The Stata Journal (2017)
17, Number 4, pp. 985–1002

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

Jesús Otero
Universidad del Rosario
Bogotá, Colombia
jesus.otero@urosario.edu.co

Christopher F. Baum
Boston College
Chestnut Hill, MA
baum@bc.edu

Abstract. In this article, we present response surface coefficients for a large range of quantiles of the Elliott, Rothenberg, and Stock (1996, *Econometrica* 64: 813–836) unit-root tests, for different combinations of number of observations, T, and lag order in the test regressions, p, where the latter can either be specified by the user or be endogenously determined. The critical values depend on the method used to select the number of lags. We present the command **ersur** and illustrate its use with an empirical example that tests the validity of the expectations hypothesis of the term structure of interest rates.

Keywords: st0508, ersur, Elliott, Rothenberg, Stock, unit-root test, Monte Carlo, response surface, critical values, lag length, p-values

1 Introduction

Since Nelson and Plosser (1982), testing for the presence of a unit root has become standard practice in the empirical analysis of economic time series. Among the tests available in the literature, the Said and Dickey (1984) unit-root test, based on extending Dickey and Fuller (1979) and commonly referred to as ADF, continues to be a favorite procedure of applied researchers. This is probably because the regression-based ADF test can be easily computed. However, a common criticism is that the ADF test exhibits disappointing power properties, as shown, for example, in the Monte Carlo simulations performed by DeJong et al. (1992).

During the last three decades, there have been three main research programs in the econometrics literature that aim to overcome the low power problem. First, some authors have continued developing more-powerful modifications of the univariate ADF test, including the generalized least squares (GLS)-ADF test of Elliott, Rothenberg, and Stock (1996), who use conditional GLS, and the ADF-max test of Leybourne (1995), who suggests taking the maximum of two ADF test statistics calculated using both forward and reversed data. Second, testing for unit roots in panel data has also been considered an alternative way to achieve power gains over unit-root tests applied to a single time series. This is because panel data, by combining information from the time-series dimension with that from the cross-section dimension, require fewer time observations for the tests to exhibit power. Among the panel unit-root tests available in the literature, perhaps those put forward by Im, Pesaran, and Shin (2003) and Pesaran (2007b) have proven to be the most popular. Third, authors such as Kapetanios, Shin, and Snell (2003) and

Kapetanios and Shin (2008) have considered tests of the unit-root hypothesis against the alternative of a globally stationary exponential smooth-transition autoregressive process.

Focusing on the first approach, Elliott, Rothenberg, and Stock (1996) propose a modified version of the ADF unit-root test—called the ERS test—that has substantially improved power in the presence of an unknown intercept or trend. Elliott, Rothenberg, and Stock further show that while the t statistic calculated from the GLS-demeaned data has an identical limiting representation to that of the conventional Dickey–Fuller t statistic when there is no intercept, the limiting representation differs in the linear trend case. To apply the test, Elliott, Rothenberg, and Stock tabulate, via stochastic simulation, asymptotic critical values (CVs) based on $T = 50$, 100, 200, and ∞ time observations.[1] In subsequent work, Cheung and Lai (1995b) examine the sensitivity of CVs to the sample size through response surface regressions that account for the effect of varying the number of observations, T, and the number of lags of the dependent variable, p. However, these CVs do not allow for their possible dependence on the criterion used to select the optimal number of lags.

In this article, we undertake an extensive set of Monte Carlo simulations, summarized through response surface regressions, to calculate finite-sample CVs and approximate p-values of the ERS unit-root test. The simulation experiments not only allow for the presence of stochastic processes with nonzero mean and nonzero trend, but also allow for the lag order to be either fixed or determined endogenously using a data-dependent procedure. We present the command **ersur**, which easily calculates the ERS test statistic, finite-sample CVs, and approximate p-values.

This article is organized as follows: Section 2 provides an overview of the ERS unit-root test. Section 3 presents the design of the Monte Carlo experiments. Section 4 reports the estimated response surfaces and describes the procedure to estimate the associated approximate p-values. Section 5 describes the `ersur` command. Section 6 illustrates the use of `ersur` with an empirical example based on interest-rate spreads. Section 7 concludes the article.

2 The Elliott, Rothenberg, and Stock test

Elliott, Rothenberg, and Stock (1996) propose a test for the null hypothesis of a unit root against the alternative of stationarity, available as the Stata command `dfgls` (see [TS] **dfgls**).[2] 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 hypothesis of stationarity $a_0 < 0$, in the following auxiliary regression:

1. Empirical applications of the ERS test include Pesaran (2007a) and Le Pen (2011) for output convergence; Pesaran et al. (2009) for purchasing power parity; and Abbott and De Vita (2012) for house price convergence.
2. The original version of `dfgls` on the Statistical Software Components archive was written by C. F. Baum and Richard Sperling for Stata 6.0.

