NAG Toolbox for MATLAB

g13af

1 Purpose

g13af is an easy-to-use version of g13ae. It fits a seasonal autoregressive integrated moving average (ARIMA) model to an observed time series, using a nonlinear least-squares procedure incorporating backforecasting. Parameter estimates are obtained, together with appropriate standard errors. The residual series is returned, and information for use in forecasting the time series is produced for use in g13ag and g13ah.

The estimation procedure is iterative, starting with initial parameter values such as may be obtained using g13ad. It continues until a specified convergence criterion is satisfied or until a specified number of iterations have been carried out. The progress of the iteration can be monitored by means of an optional printing facility.

2 Syntax

```
[par, c, s, ndf, sd, cm, st, nst, itc, isf, res, nres, ifail] =
g13af(mr, par, c, kfc, x, nppc, kpiv, nit, ires, 'npar', npar, 'nx', nx)
```

3 Description

The time series $x_1, x_2, \ldots, x_n$ supplied to the function is assumed to follow a seasonal autoregressive integrated moving average (ARIMA) model defined as follows:

$$\nabla^d \nabla_s^D x_t - c = w_t,$$

where $\nabla^d \nabla_s^D x_t$ is the result of applying non-seasonal differencing of order d and seasonal differencing of seasonality s and order D to the series x_t, as outlined in the description of g13aa. The differenced series is then of length $N = n - d'$, where $d' = d + (D \times s)$ is the generalized order of differencing. The scalar c is the expected value of the differenced series, and the series $w_1, w_2, \ldots, w_N$ follows a zero-mean stationary autoregressive moving average (ARMA) model defined by a pair of recurrence equations. These express w_t in terms of an uncorrelated series a_t, via an intermediate series e_t. The first equation describes the seasonal structure:

$$w_t = \Phi_1 w_{t-s} + \Phi_2 w_{t-2\times s} + \cdots + \Phi_P w_{t-P\times s} + e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2\times s} - \cdots - \Theta_Q e_{t-Q\times s}.$$

The second equation describes the non-seasonal structure. If the model is purely non-seasonal the first equation is redundant and e_t above is equated with w_t:

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \cdots + \phi_p e_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}.$$

Estimates of the model parameters defined by

$$\begin{array}{l}\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \\ \Phi_1, \Phi_2, \ldots, \Phi_P, \Theta_1, \Theta_2, \ldots, \Theta_Q\end{array}$$

and (optionally) c are obtained by minimizing a quadratic form in the vector $w = (w_1, w_2, \ldots, w_N)'$.

The minimization process is iterative, iterations being performed until convergence is achieved (see Section 3 of the document for g13ae for full details), or until the user-specified maximum number of iterations are completed.

The final values of the residual sum of squares and the parameter estimates are used to obtain asymptotic approximations to the standard deviations of the parameters, and the correlation matrix for the parameters. The 'state set' array of information required by forecasting is also returned.

Note: if the maximum number of iterations are performed without convergence, these quantities may not be reliable. In this case, the sequence of iterates should be checked, using the optional monitoring function, to verify that convergence is adequate for practical purposes.

4 References

Box G E P and Jenkins G M 1976 *Time Series Analysis: Forecasting and Control* (Revised Edition) Holden–Day

Marquardt D W 1963 An algorithm for least-squares estimation of nonlinear parameters *J. Soc. Indust. Appl. Math.* **11** 431

5 Parameters

5.1 Compulsory Input Parameters

1: **mr(7) – int32 array**

The orders vector (p, d, q, P, D, Q, s) of the ARIMA model whose parameters are to be estimated. p, q, P and Q refer respectively to the number of autoregressive (ϕ), moving average (θ), seasonal autoregressive (Φ) and seasonal moving average (Θ) parameters. d, D and s refer respectively to the order of non-seasonal differencing, the order of seasonal differencing and the seasonal period.

Constraints:

p, d, q, P, D, Q, $s \geq 0$;
$p + q + P + Q > 0$;
$s \neq 1$;
if $s = 0$, $P + D + Q = 0$;
if $s > 1$, $P + D + Q > 0$;
$d + s \times (P + D) \leq n$;
$p + d - q + s \times (P + D - Q) \leq n$.

2: **par(npar) – double array**

The initial estimates of the p values of the ϕ parameters, the q values of the θ parameters, the P values of the Φ parameters and the Q values of the Θ parameters, in that order.

