NAG Toolbox for MATLAB

g13ba

1 Purpose

g13ba filters a time series by an ARIMA model.

2 Syntax

```
[b, ifail] = g13ba(y, mr, par, cy, nb, 'ny', ny, 'nmr', nmr, 'npar',
npar)
```

3 Description

From a given series $y_1, y_2, \ldots, y_n$, a new series $b_1, b_2, \ldots, b_n$ is calculated using a supplied (filtering) ARIMA model. This model will be one which has previously been fitted to a series x_t with residuals a_t. The equations defining b_t in terms of y_t are very similar to those by which a_t is obtained from x_t. The only dissimilarity is that no constant correction is applied after differencing. This is because the series y_t is generally distinct from the series x_t with which the model is associated, though y_t may be related to x_t. Whilst it is appropriate to apply the ARIMA model to y_t so as to preserve the same relationship between b_t and a_t as exists between y_t and x_t, the constant term in the ARIMA model is inappropriate for y_t. The consequence is that b_t will not necessarily have zero mean.

The equations are precisely:

$$w_t = \nabla^d \nabla_s^D y_t, \tag{1}$$

the appropriate differencing of y_t; both the seasonal and non-seasonal inverted autoregressive operations are then applied,

$$u_t = w_t - \Phi_1 w_{t-s} - \cdots - \Phi_P w_{t-s\times P} \tag{2}$$

$$v_t = u_t - \phi_1 u_{t-1} - \cdots - \phi_p u_{t-p} \tag{3}$$

followed by the inverted moving average operations

$$z_t = v_t + \Theta_1 z_{t-s} + \cdots + \Theta_Q z_{t-s\times Q} \tag{4}$$

$$b_t = z_t + \theta_1 b_{t-1} + \cdots + \theta_q b_{t-q}. \tag{5}$$

Because the filtered series value b_t depends on present and past values $y_t, y_{t-1}, \ldots$, there is a problem arising from ignorance of $y_0, y_{-1}, \ldots$ which particularly affects calculation of the early values $b_1, b_2, \ldots$, causing 'transient errors'. The function allows two possibilities.

(i) The equations (1), (2) and (3) are applied from successively later time points so that all terms on their right-hand sides are known, with v_t being defined for $t = (1 + d + s \times D + s \times P), \ldots, n$. Equations (4) and (5) are then applied over the same range, taking any values on the right-hand side associated with previous time points to be zero.

This procedure may still however result in unacceptably large transient errors in early values of b_t.

(ii) The unknown values $y_0, y_{-1}, \ldots$ are estimated by backforecasting. This requires that an ARIMA model distinct from that which has been supplied for filtering, should have been previously fitted to y_t.

For efficiency, you are asked to supply both this ARIMA model for y_t and a limited number of backforecasts which are prefixed to the known values of y_t. Within the function further backforecasts of y_t, and the series w_t, u_t, v_t in (1), (2) and (3) are then easily calculated, and a set of linear equations solved for backforecasts of z_t, b_t for use in (4) and (5) in the case that $q + Q > 0$.

Even if the best model for y_t is not available, a very approximate guess such as

$$y_t = c + e_t$$

or

$$\nabla y_t = e_t$$

can help to reduce the transients substantially.

The backforecasts which need to be prefixed to y_t are of length $Q'_y = q_y + s_y \times Q_y$, where q_y and Q_y are the non-seasonal and seasonal moving average orders and s_y the seasonal period for the ARIMA model of y_t. Thus you need not carry out the backforecasting exercise if $Q'_y = 0$. Otherwise, the series $y_1, y_2, \ldots, y_n$ should be reversed to obtain $y_n, y_{n-1}, \ldots, y_1$ and g13aj should be used to forecast Q'_y values, $\hat{y}_0, \ldots, \hat{y}_{1-Q'_y}$. The ARIMA model used is that fitted to y_t (as a forward series) except that, if $d_y + D_y$ is odd, the constant should be changed in sign (to allow, for example, for the fact that a forward upward trend is a reversed downward trend). The ARIMA model for y_t supplied to the filtering function must however have the appropriate constant for the forward series.

