Response surface models for the Elliott, Rothenberg, Stock DF-GLS unit-root test

Christopher F Baum   Jesús Otero

Stata Conference, Baltimore, July 2017

Introduction

- The importance of testing for unit roots in economic time series dates back to the concept of spurious regressions developed by Granger and Newbold (J.Metrics., 1974) and the findings of Nelson and Plosser (J.Mon.Ec., 1982) for a large set of macroeconomic series.
- The Said–Dickey (Biometrika, 1984) "Augmented Dickey–Fuller test" has been widely used (cf. Stata's `dfuller`) as well as other 'first-generation' alternatives such as the Phillips–Perron test (Stata's `pperron`; Biometrika, 1988).
- These 'first-generation' alternatives, with a null hypothesis of $I(1)$, or a unit root, are known to have low power, particularly in smaller samples.

Introduction

- The importance of testing for unit roots in economic time series dates back to the concept of spurious regressions developed by Granger and Newbold (J.Metrics., 1974) and the findings of Nelson and Plosser (J.Mon.Ec., 1982) for a large set of macroeconomic series.
- The Said–Dickey (Biometrika, 1984) "Augmented Dickey–Fuller test" has been widely used (cf. Stata's `dfuller`) as well as other 'first-generation' alternatives such as the Phillips–Perron test (Stata's `pperron`; Biometrika, 1988).
- These 'first-generation' alternatives, with a null hypothesis of $I(1)$, or a unit root, are known to have low power, particularly in smaller samples.

Introduction

- The importance of testing for unit roots in economic time series dates back to the concept of spurious regressions developed by Granger and Newbold (J.Metrics., 1974) and the findings of Nelson and Plosser (J.Mon.Ec., 1982) for a large set of macroeconomic series.
- The Said–Dickey (Biometrika, 1984) "Augmented Dickey–Fuller test" has been widely used (cf. Stata's `dfuller`) as well as other 'first-generation' alternatives such as the Phillips–Perron test (Stata's `pperron`; Biometrika, 1988).
- These 'first-generation' alternatives, with a null hypothesis of $I(1)$, or a unit root, are known to have low power, particularly in smaller samples.

- Several approaches to dealing with the problem of low power have appeared in the econometric literature.
- Modification of the ADF test by Elliott, Rothenberg, Stock (ERS: Econometrica, 1996) leads to the DF-GLS (generalized least squares) test, while Leybourne (OBES, 1995) proposes the ADFmax test, involving forward and reversed regressions.
- Testing for unit roots with panel data requires fewer time series observations to achieve power (cf. Stata's `xtunitroot`).
- Tests with the null hypothesis of $I(0)$, such as Kwiatkowski, Phillips, Schmidt, Shin (J.Metrics, 1992; SSC `kpss`) can be used to confirm the verdict of D-F style tests.

- Several approaches to dealing with the problem of low power have appeared in the econometric literature.
- Modification of the ADF test by Elliott, Rothenberg, Stock (ERS: Econometrica, 1996) leads to the DF-GLS (generalized least squares) test, while Leybourne (OBES, 1995) proposes the ADFmax test, involving forward and reversed regressions.
- Testing for unit roots with panel data requires fewer time series observations to achieve power (cf. Stata's `xtunitroot`).
- Tests with the null hypothesis of $I(0)$, such as Kwiatkowski, Phillips, Schmidt, Shin (J.Metrics, 1992; SSC `kpss`) can be used to confirm the verdict of D-F style tests.

- Several approaches to dealing with the problem of low power have appeared in the econometric literature.
- Modification of the ADF test by Elliott, Rothenberg, Stock (ERS: Econometrica, 1996) leads to the DF-GLS (generalized least squares) test, while Leybourne (OBES, 1995) proposes the ADFmax test, involving forward and reversed regressions.
- Testing for unit roots with panel data requires fewer time series observations to achieve power (cf. Stata's `xtunitroot`).
- Tests with the null hypothesis of $I(0)$, such as Kwiatkowski, Phillips, Schmidt, Shin (J.Metrics, 1992; SSC `kpss`) can be used to confirm the verdict of D-F style tests.

