of heteroskedasticity is the two-sided $F_{h,h}$ -test,

$$H(h) = \sum_{t=T-h+1}^T v_t^2 / \sum_{t=d+1}^{d+1+h} v_t^2$$

where h is the closest integer to $(T-d)/3$. An increase of the variance in the last third of the sample will cause $H(h)$ to be large, while the reverse will cause it to be small. The heteroskedasticity test appears as part of the Summary statistics.

10.7 Auxiliary residuals

Auxiliary residuals are designed to detect unusual movements in a fitted time series model. These residuals are the smoothed estimates of the disturbances associated with the components irregular, level and slope. Koopman, Shephard and Doornik (1999) sets out how to calculate the standardised auxiliary residuals. It is very useful to plot them and look for any irregularities. If requested, all residuals which have absolute value exceeding two are written to the Results window.

As well as graphical representations of the auxiliary residuals, STAMP reports the standard distribution statistics for these residuals. Of course, these statistics cannot be used as standard moment tests because of the serial correlation in the auxiliary residuals. However these statistics can be corrected for serial correlation as in Harvey and Koopman (1992). In STAMP, the correction factors are calculated from the first $\max(\sqrt{T}, 20)$ sample autocorrelations of the auxiliary residuals.

10.8 Predictive testing

When the final date of the *state sample* is smaller than the final date of the *data sample*; that is, there are L observations, y_t , $t = T + 1, \dots, T + L$ ‘outside the sample’, post-sample predictive testing may be carried out using all or some of these observations. However, when these final dates are the same, predictive testing within the sample is the only option; that is, predicting y_t for $t = T - l + 1, \dots, T$. Of course, in the first situation, predictive testing within the sample is also possible.

When explanatory variables and/or interventions are specified in the model, the residuals used in predictive testing within the sample are the generalised recursive residuals defined in Harvey (1989). These residuals are denoted by w_t for $t = d^* + 1, \dots, T$ where $d^* = d + k$ and k is the number of explanatory variables in the model (excluding any variables which do not appear after time $T - L$). Note that w_t is zero when an intervention enters into the model. The error bands for the residuals are based on twice the RMSE. The CUSUM is calculated for these residuals if requested.

The Chow test within the sample is constructed as

$$Chow = \frac{T - l - d^*}{l} \sum_{t=T-l+1}^T w_t^2 / \sum_{t=d^*+1}^{T-l} w_t^2.$$

which is approximately distributed as $F(l, T - l - d^*)$.

When post-sample predictive testing is carried out, the estimates of the coefficients of the explanatory variables and/or interventions remain the same as they were at the end of the state sample. The corresponding standardised residuals are denoted by v_t for $t = T + 1, \dots, L$. The post-sample predictive ‘failure’ test statistic is then

$$pft = \sum_{j=1}^L v_{T+j}^2,$$

which is approximately distributed as χ_L^2 .

The CUSUM t-test is also outputted in the form

$$cusumt = L^{-1/2} \sum_{j=1}^L v_{T+j},$$

which is approximately distributed as a *t* distribution with $T - L - d^*$ degrees of freedom.

The extrapolative residuals, which are defined in Harvey (1989), can be used to assess the predictive performance of the model and make comparisons with rival models. If $\hat{v}_{T+j|T}$, $j = 1, \dots, L$, denotes the j -step ahead extrapolative residual in the post-sample period, the *extrapolative sum of squares* is given by

$$ess(T, L) = \sum_{j=1}^L \hat{v}_{T+j|T}^2.$$

This statistic can also be generated using residuals within the sample. A similar statistic based on the sum of absolute values is defined as

$$esa(T, L) = \sum_{j=1}^L |\hat{v}_{T+j|T}|.$$

10.9 Forecast

The forecasts of the series y_t and the components are generated as described in Koopman, Shephard and Doornik (1999). The forecast error bounds are based on one RMSE.

For example, the forecasts of a specific series \hat{y}_{T+j} are obtained from a particular element of $\hat{\mathbf{y}}_{T+j|T}$, for $j = 1, \dots, L$, and the RMSE \hat{r}_{T+j} is the square root of the corresponding diagonal element of the matrix $\hat{\sigma}^2 \hat{\mathbf{F}}_{T+j|T}$. When the data are in logs, the anti-log transformation (\exp) can be calculated for the forecasts. Their corresponding RMSEs are then based on

$$\hat{r}_{T+j}^* = \exp(\hat{y}_{T+j} + \hat{r}_{T+j}) - \exp(\hat{y}_{T+j}), \quad j = 1, \dots, L.$$

The anti-log forecasts adjusted for bias are calculated by

$$\hat{y}_{T+j}^* = \exp(\hat{y}_{T+j} + 0.5\hat{r}_{T+j}^2), \quad j = 1, \dots, L.$$

STAMP is mostly menu-driven for ease of use. To add flexibility, certain functions can be accessed through entering commands. The syntax of these commands is described in this chapter. They can be added to the general list of Batch commands provided by OxMetrics.

