Properties of estimated characteristic roots

Bent Nielsen
Department of Economics, University of Oxford
& Nuffield College, Oxford OX1 1NF, UK. Email: bent.nielsen@nuffield.ox.ac.uk

Heino Bohn Nielsen
Department of Economics, University of Copenhagen
Studiestræde 6, DK-1455 Copenhagen K. Email: heino.bohn.nielsen@econ.ku.dk

30 May 2008

Summary: Estimated characteristic roots in stationary autoregressions are shown to give rather noisy information about their population equivalents. This is remarkable given the central role of the characteristic roots in the theory of autoregressive processes. In the asymptotic analysis the problems appear when multiple roots are present as this imply a non-differentiability so the δ-method does not apply, convergence rates are slow, and the asymptotic distribution is non-normal. In finite samples this has a considerable influence on the finite sample distribution unless the roots are far apart. With increasing order of the autoregressions it becomes increasingly difficult to place the roots far apart giving a very noisy signal from the characteristic roots.

Keywords: Autoregression; Characteristic root.
JEL: C22.

1 Introduction

Estimated characteristic roots in stationary autoregressions are shown to give rather noisy information about their population equivalents. This is remarkable given the central role of the characteristic roots in the theory of autoregressive processes (see for instance Brockwell and Davis, 1996, or Hamilton, 1994). In the asymptotic analysis the problems appear when multiple roots are present as this imply a non-differentiability so the δ-method does not apply, convergence rates are slow, and the asymptotic distribution is non-normal. In finite samples this has a considerable influence on the finite sample distribution unless the roots are far apart. With increasing order of the autoregressions it becomes increasingly difficult to place the roots sufficiently far apart, hence very noisy information about the characteristic roots results.

Computations were done using Ox (Doornik, 2002). The first author gratefully acknowledges financial support from ESRC grant RES-000-27-0179.

Asymptotic theory for the estimated roots has previously been established by Wymer (1972) and Johansen (2003). Starting from the long established asymptotic theory for the autoregressive coefficients they derive an asymptotic theory for the characteristic roots using the δ-method. They carefully avoid cases of multiple roots to ensure that the transformation from the autoregressive coefficients to the roots is locally differentiable. Here, it is established that convergence rates are slow when there are multiple roots. This implies that the δ-method does not give accurate finite sample approximation when the roots are close.

The difficulties in interpreting characteristic roots have been studied recently by Granger and Jeon (2006). They argue that when empirical models have complex characteristic roots in a torus around the origin it can be consistent either with seasonal behaviour or with overfitting. This phenomena can, however, also be consistent with non-seasonal behaviour, which makes it even more difficult to interpret estimated characteristic roots. Thus, while autoregressive processes with known coefficients are interpreted through their characteristic roots it appears that estimated characteristic roots should only be interpreted with caution.

Just as the problem of finding the roots of polynomials, the problem of analysing the distribution of characteristic roots analytically is rather difficult for higher order processes. A complete asymptotic theory with accurate finite sample properties is therefore only established in the tractable second order case. This analysis gives enough information to understand the difficulties in the higher order case and to explain the observations made by Granger and Jeon (2006).

Notation: the mapping $x^{1/k}$ is a complex multifunction, see Priestley (1990, page 91f). Thus, for any real number x let $x^{1/k}$ be a uniquely chosen branch of $x^{1/k}$. In particular, when x is non-negative then $x^{1/k}$ is chosen as the positive and real solution.

2 The second order autoregression

The second order autoregressive case is analysed analytically. The analysis is done in two steps, using first the δ-method and then a direct mapping. The results are illustrated by simulations.

2.1 The model and asymptotics for the autoregressive coefficients

Consider a time series $X_{-1}, X_0, \ldots, X_T$ and the statistical model given by

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{2.1}$$

Here the initial values X_{-1}, X_0 are conditioned upon, the innovations $\varepsilon_1, \ldots, \varepsilon_T$ are independently, identically normal $\mathsf{N}(0, \omega^2)$-distributed. The parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and ω^2 vary freely so $\omega^2 > 0$.