3: **c – double scalar**

If **kfc** $= 0$, **c** must contain the expected value, c, of the differenced series.

If **kfc** $= 1$, **c** must contain an initial estimate of c.

Therefore, if **c** and **kfc** are both zero on entry, there is no constant correction.

4: **kfc – int32 scalar**

The value of 0 if the constant is to remain fixed, and 1 if it is to be estimated.

Constraint: **kfc** $= 0$ or 1.

5: **x(nx) – double array**

The n values of the original, undifferenced time series.

6: **nppc – int32 scalar**

the number of ϕ, θ, Φ, Θ and c parameters to be estimated. **nppc** $= p + q + P + Q + 1$ if the constant is being estimated and **nppc** $= p + q + P + Q$ if not.

Constraint: **nppc** $=$ **npar** $+$ **kfc**.

7: **kpiv – int32 scalar**

Must be nonzero if the progress of the optimization is to be monitored using the built-in printing facility. Otherwise **kpiv** must contain zero. If selected, monitoring output will be sent to the current advisory message unit defined by x04ab. For each iteration, the heading

```
G13AFZ MONITORING OUTPUT - ITERATION n
```

followed by the parameter values, and residual sum of squares, are printed. In certain implementations, **g13afz** may be renamed as AFZG13.

8: **nit – int32 scalar**

The maximum number of iterations to be performed.

Constraint: $\mathbf{nit} \geq 0$.

9: **ires – int32 scalar**

Constraint: $\mathbf{ires} \geq 15 \times Q' + 11n + 13 \times \mathbf{nppc} + 8 \times P' + 12 + 2 \times (Q' + \mathbf{nppc})^2$, where $P' = p + (P \times s)$ and $Q' = q + (Q \times s)$.

5.2 Optional Input Parameters

1: **npar – int32 scalar**

Default: The dimension of the array **par**.

The total number of ϕ, θ, Φ, and Θ parameters to be estimated.

Constraint: $\mathbf{npar} = p + q + P + Q$.

2: **nx – int32 scalar**

Default: The dimension of the arrays **x**, **st**. (An error is raised if these dimensions are not equal.)

n, the length of the original, undifferenced time series.

5.3 Input Parameters Omitted from the MATLAB Interface

ldcm

5.4 Output Parameters

1: **par(npar) – double array**

Contains the latest values of the estimates of these parameters.

2: **c – double scalar**

If $\mathbf{kfc} = 0$, **c** is unchanged.

If $\mathbf{kfc} = 1$, **c** contains the latest estimate of c.

3: **s – double scalar**

The residual sum of squares after the latest series of parameter estimates has been incorporated into the model. If the function exits with a faulty input parameter, **s** contains zero.

4: **ndf – int32 scalar**

The number of degrees of freedom associated with **s**.

5: **sd(nppc) – double array**

The standard deviations corresponding to the parameters in the model (p autoregressive, q moving average, P seasonal autoregressive, Q seasonal moving average and c, if estimated, in that order). If the function exits with **ifail** containing a value other than 0 or 9, or if the required number of iterations is zero, the contents of **sd** will be indeterminate.

6: **cm(ldcm,nppc) – double array**

The correlation coefficients associated with each pair of the **nppc** parameters. These are held in the first **nppc** rows and the first **nppc** columns of **cm**. These correlation coefficients are indeterminate if **ifail** contains on exit a value other than 0 or 9, or if the required number of iterations is zero.

7: **st(nx) – double array**

The value of the state set in its first **nst** elements. If the function exits with **ifail** containing a value other than 0 or 9, the contents of **st** will be indeterminate.

8: **nst – int32 scalar**

The size of the state set. $\mathbf{nst} = P \times s + D \times s + d + q + \max(p, Q \times s)$.

nst should be used subsequently in g13ag and g13ah as the dimension of **st**.

9: **itc – int32 scalar**

The number of iterations performed.

10: **isf(4) – int32 array**

The first four elements of **isf** contain success/failure indicators, one for each of the four types of parameter in the model (autoregressive, moving average, seasonal autoregressive, seasonal moving average), in that order.

Each indicator has the interpretation:

-2 On entry parameters of this type have initial estimates which do not satisfy the stationarity or invertibility test conditions.

-1 The search procedure has failed to converge because the latest set of parameter estimates of this type is invalid.