STAMP allows models to be formulated, estimated and evaluated through batch commands. Such commands are entered in OxMetrics. Certain commands are interpreted by OxMetrics, such as those for loading and saving data, as well as blocks of algebra code. The remaining commands are then passed on to the active module, which is STAMP in this case. This section gives an alphabetical list of the STAMP batch language statements. There are two types of batch commands: function calls (with or without arguments) terminated by a semicolon, and commands, which are followed by statements between curly brackets.

Anything between /* and */ is considered comment. Note that this comment cannot be nested. Everything following // up to the end of the line is also comment.

OxMetrics allows you to save the current model as a batch file, and to rerun saved batch files. If a model has been created interactively, it can be saved as a batch file for further editing or easy recall in a later session. This is also the most convenient way to create a batch file.

If an error occurs during processing, the batch run will be aborted and control returned to OxMetrics. A warning or out of memory message will have to be accepted by the user (press Enter), upon which the batch run will resume.

Batch is activated in OxMetrics via **Model/Batch** or just by pressing **Alt+b**. The batch dialog appears with default code displayed. The code reflects the current model in STAMP. For example,



In the following list, function arguments are indicated by *words*, whereas the areas where statement blocks are expected are indicated by Examples follow the list of descriptions. For terms in double quotes, the desired term must be substituted and provided together with the quotes.

`estimate();`

Starts the estimation process as described in §8.2.2 and §9.6.

`forecast(nfor, incr, option);`

Prints *nfor* ahead forecasts of dependent variables. When explanatory variables are included in the model, the *incr* value is the increment value for all exogenous variables. Otherwise, no value for *incr* should be given. When a correction for bias is required due to the log-transformation, the *option* must be set to *logbias*. Otherwise, no value is required for *option*.

`intervention(component, year, period);`

The *component* argument specifies the type of intervention to be included in the model:

- include outlier: `outlier`;
- include structural break: `level`;

- include slope change: `slope`;
 - no intervention in the model: `off`.
- The *year* and *period* indicates the start of the intervention. When *year* and *period* refer to a non-existing date within the sample, the intervention will not enter the model.
- `system { Y=...; X=...; }`
- Formulate the model, consisting of the following components:
- Y dependent variables;
 - X explanatory (including lagged dependent) variables;
- The variables listed are separated by commas, the base names (that is, name excluding lag length) for Y and X must be in the database. If the variable names are not a valid token, the name must be enclosed in double quotes.
- `store(type, variable);`
- Use this command to store prediction residuals, smoothed components and smoothed auxiliary residuals into the database, the default name is used. The *type* must be one of: `residuals`, `components`, `auxresiduals`. The *variable* must be one of: `irreg`, `level`, `slope`.
- When *type* is `residuals`, no *variable* need to be specified.
- When *type* is `components`, the *variable* can also be one of: `trend`, `season`, `cycle`, `dettrend`, `seasadj`, `trendx`. The variable *cycle* refers to the sum of smoothed components associated with autoregressive and the three cycles. The variable *trend* is the same as the smoothed level and *trendx* refers to the sum of smoothed level and all estimated explanatory variables.
- `setcmp(cmp, order, variance, par1, par2);`
- Introduces an unobserved component in the model. The possibilities are:
- `setcmp(level, -, 1.0, -, -)`; where `-` refers to "not a relevant number" and 1.0 is the variance.
 - `setcmp(slope, 2, 1.0, -, -)`; where 2 is the order of the trend.
 - `setcmp(seasonal, 4, 1.0, -, -)`; where 4 is the length of the seasonal.
 - `setcmp(irregular, -, 1.0, -, -)`;
 - `setcmp(cycle, 2, 1.0, 5, 0.9)`; where 2 is the order of the cycle, 5 is the period and 0.9 is the discounting factor of the cycle.
 - `setcmp(ar, 1, 1.0, 0.8, -)`; where 1 is the order of the autoregressive process, AR(1) and 0.8 is the autoregressive coefficient.
 - `setcmp(ar, 2, 1.0, 0.8, 0.2)`; where 2 is the order of the autoregressive process, AR(2), 0.8 is the first autoregressive coefficient and 0.2 is the second autoregressive coefficient.
- `setmodel();`

