Tatra Mt. Math. Publ. **17** (1999), 273–282

TATRA MOUNTAINS
Mathematical Publications

LINEAR MODEL WITH AUTOREGRESSIVE ERRORS: ESTIMATION AND TESTING OF THE PARAMETERS

Viktor Witkovský

ABSTRACT. The paper gives a brief overview of the classical procedures for estimation and testing of the parameters in the linear model with autoregressive errors. The estimator of the autoregressive coefficient is an important part of the test statistics used for testing the unit root hypothesis. The statistical properties of the selected estimators and test statistics are compared by simulations.

1. Introduction

We compare two estimators of the autoregressive coefficient: a) the traditional OLSE (Ordinary Least Squares Estimator), and b) the LMINQE (Linearized Minimum Norm Quadratic Estimator) which is based on the results of the variance components estimation theory, see, e.g., Rao and Kleffe ([6]), Azaïs *et al.* ([1]), and Witkovský ([7]). Classical unit root test-statistics are constructed by help of such estimators, e.g., Dickey-Fuller t-, F-, and K-statistic. See Dickey and Fuller ([2]) and Fuller ([3]). The distributional properties of the estimators are very similar for sample size 100, so, because of simplicity, the traditional OLSE is superior.

2. The model

Let $\{y_t\}$ denote the sequence of observations measured in discrete time $t = 0, \pm 1, \pm 2, \dots$. We will consider the model

2000 Mathematics Subject Classification: 62J05, 62F10.
Keywords: linear model, autoregressive errors, unit root test, Dickey-Fuller test, minimum norm quadratic estimation.

This work was supported by the Slovak Grant Agency, grants 1/4196/97, and 95/5305/468.

$$
\begin{aligned}
y_t &= \alpha_0 + \alpha_1 t + x_t\,,\\
x_t &= \varrho x_{t-1} + u_t\,,
\end{aligned}
\tag{1}
$$

where $\alpha = (\alpha_0, \alpha_1)'$ is a vector parameter and $\{x_t\}$ represents the autoregressive errors. We will assume that the forcing process $\{u_t\}$ is a weakly stationary autoregressive process of the order p and $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$. In particular, we assume that $A(L)u_t = \varepsilon_t$, (L denotes the lag operator), where $A(L) = 1 - a_1L - a_2L^2 - \cdots - a_pL^p$, and the roots λ_j, $j = 1, \dots, p$, of the characteristic polynomial $A(z) = \Pi_{j=1}^p(1 - \lambda_j z)$ fulfill the inequality restrictions $|\lambda_j| < 1$.

Under the given specifications we get

$$
(1 - \varrho L)A(L)x_t = \varepsilon_t\,. \tag{2}
$$

So, the autoregressive coefficient ϱ is the root of the characteristic polynomial of the process $\{x_t\}$.

The sequence of observations $\{y_t\}$ is said to be TS (trend-stationary) if $\{x_t\}$ is a stationary process (of the order $p+1$). On the other hand, $\{y_t\}$ is said to be DS (difference-stationary) if $\{x_t\}$ is a non-stationary process, but $\{\Delta x_t\}$, $\Delta x_t = x_t - x_{t-1}$, is a stationary process of the order p. Note, that in the DS case, differencing of $\{y_t\}$ is indispensable to get a (trend-)stationary process. In the model (1), the value of the autoregressive coefficient (the dominating root) ϱ is an indicator of DS vs. TS. The problem of DS vs. TS is also known as *the unit root problem.*

In the following we will consider $p+1+T$ observations: $y_{-p}, \dots, y_0, y_1, \dots, y_T$. Note, that if $\{u_t\}$ is a stationary AR(p) process, then the conditional distribution is determined by the distribution of $\varepsilon_1, \dots, \varepsilon_T$. Under these assumptions the model (1) can be rewritten equivalently (pre-multiply the equations in (1) by $(1 - \varrho L)$) as

$$
y_t = \gamma + \delta t + \varrho y_{t-1} + u_t\,, \qquad t = 1, \dots, T\,, \tag{3}
$$

where $\gamma = \alpha_0(1 - \varrho) + \alpha_1\varrho$ and $\delta = \alpha_1(1 - \varrho)$. If $\varrho = 1$, then α_0 cannot be determined from $(\gamma, \delta, \varrho)'$, and $\alpha_1 = \gamma$.