The asymptotic theory for the least squares estimator $\widehat{\boldsymbol{\beta}}$ has long been established in the stationary case, see Anderson (1971, page 200), and shows that

$$T^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathsf{D}} \mathsf{N}_2(0, \Sigma^{-1}). \tag{2.2}$$

Here, the variance matrix solves the equation $\Sigma = \mathbf{A}\Sigma\mathbf{A}' + BB'$, where $B = (1, 0)'$ and $\mathbf{A} = (\boldsymbol{\beta}, B)'$ is the companion matrix, and it is given by

$$\Sigma^{-1} = (1 + \beta_2) \begin{pmatrix} 1 - \beta_2 & -\beta_1 \\ -\beta_1 & 1 - \beta_2 \end{pmatrix}.$$

2.2 Analysing the characteristic roots by the δ-method

The characteristic polynomial is

$$p(z) = z^2 - \beta_1 z - \beta_2,$$

with characteristic roots, λ_1 and λ_2, solving the equation $p(z) = 0$. The roots are given as $(\beta_1 \pm D^{1/2})/2$ where $D = \beta_1^2 + 4\beta_2$ is the discriminant. If $D > 0$ the roots are real, if $D < 0$ the roots are complex, and if $D = 0$ there is a double root.

When $D \neq 0$ the mapping from $\boldsymbol{\beta}$ to the roots is differentiable. Thus the asymptotic distribution can be found using the δ-method. This was done by Wymer (1972) and more generally by Johansen (2003). In the second order case the argument can be made using a composite mapping $\varphi \circ \psi$. The first mapping finds the discriminant,

$$\begin{array}{rccc} \psi\colon & (\beta_1, \beta_2) & \mapsto & (\beta_1, D), \\ & \mathsf{R}^2 & \mapsto & \mathsf{R}^2, \end{array}$$

and is everywhere differentiable. The δ-method gives the following result.

Theorem 1 *Suppose $|\lambda_j| < 1$. Let J_ψ be the Jacobian of the mapping ψ. Then*

$$T^{1/2}(\widehat{\beta}_1 - \beta_1, \widehat{D} - D)' \xrightarrow{\mathsf{D}} \mathsf{N}(0, H),$$

where the covariance matrix is given by

$$H = J_\psi \Sigma^{-1} J_\psi' = \begin{Bmatrix} \left(1 - \beta_2^2\right) & -2\beta_1 (1 + \beta_2)^2 \\ -2\beta_1 (1 + \beta_2)^2 & 16\left(1 - \beta_2^2\right) - 4\beta_1^2 (1 + \beta_2)(3 + \beta_2) \end{Bmatrix}.$$

The second mapping is only differentiable when $D \neq 0$, and even so it depends on the sign of D. When $D > 0$ the roots are real so the mapping is

$$\begin{array}{rccc} \varphi_{\mathsf{R}^2}: & (\beta_1, D) & \mapsto & (\beta_1 + D^{1/2}, \beta_1 - D^{1/2})/2 = (\lambda_1, \lambda_2) \\ & \mathsf{R} \times \mathsf{R}_+ & \mapsto & \mathsf{R}^2, \end{array}$$

Experiment	$(\widehat{\lambda}_1, \widehat{\lambda}_2)$ Theorem 2	$(\widehat{\beta}_1, \widehat{\beta}_2)$ Eq. (2.2)	$(\widehat{\beta}_1, \widehat{D})$ Theorem 1	$\text{sign}(\widehat{D})$ Eq. (2.3)	$(\widehat{\lambda}_1, \widehat{\lambda}_2)$ Theorem 3 Remark 4
A: $(-0.5, 0.5)$	0.92	0.94	0.94	0.99	0.94
B: $(-0.35, 0.35)$	0.88	0.95	0.95	0.88	0.95
C: $(0.5, 0.8)$	0.65	0.91	0.90	0.65	0.92
D: $(0.7 \pm 0.4i)$	0.97	0.97	0.96	1.00	0.96
E: $(0.7 \pm 0.25i)$	0.90	0.92	0.93	0.93	0.93
F: $(0.7 \pm 0.15i)$	0.79	0.91	0.90	0.77	0.915

Table 1: Coverage probabilities with nominal level 0.95, based on $N = 100$ repetitions, with Monte Carlo standard deviation of 0.07.

recalling that $x^{1/2}$ is chosen to be positive and real when x is positive, whereas when $D < 0$ the roots are complex, $\text{Re}(\lambda) \pm \text{Im}(\lambda)$, so the relevant mapping is

$$\begin{array}{rccl} \varphi_{\mathsf{C}}: & (\beta_1, D) & \mapsto & \{\beta_1, (-D)^{1/2}\}/2 = \{\text{Re}(\lambda), \text{Im}(\lambda)\}, \\ & \mathsf{R} \times \mathsf{R}_- & \mapsto & \mathsf{R}^2. \end{array}$$

Applying the δ-method to these mappings gives the following result, which is a special case of Johansen (2003, Theorem 3).