This call is required to activate all specified unobserved components in the model.
 This call must always come after the last setcmp statement.

```
teststate();
Prints the estimated final state and the estimated regression and intervention vector together with the corresponding test statistics. The output depends on the specified model.

testsummary();
Prints the estimation and diagnostic summary reports.

We finish with an annotated example using most commands. To run this file, we assume that OxMetrics is loaded with SEATBQ.IN7, and that STAMP has been started.

// Batch code for UC( 1 )
module("STAMP");
package("UCstamp");
usedata("SEATBQ.IN7");
system{
  Y = Drivers;
  X = Kms, Petrol;
}
setcmp(level, 0, 0.1, 0, 0);
setcmp(seasonal, 4, 0.1, 0, 0);
setcmp(irregular, 0, 1, 0, 0);
intervention(Level, 83, 1);
setmodel();
estimate("Maximum Likelihood", 69, 1, 84, 4);
```

This Batch code produces the output:

```
UC( 1 ) Modelling Drivers by Maximum Likelihood (using SEATBQ.IN7)
The selection sample is: 69(1) - 84(4)
The model is: Y = Level + Seasonal + Irregular + Expl vars + Interv
Log-Likelihood is 136.987 (-2 LogL = -273.975).
Prediction error variance is 0.006686682
```

Summary statistics	
std.error	0.082854
Normality	5.2573
H(19)	0.56844
r(1)	0.16536
r(6)	0.046724
DW	1.6354
Q(6,4)	10.348
Rs^2	0.49962

Variances of disturbances.

Component	Value	(q-ratio)
Level	0.00035069	(0.0715)

```
Seasonal          1.2587e-005 ( 0.0026)
Irregular        0.0049063 ( 1.0000)

State vector analysis at period 84(4)
- level is 4.65234 with stand.err 1.7011.
- joint seasonal chi2 test is 72.2287 with 3 df.
- seasonal effects are

  period      value    stand. err      t-value    prob
    1   -0.073881  0.028365
    2   -0.142151  0.028420
    3   -0.013761  0.02807
    4   0.229794  0.031621

Regression effects in final state at time 84(4)

  Coefficient  stand.err  t-value  prob
Kms          -0.21887  0.05351 -4.09054 [0.00014]
Petrol        0.22469  0.17688  1.27030 [0.20914]
Level break 83. 1 -0.26824  0.12211 -2.19677 [0.03212]
```

References

- Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.
- Ansley, C. F. and Kohn, R. (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *J. Statistical Computation and Simulation*, **24**, 99–106.
- Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business and Economic Statistics*, **11**, 81–92.
- Baxter, M. and King, R. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Rev. Economics and Statistics*, **81**, 575–93.
- Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, **62**, 243–50.
- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.
- Bruce, A. G. and Martin, R. D. (1989). Leave-k-out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society, Series B*, **51**, 363–424.
- Bryan, A. C. and Cecchetti (1994). Measuring core inflation. In Mankiw, N. G. (ed.), *Monetary Policy*, pp. 195–215. Chicago: University of Chicago Press.
- Bryson, A. E. and Ho, Y. C. (1969). *Applied Optimal Control*. Massachusetts: Blaisdell.
- Busetti, F. (2006). Tests of seasonal integration and cointegration in multivariate unobserved components models. *J. Applied Econometrics*, **21**, 419–38.
- Carvalho, V., Harvey, A. C. and Trimbur, T. (2007). A note on common cycles, common trends and convergence. *J. Business and Economic Statist.*, **23**, 12–20.
- Chan, W. Y. T. and Wallis, K. F. (1978). Multiple time series modelling: another look at the mink-muskat interactions. *Applied Statistics*, **27**, 168–75.
- Christiano, L. and Fitzgerald, T. (2003). The band-pass filter. *International Economic Review*, **44**, 435–65.
- Chu-Chun-Lin, S. and de Jong, P. (1993). A note on fast smoothing. Unpublished: University of British Columbia.
- Clark, P. K. (1989). Trend reversion in real output and unemployment. *Journal of Econometrics*, **40**, 15–32.
- Cogley, T. and Sargent, T. (2007). Inflation-Gap persistence in the US. Mimeo.
- Commandeur, J. J. F. and Koopman, S. J. (2007). *An Introduction to State Space Time Series Analysis*. Oxford: Oxford University Press.
- Cramer, J. S. (1986). *Econometric Applications of Maximum Likelihood Methods*. Cambridge: Cambridge University Press.
- de Jong, P. (1988). The likelihood for a state space model. *Biometrika*, **75**, 165–169.

