

dfuller — Augmented Dickey–Fuller unit-root test

Description

`dfuller` performs the augmented Dickey–Fuller test that a variable follows a unit-root process. The null hypothesis is that the variable contains a unit root, and the alternative is that the variable was generated by a stationary process. You may optionally exclude the constant, include a trend term, and include lagged values of the difference of the variable in the regression.

Quick start

Augmented Dickey–Fuller test for presence of a unit root in `y` using `tsset` data

```
dfuller y
```

As above, but with a trend term

```
dfuller y, trend
```

Augmented Dickey–Fuller test for presence of a unit root in `y` with a drift term

```
dfuller y, drift
```

As above, but include 3 lagged differences and display the regression table

```
dfuller y, drift lags(3) regress
```

Menu

Statistics > Time series > Tests > Augmented Dickey-Fuller unit-root test

Syntax

`dfuller` *varname* [*if*] [*in*] [, *options*]

options	Description
Main	
`noconstant`	suppress constant term in regression
`trend`	include trend term in regression
`drift`	include drift term in regression
`regress`	display regression table
`lags(#)`	include # lagged differences

You must `tsset` your data before using `dfuller`; see [TS] **tsset**.
varname may contain time-series operators; see **[U] 11.4.4 Time-series varlists**.

Options

Main

`noconstant` suppresses the constant term (intercept) in the model and indicates that the process under the null hypothesis is a random walk without drift. `noconstant` cannot be used with the `trend` or `drift` option.

`trend` specifies that a trend term be included in the associated regression and that the process under the null hypothesis is a random walk, perhaps with drift. This option may not be used with the `noconstant` or `drift` option.

`drift` indicates that the process under the null hypothesis is a random walk with nonzero drift. This option may not be used with the `noconstant` or `trend` option.

`regress` specifies that the associated regression table appear in the output. By default, the regression table is not produced.

`lags(#)` specifies the number of lagged difference terms to include in the covariate list.

Remarks and examples

stata.com

Dickey and Fuller (1979) developed a procedure for testing whether a variable has a unit root or, equivalently, that the variable follows a random walk. Hamilton (1994, 528–529) describes the four different cases to which the augmented Dickey–Fuller test can be applied. The null hypothesis is always that the variable has a unit root. They differ in whether the null hypothesis includes a drift term and whether the regression used to obtain the test statistic includes a constant term and time trend. Becketti (2013, chap. 9) provides additional examples showing how to conduct these tests.

The true model is assumed to be

$$y_t = \alpha + y_{t-1} + u_t$$

where u_t is an independent and identically distributed zero-mean error term. In cases one and two, presumably $\alpha = 0$, which is a random walk without drift. In cases three and four, we allow for a drift term by letting α be unrestricted.

The Dickey–Fuller test involves fitting the model

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t$$

by ordinary least squares (OLS), perhaps setting $\alpha = 0$ or $\delta = 0$. However, such a regression is likely to be plagued by serial correlation. To control for that, the augmented Dickey–Fuller test instead fits a model of the form

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_k \Delta y_{t-k} + \epsilon_t \tag{1}$$

where k is the number of lags specified in the `lags()` option. The `noconstant` option removes the constant term α from this regression, and the `trend` option includes the time trend δt, which by default is not included. Testing $\beta = 0$ is equivalent to testing $\rho = 1$, or, equivalently, that y_t follows a unit root process.

In the first case, the null hypothesis is that y_t follows a random walk without drift, and (1) is fit without the constant term α and the time trend δt. The second case has the same null hypothesis as the first, except that we include α in the regression. In both cases, the population value of α is zero under the null hypothesis. In the third case, we hypothesize that y_t follows a unit root with drift, so that the population value of α is nonzero; we do not include the time trend in the regression. Finally, in the fourth case, the null hypothesis is that y_t follows a unit root with or without drift so that α is unrestricted, and we include a time trend in the regression.

The following table summarizes the four cases.