The series $\hat{y}_{1-Q'_y}, \ldots, \hat{y}_0, y_1, \ldots, y_n$ is then supplied to the function, and a corresponding set of values returned for b_t.

4 References

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

5 Parameters

5.1 Compulsory Input Parameters

1: **y(ny) – double array**

The Q'_y backforecasts, starting with backforecast at time $1 - Q'_y$ to backforecast at time 0, followed by the time series starting at time 1, where $Q'_y = \mathbf{mr}(10) + \mathbf{mr}(13) \times \mathbf{mr}(14)$. If there are no backforecasts, either because the ARIMA model for the time series is not known, or because it is known but has no moving average terms, then the time series starts at the beginning of **y**.

2: **mr(nmr) – int32 array**

The orders vector for the filtering model, followed by the orders vector for the ARIMA model for the time series if the latter is known. The orders appear in the standard sequence (p, d, q, P, D, Q, s) as given in the G13 Chapter Introduction. If the ARIMA model for the time series is supplied, then the function will assume that the first Q'_y values of the array **y** are backforecasts.

Constraints:

$\mathbf{mr}(1) + \mathbf{mr}(3) + \mathbf{mr}(4) + \mathbf{mr}(6) > 0$;
$\mathbf{mr}(k) \geq 0$, for $k = 1, 2, \ldots, 7$;
if $\mathbf{mr}(7) = 0$, $\mathbf{mr}(4) + \mathbf{mr}(5) + \mathbf{mr}(6) = 0$;
if $\mathbf{mr}(7) \neq 0$, $\mathbf{mr}(4) + \mathbf{mr}(5) + \mathbf{mr}(6) \neq 0$;
$\mathbf{mr}(7) \neq 1$;
$\mathbf{mr}(k) \geq 0$, for $k = 8, 9, \ldots, 14$;
if $\mathbf{mr}(14) = 0$, $\mathbf{mr}(11) + \mathbf{mr}(12) + \mathbf{mr}(13) = 0$;
if $\mathbf{mr}(14) \neq 0$, $\mathbf{mr}(11) + \mathbf{mr}(12) + \mathbf{mr}(13) \neq 0$;
$\mathbf{mr}(14) \neq 1$.

3: **par(npar) – double array**

The parameters of the filtering model, followed by the parameters of the ARIMA model for the time series, if supplied. Within each model the parameters are in the standard order of non-seasonal AR and MA followed by seasonal AR and MA.

4: **cy – double scalar**

If the ARIMA model is known (i.e., **nmr** $= 14$), **cy** must specify the constant term of the ARIMA model for the time series. If this model is not known (i.e., **nmr** $= 7$), then **cy** is not used.

5: **nb – int32 scalar**

In addition to holding the returned filtered series, **b** is also used as an intermediate work array if the ARIMA model for the time series was known.

Constraints:

if **nmr** $= 14$, $\mathbf{nb} \geq \mathbf{ny} + \max(K_3, K_1 + K_2)$;
if **nmr** $= 7$, $\mathbf{nb} \geq \mathbf{ny}$.

Where

$K_1 = \mathbf{mr}(1) + \mathbf{mr}(4) \times \mathbf{mr}(7)$;
$K_2 = \mathbf{mr}(2) + \mathbf{mr}(5) \times \mathbf{mr}(7)$;
$K_3 = \mathbf{mr}(3) + \mathbf{mr}(6) \times \mathbf{mr}(7)$.

5.2 Optional Input Parameters

1: **ny – int32 scalar**

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

the total number of backforecasts and time series data points in array **y**.

Constraint: $\mathbf{ny} \geq \max\left(1 + Q_y', \mathbf{npar}\right)$.

2: **nmr – int32 scalar**

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

the number of values specified in the array **mr**. It takes the value 7 if no ARIMA model for the time series is supplied but otherwise it takes the value 14. Thus **nmr** acts as an indicator as to whether backforecasting can be carried out.