Theorem 2 *Suppose* $|\lambda_j| < 1$.
Let $D > 0$. *Then as* $T \to \infty$,

$$T^{1/2} \begin{pmatrix} \widehat{\lambda}_1 - \lambda_1 \\ \widehat{\lambda}_2 - \lambda_2 \end{pmatrix} \xrightarrow{\mathsf{D}} \mathsf{N}_2\left(0, J_+ H J_+'\right), \qquad J_+ = \frac{1}{2} \begin{pmatrix} 1 & D^{-1/2} \\ 1 & -D^{-1/2} \end{pmatrix}.$$

Let $D < 0$. *Then as* $T \to \infty$,

$$T^{1/2} \left\{ \begin{array}{c} \text{Re}(\widehat{\lambda}) - \text{Re}(\lambda) \\ \text{Im}(\widehat{\lambda}) - \text{Im}(\lambda) \end{array} \right\} \xrightarrow{\mathsf{D}} \mathsf{N}_2\left(0, J_- H J_-'\right), \qquad J_- = \frac{1}{2} \left\{ \begin{array}{cc} 1 & 0 \\ 0 & (-D)^{-1/2} \end{array} \right\}.$$

The asymptotic distribution approximation given by Theorem 2 can be quite poor in finite samples. Table 1 reports the result from a simulation study with sample length $T = 100$ and six different pairs of characteristic roots. The coverage probability of a 95% sampling region based on Theorem 2 is reported in column 2. It is seen to deteriorate as the distance between the population roots decreases. Columns 3 and 4 indicate that the first transformation ψ preserves the coverage probability of $\widehat{\boldsymbol{\beta}}$, so the problem lies with the second transformation φ. Johansen (2003) pointed

	β_2	0	0.01	0.04	0.09	0.16	0.25
	Roots	0	± 0.1	± 0.2	± 0.3	± 0.4	± 0.5
Asymptotic	$T = 10^2$	0.5	0.54	0.66	0.82	0.95	0.995
	$T = 10^3$	0.5	0.62	0.90	0.998		
	$T = 10^4$	0.5	0.84				
Simulated	$T = 10^2$	0.469	0.508	0.624	0.787	0.929	0.990
	$T = 10^3$	0.491	0.614	0.893	0.997		
	$T = 10^4$	0.497	0.840				

Table 2: Probability that $\widehat{D}$ and D have the same sign when $\beta_1 = 0$. Upper panel: asymptotic approximation to probability based on (2.3). Lower panel: simulated probability based on 10^5 repetitions, and zero initial values.

out that in finite samples problems can arise because the parabola determined by the characteristic polynomial may be perturbed in such a way that for instance the sample version of a parabola that intersects with the x-axis may not intersect with the x-axis. This phenomena would arise if the estimated discriminant, $\widehat{D}$, has the wrong sign. Column 5 of Table 1 shows this probability computed from the asymptotic normal approximation of Theorem 1, that is

$$\mathsf{P}\{\text{sign}(\widehat{D}) = \text{sign}\,(D)\} \approx \Phi(|\nu_T|) \qquad \text{where} \qquad \nu_T = -T^{1/2} H_{22}^{-1/2} D. \tag{2.3}$$

These probabilities match the coverage probability of Theorem 2 to some extent.

The sign of the discriminant is therefore the key to the quality of the approximation in Theorem 2. Table 2 therefore reports a detailed simulation study of the distribution of $\widehat{D}$ evaluating the variation with the parameters and the quality of the asymptotic approximation provided by equation (2.3). Different sample lengths are chosen while the population roots, for simplicity, are located symmetrically around zero so $\beta_1 = 0$ and $D = 4\beta_2$. In particular, when $D = 0$ there is a double root in zero and formula (2.3) shows the probability of observing real roots is about 50% for all sample lengths.