- Several approaches to dealing with the problem of low power have appeared in the econometric literature.
- Modification of the ADF test by Elliott, Rothenberg, Stock (ERS: Econometrica, 1996) leads to the DF-GLS (generalized least squares) test, while Leybourne (OBES, 1995) proposes the ADFmax test, involving forward and reversed regressions.
- Testing for unit roots with panel data requires fewer time series observations to achieve power (cf. Stata's `xtunitroot`).
- Tests with the null hypothesis of *I*(0), such as Kwiatkowski, Phillips, Schmidt, Shin (J.Metrics, 1992; SSC `kpss`) can be used to confirm the verdict of D-F style tests.

- We have produced response surface estimates of critical values, for a large range of quantiles, different combinations of the number of observations, and the lag order in the test regressions for the ERS DF-GLS and ADFmax unit root tests.
- The DF-GLS test of Baum and Sperling appeared in Stata 6.0 (Stata Tech.Bull. 57,58) and was added to official Stata as `dfgls`. Our version of that test, accessing the response surface estimates, is now available from the SSC Archive as `ersur`.
- The ADFmax test with response surface estimates is now available from the SSC Archive as `adfmaxur`.

- We have produced response surface estimates of critical values, for a large range of quantiles, different combinations of the number of observations, and the lag order in the test regressions for the ERS DF-GLS and ADFmax unit root tests.
- The DF-GLS test of Baum and Sperling appeared in Stata 6.0 (Stata Tech.Bull. 57,58) and was added to official Stata as `dfgls`. Our version of that test, accessing the response surface estimates, is now available from the SSC Archive as `ersur`.
- The ADFmax test with response surface estimates is now available from the SSC Archive as `adfmaxur`.

- We have produced response surface estimates of critical values, for a large range of quantiles, different combinations of the number of observations, and the lag order in the test regressions for the ERS DF-GLS and ADFmax unit root tests.
- The DF-GLS test of Baum and Sperling appeared in Stata 6.0 (Stata Tech.Bull. 57,58) and was added to official Stata as `dfgls`. Our version of that test, accessing the response surface estimates, is now available from the SSC Archive as `ersur`.
- The ADFmax test with response surface estimates is now available from the SSC Archive as `adfmaxur`.

Assuming the presence of a nonzero trend in the underlying data, the ERS test is based on the *t* statistic that tests the null hypothesis that $a_0 = 0$, against the alternative of stationarity $a_0 < 0$, in the auxiliary regression:

$$\Delta y_t^d = a_0 y_{t-1}^d + b_1 \Delta y_{t-1}^d + ... + b_p \Delta y_{t-p}^d + \varepsilon_t, \tag{1}$$

where *p* lags of the dependent variable are included to account for residual serial correlation, and y_t^d is the GLS-detrended version of the original series y_t, that is:

$$y_t^d = y_t - \hat{\beta}_0 - \hat{\beta}_1 t$$

The coefficients defining the detrended series y_t^d, $\hat{\beta}_0$ and $\hat{\beta}_1$, are obtained through an ordinary least squares regression of $\bar{y}$ against $\bar{w}$, where:

$$\begin{aligned} \bar{y} &= \left[y_1, (1-\bar{\rho}L)\, y_2, ..., (1-\bar{\rho}L)\, y_T\right], \\ \bar{w} &= \left[w_1, (1-\bar{\rho}L)\, w_2, ..., (1-\bar{\rho}L)\, w_T\right], \\ \bar{\rho} &= 1 + \frac{\bar{c}}{T}, \end{aligned}$$

and $w_t = (1, t)$ contains the deterministic components. ERS recommend to set $\bar{c} = -13.5$ in order to obtain the highest power of the test. A similar procedure is followed in a model with no trend, in which GLS demeaning is applied with $\bar{c} = -7$.

Cheung and Lai (OBES, 1995) present response surface estimates for values of T and exogenously determined p, the lag order, for both the GLS-detrended and GLS-demeaned versions of the ERS test.

