NAG Library Routine Document

G01EWF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G01EWF returns the probability associated with the lower tail of the distribution for the Dickey–Fuller unit root test statistic.

2 Specification

```
FUNCTION G01EWF (METHOD, TYPE, N, TS, NSAMP, STATE, IFAIL)
REAL (KIND=nag_wp) G01EWF

INTEGER            METHOD, TYPE, N, NSAMP, STATE(*), IFAIL
REAL (KIND=nag_wp) TS
```

3 Description

If the root of the characteristic equation for a time series is one then that series is said to have a unit root. Such series are nonstationary. G01EWF is designed to be called after G13AWF and returns the probability associated with one of three types of (augmented) Dickey–Fuller test statistic: τ, τ_μ or τ_τ, used to test for a unit root, a unit root with drift or a unit root with drift and a deterministic time trend, respectively. The three types of test statistic are constructed as follows:

1. To test whether a time series, y_t, for $t = 1, 2, \ldots, n$, has a unit root the regression model

$$\nabla y_t = \beta_1 y_{t-1} + \sum_{i=1}^{p-1} \delta_i \nabla y_{t-i} + \epsilon_t$$

is fit and the test statistic τ constructed as

$$\tau = \frac{\hat{\beta}_1}{\sigma_{11}}$$

where ∇ is the difference operator, with $\nabla y_t = y_t - y_{t-1}$, and where $\hat{\beta}_1$ and σ_{11} are the least squares estimate and associated standard error for β_1 respectively.

2. To test for a unit root with drift the regression model

$$\nabla y_t = \beta_1 y_{t-1} + \sum_{i=1}^{p-1} \delta_i \nabla y_{t-i} + \alpha + \epsilon_t$$

is fit and the test statistic τ_μ constructed as

$$\tau_\mu = \frac{\hat{\beta}_1}{\sigma_{11}}.$$

3. To test for a unit root with drift and deterministic time trend the regression model

$$\nabla y_t = \beta_1 y_{t-1} + \sum_{i=1}^{p-1} \delta_i \nabla y_{t-i} + \alpha + \beta_2 t + \epsilon_t$$

is fit and the test statistic τ_τ constructed as

$$\tau_\tau = \frac{\hat{\beta}_1}{\sigma_{11}}.$$

All three test statistics: τ, τ_μ and τ_τ can be calculated using G13AWF.

The probability distributions of these statistics are nonstandard and are a function of the length of the series of interest, n. The probability associated with a given test statistic, for a given n, can therefore only be calculated by simulation as described in Dickey and Fuller (1979). However, such simulations require a significant number of iterations and are therefore prohibitively expensive in terms of the time taken. As such G01EWF also allows the probability to be interpolated from a look-up table. Two such tables are provided, one from Dickey (1976) and one constructed as described in Section 9. The three different methods of obtaining an estimate of the probability can be chosen via the METHOD parameter. Unless there is a specific reason for choosing otherwise, METHOD $= 1$ should be used.

4 References

Dickey A D (1976) Estimation and hypothesis testing in nonstationary time series *PhD Thesis* Iowa State University, Ames, Iowa

Dickey A D and Fuller W A (1979) Distribution of the estimators for autoregressive time series with a unit root *J. Am. Stat. Assoc.* **74 366** 427–431

5 Parameters

1: METHOD – INTEGER *Input*

On entry: the method used to calculate the probability.

METHOD $= 1$
The probability is interpolated from a look-up table, whose values were obtained via simulation.

METHOD $= 2$
The probability is interpolated from a look-up table, whose values were obtained from Dickey (1976).

METHOD $= 3$
The probability is obtained via simulation.

The probability calculated from the look-up table should give sufficient accuracy for most applications.

Suggested value: METHOD $= 1$.

Constraint: METHOD $= 1$, 2 or 3.

2: TYPE – INTEGER *Input*

On entry: the type of test statistic, supplied in TS.

Constraint: TYPE $= 1$, 2 or 3.

3: N – INTEGER *Input*

On entry: n, the length of the time series used to calculate the test statistic.

Constraints:

if METHOD $\neq 3$, N > 0;
if METHOD $= 3$ and TYPE $= 1$, N > 2;
if METHOD $= 3$ and TYPE $= 2$, N > 3;
if METHOD $= 3$ and TYPE $= 3$, N > 4.

4: TS – REAL (KIND=nag_wp) *Input*

On entry: the Dickey–Fuller test statistic for which the probability is required. If

TYPE $= 1$
TS must contain τ.