If the order, say p, of the stationary autoregressive process $\{u_t\}$ is known, the model (1) can be transformed to

$$
(1 - \varrho L)A(L)\bigl[y_t - (\alpha_0 + \alpha_1 t)\bigr] = \varepsilon_t\,. \tag{4}
$$

Using one-to-one mapping between $(\varrho, a_1, \dots, a_p)'$ and $(\varrho^*, a_1^*, \dots, a_p^*)$ given by the relation

$$
(1 - \varrho L)A(L) = (1 - \varrho^* L) - (1 - L)A^*(L)\,, \tag{5}
$$

where $A(L)^* = a_1^* L + \cdots + a_p^* L^p$, we get

$$\big[(1-\varrho^* L) - (1-L)A^*(L)\big]\,\big[y_t - (\alpha_0 + \alpha_1 t)\big] = \varepsilon_t\,,$$
$$(y_t - \varrho^* y_{t-1}) - (a_1^* \Delta y_{t-1} + \cdots + a_p^* \Delta y_{t-p}) = \alpha_0^* + \alpha_1^* t + \varepsilon_t\,, \tag{6}$$

where α_0^* and α_1^* denote the transformed coefficients α_0 and α_1. In particular, $\alpha_0^* = \alpha_0(1-\varrho)(1-a_1-\cdots-a_p)+\alpha_1\varrho(1-a_1-\cdots-a_p)+\alpha_1(a_1+2a_2+\cdots+pa_p)$ and $\alpha_1^* = \alpha_1(1-\varrho)(1-a_1-\cdots-a_p)$ and $1-\varrho^* = (1-\varrho)(1-a_1-\cdots-a_p)$. So, ϱ^* is an indicator of DS vs. TS. The transformation (5) is known as the *Fuller transform.* The transformed model (6) is usually presented as

$$\Delta y_t = \alpha_0^* + \alpha_1^* t - (1-\varrho^*)y_{t-1} + a_1^* \Delta y_{t-1} + \cdots + a_p^* \Delta y_{t-p} + \varepsilon_t\,, \tag{7}$$

for $t = 1,\ldots,T$.

Starting with the initial value $y_0 = \alpha_0 + x_0$, by subsequent iterations, we get the other representation of the model (1):

$$y_t = \alpha_0(1-\varrho^t) + \alpha_1 t + \varrho^t y_0 + \sum_{i=0}^{t-1} \varrho^i u_{t-i}\,, \qquad t = 1,\ldots,T\,. \tag{8}$$

Let $y_t^* = y_t - \varrho^t y_0$, $y_\varrho^* = (y_1^*,\ldots,y_T^*)'$, $\nu_t = \sum_{i=0}^{t-1} \varrho^i u_{t-i}$ and $\nu = (\nu_1,\ldots,\nu_T)'$. Immediately, we get the expectation

$$E(y_\varrho^*) = Z_\varrho \alpha\,, \tag{9}$$

where the ith row of Z_ϱ is $(1-\varrho^i, i)$, and the variance-covariance matrix is

$$\mathrm{Var}(y_\varrho^*) = \mathrm{Var}(\nu) = V(\varrho, a, \sigma_\varepsilon^2) = \Phi_\varrho' \Gamma_{a,\sigma_\varepsilon^2} \Phi_\varrho\,, \tag{10}$$

where $a = (a_1,\ldots,a_p)'$, $\Gamma_{a,\sigma_\varepsilon^2}$ is the variance-covariance matrix of the vector u ($\{u_t\}$ is a stationary AR(p) process), and Φ_ϱ^{-1} is a transformation matrix with diagonal elements equal to 1 and subdiagonal elements equal to $-\varrho$, otherwise zeros.

The variance-covariance matrix $V(\varrho, a, \sigma_\varepsilon^2)$ is n.n.d. (nonnegative definite), and as a function of its parameters belongs to the class of at least twice differentiable functions.