The coefficients defining the detrended series y_t^d, $\hat{\beta}_0$ and $\hat{\beta}_1$, are obtained through an ordinary least squares regression of $\bar{y}$ against $\bar{w}$, where:

$$\begin{aligned}
\bar{y} &= \left[y_1, (1-\bar{\rho}L)\, y_2, ..., (1-\bar{\rho}L)\, y_T\right], \\
\bar{w} &= \left[w_1, (1-\bar{\rho}L)\, w_2, ..., (1-\bar{\rho}L)\, w_T\right], \\
\bar{\rho} &= 1 + \frac{\bar{c}}{T},
\end{aligned}$$

and $w_t = (1, t)$ contains the deterministic components. ERS recommend to set $\bar{c} = -13.5$ in order to obtain the highest power of the test. A similar procedure is followed in a model with no trend, in which GLS demeaning is applied with $\bar{c} = -7$.

Cheung and Lai (OBES, 1995) present response surface estimates for values of T and exogenously determined p, the lag order, for both the GLS-detrended and GLS-demeaned versions of the ERS test.

The Monte Carlo experiment

- Response surface estimates are generated with a Monte Carlo simulation experiment similar to that used by Otero and Smith (Comp.Stat., 2012).
- Assume that y_t is a unit root process with standard Normal errors and a sample of $T + 1$ observations, with T ranging from 18 to 2000.
- The number of lagged differences of y_t, p, varies between 0 and 8, with 8 lags used for $T \geq 36$.
- The experiment defines 456 combinations of T and p, and involves 50,000 Monte Carlo replications.

The Monte Carlo experiment

- Response surface estimates are generated with a Monte Carlo simulation experiment similar to that used by Otero and Smith (Comp.Stat., 2012).
- Assume that y_t is a unit root process with standard Normal errors and a sample of $T+1$ observations, with T ranging from 18 to 2000.
- The number of lagged differences of y_t, p, varies between 0 and 8, with 8 lags used for $T \geq 36$.
- The experiment defines 456 combinations of T and p, and involves 50,000 Monte Carlo replications.

The Monte Carlo experiment

- Response surface estimates are generated with a Monte Carlo simulation experiment similar to that used by Otero and Smith (Comp.Stat., 2012).
- Assume that y_t is a unit root process with standard Normal errors and a sample of $T + 1$ observations, with T ranging from 18 to 2000.
- The number of lagged differences of y_t, p, varies between 0 and 8, with 8 lags used for $T \geq 36$.
- The experiment defines 456 combinations of T and p, and involves 50,000 Monte Carlo replications.

The Monte Carlo experiment

- Response surface estimates are generated with a Monte Carlo simulation experiment similar to that used by Otero and Smith (Comp.Stat., 2012).
- Assume that y_t is a unit root process with standard Normal errors and a sample of $T+1$ observations, with T ranging from 18 to 2000.
- The number of lagged differences of y_t, p, varies between 0 and 8, with 8 lags used for $T \geq 36$.
- The experiment defines 456 combinations of T and p, and involves 50,000 Monte Carlo replications.

- Critical values are computed for each of 221 significance levels: $0.0001, 0.0002, \ldots, 0.9998, 0.9999$ for both the detrended and demeaned cases.
- Response surface models are then estimated for each significance level.
- The functional form of these models follows MacKinnon (1991), Cheung and Lai (JBES, 1995; OBES, 1995) and Harvey and Van Dijk (CSDA, 2006), in which the critical values are regressed on an intercept term and power functions of $\frac{1}{T}$ and $\frac{p}{T}$.

- Critical values are computed for each of 221 significance levels: $0.0001, 0.0002, \ldots, 0.9998, 0.9999$ for both the detrended and demeaned cases.
- Response surface models are then estimated for each significance level.
- The functional form of these models follows MacKinnon (1991), Cheung and Lai (JBES, 1995; OBES, 1995) and Harvey and Van Dijk (CSDA, 2006), in which the critical values are regressed on an intercept term and power functions of $\frac{1}{T}$ and $\frac{p}{T}$.

- Critical values are computed for each of 221 significance levels: $0.0001, 0.0002, \ldots, 0.9998, 0.9999$ for both the detrended and demeaned cases.
- Response surface models are then estimated for each significance level.
- The functional form of these models follows MacKinnon (1991), Cheung and Lai (JBES, 1995; OBES, 1995) and Harvey and Van Dijk (CSDA, 2006), in which the critical values are regressed on an intercept term and power functions of $\frac{1}{T}$ and $\frac{p}{T}$.