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

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,

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

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are obtained through an ordinary least-squares (OLS) regression of $\overline{y}$ against $\overline{w}$, where

$$\begin{aligned}
&\overline{y} = \{y_1, (1-\overline{\rho}L)\, y_2, \ldots, (1-\overline{\rho}L)\, y_T\} \\
&\overline{w} = \{w_1, (1-\overline{\rho}L)\, w_2, \ldots, (1-\overline{\rho}L)\, w_T\} \\
&\overline{\rho} = 1 + \tfrac{\overline{c}}{T}
\end{aligned}$$

and $w_t = (1, t)$ contains the deterministic components.

Elliott, Rothenberg, and Stock (1996) recommend setting $\overline{c} = -13.5$ to obtain the best results in terms of the power of the test. The CVs of the test for trended data were tabulated by Elliott, Rothenberg, and Stock in table 1 for $T = 50$, 100, 200, and ∞. Cheung and Lai (1995b) present response surface coefficients that allow for changing T and exogenously determined p.

In the model with no trend, the GLS-demeaned version of the original series y_t is obtained as

$$y_t^d = y_t - \widehat{\beta}_0$$

where $\widehat{\beta}_0$ is obtained through an OLS regression of $\overline{y}$ against $\overline{w}$, where

$$\begin{aligned}
&\overline{y} = \{y_1, (1-\overline{\rho}L)\, y_2, \ldots, (1-\overline{\rho}L)\, y_T\} \\
&\overline{w} = \{w_1, (1-\overline{\rho}L)\, w_2, \ldots, (1-\overline{\rho}L)\, w_T\} \\
&\overline{\rho} = 1 + \tfrac{\overline{c}}{T}
\end{aligned}$$

and $w_t = (1)$ contains the deterministic component. Elliott, Rothenberg, and Stock (1996) recommend setting $\overline{c} = -7$. The CVs correspond to those originally tabulated by Dickey and Fuller (1979) for the model with no constant; see also MacKinnon (1991) and Cheung and Lai (1995a).

3 Monte Carlo experiment design

The design of the Monte Carlo simulation experiment follows Otero and Smith (2012). Assume that y_t is generated by an autoregressive process of order 1:

$$y_t = y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, 1)$ and $t = 1, \ldots, T+1$. Simulation experiments are carried out for a total of 56 different sample sizes, with $T = 18(2)62$, $65(5)100$, $110(10)200$, $220(20)300$, $350(50)500$, $600(100)800$, 1000, 1400, and 2000, where, for example, $18(2)62$ means that all samples from $T = 18$ to $T = 62$ increasing in steps of 2 are accounted for. The same notation is used later when listing significance levels. The time series y_t is generated by setting an initial value $y_{-99} = 0$, and then the first 100 observations are discarded. Each experiment consists of 50,000 Monte Carlo replications. The number of lagged differences of the dependent variable, p, is set equal to $p = 0, 1, \ldots, 8$. For $T \leq 20$, $p \leq 1$ is used; for $22 \leq T \leq 24$, $p \leq 2$ is used; for $26 \leq T \leq 28$, $p \leq 3$ is used; for $30 \leq T \leq 32$, $p \leq 4$ is used; for $34 \leq T \leq 36$, $p \leq 6$ is used; and for $T > 36$, all values of p are used. Overall, there will be 456 different pairings of T and p.

To account for sampling variability, the setup outlined above is repeated 50 times, implying that there will be 50 CVs of the test for each combination of number of observations, T, and lag truncation, p. Following earlier work by MacKinnon (1991), CVs are calculated at each of 221 significance levels ($l = 0.0001$, 0.0002, 0.0005, 0.001(0.001)0.01, 0.015(0.005)0.990, 0.991(0.001)0.999, 0.9995, 0.9998, and 0.9999) of the ERS t statistic for two cases, namely, a nonzero mean process (demeaned data) and a nonzero trend process (detrended data).

Using the simulated CVs, we subsequently fit response surface models at each of the $l = 221$ significance levels. The choice of the response surface functional form follows authors such as MacKinnon (1991), Cheung and Lai (1995a,b), and Harvey and van Dijk (2006), in which the CVs are regressed on an intercept term and on power functions of $1/T$ and p/T. The functional form that is finally selected is

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

where $\mathrm{CV}^l_{T,p}$ is the CV 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.[3] It is worth noticing that the functional form in (2) is such that the larger the number of observations, T, the weaker the CVs' dependence on the lag truncation, p. In addition, as $T \to \infty$, the intercept term θ^l_∞ can be thought of as an estimate of the corresponding asymptotic CV.