- de Jong, P. (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association*, **84**, 1085–8.
- de Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics*, **19**, 1073–1083.
- de Jong, P. and Chu-Chun-Lin, S. (1994). Fast likelihood evaluation and prediction for nonstationary state space models. *Biometrika*, **81**, 133–142.
- den Butter, F. A. G. and Mourik, T. J. (1990). Seasonal adjustment using structural time series models: an application and comparison with the X-11 method. *Journal of Business and Economic Statistics*, **8**, 385–94.
- Doornik, J. A. and Hansen, H. (1994). An omnibus test of univariate and multivariate normality. Unpublished paper: Nuffield College, Oxford.
- Durbin, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika*, **56**, 1–15.
- Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, **37**, 409–428.
- Engle, R. F. and Kozicki, S. (1993). Testing for common features (with discussion). *Journal of Business and Economic Statistics*, **11**, 369–395.
- Fuller, W. A. (1996). *Introduction to Time Series* 2nd edition. New York: John Wiley.
- Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic volatility. In Rao, C. R. and Maddala, G. S. (eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
- Gill, P. E., Murray, W. and Wright, M. H. (1981). *Practical Optimization*. New York: Academic Press.
- Gomez, V. (2001). The use of Butterworth filters for trend and cycle estimation in economic time series. *J. Business and Economic Statist.*, **19**, 365–73.
- Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise (with discussion). *J. Business and Economic Statist.*, **24**, 127–218.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Harvey, A. C. (1993). *Time Series Models* 2nd edition. Hemel Hempstead: Harvester Wheatsheaf.
- Harvey, A. C. (2001). Testing in unobserved components models. *J. of Forecasting*, **20**, 1–19.
- Harvey, A. C. (2006). Forecasting with unobserved components time series models. In Elliot, G., Granger, C. W. and Timmermann, A. (eds.), *Handbook of Economic Forecasting*. Amsterdam: North Holland.
- Harvey, A. C. and Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society, Series B*, **149**, 187–227.
- Harvey, A. C. and Jaeger, A. (1993). Detrending, stylised facts and the business cycle. *Journal of Applied Econometrics*, **8**, 231–47.
- Harvey, A. C. and Koopman, S. J. (1992). Diagnostic checking of unobserved components time series models. *Journal of Business and Economic Statistics*, **10**, 377–389.
- Harvey, A. C. and Koopman, S. J. (2000). Signal extraction and the formulation of unobserved components models. *Econometrics Journal*, **3**, 84–107.
- Harvey, A. C. and Proietti, T. (2005). *Readings in Unobserved Components Models*. Oxford: Oxford University Press.
- Harvey, A. C., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, **61**, 247–264.
- Harvey, A. C. and Shephard, N. (1993). Structural time series models. In Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds.), *Handbook of Statistics, Volume II*. Amsterdam: Elsevier Science Publishers B. V.
- Harvey, A. C. and Trimbur, T. (2003). Generalised model-based filters for extracting trends and cycles in economic time series. *Rev. Economics and Statistics*, **85**, 244–55.
- Harvey, A. C., Trimbur, T. and van Dijk, H. K. (2007). Trends and cycles in economic time series: a Bayesian approach. *J. Econometrics*, **140**, 618–49.
- Hodrick, R. J. and Prescott, E. C. (1980). Postwar U.S. business cycles: an empirical investigation. Discussion paper 451, Carnegie-Mellon University.
- Hull, J. and White, A. (1987). The pricing of options with stochastic volatilities. *Journal of Finance*, **42**, 281–300.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990). Seasonal integration and cointegration. *J. Econometrics*, **44**, 215–38.
- Kane, R. P. and Trivedi, N. B. (1986). Are droughts predictable. *Climate Change*, **8**, 208–23.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with AR(1) models. *Review of Economic Studies*, **65**, 361–393.
- King, R., Plosser, C., Stock, J. H. and Watson, M. W. (1991). Stochastic trends and economic fluctuations. *American Economic Review*, **81**, 819–840.
- Kitagawa, G. and Gersch, W. (1984). A smoothness prior — state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, **79**, 378–89.
- Kohn, R. and Ansley, C. F. (1989). A fast algorithm for signal extraction, influence and cross-validation. *Biometrika*, **76**, 65–79.
- Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika*, **80**, 117–126.
- Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. (2007). *STAMP 8.0: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.
- Koopman, S. J. and Shephard, N. (1992). Exact score for time series models in state space form. *Biometrika*, **79**, 823–6.
- Koopman, S. J., Shephard, N. and Doornik, J. A. (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal*, **2**, 113–66. <http://www.ssfpack.com/>.
- Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **66**, 67–72.
- Mahieu, R. and Schotman, P. (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, **16**, 333–59.
- Maravall, A. (1985). On structural time series models and the characterization of components. *Journal of Business and Economic Statistics*, **3**, 350–355.