The chosen functional form is:

$$CV^{l}_{T,p} = \theta^{l}_{\infty} + \sum_{i=1}^{4} \theta^{l}_{i} \left(\frac{1}{T}\right)^{i} + \sum_{i=1}^{4} \phi^{l}_{i} \left(\frac{p^{i}}{T}\right) + \epsilon^{l}, \tag{2}$$

where $CV^{l}_{T,p}$ is the critical value estimate at significance level l, T refers to the number of observations on Δy_t, which is one less than the total number of available observations, and p is the number of lags of the dependent variable that are included to account for residual serial correlation.

Note that the larger the number of observations, T, the weaker is the dependence of the critical values on the lag truncation p. Also, as $T \rightarrow \infty$ the intercept term, θ^{l}_{∞}, can be thought of as an estimate of the corresponding asymptotic critical value.

The chosen functional form is:

$$CV^l_{T,p} = \theta^l_\infty + \sum_{i=1}^{4} \theta^l_i \left(\frac{1}{T}\right)^i + \sum_{i=1}^{4} \phi^l_i \left(\frac{p^i}{T}\right) + \epsilon^l, \tag{2}$$

where $CV^l_{T,p}$ is the critical value estimate at significance level l, T refers to the number of observations on Δy_t, which is one less than the total number of available observations, and p is the number of lags of the dependent variable that are included to account for residual serial correlation.

Note that the larger the number of observations, T, the weaker is the dependence of the critical values on the lag truncation p. Also, as $T \to \infty$ the intercept term, θ^l_∞, can be thought of as an estimate of the corresponding asymptotic critical value.

Dealing with endogenous lag order

- The tabulated response surface values can be used to obtain critical values for any given T and fixed lag order.
- In practice the lag order, p, is rarely fixed by the user, but rather chosen endogenously using a data-dependent procedure such as the information criteria of Akaike and Schwarz, AIC and SIC respectively.
- The optimal number of lags is determined by varying p in the regression (1) between $p_{\max}$ and 0 lags, and choosing the best model according to the information criterion that is being employed.
- These criteria are implemented in `dfgls` as `Min MAIC` and `Min SC`, respectively.

Dealing with endogenous lag order

- The tabulated response surface values can be used to obtain critical values for any given T and fixed lag order.
- In practice the lag order, p, is rarely fixed by the user, but rather chosen endogenously using a data-dependent procedure such as the information criteria of Akaike and Schwarz, AIC and SIC respectively.
- The optimal number of lags is determined by varying p in the regression (1) between $p_{\max}$ and 0 lags, and choosing the best model according to the information criterion that is being employed.
- These criteria are implemented in `dfgls` as `Min MAIC` and `Min SC`, respectively.

Dealing with endogenous lag order

- The tabulated response surface values can be used to obtain critical values for any given T and fixed lag order.
- In practice the lag order, p, is rarely fixed by the user, but rather chosen endogenously using a data-dependent procedure such as the information criteria of Akaike and Schwarz, AIC and SIC respectively.
- The optimal number of lags is determined by varying p in the regression (1) between $p_{\max}$ and 0 lags, and choosing the best model according to the information criterion that is being employed.
- These criteria are implemented in `dfgls` as `Min MAIC` and `Min SC`, respectively.

Dealing with endogenous lag order

- The tabulated response surface values can be used to obtain critical values for any given T and fixed lag order.
- In practice the lag order, p, is rarely fixed by the user, but rather chosen endogenously using a data-dependent procedure such as the information criteria of Akaike and Schwarz, AIC and SIC respectively.
- The optimal number of lags is determined by varying p in the regression (1) between $p_{\max}$ and 0 lags, and choosing the best model according to the information criterion that is being employed.
- These criteria are implemented in `dfgls` as `Min MAIC` and `Min SC`, respectively.

- We also consider another data-dependent procedure to optimally select p, which is commonly referred to as the general-to-specific (GTS) algorithm of Campbell and Perron (1991), Hall (JBES, 1994) and Ng and Perron (JASA, 1995).
- This algorithm starts by setting some upper bound on p, let us say $p_{\max}$, where $p_{\max} = 0, 1, 2, ..., 8$, estimating equation (1) with $p = p_{\max}$, and testing the statistical significance of $b_{p_{\max}}$.
- If this coefficient is statistically significant, for instance using a significance level of 5% (denoted GTS_5) or 10% (denoted GTS_{10}), one chooses $p = p_{\max}$. Otherwise, the order of the estimated autoregression in (1) is reduced by one until the coefficient on the last included lag is statistically different from zero.
- The GTS_{10} criterion is implemented in `dfgls` as `Opt Lag (Ng-Perron seq t)`.