4 Main results

Tables 1 and 2 report response surface regression estimates for 3 of the 221 significance levels, namely, $l = 0.01$, 0.05, and 0.10 for demeaned and detrended data, respectively. These estimates can be used to obtain CVs for any given T and fixed lag order p. However, in practice, the lag order p is rarely fixed by the user and instead is chosen endogenously using a data-dependent procedure. Thus, we also use information criteria such as Akaike and Schwarz, which we denote as AIC and SIC, respectively. Here the optimal number of lags is determined by varying p in regression (1) between $p_{\max}$

3. Experimenting with even higher powered terms generally yielded coefficients that were not statistically different from 0 at the 1% significance level nor led to any noticeable increase in the $\overline{R}^2$ for these models.

and $p_{\min} = 0$ lags, and choosing the best model according to the information criterion being used. We also consider another data-dependent procedure, which is commonly referred to as the general-to-specific (GTS) algorithm, to optimally select p. This algorithm, advocated by Campbell and Perron (1991), Hall (1994), and Ng and Perron (1995), starts by setting some upper bound on p, say, $p_{\max}$, where $p_{\max} = 0, 1, 2, \ldots, 8$, estimating (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 1 until the coefficient on the last included lag is statistically different from 0. Finally, 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.

Table 1. Response surface estimates for demeaned data

Lags	l	Intercept	(Std. Err.)	$1/T$	$1/T^2$	$1/T^3$	$1/T^4$	p/T	p^2/T	p^3/T	p^4/T	R^2
Fixed	0.01	−2.569	(0.0004)	−18.779	151.9	−814.6	−9349.3	−0.306	1.059	−0.214	0.013	0.979
	0.05	−1.942	(0.0002)	−22.761	427.3	−6357.6	33186.4	0.057	0.764	−0.155	0.010	0.990
	0.10	−1.617	(0.0002)	−25.453	555.3	−8898.6	53120.1	0.237	0.618	−0.128	0.008	0.994
AIC	0.01	−2.568	(0.0004)	−19.642	182.4	−939.8	−9367.2	−5.556	1.474	−0.205	0.011	0.994
	0.05	−1.942	(0.0002)	−21.823	369.8	−5570.6	33264.7	−3.440	1.033	−0.144	0.007	0.997
	0.10	−1.617	(0.0001)	−24.485	501.9	−8301.2	54686.0	−2.388	0.804	−0.115	0.006	0.998
SIC	0.01	−2.569	(0.0004)	−15.966	39.0	−1251.2	19847.5	−4.668	1.619	−0.227	0.011	0.991
	0.05	−1.941	(0.0002)	−21.573	414.6	−8169.9	62936.0	−2.453	1.036	−0.155	0.008	0.997
	0.10	−1.617	(0.0001)	−24.784	563.9	−10656.2	77980.7	−1.493	0.748	−0.115	0.006	0.998
GTS_5	0.01	−2.566	(0.0003)	−21.472	395.0	−7654.8	52031.5	−3.304	0.735	−0.104	0.006	0.993
	0.05	−1.940	(0.0002)	−23.286	469.9	−7592.0	45976.4	−1.219	0.251	−0.037	0.002	0.997
	0.10	−1.616	(0.0001)	−25.483	551.6	−8809.1	54251.6	−0.472	0.108	−0.018	0.001	0.998
GTS_{10}	0.01	−2.565	(0.0004)	−22.210	419.0	−7449.8	44977.9	−4.669	1.120	−0.152	0.008	0.993
	0.05	−1.940	(0.0002)	−22.503	401.2	−5768.6	30522.2	−2.341	0.545	−0.074	0.004	0.997
	0.10	−1.617	(0.0001)	−24.387	466.4	−6762.8	38261.7	−1.372	0.334	−0.046	0.003	0.998