- Mills, T. C. (1993). *Time Series Methods for Finance*. Cambridge: Cambridge University Press.
- Morrettin, P. A., Mesquita, A. R. and Rocha, J. G. C. (1985). Rainfall in Fortaleza in Brazil revisited. In Anderson, O. D., Robinson, E. A. and Ord, J. K. (eds.), *Time Series Analysis: Theory and Practice 6*. Amsterdam: North Holland.
- Newton, J. H., North, G. R. and Crowley, T. J. (1991). Forecasting global ice volume, *Journal of Time Series Analysis*, **12**, 255–265.
- Nyblom, J. and Harvey, A. C. (2001). Testing against smooth stochastic trends. *J. Applied Econometrics*, **16**, 415–29.
- Nyblom, J. and Makelainen, T. (1983). Comparison of tests of for the presence of random walk coefficients in a simple linear models, *Journal of the American Statistical Association*, **78**, 856–64.
- Prest, A. R. (1949). Some experiments with demand analysis, *Review of Economics and Statistics*, **31**, 33–49.
- Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, **87**, 271–301.
- Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components, *Journal of the American Statistical Association*, **88**, 590–595.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V. and Barndorff-Nielsen, O. E. (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Shumway, R. H. and Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, **3**, 253–264.
- Thisted, R. A. (1988). *Elements of Statistical Computing*. New York: Chapman & Hall.
- Tunnicliffe-Wilson, G. (1989). On the use of marginal likelihood in time series model estimation, *Journal of the Royal Statistical Society, Series B*, **51**, 15–27.
- Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression, *Journal of Econometrics*, **23**, 385–400.
- West, M. and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* 2 edition. New York: Springer-Verlag.
- Anderson, B. D. O. 171, 182
- Ansley, C. F. 187, 189, 193
- Balkie, N. S. 16, 69
- Baxter, M. 64, 110
- Bowman, K. O. 211
- Box, G. E. P. 18, 19, 53, 211
- Bruce, A. G. 19
- Bryan, A. C. 116
- Bryson, A. E. 187
- Busetti, F. 101
- Carvalho, V. 101, 124
- Cecchetti, 116
- Chan, W. Y. T. 16, 91
- Chib, S. 125
- Christiano, L. 110
- Chu-Chun-Lin, S. 178, 183, 188
- Clark, P. K. 106
- Cogley, T. 116
- Commander, J. J. F. 4, 31
- Cramer, J. S. 199
- Crowley, T. J. 19, 67
- de Jong, P. 171, 177, 178, 183, 184, 187–189
- den Butter, F. A. G. 128
- Doornik, J. A. xix, 4, 13, 203, 205–207, 209, 210, 212, 213
- Durbin, J. 4, 16, 19, 75, 103, 136, 210, 211
- Engle, R. F. 101, 188
- Fitzgerald, T. 110
- Fuller, W. A. 78, 87
- Gersch, W. 128
- Ghysels, E. 125
- Gill, P. E. 199
- Gomez, V. 55
- Granger, C. W. J. 101
- Ljung, G. M. 211
- Lunde, A. 133
- Mahieu, R. 125
- Makelainen, T. 56
- Maravall, A. 128
- Martin, R. D. 19
- Mesquita, A. R. 65
- Mills, T. C. 15
- Moore, J. B. 171, 182
- Morrettin, P. A. 65
- Mourik, T. J. 128
- Murray, W. 199
- Newton, J. H. 19, 67
- North, G. R. 19, 67
- Nyblom, J. 56, 95
- Peters, S. 18, 87
- Hansen, P. R. 133
- Hansen, H. 203
- Plosser, C. 17
- Prescott, E. C. 128
- Harrison, J. 4, 19, 177

Subject Index

- Prest, A. R. 17
 Proietti, T. 4
 Renault, E. 125
 Rocha, J. G. C. 65
 Ruiz, E. 102, 125
 Sandmann, G. 125
 Sargent, T. 116
 Schornan, P. 125
 Shenton, L. R. 211
 Shephard, N. xix, 4, 13, 102, 125, 178, 189,
 205–207, 209, 210, 212, 213
 Shumway, R. H. 188
 Stock, J. H. 17
 Stofer, D. 188
 Thisted, R. A. 199
 Trimbur, T. 63, 64, 101, 106, 110, 120, 124
 Trivedi, N. B. 16, 65
 Tunnicliffe-Wilson, G. 178
 van Dijk, H. K. 120, 124
 Wallis, K. F. 16, 91
 Watson, G. S. 210
 Watson, M. W. 17, 188
 West, M. 4, 19, 177
 White, A. 125
 Wren-Lewis, S. 18, 87
 Wright, M. H. 199
 Yoo, B. S. 101