2.3 Direct mapping of sampling regions

In order to get an accurate distribution approximation for the characteristic roots the δ-method is abandoned. Instead the sampling regions of Theorem 1 are transformed directly using the mappings φ_{R^2} and φ_{C}. This reveals a $T^{1/4}$-convergence rate when $D = 0$ which will influence the finite sample approximation when D is close to zero.

Based on Theorem 1 a sampling region for $(\widehat{\beta}_1, \widehat{D})$ with coverage level α, of 95%

say, can be constructed as the ellipsoid

$$\partial S\left(\gamma_2\right)=\left\{\left(\begin{array}{c}\beta_1\\ D\end{array}\right)+\frac{\gamma_2}{T^{1/2}}\left(\begin{array}{cc}H_{11\cdot 2}^{1/2} & H_{12}H_{22}^{-1/2}\\ 0 & H_{22}^{1/2}\end{array}\right)\left(\begin{array}{c}\cos\theta\\ \sin\theta\end{array}\right)\middle| \, 0\leq\theta<2\pi\right\}, \quad (2.4)$$

where γ_2^2 is the α-quantile of the χ_2^2-distribution and $H_{11\cdot 2}=H_{11}-H_{12}H_{22}^{-1}H_{21}$. A direct application of the φ_{R^2} and φ_{C} then results in the sampling regions

$$S_{\mathsf{R}^2}=\varphi_{\mathsf{R}^2}\{S\cup(\widehat{D}\geq 0)\},\qquad S_{\mathsf{C}}=\varphi_{\mathsf{C}}\{S\cup(\widehat{D}\leq 0)\}.$$

The interiors of these sets are disjoint and the coverage of their union set $S_{\mathsf{R}^2}\cup S_{\mathsf{C}}$ is the same as that of S, that is $\mathsf{P}(S_{\mathsf{R}^2}\cup S_{\mathsf{C}})=\mathsf{P}(S)$.

The key to describing the sets S_{R^2} and S_{C} is to evaluate the event $(\widehat{D}=0)$. This is included in S whenever $D+\gamma_2(H_{22}/T)^{1/2}\sin\theta$ equals zero for some $-\pi/2\leq\theta\leq\pi/2$. Thus, recall the notation $\nu_T=-(T/H_{22})^{1/2}D$ from (2.3) and denote

$$\widetilde{\theta}=\left\{\begin{array}{ll}\arcsin\left(\nu_T/\gamma_2\right) & \text{if } |\nu_T|<\gamma_2,\\ \operatorname{sign}\left(\nu_T\right)(\pi/2) & \text{if } |\nu_T|\geq\gamma_2.\end{array}\right. \quad (2.5)$$

When $|\nu_T|<\gamma_2$ both of the sets S_{R^2} and S_{C} are non-empty, whereas, for instance S_{C} is empty if $\nu_T>\gamma_2$. The sets S_{R^2} and S_{C} can be described as follows.

Theorem 3 *Recall $\widetilde{\theta}$ defined in* (2.5) *and define the functions*

$$\begin{aligned}x\left(\gamma,\theta\right) &= \frac{\beta_1}{2}+\frac{\gamma}{2T^{1/2}}(H_{11\cdot 2}^{1/2}\cos\theta+H_{12}H_{22}^{-1/2}\sin\theta),\\ y\left(\gamma,\theta\right) &= \frac{D}{4}+\frac{\gamma}{4T^{1/2}}H_{22}^{1/2}\sin\theta.\end{aligned}$$

Then sets $S_{\mathsf{R}^2}(\gamma_2)$ and $S_{\mathsf{C}}(\gamma_2)$ are then given by

$$\begin{aligned}\partial S_{\mathsf{R}^2}(\gamma) &= \{[x\left(\gamma,\theta\right)+\{y\left(\gamma,\theta\right)\}^{1/2},x\left(\gamma,\theta\right)-\{y\left(\gamma,\theta\right)\}^{1/2}] \text{ for } \widetilde{\theta}\leq\theta\leq\pi-\widetilde{\theta}\}\\ &\quad\cup\{(z,z) \text{ where } z=\xi x(\gamma,\widetilde{\theta})+(1-\xi)\,x(\gamma,-\pi-\widetilde{\theta}) \text{ for } 0<\xi<1\},\\ \partial S_{\mathsf{C}}(\gamma) &= \{[x\left(\gamma,\theta\right),\{-y\left(\gamma,\theta\right)\}^{1/2}] \text{ for } \widetilde{\theta}\leq\theta\leq-\pi-\widetilde{\theta}\}\\ &\quad\cup\{(z,0) \text{ where } z=\xi x(\gamma,\widetilde{\theta})+(1-\xi)\,x(\gamma,\pi-\widetilde{\theta}) \text{ for } 0<\xi<1\}.\end{aligned}$$