Table 2. Response surface estimates for detrended data

Lags	l	Intercept	(Std. Err.)	$1/T$	$1/T^2$	$1/T^3$	$1/T^4$	p/T	p^2/T	p^3/T	p^4/T	R^2
Fixed	0.01	−3.405	(0.0004)	−23.650	251.5	−4110.3	9281.0	0.404	1.327	−0.269	0.017	0.984
	0.05	−2.844	(0.0003)	−23.823	477.7	−8786.7	50347.0	0.532	1.060	−0.215	0.014	0.985
	0.10	−2.555	(0.0003)	−24.288	563.5	−10354.1	63554.9	0.584	0.922	−0.187	0.012	0.985
AIC	0.01	−3.399	(0.0005)	−32.582	807.0	−14439.0	72508.5	−6.624	1.676	−0.233	0.012	0.992
	0.05	−2.844	(0.0002)	−23.279	368.0	−4832.1	15025.6	−5.365	1.416	−0.193	0.010	0.997
	0.10	−2.557	(0.0002)	−22.079	377.0	−5816.4	32017.2	−4.522	1.270	−0.171	0.009	0.998
SIC	0.01	−3.412	(0.0004)	−12.420	−627.9	15994.2	−127428.4	−6.018	1.812	−0.233	0.011	0.995
	0.05	−2.847	(0.0002)	−16.716	48.8	−2174.0	30804.8	−4.392	1.694	−0.240	0.012	0.997
	0.10	−2.558	(0.0002)	−18.829	262.8	−6505.4	60506.8	−3.319	1.487	−0.222	0.011	0.998
GTS_5	0.01	−3.403	(0.0004)	−29.021	739.2	−17781.2	132079.9	−5.268	1.261	−0.173	0.009	0.993
	0.05	−2.842	(0.0002)	−25.286	632.8	−13630.8	97408.7	−2.697	0.583	−0.082	0.005	0.996
	0.10	−2.555	(0.0002)	−24.703	615.9	−12227.4	84332.6	−1.469	0.287	−0.043	0.003	0.996
GTS_{10}	0.01	−3.403	(0.0004)	−32.813	982.4	−22381.6	160071.8	−6.414	1.631	−0.217	0.011	0.992
	0.05	−2.841	(0.0003)	−26.826	714.7	−14731.7	100113.5	−4.423	1.090	−0.145	0.008	0.995
	0.10	−2.554	(0.0002)	−25.201	633.2	−12149.3	79983.8	−3.213	0.758	−0.099	0.005	0.996

We estimate 2,210 response surface regressions: 2 models multiplied by 5 criteria to select p multiplied by the 221 significance levels. The chosen functional form performs very well, with an average coefficient of determination of 0.994; in only 36 (out of 2,210) cases, it was below 0.95.

Tables 3 and 4 report the CVs estimated from the response surface models for selected values of T and p. For comparison purposes, we also include the CVs of the ERS test. As can be seen from the tables, for $T = 1000$ the implied asymptotic CVs from the response surface models fit in this article are close to those obtained by Elliott, Rothenberg, and Stock (1996). Interestingly, the implied CVs 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. In particular, for a given T, the implied CVs from the response surfaces decrease (in an absolute sense) in p when the augmentation order is fixed by the user, while they increase (in an absolute sense) in $p_{\max}$ when it is optimally determined using any of the data-dependent procedures being considered.

Table 3. Lag order and finite-sample CVs for demeaned data

Criterion to choose lag	p/T	$l = 0.01$ 100	$l = 0.01$ 200	$l = 0.01$ 1,000	$l = 0.05$ 100	$l = 0.05$ 200	$l = 0.05$ 1,000	$l = 0.10$ 100	$l = 0.10$ 200	$l = 0.10$ 1,000
Fixed	0	−2.74	−2.66	−2.59	−2.13	−2.05	−1.96	−1.82	−1.73	−1.64
	2	−2.72	−2.65	−2.59	−2.11	−2.04	−1.96	−1.80	−1.72	−1.64
	4	−2.69	−2.63	−2.58	−2.08	−2.02	−1.96	−1.78	−1.71	−1.64
	6	−2.67	−2.62	−2.58	−2.06	−2.01	−1.96	−1.76	−1.70	−1.64
AIC	0	−2.75	−2.66	−2.59	−2.13	−2.04	−1.96	−1.82	−1.73	−1.64
	2	−2.81	−2.70	−2.59	−2.17	−2.06	−1.97	−1.84	−1.74	−1.64
	4	−2.84	−2.71	−2.60	−2.17	−2.07	−1.97	−1.85	−1.74	−1.64
	6	−2.86	−2.72	−2.60	−2.18	−2.07	−1.97	−1.84	−1.74	−1.64
SIC	0	−2.73	−2.65	−2.58	−2.12	−2.04	−1.96	−1.82	−1.73	−1.64
	2	−2.77	−2.67	−2.59	−2.14	−2.05	−1.96	−1.83	−1.73	−1.64
	4	−2.77	−2.67	−2.59	−2.13	−2.05	−1.96	−1.82	−1.73	−1.64
	6	−2.77	−2.67	−2.59	−2.13	−2.04	−1.96	−1.81	−1.72	−1.64
GTS_5	0	−2.75	−2.66	−2.59	−2.13	−2.05	−1.96	−1.82	−1.73	−1.64
	2	−2.79	−2.69	−2.59	−2.15	−2.05	−1.96	−1.83	−1.73	−1.64
	4	−2.82	−2.70	−2.59	−2.16	−2.06	−1.97	−1.83	−1.74	−1.64
	6	−2.83	−2.71	−2.60	−2.17	−2.06	−1.97	−1.84	−1.74	−1.64
GTS_{10}	0	−2.75	−2.67	−2.59	−2.13	−2.04	−1.96	−1.82	−1.73	−1.64
	2	−2.81	−2.70	−2.59	−2.16	−2.06	−1.97	−1.84	−1.74	−1.64
	4	−2.84	−2.71	−2.60	−2.17	−2.07	−1.97	−1.84	−1.74	−1.64
	6	−2.85	−2.72	−2.60	−2.18	−2.07	−1.97	−1.85	−1.74	−1.64
ERS	0	−2.59	−2.58	−2.57	−1.94	−1.94	−1.94	−1.61	−1.62	−1.62