Constraint: **nmr** $= 7$ or 14.

3: **npar – int32 scalar**

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

The total number of parameters held in array **par**.

Constraints:

if **nmr** $= 7$, $\mathbf{npar} = \mathbf{mr}(1) + \mathbf{mr}(3) + \mathbf{mr}(4) + \mathbf{mr}(6)$;
if **nmr** $= 14$, $\mathbf{npar} = \mathbf{mr}(1) + \mathbf{mr}(3) + \mathbf{mr}(4) + \mathbf{mr}(6) +$
$\mathbf{mr}(8) + \mathbf{mr}(10) + \mathbf{mr}(11) + \mathbf{mr}(13)$.

Note: the first constraint (i.e., $\mathbf{mr}(1) + \mathbf{mr}(3) + \mathbf{mr}(4) + \mathbf{mr}(6) > 0$) on the orders of the filtering model, in parameter **mr**, ensures that $\mathbf{npar} > 0$.

5.3 Input Parameters Omitted from the MATLAB Interface

wa, nwa

5.4 Output Parameters

1: **b(nb) – double array**

The filtered output series. If the ARIMA model for the time series was known, and hence Q_y' backforecasts were supplied in **y**, then **b** contains Q_y' 'filtered' backforecasts followed by the filtered series. Otherwise, the filtered series begins at the start of **b** just as the original series began at the start of **y**. In either case, if the value of the series at time t is held in $\mathbf{y}(t)$, then the filtered value at time t is held in $\mathbf{b}(t)$.

2: **ifail – int32 scalar**

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

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, **nmr** $\neq 7$ and **nmr** $\neq 14$.

ifail $= 2$

On entry, the orders vector **mr** does not satisfy the constraints given in Section 5.

ifail $= 3$

On entry, **npar** is inconsistent with the contents of **mr** (see Section 5).

ifail $= 4$

On entry, **ny** is too small to successfully carry out the requested filtering, (see Section 5).

ifail $= 5$

On entry, the work array **wa** is too small.

ifail $= 6$

On entry, the array **b** is too small.

ifail $= 7$

The orders vector for the filtering model is invalid.

ifail $= 8$

The orders vector for the ARIMA model is invalid. (Only occurs if **nmr** $= 14$.)

ifail $= 9$

The initial values of the filtered series are indeterminate for the given models.

ifail $= -999$

Internal memory allocation failed.

7 Accuracy

Accuracy and stability are high except when the MA parameters are close to the invertibility boundary.

8 Further Comments

If an ARIMA model is supplied, a local workspace array of fixed length is allocated internally by g13ba. The total size of this array amounts to K integer elements, where K is the expression defined in the description of the parameter **wa**.

The time taken by g13ba is approximately proportional to

$$\mathbf{ny} \times (\mathbf{mr}(1) + \mathbf{mr}(3) + \mathbf{mr}(4) + \mathbf{mr}(6)),$$

with an appreciable fixed increase if an ARIMA model is supplied for the time series.