0 No parameter of this type is in the model.

1 Valid final estimates for parameters of this type have been obtained.

11: **res(ires) – double array**

The first **nres** elements of **res** contain the model residuals derived from the differenced series. If the function exits with **ifail** holding a value other than 0 or 9, these elements of **res** will be indeterminate. The rest of the array **res** is used as workspace.

12: **nres – int32 scalar**

The number of model residuals returned in **res**.

13: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Note: g13af may return useful information for one or more of the following detected errors or warnings.

ifail = 1

On entry, $\mathbf{npar} \neq p + q + P + Q$,
or the orders vector **mr** is invalid (check the constraints in Section 5),
or $\mathbf{kfc} \neq 0$ or 1,
or $\mathbf{nppc} \neq \mathbf{npar} + \mathbf{kfc}$.

ifail $= 2$

On entry, **nx** $- d - D \times s \leq$ **npar** $+$ **kfc**, i.e., the number of terms in the differenced series is not greater than the number of parameters in the model. The model is over-parameterised.

ifail $= 3$

On entry, **nit** < 0.

ifail $= 4$

On entry, the required size of the state set array **st** is greater than **nx**. This occurs only for very unusual models with long seasonal periods or large numbers of parameters. First check that the orders vector **mr** has been set up as intended. If it has, change to g13ae with **st** dimensioned at least (**nst**), where **nst** is the value returned by g13af, or computed using the formula in Section 5 of this document.

ifail $= 5$

On entry, the workspace array **res** is too small. Check the value of **ires** against the constraints in Section 5.

ifail $= 6$

On entry, **ldcm** $<$ **nppc**.

ifail $= 7$

The search procedure in the algorithm has failed. This may be due to a badly conditioned sum of squares function, or the default convergence criterion may be too strict. Use g13ae with a less strict convergence criterion.

Some output parameters may contain meaningful values; see Section 5 for details.

ifail $= 8$

The inversion of the Hessian matrix in the calculation of the covariance matrix of the parameter estimates has failed.

Some output parameters may contain meaningful values; see Section 5 for details.

ifail $= 9$

This indicates a failure in f03af which is used to solve the equations giving the latest estimates of the backforecasts.

Some output parameters may contain meaningful values; see Section 5 for details.

ifail $= 10$

Satisfactory parameter estimates could not be obtained for all parameter types in the model. Inspect array **isf** for further information on the parameter type(s) in error.

ifail $= 11$

An internal error has arisen in partitioning **res** for use by g13ae. This error should not occur; report it to NAG via your site representative.

7 Accuracy

The computations are believed to be stable.

8 Further Comments

The time taken by g13af is approximately proportional to $\mathbf{nx} \times \mathbf{itc} \times (q + Q \times s + \mathbf{nppc})^2$.

9 Example

```
mr = [int32(1);
     int32(1);
     int32(2);
     int32(0);
     int32(0);
     int32(0);
     int32(0)];
par = [0;
     0;
     0];
c = 0;
kfc = int32(1);
x = [-217;
     -177;
     -166;
     -136;
     -110;
     -95;
     -64;
     -37;
     -14;
     -25;
     -51;
     -62;
     -73;
     -88;
     -113;
     -120;
     -83;
     -33;
     -19;
     21;
     17;
     44;
     44;
     78;
     88;
     122;
     126;
     114;
     85;
     64];
nppc = int32(4);
kpiv = int32(0);
nit = int32(50);
ires = int32(504);
[parOut, cOut, s, ndf, sd, cm, st, nst, itc, isf, res, nres, ifail] = ...
    g13af(mr, par, c, kfc, x, nppc, kpiv, nit, ires)
```

```
parOut =
   -0.0543
   -0.5548
   -0.6734
cOut =
    9.9848
s =
   9.3972e+03
ndf =
          25
sd =
    0.3457
    0.2636
    0.1665
```

```
    7.4170
cm =
    1.0000    0.8072    0.3548   -0.0404
    0.8072    1.0000    0.4681   -0.0491
    0.3548    0.4681    1.0000   -0.0376
   -0.0404   -0.0491   -0.0376    1.0000
st =
   64.0000
  -30.9848
  -20.6573
   -2.2555
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
nst =
           4
itc =
          25
isf =
           1
           1
           0
           0
res =
     array elided
nres =
          29
ifail =
           0
```