Note: ERS denotes the MacKinnon (1991) CVs for the no-constant case.

Table 4. Lag order and finite-sample CVs for detrended data

Criterion to		$l = 0.01$			$l = 0.05$			$l = 0.10$		
choose lag	p/T	100	200	1,000	100	200	1,000	100	200	1,000
Fixed	0	−3.62	−3.52	−3.43	−3.04	−2.95	−2.87	−2.75	−2.66	−2.58
	2	−3.58	−3.50	−3.42	−3.00	−2.93	−2.86	−2.72	−2.65	−2.58
	4	−3.52	−3.47	−3.42	−2.95	−2.91	−2.86	−2.67	−2.62	−2.57
	6	−3.48	−3.45	−3.41	−2.92	−2.89	−2.85	−2.64	−2.61	−2.57
AIC	0	−3.66	−3.54	−3.43	−3.04	−2.95	−2.87	−2.75	−2.66	−2.58
	2	−3.74	−3.58	−3.44	−3.11	−2.98	−2.87	−2.80	−2.68	−2.58
	4	−3.77	−3.60	−3.44	−3.13	−2.99	−2.88	−2.81	−2.69	−2.59
	6	−3.80	−3.61	−3.44	−3.14	−3.00	−2.88	−2.82	−2.69	−2.59
SIC	0	−3.58	−3.49	−3.43	−3.01	−2.93	−2.86	−2.73	−2.65	−2.58
	2	−3.65	−3.52	−3.43	−3.05	−2.95	−2.87	−2.75	−2.66	−2.58
	4	−3.66	−3.52	−3.43	−3.04	−2.94	−2.87	−2.73	−2.65	−2.58
	6	−3.65	−3.52	−3.43	−3.03	−2.94	−2.87	−2.72	−2.64	−2.58
GTS_5	0	−3.64	−3.53	−3.43	−3.04	−2.95	−2.87	−2.75	−2.66	−2.58
	2	−3.70	−3.57	−3.44	−3.08	−2.97	−2.87	−2.77	−2.67	−2.58
	4	−3.73	−3.58	−3.44	−3.10	−2.98	−2.87	−2.78	−2.68	−2.58
	6	−3.75	−3.59	−3.44	−3.11	−2.99	−2.87	−2.79	−2.69	−2.58
GTS_{10}	0	−3.65	−3.54	−3.43	−3.05	−2.96	−2.87	−2.75	−2.67	−2.58
	2	−3.73	−3.58	−3.44	−3.11	−2.99	−2.87	−2.79	−2.69	−2.58
	4	−3.76	−3.60	−3.45	−3.13	−3.00	−2.87	−2.81	−2.69	−2.58
	6	−3.77	−3.61	−3.45	−3.14	−3.00	−2.88	−2.82	−2.70	−2.58
ERS	0	−3.58	−3.46	−3.48	−3.03	−2.93	−2.89	−2.74	−2.64	−2.57

Note: ERS denotes the CVs simulated by Elliott, Rothenberg, and Stock (1996) in table 1.