3. Estimation and testing

The different representations of the model (1) lead to different estimators of the autoregressive coefficient ϱ. The test statistics for testing DS vs. TS are based on such estimators.

3.1 The ordinary least squares estimator.

Consider the model (3). Let X denote the appropriate design matrix and let $\hat{\beta} = (X'X)^{-1}X'y$ denote the ordinary least squares estimator (OLSE) of $\beta = (\gamma, \delta, \varrho)'$. Then $\hat{\beta}_3 = \hat{\varrho}$ is the OLSE of ϱ. Similarly, if the order of the AR(p) is known we can use the Fuller transformation and the model representation (7). Let X^* denote the appropriate design matrix and $\hat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y$ denote the OLSE of β^* in the model (7). Then $\hat{\varrho}^* = 1 - \hat{\beta}_3^*$ is the OLSE of ϱ^*.

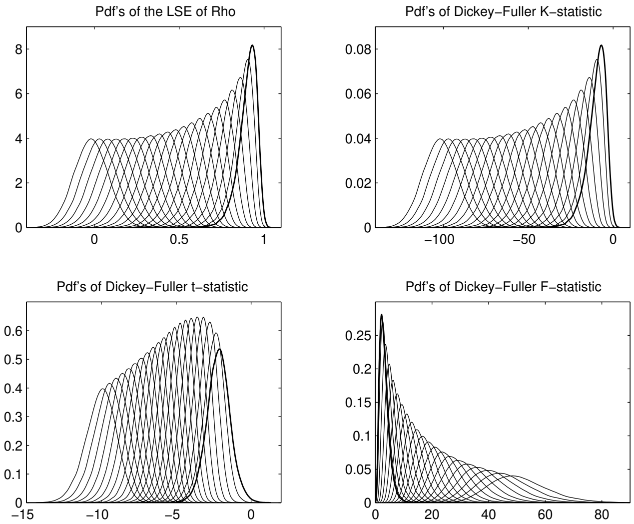

FIGURE 1. Estimated quantile functions of the OLSE of the parameter ϱ, sample size T=100. The quartiles corresponds to the levels $\alpha = [0.01, 0.025, 0.05, 0.1, \ldots, 0.9, 0.95, 0.975, 0.99]$. Estimated functions are based on $N = 50000$ simulations calculated for each true value of $\varrho \in \langle 0 : 0.05 : 1 \rangle$. The auxiliary diagonal is plotted in order to detect the bias of the estimator.

Under the TS alternative a special type of consistency of the OLSE $\hat{\beta}$ in the model (3) can be established, for proof see, e.g., Hatanaka ([4]).

THEOREM 1. *Let us consider the model* (1). *Let* $\{x_t\}$ *be stationary autoregressive process with zero mean, variance* σ_x^2, *and with the serial correlation coefficients* $\mathrm{Corr}(x_t, x_{t-i}) = \varrho_i$, $i = 1, \ldots$ *Then the* OLSE $\hat{\beta} = (X'X)^{-1}X'y$ *of* β *in the model* (3) *is a consistent estimator of* $\tilde{\beta} = \big(\alpha_0(1-\varrho_1)+\alpha_1\varrho_1, \alpha_1(1-\varrho_1), \varrho_1\big)'$, *i.e.,* $\operatorname*{plim}_{T\to\infty} \hat{\beta} = \tilde{\beta}$.

3.2 Minimum norm quadratic estimator.

Following the MINQE method to estimation variance components developed by Rao ([5]) and modified by Azaïs *et al.* ([1]), it is possible to derive the estimators, LMINQE's (Linearized minimum norm quadratic estimator), of variance and covariance components in the model (9). The following theorem gives the explicit form of the estimator under the assumption that the process $\{u_t\}$ is a white noise, i.e., $u_t \equiv \varepsilon_t$.