- Amplitude 62, 91, 205, 207,
Also see Cycle
ARCH *see* Autoregressive conditional heteroskedasticity, Stochastic volatility
ARIMA *see* Autoregressive integrated moving average
 Augmented disturbance smoothing 188
 Augmented Kalman filter 182, 184, 186, 198
 Autocorrelation function 61, 71, 90,
Also see Correlogram
Autoregression 66, 92
Autoregressive 177
 Autoregressive conditional heteroskedasticity (ARCH) 125
 Autoregressive integrated moving average (ARIMA) 53, 69, 87, 207
 Auxiliary residuals 8, 44, 70, 162, 187, 190,
 212, 217
 — graphics 71, 87, 162
 Bandpass filter 63
 Basic structural model 34, 57, 76, 88, 128,
 141
 Multivariate 89
 Batch 9, 215
 — file (FL) 215
 Language 215
 Variable names 217
 Bayesian 4
 Bowman-Shenton statistic 155, 211
 Box-Jenkins method 49,
Also see ARIMA
 Box-P-Ljung statistic 36, 80, 154, 203, 211
 Broyden-Fletcher-Goldfarb-Shanno (BFGS) estimation 149, 193, 194, 200, 201
 Bureau of the Census X-11 66
 Cholesky decomposition 94, 95, 102, 174,
 Fixed 49, 172, 176, 180, 191
 Level 78, 205, 207, 209, 212
 Local level 140
 Slope 78, 209, 212
 Citation 13
 Coefficient of Determination 203, 208
 Cointegrating vector 97
 Cointegration 93, 95, 114
 Seasonal 101
 Common cycles 101, 176
 Common factors 88, 93, 174, 176, 180
 Balanced levels 97
 Common cycles 101
 Common seasonals 101
 Common trends 99
 Common seasonals 101, 176
 Common trends 94, 99, 174–176
 Components 217
 — graphics 39, 51, 70, 81, 91, 131, 153,
 159
 Autoregressive 5, 6, 55, 63, 66, 81, 87, 92,
 106, 141, 171, 172, 174, 191, 193, 202,
 204, 206
 Cycle 6, 87, 90, 113, 141, 179, 206, 207,
 217
 Similar cycles 120, 171
 Stochastic 66, 172, 204
 Trigonometric 106
 Irregular 6, 34, 78, 111
 Seasonal 6, 56, 89, 101, 113, 118, 119,
 128, 160, 171, 180, 206, 210, 217
 Dummy variable 206
 Fixed 49, 172, 176, 180, 191
 Stochastic 34
 Trigonometric 5, 89, 101, 171, 205
 Selection 33, 49, 140
 Trend 6, 56, 160, 172, 173, 178, 180, 209,
 217