MacKinnon (1994, 1996) points out that the residuals of the estimated response surfaces are expected to exhibit heteroskedasticity. Thus, to evaluate the robustness of the OLS results, we also considered estimation using the generalized method of moments procedure outlined by MacKinnon, which in the context of our simulation exercise averages the CVs across the 50 replications for each combination of T and p, and scaling all the variables in (2) by the standard error in these replications. Then OLS can be used to estimate the resulting equation using the rescaled variables. The results of applying this generalized method of moments procedure produces qualitatively similar results to those obtained with OLS, so they are not reported here.

To obtain approximate p-values of the ERS statistic, we follow MacKinnon (1994, 1996) by estimating the regression

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

where Φ^{-1} is the inverse of the cumulative standard normal distribution at each of the 221 quantiles, and $\widehat{\mathrm{CV}^l}$ is the fitted value from (2) at the l quantile. Following Harvey and van Dijk (2006), (3) is estimated by OLS using 15 observations, made up of the actual quantile and the 7 quantile observations on either side of the desired quantile.[4] Approximate p-values of the ERS test statistic can then be obtained as

$$p\text{-value} = \Phi\left[\widehat{\gamma}_0^l + \widehat{\gamma}_1^l \mathrm{ERS}\left(p\right) + \widehat{\gamma}_2^l \left\{\mathrm{ERS}\left(p\right)\right\}^2\right]$$

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

Finally, it is worth noting that in all the Monte Carlo simulation experiments, the error term was assumed to be white noise, which to some extent might be regarded as a weakness, because the error process can be quite general. To assess whether the estimated response surfaces continue to be reliable as we diverge from this white noise error specification, we carried out an additional set of simulations in which we computed the 1%, 5%, and 10% estimated CVs of the ERS statistics (for demeaned and detrended data) for two alternative specifications of the error process. The first follows a first-order autoregressive [AR(1)] process with an autoregressive coefficient of 0.9, while the second follows a first-order moving-average [MA(1)] process with a negative coefficient of -0.5. The sample sizes considered in these additional Monte Carlo simulations were $T = 100$, 200, and 400 observations, and the number of lags of the dependent variable in the test regression was assumed to be exogenously determined varying between 0 and 8 lags.

Taking the AR and MA error specifications as correct, we then proceeded to calculate, in percentage terms, how the CVs vary from those computed under the assumption of white noise errors. Averaging across T for the two model specifications, our findings (for brevity, not reported here but available upon request) indicate that if errors are AR(1) with an autoregressive coefficient equal to 0.9, the zero-lag CVs for white noise are seriously biased, but for lags 1 to 8, the percentage bias is never more than 1.5%

4. For $l \leq 0.004$ and $l \geq 0.996$, we use the actual quantile and the 14 observations closest to the desired quantile, because there will not be 7 observations on either side.

(and is generally lower). For the MA(1) specification with a coefficient equal to -0.5, the bias for lags equal to 0, 1, and 2 is sizable but less than 10% for higher lags.

These additional results highlight the importance of capturing serial correlation patterns that may be present in the error term of the series under consideration, while indicating that, for reasonable choices of the lag parameter, the CVs reported by our command are robust to misspecification of the error process.

5 The ersur command

The command `ersur` calculates the ERS test statistic; its associated finite-sample CVs for $l = 0.01$, 0.05, and 0.10; and its approximate p-values. The estimation of CVs and approximate p-values permits different combinations of number of observations, T, and lag order in the test regression, p, where the latter can be either specified by the user or optimally selected using a data-dependent procedure.

5.1 Syntax

Before using the command `ersur`, and similar to other Stata time-series commands, it is necessary to `tsset` the data. Then,

`ersur` *varname* [*if*] [*in*] [, `noprint` `maxlag(`*integer*`)` `trend`]

varname may not contain gaps. *varname* may contain time-series operators. The command may be applied to one unit of a panel.

5.2 Options

`noprint` specifies that the results be returned but not printed.

`maxlag(`*integer*`)` sets the number of lags to be included in the test regression to account for residual serial correlation. By default, `ersur` sets the number of lags following Schwert (1989), with the formula `maxlag()` $= \text{int}\{12(T/100)^{0.25}\}$, where T is the total number of observations.

`trend` specifies the modeling of intercepts and trends. By default, `ersur` considers *varname* to be a nonzero mean stochastic process; in this case, Elliott, Rothenberg, and Stock (1996) recommend demeaning the data using GLS. If the `trend` option is specified, `ersur` assumes that *varname* is a nonzero trend stochastic process, in which case Elliott, Rothenberg, and Stock recommend detrending the data using GLS.