THEOREM 2. *Let us consider the model* (9). *Let* $e = (e_1, \dots, e_T)'$ *denote the vector of residuals* $e = Y - Z(Z'V_0^{-1}Z)^{-}Z'V_0^{-1}Y$, *where* $Y = y^*_{\varrho_0}$, $Z = Z_{\varrho_0}$, *and* $V_0 = V(\varrho_0, 0, \sigma^2_{\varepsilon 0})$, *with* $|\varrho_0| \leq 1$ *and* $\sigma^2_{\varepsilon 0} > 0$ *denoting the prior values of the parameters. The explicit form of* LMINQE's *of* ϱ *and* σ^2_ε *is then given by*

$$\tilde{\varrho} = \varrho_0 + \frac{1}{\sigma^2_{\varepsilon 0}\left(\sum\limits_{t=0}^{T-1}(T-1-t)\varrho_0^{2t}\right)}\left\{\sum_{t=1}^{T-1} e_t e_{t+1} - \varrho_0 \sum_{t=1}^{T-1} e_t^2\right\} \tag{11}$$

$$\tilde{\sigma_\varepsilon}^2 = \frac{1}{T}\left\{(1+\varrho_0^2)\sum_{t=1}^{T-1} e_t^2 - 2\varrho_0 \sum_{t=1}^{T-2} e_t e_{t+1} + e_T^2 - \varrho_0 e_{T-1} e_T\right\}. \tag{12}$$

P r o o f . Let us consider the model (9). To get the LMINQE's of ϱ and σ^2_ε we should derive the minimum norm quadratic estimators of the parameters of the locally linearized model, i.e., of the model with linearized covariance structure (with respect to the parameters). For that we will use a Taylor expansion of the covariance matrix $V(\varrho, 0, \sigma^2_\varepsilon)$, locally at the point $(\varrho_0, 0, \sigma^2_{\varepsilon 0})$:

$$V(\varrho, 0, \sigma^2_\varepsilon) \approx V_0 + (\varrho - \varrho_0)V_1 + (\sigma^2_\varepsilon - \sigma^2_{\varepsilon 0})V_2\,, \tag{13}$$

where $V_0 = V(\varrho_0, 0, \sigma^2_{\varepsilon 0})$, and

$$V_1 = \left.\frac{\partial V(\varrho, 0, \sigma^2_\varepsilon)}{\partial \varrho}\right|_{\varrho_0, 0, \sigma^2_{\varepsilon 0}}, \qquad V_2 = \left.\frac{\partial V(\varrho, 0, \sigma^2_\varepsilon)}{\partial \sigma^2_\varepsilon}\right|_{\varrho_0, 0, \sigma^2_{\varepsilon 0}}. \tag{14}$$

Under the given assumptions we have $V_0 = \sigma^2_{\varepsilon 0}$, so we get

$$V(\varrho, 0, \sigma^2_\varepsilon) \approx (\varrho - \varrho_0)V_1 + \sigma^2_\varepsilon V_2\,. \tag{15}$$

The invariant MINQE of the parameters $\varrho - \varrho_0$ and σ^2_ε in the linearized model is given as

$$\begin{pmatrix} \widehat{\varrho - \varrho_0} \\ \sigma^2_\varepsilon \end{pmatrix} = K^{-1}\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \tag{16}$$

where K denotes the criterion matrix of the invariant MINQE and q_1 and q_2 are the MINQE-quadratics, for more details on MINQE see Rao and Kleffe

([6]). Here we have

$$K = \begin{pmatrix} 2\sum_{t=0}^{T-1}(T-1-t)\varrho_0^{2t} & 0 \\ 0 & T/(\sigma_{\varepsilon 0}^2)^2 \end{pmatrix}, \tag{17}$$

$$q_1 = \frac{2}{\sigma_{\varepsilon 0}^2}\left\{-\varrho_0 \sum_{t=1}^{T-1} e_t^2 + \sum_{t=1}^{T-1} e_t e_{t+1}\right\}, \tag{18}$$

$$q_2 = \frac{1}{(\sigma_{\varepsilon 0}^2)^2}\left\{(1+\varrho_0^2)\sum_{t=1}^{T-1} e_t^2 - 2\varrho_0 \sum_{t=1}^{T-2} e_t e_{t+1} + e_T^2 - \varrho_0 e_{T-1} e_T\right\}. \tag{19}$$

Finally, the LMINQE's are given by the relations $\tilde{\varrho} = \varrho_0 + (\widehat{\varrho - \varrho_0})$ and $\tilde{\sigma}_\varepsilon^2 = \widehat{\sigma}_\varepsilon^2$. □