Case	Process under null hypothesis	Regression restrictions	`dfuller` option
1	Random walk without drift	$\alpha = 0$, $\delta = 0$	`noconstant`
2	Random walk without drift	$\delta = 0$	(default)
3	Random walk with drift	$\delta = 0$	`drift`
4	Random walk with or without drift	(none)	`trend`

Except in the third case, the t-statistic used to test H_0: $\beta = 0$ does not have a standard distribution. Hamilton (1994, chap. 17) derives the limiting distributions, which are different for each of the three other cases. The critical values reported by `dfuller` are interpolated based on the tables in Fuller (1996). MacKinnon (1994) shows how to approximate the p-values on the basis of a regression surface, and `dfuller` also reports that p-value. In the third case, where the regression includes a constant term and under the null hypothesis the series has a nonzero drift parameter α, the t statistic has the usual t distribution; `dfuller` reports the one-sided critical values and p-value for the test of H_0 against the alternative H_a: $\beta < 0$, which is equivalent to $\rho < 1$.

Deciding which case to use involves a combination of theory and visual inspection of the data. If economic theory favors a particular null hypothesis, the appropriate case can be chosen based on that. If a graph of the data shows an upward trend over time, then case four may be preferred. If the data do not show a trend but do have a nonzero mean, then case two would be a valid alternative.

▷ Example 1

In this example, we examine the international airline passengers dataset from Box, Jenkins, and Reinsel (2008, Series G). This dataset has 144 observations on the monthly number of international airline passengers from 1949 through 1960. Because the data show a clear upward trend, we use the `trend` option with `dfuller` to include a constant and time trend in the augmented Dickey–Fuller regression.

```
. use http://www.stata-press.com/data/r15/air2
(TIMESLAB: Airline passengers)

. dfuller air, lags(3) trend regress

Augmented Dickey-Fuller test for unit root         Number of obs   =       140

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -6.936            -4.027            -3.445            -3.145
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
D.air        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         air |
         L1. |  -.5217089   .0752195    -6.94   0.000      -.67048   -.3729379
         LD. |   .5572871   .0799894     6.97   0.000      .399082    .7154923
        L2D. |    .095912   .0876692     1.09   0.276    -.0774825    .2693065
        L3D. |     .14511   .0879922     1.65   0.101    -.0289232    .3191433
      _trend |   1.407534   .2098378     6.71   0.000     .9925118    1.822557
       _cons |   44.49164    7.78335     5.72   0.000     29.09753    59.88575
------------------------------------------------------------------------------
```

Here we can overwhelmingly reject the null hypothesis of a unit root at all common significance levels. From the regression output, the estimated β of -0.522 implies that $\rho = (1 - 0.522) = 0.478$. Experiments with fewer or more lags in the augmented regression yield the same conclusion.

◁

▷ Example 2

In this example, we use the German macroeconomic dataset to determine whether the log of consumption follows a unit root. We will again use the `trend` option, because consumption grows over time.

```
. use http://www.stata-press.com/data/r15/lutkepohl2
(Quarterly SA West German macro data, Bil DM, from Lutkepohl 1993 Table E.1)
. tsset qtr
        time variable:  qtr, 1960q1 to 1982q4
                delta:  1 quarter
. dfuller ln_consump, lags(4) trend
Augmented Dickey-Fuller test for unit root         Number of obs   =        87
                                 ---------- Interpolated Dickey-Fuller ---------
                  Test          1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
-------------------------------------------------------------------------------
 Z(t)             -1.318            -4.069            -3.463            -3.158
-------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8834
```

As we might expect from economic theory, here we cannot reject the null hypothesis that log consumption exhibits a unit root. Again using different numbers of lag terms yield the same conclusion.