5.3 Stored results

ersur stores the following in r():

Scalars			
r(N)	number of observations	r(maxp)	last time period used in the test regression
r(minp)	first time period used in the test regression		
Macros			
r(varname)	variable name	r(tsfmt)	time-series format of the time variable
r(treat)	demeaned or detrended, depending on the trend option		
Matrices			
r(results)	results matrix, 5 x 6		

6 Empirical application

The expectations theory of the term structure of interest rates implies that interest rates of different maturities maintain a long-run equilibrium relationship, so the interest-rate spread does not exhibit a tendency to grow systematically over time; see, for instance, Campbell and Shiller (1987), Stock and Watson (1988), and Hall, Anderson, and Granger (1992) for early applications. This is essentially a question of whether interest rate spreads, defined as the differences between long-term and short-term interest rates, may be characterized as stationary stochastic processes.

In this section, we illustrate the use of the ersur command to address this question. We use monthly data on the United States Treasury interest rate series at nine maturities over the sample period 1993m10 to 2013m3, which yields a total of 234 time observations for each series. The specific maturities considered in the analysis correspond to the 3-month, 6-month, 1-year, 3-year, 5-year, 7-year, 10-year, 20-year, and 30-year constant maturity rates, as retrieved from the Federal Reserve Economic Data provided by the Economic Research Division of the Federal Reserve Bank of St. Louis.[5] The interest rates are denoted r_3, r_6, r_{12}, r_{36}, r_{60}, r_{84}, r_{120}, r_{240}, and r_{360}.

We begin by loading the dataset and declaring that it has a time-series format:

```
. use usrates

. tsset date, monthly
        time variable:  date, 1993m10 to 2013m3
                delta:  1 month
```

Next suppose we want to test whether the interest rate spread between r_6 and r_3 (which we shall denote as s_6) contains a unit root against the alternative, that it is a stationary process. Given that s_6 has a nonzero mean, the relevant ERS statistic is based on GLS demeaned data, the default for ersur. Setting $p = 3$ lags, the results of applying the command ersur are as follows:

5. The task of downloading the time series from the Federal Reserve Economic Data database was greatly simplified using the command freduse; see Drukker (2006).

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

Table 5 summarizes the results of applying the ERS test to $s_6 = r_6 - r_3$, $s_{12} = r_{12} - r_3$, and so on, until $s_{360} = r_{360} - r_3$, where the interest rate spreads have been previously demeaned using GLS. We set $p = 3$ when the lag length is fixed and $p_{\max} = 3$ when it is optimally determined. All in all, the ERS test results support the validity of the term structure of interest rates. However, in the case of the longest maturity differential between short- and long-run rates, that is, s_{360}, we fail to reject the presence of a unit root in the corresponding spread at the 5% significance level.

Table 5. Applying the ERS test to interest rate differentials

Variable		Fixed	AIC	SIC	GTS_5	GTS_{10}
s_6	Lags	3	0	0	0	2
	ERS test	−3.780	−4.298	−4.298	−4.298	−3.562
	p-value	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
s_{12}	Lags	3	3	1	3	3
	ERS test	−4.013	−4.013	−3.904	−4.013	−4.013
	p-value	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
s_{24}	Lags	3	1	1	1	3
	ERS test	−3.574	−3.504	−3.504	−3.504	−3.574
	p-value	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]
s_{36}	Lags	3	1	1	1	3
	ERS test	−3.262	−3.078	−3.078	−3.078	−3.262
	p-value	[0.001]	[0.003]	[0.003]	[0.003]	[0.002]
s_{60}	Lags	3	3	1	3	3
	ERS test	−2.739	−2.739	−2.528	−2.739	−2.739
	p-value	[0.007]	[0.009]	[0.014]	[0.008]	[0.009]
s_{84}	Lags	3	3	1	3	3
	ERS test	−2.485	−2.485	−2.278	−2.485	−2.485
	p-value	[0.015]	[0.017]	[0.028]	[0.017]	[0.017]
s_{120}	Lags	3	3	1	3	3
	ERS test	−2.258	−2.258	−2.065	−2.258	−2.258
	p-value	[0.028]	[0.030]	[0.046]	[0.030]	[0.030]
s_{240}	Lags	3	3	3	3	3
	ERS test	−2.061	−2.061	−2.061	−2.061	−2.061
	p-value	[0.045]	[0.049]	[0.047]	[0.048]	[0.048]
s_{360}	Lags	3	3	3	3	3
	ERS test	−1.983	−1.983	−1.983	−1.983	−1.983
	p-value	[0.054]	[0.058]	[0.056]	[0.057]	[0.058]

7 Conclusions

We fit response surface models for the CVs of the Elliott, Rothenberg, and Stock (1996) unit-root test. The models are fit as a function of the number of observations, T, and lags of the dependent variable in the test regressions, p, for 221 significance levels. The lag length can be determined either exogenously by the user or endogenously using a data-dependent procedure. The results suggest that the method used to select the order of the augmentation affects the finite-sample CVs.