3.3 Test statistics for testing DS against TS

The OLSE of ϱ is suitable for constructing the test statistic for testing DS hypothesis against the TS alternative. Note that the OLSE ($\hat{\varrho}$ and/or $\hat{\varrho}^*$) is an indicator of DS vs. TS. So, the K-statistic defined by $K = T(\hat{\varrho} - 1)$ is the simplest statistic used to test DS against TS. We will use the notation K and K^*, respectively, to denote the statistics based on the model (3) and (7), respectively.

The other statistic used to test DS vs. TS hypothesis is the Dickey-Fuller t-statistic that is formally defined as the usual t-statistic in the ordinary linear regression model, i.e., $t = (\hat{\varrho} - 1)/s_{\hat{\varrho}}$ with $s_{\hat{\varrho}}^2 = \{s^2(X'X)^{-1}\}_{33}$, and $s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(T-k)$, where $k = 3$ is the rank of X. The test criterion based on the distribution of the t-statistic is called the Dickey-Fuller test. The test criterion based on t^*-statistic of the model (7) is called the augmented Dickey-Fuller test of DS against TS.

Under the null (DS) hypothesis it holds that $\varrho = 1$. This implies $(\delta, \varrho)' = (0,1)'$ in the model (3) and $(\alpha_1^*, 1 - \varrho^*)' = (0,0)'$ in the model (7). The Dickey-Fuller F-statistic (augmented Dickey-Fuller F^*-statistic) is defined as $F = \left[(\mathrm{RSS}_R - \mathrm{RRS}_U)/r\right]/\left[\mathrm{RSS}_U/(T-k)\right]$, where RSS_R and RSS_U denote the sum of the squared residuals from the restricted and unrestricted models, k is the rank of the design matrix (X or X^*) and $r = 2$ denotes the number of restricted parameters.

The limiting distributions of the above statistics (under null hypothesis) are functionals of the standard Wiener process. Let $w(t)$ denote the standard Wiener process, i.e., a process defined on a continuous time $0 \le t \le 1$ that satisfies:

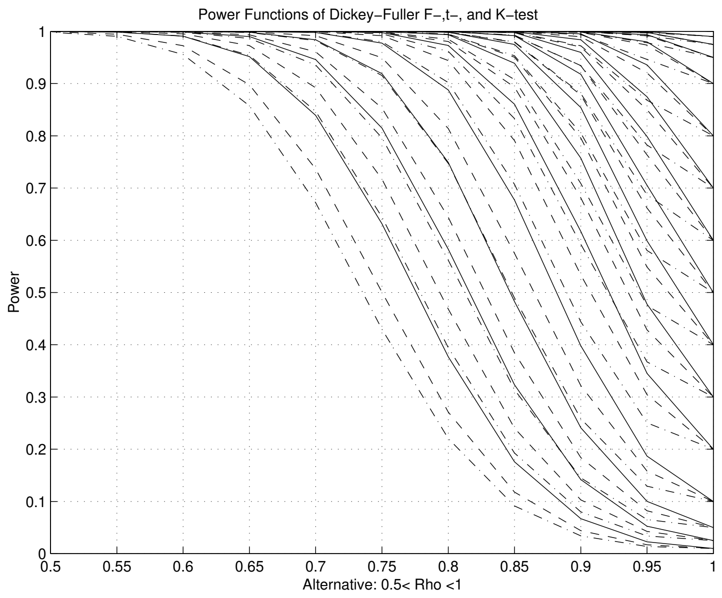

FIGURE 2. Estimated probability density functions of the OLSE of ϱ and Dickey-Fuller K-, t-, and F-statistic, sample size T=100. Estimated functions are based on N=50000 simulations calculated for each true value of $\varrho \in \langle 0 : 0.05 : 1\rangle$.