- We also consider another data-dependent procedure to optimally select p, which is commonly referred to as the general-to-specific (GTS) algorithm of Campbell and Perron (1991), Hall (JBES, 1994) and Ng and Perron (JASA, 1995).
- This algorithm starts by setting some upper bound on p, let us say $p_{\max}$, where $p_{\max} = 0, 1, 2, ..., 8$, estimating equation (1) with $p = p_{\max}$, and testing the statistical significance of $b_{p_{\max}}$.
- If this coefficient is statistically significant, for instance using a significance level of 5% (denoted GTS_5) or 10% (denoted GTS_{10}), one chooses $p = p_{\max}$. Otherwise, the order of the estimated autoregression in (1) is reduced by one until the coefficient on the last included lag is statistically different from zero.
- The GTS_{10} criterion is implemented in `dfgls` as `Opt Lag (Ng-Perron seq t)`.

- We also consider another data-dependent procedure to optimally select p, which is commonly referred to as the general-to-specific (GTS) algorithm of Campbell and Perron (1991), Hall (JBES, 1994) and Ng and Perron (JASA, 1995).
- This algorithm starts by setting some upper bound on p, let us say $p_{\max}$, where $p_{\max} = 0, 1, 2, ..., 8$, estimating equation (1) with $p = p_{\max}$, and testing the statistical significance of $b_{p_{\max}}$.
- If this coefficient is statistically significant, for instance using a significance level of 5% (denoted GTS_5) or 10% (denoted GTS_{10}), one chooses $p = p_{\max}$. Otherwise, the order of the estimated autoregression in (1) is reduced by one until the coefficient on the last included lag is statistically different from zero.
- The GTS_{10} criterion is implemented in `dfgls` as `Opt Lag (Ng-Perron seq t)`.

- We also consider another data-dependent procedure to optimally select p, which is commonly referred to as the general-to-specific (GTS) algorithm of Campbell and Perron (1991), Hall (JBES, 1994) and Ng and Perron (JASA, 1995).
- This algorithm starts by setting some upper bound on p, let us say $p_{\max}$, where $p_{\max} = 0, 1, 2, ..., 8$, estimating equation (1) with $p = p_{\max}$, and testing the statistical significance of $b_{p_{\max}}$.
- If this coefficient is statistically significant, for instance using a significance level of 5% (denoted GTS$_5$) or 10% (denoted GTS$_{10}$), one chooses $p = p_{\max}$. Otherwise, the order of the estimated autoregression in (1) is reduced by one until the coefficient on the last included lag is statistically different from zero.
- The GTS$_{10}$ criterion is implemented in `dfgls` as `Opt Lag (Ng-Perron seq t)`.

- For AIC, SIC, GTS_5, and GTS_{10} the same 221 quantiles of the empirical small sample distribution are recorded as before, but the response surface regressions given in (2) are estimated using $p_{\max}$ instead of p lags.
- We estimate 2210 response surface regressions: two models multiplied by five criteria to select p multiplied by the 221 significance levels. The chosen functional form performs very well, with an average coefficient of determination was 0.994; in only 36 (out of 2210) cases it was below 0.95.
- For $T = 1000$ the implied asymptotic critical values from the response surface models estimated in this paper are close to those obtained by ERS.
- The implied critical values also exhibit dependence on the method used to select the lag length, and in some cases the differences may be noticeable, especially when T and l are small and p is large.

- For AIC, SIC, GTS_5, and GTS_{10} the same 221 quantiles of the empirical small sample distribution are recorded as before, but the response surface regressions given in (2) are estimated using $p_{\max}$ instead of p lags.
- We estimate 2210 response surface regressions: two models multiplied by five criteria to select p multiplied by the 221 significance levels. The chosen functional form performs very well, with an average coefficient of determination was 0.994; in only 36 (out of 2210) cases it was below 0.95.
- For $T = 1000$ the implied asymptotic critical values from the response surface models estimated in this paper are close to those obtained by ERS.
- The implied critical values also exhibit dependence on the method used to select the lag length, and in some cases the differences may be noticeable, especially when T and l are small and p is large.