- Stochastic 34, 140, 171, 175, 179,
— 180
Concentrated likelihood 184
Control group 103
Convergence 154, 194, 200, 201, 203
Correlation matrix 6, 88, 95, 113, 123, 183
Correlogram 42, 53, 61, 161, 210
Cumulative periodogram 211
CUSUM 211, 213
CUSUMQ 211
Cycle 60, 62, 63, 65, 66, 171–174, 176, 179,
— 180, 191, 192, 204–207, 210
— higher order 63
— plus noise 63
— plus noise model 60
Multiple cycle 64
Stochastic — 60, 62, 63, 66, 172
Data — loading 21
— storage 10
— transformations 25
Spreadsheet files 10
Data set 10
Airline — (AIRLINE) 18
Brazilian rainfall — (RAINBRAZ) 16,
60, 64
Employment and output — (EMPL) 18,
87
Energy — (ENERGY) 14, 21, 31, 47, 87
Exchange rate — (EXCH) 15, 125
Ice volume — (ICEVOL) 18, 67
Interest rate — (INTEREST) 15, 187
Latin American exports — (LAXQ) 19,
87
Mink and Muskrat — (MINKMUSK) 16,
90
Nile — (NILE) 16, 69–71, 86, 87
Purse — (PURSE) 19
Seat belt monthly — (SEATBELT) 16, 75,
87
Seat belt quarterly — (SEATBQ) 16, 95,
103, 136
Spirits — (SPIRIT) 17, 78, 83, 87
Telephone calls — (TELEPHON) 17
Trade price — (TRADEPRICE, INT) 133
UK consumption, income and prices —
(UKCYP) 17, 57, 89
- US income, consumption, investment,
money and prices — (USmacro07) 17,
107, 113
US income, consumption, investment,
money and prices — (USYCIMP) 17,
50, 64
Detrending 9, 91, 127, 128, 135, 160, 187,
189, 209, 217
Diagnostic checking 9, 38, 42, 49, 153, 154,
161, 181, 187, 203, 207
Dialog 11
Multiple selection list box 11, 24
Differencing 172
First difference 87
Diffuse
— distributions 188, 198
— initial conditions 177, 178
— likelihood 184, 189
Disturbance smoothing 187
Documentation 3, 12
Doornik-Hansen statistic 203
Drift 140, 172, 208
Dummy variable 68
Durbin-Watson statistic 78, 80, 203
EM algorithm 187–189
Estimated variances 36, 80, 89, 190
Excel 8, 10, 21
Explanatory variables 6, 75, 77, 82, 102,
138, 165, 173, 177, 206, 207, 212,
217
Extrapolation 164
Extrapolative residuals 213
Factor loadings 94, 97, 100, 174, 192, 197
Factor rotations 102
Failure chi2 test 76
Filtering 5, 111, 186, 191, 198, 209
Final state vector 38, 91, 100, 204, 205
Forecasting 5, 6, 8, 9, 38, 47, 59, 70, 83,
153, 164, 190, 216
— explanatory variables 84
incremental change 84
manual input 85
model-based 86
- Gain functions 8, 110
- Generalised least squares 181, 183, 186
Goodness of fit 36, 49, 51, 80, 154, 155,
203–205, 207, 208
- Graph
— printing 25
— saving 25
- Help 11
Hessian 193, 201, 205
Heteroskedasticity 80, 155, 203, 212
Stochastic volatility 125
Higher order cycle 6, 63, 106
Hodrick-Prescott filter 128
IN7/BN7 files 10, 21
Innovations 181, 183, 186, 204, 207, 210
Interventions 6, 68, 71, 86, 87, 102, 103,
141, 173, 216
— selection 69, 143
- Kalman filter 7, 53, 125, 181–183, 198, 207
Initialisation 182
Steady-state 182, 184, 185, 203, 207
- Lags 82, 118, 138
Likelihood kernel 194, 196, 202, 203
Likelihood ratio 99, 101
Line search 200
List box
Multiple-selection — 11, 24
Local level model 50, 94, 97, 173, 174, 184
Multivariate 89
Local linear trend 53, 55, 86, 106, 175, 179
Lotus 8, 10, 21
Maximum likelihood estimation 7, 34, 107,
125, 148, 183, 184, 188, 191, 199,
202
Measurement equation 86, 177, 178, 187
Model
Missing values 5, 130
- Nile 16, 69–71, 86, 87
Seat belt monthly — (SEATBELT) 16, 75,
87
Seat belt quarterly — (SEATBQ) 16, 95,
103, 136
Spirits — (SPIRIT) 17, 78, 83, 87
Telephone calls — (TELEPHON) 17
Trade price — (TRADEPRICE, INT) 133
UK consumption, income and prices —
(UKCYP) 17, 57, 89
- Variance matrix structures 89
More written output 38, 57, 153, 156
Multi-step prediction 45, 50
Multivariate 186
— diffuse likelihood 184
— model 3, 5, 88–105, 119, 173, 175,
180, 184, 191, 196
- Newton-Raphson 200
Non-parametric 212
Nonlinear 194, 199, 200
Normality test 190
Numerical optimisation 9, 36, 141, 149, 193,
199, 202, 203
- Outliers 5, 7, 68, 70, 81, 87, 155, 162, 187,
190, 216
- OxMetrics 3, 4, 20, 27, 153, 215
— graphics 23
Modules menu 20
Calculator 25
Graphics 31
Model/Tail Probability 36
OxMetrics Data File (INT7/BN7) 10
- P-value 36
PCGive 19
Periodogram 65, 211
Portmanteau statistic 211
Post-intervention predictive test 76, 87
Post-sample predictive test 48, 212, 213
Prediction 5, 44, 50, 76, 81, 163, 165,
181–183, 185, 204, 207, 208, 210,
217
— graphics 44, 76, 163, 190
Multi-step 45, 164
Prediction error decomposition 181, 183
Prediction error variance 155, 185, 203, 208
Predictive failure test 163
- Q statistic *see* Box-Ljung test
Q-ratio 7, 53, 91, 107, 184, 197, 204
Quasi Newton 193, 195, 200
Quasi-maximum likelihood estimation 125
- Random walk 5, 52, 97, 99, 106, 116, 119,
133, 140, 143, 172, 208
—plus white noise 50, 71
- Rank 94