a) $w(0) = 0$, b) $w(t) \sim N(0,t)$ for all $0 \le t \le 1$, and c) $w(t_2) - w(t_1)$ is independent with $w(t) - w(t_3)$ for all $0 \le t_1 < t_2 < t_3 \le t \le 1$. From that we get $\mathrm{Cov}\big(w(t_1), w(t_2)\big) = \min(t_1, t_2)$. If $[rT]$ denotes the integer part of rT, $0 \le r \le 1$, and $X_T(r) = \Big(1/\sqrt{T\sigma_\varepsilon^2}\Big) \sum_{t=1}^{[rT]} \varepsilon_t$, where ε_t are i.i.d. random variables with $\varepsilon_t \sim IID(0,\sigma^2)$, then the following holds true as $T \to \infty$: a) $X_T(r) \xrightarrow{D} w(r)$, b) $\big(X_T(r_1), \dots, X_T(r_m)\big) \xrightarrow{D} \big(w(r_1), \dots, w(r_m)\big)$ for any (countable) set of numbers $0 \le r_i \le 1$, and c) if $f(\cdot)$ is an appropriate continuous mapping then $f\big(X_T(r)\big) \xrightarrow{D} f\big(w(r)\big)$. For more details see, e.g., Hatanaka ([4]).

THEOREM 3. *Let us consider the model* (3). *Moreover, let us assume that* $\{u_t\}$ *is an* AR(0), *i.e.,* $u_t \equiv \varepsilon_t$. *Then, under the* DS *hypothesis (true* $\varrho = 1$*), the test statistics* K, t, *and* F, *respectively, converge in distribution to the following random variables (functions of the standard Wiener process):*

$$K \xrightarrow{D} \left(\int \tilde{w}(r)^2 \, dr\right)^{-1} \int \tilde{w}(r)\, dw(r)\,, \tag{20}$$

$$t \xrightarrow{D} \left(\int \tilde{w}(r)^2 \, dr\right)^{-1/2} \int \tilde{w}(r)\, dw(r)\,, \tag{21}$$

$$F \xrightarrow{D} \frac{1}{2}\left[12\left(\int (r-1/2)\,dw(r)\right)^2 + \left(\int \tilde{w}(r)^2\,dr\right)^{-1}\left(\int \tilde{w}(r)\,dw(r)\right)^2\right], \tag{22}$$

where $\tilde{w}(r) \equiv w(r) - \int w(s)\,ds - 12(r-1/2)\int (s-1/2)w(s)\,ds$.

The asymptotic distributions of t^* and F^*-statistic are the same as the asymptotic distributions of t and F-statistic, i.e., the distributions does not depend on the nuisance parameters $(a_1, \ldots, a_p)'$. On the other hand, this is not true for K^*-statistic which depends on the nuisance parameters if the true model is (7) instead of (3).

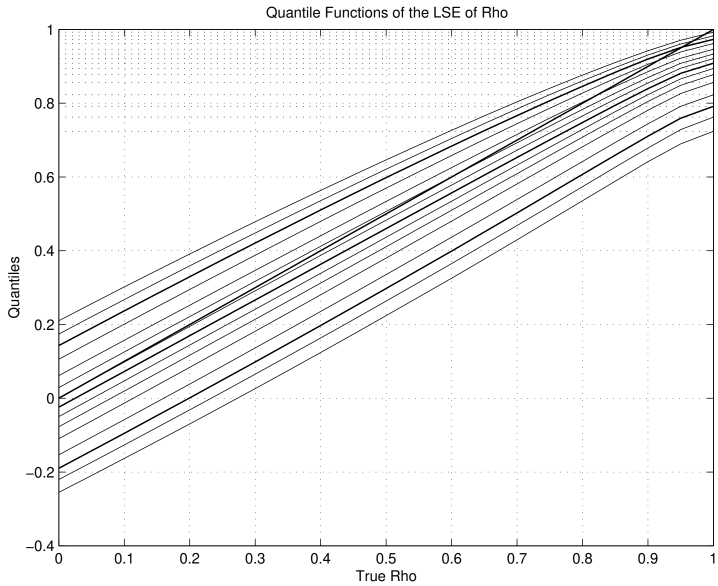

FIGURE 3. Estimated power functions of the Dickey-Fuller K-test (solid line), t-test (dashed line), and F-test (dash-dotted line), sample size T=100. Estimated functions are based on N=50000 simulations calculated for each true value of $\varrho \in \langle 0.5 : 0.05 : 1 \rangle$.