- For AIC, SIC, GTS_5, and GTS_{10} the same 221 quantiles of the empirical small sample distribution are recorded as before, but the response surface regressions given in (2) are estimated using $p_{\max}$ instead of p lags.
- We estimate 2210 response surface regressions: two models multiplied by five criteria to select p multiplied by the 221 significance levels. The chosen functional form performs very well, with an average coefficient of determination was 0.994; in only 36 (out of 2210) cases it was below 0.95.
- For $T = 1000$ the implied asymptotic critical values from the response surface models estimated in this paper are close to those obtained by ERS.
- The implied critical values also exhibit dependence on the method used to select the lag length, and in some cases the differences may be noticeable, especially when T and l are small and p is large.

- For AIC, SIC, GTS_5, and GTS_{10} the same 221 quantiles of the empirical small sample distribution are recorded as before, but the response surface regressions given in (2) are estimated using p_{max} instead of p lags.
- We estimate 2210 response surface regressions: two models multiplied by five criteria to select p multiplied by the 221 significance levels. The chosen functional form performs very well, with an average coefficient of determination was 0.994; in only 36 (out of 2210) cases it was below 0.95.
- For $T = 1000$ the implied asymptotic critical values from the response surface models estimated in this paper are close to those obtained by ERS.
- The implied critical values also exhibit dependence on the method used to select the lag length, and in some cases the differences may be noticeable, especially when T and l are small and p is large.

To obtain *p*-values of the ERS statistic, we follow MacKinnon (JBES, 1994; App.Econometrics, 1996) by estimating the regression:

$$\Phi^{-1}(l) = \gamma_0^l + \gamma_1^l \widehat{CV^l} + \gamma_2^l \left(\widehat{CV^l}\right)^2 + \upsilon^l, \tag{3}$$

where Φ^{-1} is the inverse of the cumulative standard normal distribution at each of the 221 quantiles, and $\widehat{CV^l}$ is the fitted value from (2) at the l quantile. Following Harvey and van Dijk (CSDA, 2006), equation (3) is estimated by OLS using 15 observations. Approximate *p*-values of the ERS test statistic can then be obtained as:

$$pvalue = \Phi\left(\widehat{\gamma}_0^l + \widehat{\gamma}_1^l ERS\,(p) + \widehat{\gamma}_2^l \left(ERS\,(p)\right)^2\right), \tag{4}$$

where $\widehat{\gamma}_0^l$, $\widehat{\gamma}_1^l$ and $\widehat{\gamma}_2^l$ are the OLS parameter estimates from (3).

Implementation issues

- The result of the Monte Carlo experiment is a 221 $\times$ 91 matrix, with the rows indexed by the quantile and the columns representing combinations of lag method, demeaned or detrended, and sample size.
- Although it would be possible to include this as a Stata matrix coded into the ado-file, that appeared to be a very inelegant solution.
- Accordingly, the matrix was stored as a binary matrix using Mata's `fputmatrix()` function, and the ado-file uses Mata's `fopen()` and `fgetmatrix()` functions to retrieve it from the `PLUS` directory.
- The lookup routine, and the regression referenced in equation 4, is implemented as a Mata function contained in `ersur.ado`.

Implementation issues

- The result of the Monte Carlo experiment is a 221 $\times$ 91 matrix, with the rows indexed by the quantile and the columns representing combinations of lag method, demeaned or detrended, and sample size.
- Although it would be possible to include this as a Stata matrix coded into the ado-file, that appeared to be a very inelegant solution.
- Accordingly, the matrix was stored as a binary matrix using Mata's `fputmatrix()` function, and the ado-file uses Mata's `fopen()` and `fgetmatrix()` functions to retrieve it from the `PLUS` directory.
- The lookup routine, and the regression referenced in equation 4, is implemented as a Mata function contained in `ersur.ado`.