9 Example

```
y = [49.98070870808541;
     52.67136711304528;
     53.8;
     53.6;
     53.5;
     53.5;
     53.4;
     53.1;
     52.7;
     52.4;
     52.2;
     52;
     52;
     52.4;
     53;
     54;
     54.9;
     56;
     56.8;
     56.8;
     56.4;
     55.7;
     55;
     54.3;
     53.2;
     52.3;
     51.6;
     51.2;
     50.8;
     50.5;
     50;
     49.2;
     48.4;
     47.9;
     47.6;
     47.5;
     47.5;
     47.6;
     48.1;
     49;
     50;
     51.1;
     51.8;
     51.9;
     51.7;
     51.2;
     50;
     48.3;
     47;
     45.8;
     45.6;
     46;
     46.9;
     47.8;
     48.2;
     48.3;
     47.9;
     47.2;
     47.2;
```

```
      48.1;
      49.4;
      50.6;
      51.5;
      51.6;
      51.2;
      50.5;
      50.1;
      49.8;
      49.6;
      49.4;
      49.3;
      49.2;
      49.3;
      49.7;
      50.3;
      51.3;
      52.8;
      54.4;
      56;
      56.9;
      57.5;
      57.3;
      56.6;
      56;
      55.4;
      55.4;
      56.4;
      57.2;
      58;
      58.4;
      58.4;
      58.1;
      57.7;
      57;
      56;
      54.7;
      53.2;
      52.1;
      51.6;
      51;
      50.5;
      50.4;
      51;
      51.8;
      52.4;
      53;
      53.4;
      53.6;
      53.7;
      53.8;
      53.8;
      53.8;
      53.3;
      53;
      52.9;
      53.4;
      54.6;
      56.4;
      58;
      59.4;
      60.2;
      60;
      59.4;
      58.4;
      57.6;
      56.9;
      56.4;
      56;
      55.7;
```

```
      55.3;
      55;
      54.4;
      53.7;
      52.8;
      51.6;
      50.6;
      49.4;
      48.8;
      48.5;
      48.7;
      49.2;
      49.8;
      50.4;
      50.7;
      50.9;
      50.7;
      50.5;
      50.4;
      50.2;
      50.4;
      51.2;
      52.3;
      53.2;
      53.9;
      54.1;
      54;
      53.6;
      53.2;
      53;
      52.8;
      52.3;
      51.9;
      51.6;
      51.6;
      51.4;
      51.2;
      50.7;
      50;
      49.4;
      49.3;
      49.7;
      50.6;
      51.8;
      53;
      54;
      55.3;
      55.9;
      55.9;
      54.6;
      53.5;
      52.4;
      52.1;
      52.3;
      53;
      53.8;
      54.6;
      55.4;
      55.9;
      55.9;
      55.2;
      54.4;
      53.7;
      53.6;
      53.6;
      53.2;
      52.5;
      52;
      51.4;
      51;
```

```
      50.9;
      52.4;
      53.5;
      55.6;
      58;
      59.5;
      60;
      60.4;
      60.5;
      60.2;
      59.7;
      59;
      57.6;
      56.4;
      55.2;
      54.5;
      54.1;
      54.1;
      54.4;
      55.5;
      56.2;
      57;
      57.3;
      57.4;
      57;
      56.4;
      55.9;
      55.5;
      55.3;
      55.2;
      55.4;
      56;
      56.5;
      57.1;
      57.3;
      56.8;
      55.6;
      55;
      54.1;
      54.3;
      55.3;
      56.4;
      57.2;
      57.8;
      58.3;
      58.6;
      58.8;
      58.8;
      58.6;
      58;
      57.4;
      57;
      56.4;
      56.3;
      56.4;
      56.4;
      56;
      55.2;
      54;
      53;
      52;
      51.6;
      51.6;
      51.1;
      50.4;
      50;
      50;
      52;
      54;
      55.1;
```

```
      54.5;
      52.8;
      51.4;
      50.8;
      51.2;
      52;
      52.8;
      53.8;
      54.5;
      54.9;
      54.9;
      54.8;
      54.4;
      53.7;
      53.3;
      52.8;
      52.6;
      52.6;
      53;
      54.3;
      56;
      57;
      58;
      58.6;
      58.5;
      58.3;
      57.8;
      57.3;
      57];
mr = [int32(3);
      int32(0);
      int32(0);
      int32(0);
      int32(0);
      int32(0);
      int32(0);
      int32(4);
      int32(0);
      int32(2);
      int32(0);
      int32(0);
      int32(0);
      int32(0)];
par = [1.97;
      -1.37;
      0.34;
      2.42;
      -2.38;
      1.16;
      -0.23;
      0.31;
      -0.47];
cy = 0;
nb = int32(301);
[b, ifail] = g13ba(y, mr, par, cy, nb)
```

```
b =
      array elided
ifail =
            0
```