The command `ersur` can easily be used to calculate the ERS test statistic, finite-sample CVs, and approximate p-values. As an empirical application, `ersur` is illustrated by examining whether the theory of the term structure of interest rates holds among a set of U.S. interest rates.

8 Acknowledgments

Jesús Otero would like to acknowledge financial support received from the Universidad del Rosario through Fondo de Investigaciones de la Universidad del Rosario. We are also grateful to an anonymous referee and participants at the 2017 Stata Conferences in Baltimore and London for useful comments and suggestions. The usual disclaimer applies.

9 References

Abbott, A., and G. De Vita. 2012. Pairwise convergence of district-level house prices in London. *Urban Studies* 49: 721–740.

Campbell, J. Y., and P. Perron. 1991. Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomics Annual 1991* 6: 141–201.

Campbell, J. Y., and R. J. Shiller. 1987. Cointegration and tests of present value models. *Journal of Political Economy* 95: 1062–1088.

Cheung, Y.-W., and K. S. Lai. 1995a. Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business and Economic Statistics* 13: 277–280.

———. 1995b. Lag order and critical values of a modified Dickey–Fuller test. *Oxford Bulletin of Economics and Statistics* 57: 411–419.

DeJong, D. N., J. C. Nankervis, N. E. Savin, and C. H. Whiteman. 1992. The power problems of unit root test in time series with autoregressive errors. *Journal of Econometrics* 53: 323–343.

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.

Drukker, D. M. 2006. Importing Federal Reserve economic data. *Stata Journal* 6: 384–386.

Elliott, G., T. J. Rothenberg, and J. H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836.

Hall, A. 1994. Testing for a unit root in time series with pretest data-based model selection. *Journal of Business and Economic Statistics* 12: 461–470.

Hall, A. D., H. M. Anderson, and C. W. J. Granger. 1992. A cointegration analysis of treasury bill yields. *Review of Economics and Statistics* 74: 116–126.

Harvey, D. I., and D. van Dijk. 2006. Sample size, lag order and critical values of seasonal unit root tests. *Computational Statistics and Data Analysis* 50: 2734–2751.

Im, K. S., M. H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74.

Kapetanios, G., and Y. Shin. 2008. GLS detrending-based unit root tests in nonlinear STAR and SETAR models. *Economics Letters* 100: 377–380.

Kapetanios, G., Y. Shin, and A. Snell. 2003. Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics* 112: 359–379.

Le Pen, Y. 2011. A pair-wise approach to output convergence between European regions. *Economic Modelling* 28: 955–964.

Leybourne, S. J. 1995. Testing for unit roots using forward and reverse Dickey–Fuller regressions. *Oxford Bulletin of Economics and Statistics* 57: 559–571.

MacKinnon, J. G. 1991. Critical values for cointegration tests. In *Long-Run Economic Relationships: Readings in Cointegration*, ed. R. F. Engle and C. W. J. Granger, 267–276. Oxford: Oxford University Press.

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

———. 1996. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* 11: 601–618.

Nelson, C. R., and C. I. Plosser. 1982. Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics* 10: 139–162.

Ng, S., and P. Perron. 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90: 268–281.

Otero, J., and J. Smith. 2012. Response surface models for the Leybourne unit root tests and lag order dependence. *Computational Statistics* 27: 473–486.

Pesaran, M. H. 2007a. A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* 138: 312–355.

———. 2007b. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22: 265–312.

Pesaran, M. H., R. P. Smith, T. Yamagata, and L. Hvozdyk. 2009. Pairwise tests of purchasing power parity. *Econometric Reviews* 28: 495–521.

Said, S. E., and D. A. Dickey. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71: 599–607.

Schwert, G. W. 1989. Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7: 147–159.

Stock, J. H., and M. W. Watson. 1988. Testing for common trends. *Journal of the American Statistical Association* 83: 1097–1107.

About the authors

Jesús Otero is a professor of economics at the Universidad del Rosario in Colombia. He received his PhD in economics from Warwick University in the United Kingdom. His field of study is applied time-series econometrics.

Christopher F. Baum is a professor of economics and social work at Boston College. He is an associate editor of the *Stata Journal* and the *Journal of Statistical Software*. Baum founded and manages the Boston College Statistical Software Components (SSC) archive at RePEc (http://repec.org). His recent research has addressed issues in social epidemiology and the labor market progress of refugees in developed economies.