- Reduced rank 6, 88, 94, 99, 192
 Regression 6, 77, 80, 87, 177, 178, 181, 182,
 — fixed 142
 — time-varying 142
 Residuals 57, 60, 70, 81, 161, 162, 181, 186,
 187, 190, 199, 207, 208, 210–213
 Auxiliary — graphics 71, 162
 Innovations 181, 183, 186, 204, 207, 210
 Results
 — storage 12
 Diagnostic summary report 36
 Estimation report 36
 Root mean square error 17, 45, 52, 181, 183,
 190, 205, 212, 214
 Sample autocorrelations 212
 Sample period 10
 Score vector 187–189
 Seasonal adjustment 59, 127, 128, 160, 187,
 189, 209, 217
 Seasonal component 56, 57, 59, 209
 Fixed 60, 63, 65
 Seasonality 56, 58, 60, 70, 101, 176, 206
 Fixed 56, 101
 Seemingly unrelated time series equations
 (SUTSE) 88, 94, 173, 174, 176,
 180
 Serial correlation 36, 56, 57, 60, 70, 72, 80,
 154, 186, 190, 203, 212
 Signal extraction 9, 40, 53, 134, 153, 159,
 187
 Signal to noise ratio 52
 Smoothing 5, 49, 53, 111, 125, 181,
 187–189, 191, 198, 209
 Component smoothing 189, 217
 Disturbance smoothing 187
 State smoothing 189
 Spectral gain function 160
 Spectrum 5, 90, 161, 211
 Spline 5, 128, 143
 Spurious cycles 129
 STAMP menus
 Intervention 69
 Matrix editor 86
 Model menu 31, 136
 Algebra 31
- Batch 215
 Calculator 25, 31
 Components 33
 Estimate 35, 141
 Formulate 11, 32, 141
 Test 11
 Progress 152
 Test menu 38, 149, 202, 204
 Components graphics 39, 153
 Forecasting 47
 More written output 38, 153
 Prediction graphics 44
 Residuals graphics 42
 Starting STAMP 20
 State space 4, 125, 171, 177, 179, 194,
 ^f
 Also see Structural time series
 model
 Measurement equation 86, 177, 178, 187
 State vector 178–183, 185–187, 189, 190,
 205
 Structural time series model 205, 208
 Transition equation 177, 187
 State vector 178–183, 185–187, 189, 190
 Stationarity 53, 116
 Cycle 60, 62, 63, 65, 66, 87, 90, 95,
 171–174, 176, 179, 180, 191, 192,
 204–207, 210
 Step length 200, 201
 Stochastic volatility 125, 135
 Structural breaks 5, 7, 16, 68, 70, 81, 178,
 187, 216
 Structural time series model 3, 56, 66, 179,
 182, 185, 205, 208
 Basic structural model 34, 57, 88, 141
 Estimation 79, 147, 154, 180, 181, 184,
 187, 191, 193–195, 197, 202, 203
 Explanatory variables 68, 77–213
 Local level 50, 94, 97, 140, 172–174, 184
 Local linear trend 53, 55, 175
 Reduced form 87, 207
 Seasonality 176, 206
 SUTSE
 see Seemingly unrelated time series
 equations
- Butterworth trend 56, 140
 Fixed 60, 63, 65, 101
 Global 67
 Level 50, 55
 Fixed 47
 Stochastic 34, 53
 Local level 50, 56, 140, 172–174, 179,
 184
 Local level with drift 56
 Local linear trend 55, 175
 Quadratic 87
 Slope 50, 55
 Stochastic 34, 53
 Smooth trend 56, 140, 179
 Stochastic 56, 66, 77, 78, 80, 100, 171,
 172
 Trend-cycle decomposition 106
 Tutorial data sets 13
 Unobserved components 4, 5, 77, 140, 173,
 177, 191, 205, 217
 VAR *see* Vector autoregression
 Variable names in batch 217
 Vector autoregression 6, 16, 92, 173, 191,
 193, 196
 Stationarity 53
 Volatility *see* Stochastic volatility
- Weight functions 5, 8, 107, 160
 — graphics 40, 110, 131
 White noise 52, 63, 106, 125, 211
 Window 202
 World Wide Web 13
- X-11 *see* Bureau of the Census X-11
- Transition equation 177, 187
 Trend 50, 56, 171, 172,
 Also see Detrending
 — plus cycle model 64