4. Simulation study

We have simulated the behavior of the OLSE, (iterated) LMINQE, and Dickey-Fuller K-, t-, and F-statistic in the model (3) for finite sample size T=100. Moreover, we have simulated the asymptotic distributions of Dickey-Fuller K-, t-, and F-test (under $H_0 : \varrho = 1$) based on a discrete approximation of the distributions derived in Theorem 3. All the reported results are based on N=50000 simulations for each true value of $\varrho \in \langle(0 : 0.05 : 1\rangle$, see Figure 1–3. Statistical properties (derived from our simulations) of the (iterated) LMINQE of ϱ are very similar to the properties of the OLSE for sample size 100. So, because of (computational) simplicity, the traditional OLSE is superior. Finally, we report the table with selected critical values of the Dickey Fuller tests (asymptotic and finite sample T=100) based on our simulations.

TABLE 1. Estimated critical values of the Dickey-Fuller F-test, t-test, and K-test based on the asymptotic distribution and on the finite sample distribution, (sample size T=100). Estimated values are based on N=50000 simulations calculated for true value $\varrho = 1$.

α	as. F	F	as. t	t	as. K	K
0.0100	8.2584	8.7767	−3.9547	−4.0704	−29.2042	−27.1759
0.0250	7.1154	7.4408	−3.6590	−3.7295	−25.0177	−23.4318
0.0500	6.2424	6.4708	−3.4047	−3.4593	−21.7279	−20.4929
0.1000	5.3153	5.4647	−3.1212	−3.1589	−18.1280	−17.3617
0.2000	4.3303	4.4188	−2.7880	−2.7995	−14.5231	−13.9830
0.3000	3.7179	3.7613	−2.5557	−2.5660	−12.2884	−11.8948
0.4000	3.2412	3.2766	−2.3565	−2.3604	−10.5782	−10.2488
0.5000	2.8367	2.8601	−2.1813	−2.1713	−9.1395	−8.8505
0.6000	2.4838	2.4751	−2.0036	−1.9868	−7.8227	−7.5851
0.7000	2.1486	2.1278	−1.8103	−1.7884	−6.5581	−6.3899
0.8000	1.7940	1.7750	−1.5847	−1.5583	−5.2714	−5.1464
0.9000	1.3956	1.3679	−1.2440	−1.2200	−3.7593	−3.7045
0.9500	1.1306	1.1149	−0.9280	−0.9054	−2.6759	−2.6008
0.9750	0.9506	0.9289	−0.6337	−0.6234	−1.7456	−1.7189
0.9900	0.7752	0.7624	−0.3039	−0.2985	−0.8029	−0.8058

REFERENCES

[1] AZAÏS, J. M.—BARDIN, A.—DHORNE, T.: *MINQE, maximum likelihood estimation and Fisher scoring algorithm for non linear variance models*, Statistics **24** (1993), 205–213.

[2] DICKEY, D. A.—FULLER, W. A.: *Distribution of the estimators for autoregressive time series with a unit root*, J. Amer. Statist. Assoc. **74** (1979), 427–431.

[3] FULLER, W. A.: *Introduction to Statistical Time Series*, J. Wiley, New York, 1976.

[4] HATANAKA, M.: *Time-Series-Based Econometrics. Unit Roots and Cointegration*, Advanced Texts in Econometrics. Oxford University Press, 1996.

[5] RAO, C. R.: *Estimation of variance and covariance components — MINQUE theory*, J. Multivariate Anal. **1** (1971), 257–275.

[6] RAO, C. R.—KLEFFE, J.: *Estimation of Variance Components and Applications*, North-Holland Series in Statistics and Probability, Vol. 3, Amsterdam, 1988.

[7] WITKOVSKÝ, V.: *On variance-covariance components estimation in linear models with AR(1) disturbances*, Acta Math. Univ. Comenian., Vol. LXV (1996), 129–139.

Received May 20, 1998

Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9
SK–842 19 Bratislava
SLOVAKIA
E-mail: umerwitk@savba.sk