◁

Stored results

`dfuller` stores the following in `r()`:

Scalars

`r(N)`	number of observations
`r(lags)`	number of lagged differences
`r(Zt)`	Dickey–Fuller test statistic
`r(cv1)`	1% critical value
`r(cv5)`	5% critical value
`r(cv10)`	10% critical value
`r(p)`	MacKinnon approximate p-value (if there is a constant or trend in associated regression)

Methods and formulas

In the OLS estimation of an AR(1) process with Gaussian errors,

$$y_t = \rho y_{t-1} + \epsilon_t$$

where ϵ_t are independent and identically distributed as $N(0, \sigma^2)$ and $y_0 = 0$, the OLS estimate (based on an n-observation time series) of the autocorrelation parameter ρ is given by

$$\widehat{\rho}_n = \frac{\displaystyle\sum_{t=1}^{n} y_{t-1} y_t}{\displaystyle\sum_{t=1}^{n} y_t^2}$$

If $|\rho| < 1$, then

$$\sqrt{n}(\widehat{\rho}_n - \rho) \rightarrow N(0, 1 - \rho^2)$$

If this result were valid when $\rho = 1$, the resulting distribution would have a variance of zero. When $\rho = 1$, the OLS estimate $\widehat{\rho}$ still converges in probability to one, though we need to find a suitable nondegenerate distribution so that we can perform hypothesis tests of H_0: $\rho = 1$. Hamilton (1994, chap. 17) provides a superb exposition of the requisite theory.

To compute the test statistics, we fit the augmented Dickey–Fuller regression

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \sum_{j=1}^{k} \zeta_j \Delta y_{t-j} + e_t$$

via OLS where, depending on the options specified, the constant term α or time trend δt is omitted and k is the number of lags specified in the `lags()` option. The test statistic for H_0: $\beta = 0$ is $Z_t = \widehat{\beta}/\widehat{\sigma}_\beta$, where $\widehat{\sigma}_\beta$ is the standard error of $\widehat{\beta}$.

The critical values included in the output are linearly interpolated from the table of values that appears in Fuller (1996), and the MacKinnon approximate p-values use the regression surface published in MacKinnon (1994).

David Alan Dickey (1945–) was born in Ohio and obtained degrees in mathematics at Miami University and a PhD in statistics at Iowa State University in 1976 as a student of Wayne Fuller. He works at North Carolina State University and specializes in time-series analysis.

Wayne Arthur Fuller (1931–) was born in Iowa, obtained three degrees at Iowa State University and then served on the faculty between 1959 and 2001. He has made many distinguished contributions to time series, measurement-error models, survey sampling, and econometrics.

References

Becketti, S. 2013. *Introduction to Time Series Using Stata*. College Station, TX: Stata Press.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control*. 4th ed. Hoboken, NJ: Wiley.

del Barrio Castro, T., A. Bodnar, and A. Sansó. 2016. The lag-length selection and detrending methods for HEGY seasonal unit-root tests using Stata. *Stata Journal* 16: 740–760.

Dickey, D. A., and W. A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431.

Fuller, W. A. 1996. *Introduction to Statistical Time Series*. 2nd ed. New York: Wiley.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

MacKinnon, J. G. 1994. Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics* 12: 167–176.

Newton, H. J. 1988. *TIMESLAB: A Time Series Analysis Laboratory*. Belmont, CA: Wadsworth.

Otero, J., and C. F. Baum. 2017. Response surface models for the Elliott, Rothenberg, and Stock unit-root test. *Stata Journal* 17: 985–1002.

Otero, J., and J. Smith. 2017. Response surface models for OLS and GLS detrending-based unit-root tests in nonlinear ESTAR models. *Stata Journal* 17: 704–722.

Rajbhandari, A. 2016. Unit-root tests in Stata. *The Stata Blog: Not Elsewhere Classified*. http://blog.stata.com/2016/06/21/unit-root-tests-in-stata/.

Also see

[TS] **tsset** — Declare data to be time-series data

[TS] **dfgls** — DF-GLS unit-root test

[TS] **pperron** — Phillips–Perron unit-root test

[XT] **xtunitroot** — Panel-data unit-root tests