TYPE = 2
TS must contain τ_μ.

TYPE = 3
TS must contain τ_τ.

If the test statistic was calculated using G13AWF the value of TYPE and N must not change between calls to G01EWF and G13AWF.

5: NSAMP – INTEGER *Input*

On entry: if METHOD = 3, the number of samples used in the simulation; otherwise NSAMP is not referenced and need not be set.

Constraint: if METHOD = 3, NSAMP > 0.

6: STATE(*) – INTEGER array *Communication Array*

Note: the actual argument supplied **must** be the array STATE supplied to the initialization routines G05KFF or G05KGF.

On entry: if METHOD = 3, STATE must contain information on the selected base generator and its current state; otherwise STATE is not referenced and need not be set.

On exit: if METHOD = 3, STATE contains updated information on the state of the generator otherwise a zero length vector is returned.

7: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, −1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

For environments where it might be inappropriate to halt program execution when an error is detected, the value −1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this parameter, the recommended value is 0. **When the value −1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: IFAIL = 0 unless the routine detects an error or a warning has been flagged (see Section 6).

6 Error Indicators and Warnings

If on entry IFAIL = 0 or −1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL = 11

On entry, METHOD = ⟨*value*⟩.
Constraint: METHOD = 1, 2 or 3.

IFAIL = 21

On entry, TYPE = ⟨*value*⟩.
Constraint: TYPE = 1, 2 or 3.

IFAIL = 31

On entry, N = ⟨*value*⟩.
Constraint: if METHOD ≠ 3, N > 0.

On entry, N = ⟨*value*⟩.
Constraint: if METHOD = 3 and TYPE = 1, N > 2.

On entry, N = ⟨*value*⟩.
Constraint: if METHOD = 3 and TYPE = 2, N > 3.

On entry, N = ⟨*value*⟩.
Constraint: if METHOD = 3 and TYPE = 3, N > 4.

IFAIL = 51

On entry, NSAMP = ⟨*value*⟩.
Constraint: if METHOD = 3, NSAMP > 0.

IFAIL = 61

On entry, METHOD = 3 and the STATE vector has been corrupted or not initialized.

IFAIL = 201

The supplied input values were outside the range of at least one look-up table, therefore extrapolation was used.

IFAIL = −99

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.8 in the Essential Introduction for further information.

IFAIL = −399

Your licence key may have expired or may not have been installed correctly.

See Section 3.7 in the Essential Introduction for further information.

IFAIL = −999

Dynamic memory allocation failed.

See Section 3.6 in the Essential Introduction for further information.

7 Accuracy

When METHOD = 1, the probability returned by this routine is unlikely to be accurate to more than 4 or 5 decimal places, for METHOD = 2 this accuracy is likely to drop to 2 or 3 decimal places (see Section 9 for details on how these probabilities are constructed). In both cases the accuracy of the probability is likely to be lower when extrapolation is used, particularly for small values of N (less than around 15). When METHOD = 3 the accuracy of the returned probability is controlled by the number of simulations performed (i.e., the value of NSAMP used).

8 Parallelism and Performance

Not applicable.

9 Further Comments

When METHOD = 1 or 2 the probability returned is constructed by interpolating from a series of look-up tables. In the case of METHOD = 2 the look-up tables are taken directly from Dickey (1976) and the interpolation is carried out using E01SAF and E01SBF . For METHOD = 1 the look-up tables were constructed as follows:

(i) A sample size, n was chosen.

(ii) 2^{28} simulations were run.

(iii) At each simulation, a time series was constructed as described in chapter five of Dickey (1976). The relevant test statistic was then calculated for each of these time series.

(iv) A series of quantiles were calculated from the sample of 2^{28} test statistics. The quantiles were calculated at intervals of 0.0005 between 0.0005 and 0.9995.

(v) A spline was fit to the quantiles using E02BEF.

This process was repeated for $\mathrm{n} = 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 5000, 10000$, resulting in 22 splines.

Given the 22 splines, and a user-supplied sample size, n and test statistic, τ, an estimated p-value is calculated as follows:

(i) Evaluate each of the 22 splines, at τ, using E02BEF. If, for a particular spline, the supplied value of τ lies outside of the range of the simulated data, then a third-order Taylor expansion is used to extrapolate, with the derivatives being calculated using E02BCF.

(ii) Fit a spline through these 22 points using E01BEF.

(iii) Estimate the p-value using E01BFF.

10 Example

See Section 10 in G13AWF.