Implementation issues

- The result of the Monte Carlo experiment is a 221 $\times$ 91 matrix, with the rows indexed by the quantile and the columns representing combinations of lag method, demeaned or detrended, and sample size.
- Although it would be possible to include this as a Stata matrix coded into the ado-file, that appeared to be a very inelegant solution.
- Accordingly, the matrix was stored as a binary matrix using Mata's `fputmatrix()` function, and the ado-file uses Mata's `fopen()` and `fgetmatrix()` functions to retrieve it from the `PLUS` directory.
- The lookup routine, and the regression referenced in equation 4, is implemented as a Mata function contained in `ersur.ado`.

Implementation issues

- The result of the Monte Carlo experiment is a 221 $\times$ 91 matrix, with the rows indexed by the quantile and the columns representing combinations of lag method, demeaned or detrended, and sample size.
- Although it would be possible to include this as a Stata matrix coded into the ado-file, that appeared to be a very inelegant solution.
- Accordingly, the matrix was stored as a binary matrix using Mata's `fputmatrix()` function, and the ado-file uses Mata's `fopen()` and `fgetmatrix()` functions to retrieve it from the `PLUS` directory.
- The lookup routine, and the regression referenced in equation 4, is implemented as a Mata function contained in `ersur.ado`.

Empirical implementation

- We illustrate the use of `ersur` to consider whether interest rate spreads on US Treasury securities are stationary stochastic processes, as implied by the expectations theory of the term structure of interest rates. Monthly interest rates for 1993m10–2013m3 are acquired from the FRED database.
- We consider whether the spread between six-month and three-month Treasury bills, `s6`, contains a unit root. The interest rate spread does not exhibit a trend, so we use the default specification of GLS demeaning.

Empirical implementation

- We illustrate the use of `ersur` to consider whether interest rate spreads on US Treasury securities are stationary stochastic processes, as implied by the expectations theory of the term structure of interest rates. Monthly interest rates for 1993m10–2013m3 are acquired from the FRED database.
- We consider whether the spread between six-month and three-month Treasury bills, `s6`, contains a unit root. The interest rate spread does not exhibit a trend, so we use the default specification of GLS demeaning.

```
. ersur s6, maxlag(3)
Elliott, Rothenberg & Stock (1996) test results for 1994m2 - 2013m3
Variable name: s6
Ho: Unit root
Ha: Stationary
GLS demeaned data
```

Criteria	Lags	ers stat.	p-value	1% cv	5% cv	10% cv
FIXED	3	-3.780	0.000	-2.630	-2.016	-1.702
AIC	0	-4.298	0.000	-2.684	-2.048	-1.725
SIC	0	-4.298	0.000	-2.656	-2.033	-1.715
GTS05	0	-4.298	0.000	-2.676	-2.042	-1.720
GTS10	2	-3.562	0.001	-2.685	-2.046	-1.723

The test results show that we can decisively reject the null hypothesis of $I(1)$ in favor of stationarity of the interest rate spread. The same test applied to spreads between 12, 36, 60, 84, 120 and 240-month rates and the three-month rate yields the same conclusion.

```
. ersur s6, maxlag(3)
Elliott, Rothenberg & Stock (1996) test results for 1994m2 - 2013m3
Variable name: s6
Ho: Unit root
Ha: Stationary
GLS demeaned data
```

Criteria	Lags	ers stat.	p-value	1% cv	5% cv	10% cv
FIXED	3	-3.780	0.000	-2.630	-2.016	-1.702
AIC	0	-4.298	0.000	-2.684	-2.048	-1.725
SIC	0	-4.298	0.000	-2.656	-2.033	-1.715
GTS05	0	-4.298	0.000	-2.676	-2.042	-1.720
GTS10	2	-3.562	0.001	-2.685	-2.046	-1.723

The test results show that we can decisively reject the null hypothesis of $I(1)$ in favor of stationarity of the interest rate spread. The same test applied to spreads between 12, 36, 60, 84, 120 and 240-month rates and the three-month rate yields the same conclusion.

- Examination of the response surface estimates shows that the method used to select the order of the augmentation of the ERS DF-GLS test affects the finite-sample critical values.
- The `ersur` command evaluates the appropriate critical values depending on the method used for lag order determination and allows the researcher to evaluate the sensitivity of her findings to the choice of method.

- Examination of the response surface estimates shows that the method used to select the order of the augmentation of the ERS DF-GLS test affects the finite-sample critical values.
- The `ersur` command evaluates the appropriate critical values depending on the method used for lag order determination and allows the researcher to evaluate the sensitivity of her findings to the choice of method.

The Leybourne ADFmax test

- Although time does not permit a detailed discussion, we have implemented a similar procedure for the Leybourne (OBES, 1995) ADFmax unit root test.
- According to Leybourne, this test exhibits greater power than the standard ADF test, so it is more likely to reject a false unit root hypothesis.
- The test involves running Dickey–Fuller regressions using forward and reverse realizations of the time series.

The Leybourne ADFmax test

- Although time does not permit a detailed discussion, we have implemented a similar procedure for the Leybourne (OBES, 1995) ADFmax unit root test.
- According to Leybourne, this test exhibits greater power than the standard ADF test, so it is more likely to reject a false unit root hypothesis.
- The test involves running Dickey–Fuller regressions using forward and reverse realizations of the time series.

The Leybourne ADFmax test

- Although time does not permit a detailed discussion, we have implemented a similar procedure for the Leybourne (OBES, 1995) ADFmax unit root test.
- According to Leybourne, this test exhibits greater power than the standard ADF test, so it is more likely to reject a false unit root hypothesis.
- The test involves running Dickey–Fuller regressions using forward and reverse realizations of the time series.

- Leybourne's critical values for the test, for models with an intercept or an intercept and trend, do not account for the possibility that the lag order in the ADF regressions may be endogenously determined.
- Otero and Smith (Comp.Stat., 2012) computed response surface regressions that allow for different specifications, sample size and lag order determination.
- These response surface estimates of critical values for the Leybourne ADFmax test are now available in our `adfmaxur` command.
- Like `ersur`, the command allows for fixed lag order or lag order chosen by SIC, AIC, GTS05 or GTS10 methods.
- As in the case of `ersur`, the finite-sample critical values are shown to be sensitive to the method used for lag order augmentation.

- Leybourne's critical values for the test, for models with an intercept or an intercept and trend, do not account for the possibility that the lag order in the ADF regressions may be endogenously determined.
- Otero and Smith (Comp.Stat., 2012) computed response surface regressions that allow for different specifications, sample size and lag order determination.
- These response surface estimates of critical values for the Leybourne ADFmax test are now available in our `adfmaxur` command.
- Like `ersur`, the command allows for fixed lag order or lag order chosen by SIC, AIC, GTS05 or GTS10 methods.
- As in the case of `ersur`, the finite-sample critical values are shown to be sensitive to the method used for lag order augmentation.

- Leybourne's critical values for the test, for models with an intercept or an intercept and trend, do not account for the possibility that the lag order in the ADF regressions may be endogenously determined.
- Otero and Smith (Comp.Stat., 2012) computed response surface regressions that allow for different specifications, sample size and lag order determination.
- These response surface estimates of critical values for the Leybourne ADFmax test are now available in our `adfmaxur` command.
- Like `ersur`, the command allows for fixed lag order or lag order chosen by SIC, AIC, GTS05 or GTS10 methods.
- As in the case of `ersur`, the finite-sample critical values are shown to be sensitive to the method used for lag order augmentation.

- Leybourne's critical values for the test, for models with an intercept or an intercept and trend, do not account for the possibility that the lag order in the ADF regressions may be endogenously determined.
- Otero and Smith (Comp.Stat., 2012) computed response surface regressions that allow for different specifications, sample size and lag order determination.
- These response surface estimates of critical values for the Leybourne ADFmax test are now available in our `adfmaxur` command.
- Like `ersur`, the command allows for fixed lag order or lag order chosen by SIC, AIC, GTS05 or GTS10 methods.
- As in the case of `ersur`, the finite-sample critical values are shown to be sensitive to the method used for lag order augmentation.

- Leybourne's critical values for the test, for models with an intercept or an intercept and trend, do not account for the possibility that the lag order in the ADF regressions may be endogenously determined.
- Otero and Smith (Comp.Stat., 2012) computed response surface regressions that allow for different specifications, sample size and lag order determination.
- These response surface estimates of critical values for the Leybourne ADFmax test are now available in our `adfmaxur` command.
- Like `ersur`, the command allows for fixed lag order or lag order chosen by SIC, AIC, GTS05 or GTS10 methods.
- As in the case of `ersur`, the finite-sample critical values are shown to be sensitive to the method used for lag order augmentation.