A number of remarks follow concerning the implications of Theorem 3.

Remark 1 ***Link with Theorem 2.*** *When there are distinct roots, so $D\neq 0$ then $\nu_T=-(T/H_{22})^{1/2}D$ diverges and $|\widetilde{\theta}|\to\pi/2$. The contours given by Theorem 3 then reduces to the ellipsoid contours of the normal distributions given by Theorem 2.*

Remark 2 ***The case of double roots.*** *When $D = 0$ then $\widetilde{\theta} = 0$. Formula (2.3) shows that asymptotically the sets S_{R^2} and S_{C} each have probability 50%. The functions x and y of Theorem 3 reduce to*

$$x(\gamma, \theta) = \frac{\gamma \cos\theta}{2T^{1/2}} \qquad |y(\gamma, \theta)|^{1/2} = \frac{(2\gamma |\sin\theta|)^{1/2}}{T^{1/4}},$$

which gives convergence rates of order $T^{-1/4}$.

Remark 3 ***Implications for finite sample behaviour.*** *The non-uniform convergence rates seen above have profound implications for the finite sample behaviour of the estimated roots. The sets given in Theorem 3 involve the square root of the function y. This can be expressed in terms of $\nu_T = -D(T/H_{22})^{1/2}$ as*

$$y(\gamma, \theta) = \frac{D}{4}(1 - \frac{\gamma}{\nu_T}\sin\theta) = \frac{H_{22}^{1/2}}{4T^{1/2}}(\gamma \sin\theta - \nu_T).$$

The order of magnitude of ν_T relates to the probability that $\widehat{D}$ has the correct sign through (2.3). When $|\nu_T|$ is much larger than γ the usual $T^{-1/2}$ asymptotics dominates in that

$$\{|y(\gamma, \theta)|\}^{1/2} = \frac{|D|^{1/2}}{2} + \mathrm{O}\left(\nu_T^{-1}\right) = \frac{|D|^{1/2}}{2} + \mathrm{O}\left(T^{-1/2}\right),$$

whereas when $|\nu_T|$ is much smaller than γ and $|\sin\theta|$ is not too small then the $T^{-1/4}$ asymptotics dominates

$$\{|y(\gamma, \theta)|\}^{1/2} \approx \frac{H_{22}^{1/4}}{2T^{1/4}}\gamma^{1/2}|\sin\theta|^{1/2}.$$

Remark 4 ***Plot of characteristic roots in complex plane.*** *Often it is of interest to plot the characteristic roots (λ_1, λ_2) in the complex plane. The sampling region will then be the union of four sets*

$$R_{\mathsf{R}}^{+}(\gamma_1) \cup R_{\mathsf{R}}^{-}(\gamma_1) \cup R_{\mathsf{C}}^{+}(\gamma_2) \cup R_{\mathsf{C}}^{-}(\gamma_2),$$

where critical values γ_1 and γ_2 of χ_1^2 and χ_2^2 distributions are needed, due to the dimension reduction when plotting both real roots on the same real axis. The sets on the real axis are derived from $S_{\mathsf{R}^2}(\gamma)$ and given by the intervals

$$R_{\mathsf{R}}^{\pm}(\gamma) = \{\min_{\widetilde{\theta}<\theta<\pi-\widetilde{\theta}} z^{\pm}(\gamma, \theta) < x < \max_{\widetilde{\theta}<\theta<\pi-\widetilde{\theta}} z^{\pm}(\gamma, \theta)\}$$

where $z^{\pm}(\gamma, \theta) = x(\gamma, \theta) \pm \{y(\gamma, \theta)\}^{1/2}$. The coverage probability of $R_{\mathsf{R}}^{+}(\gamma_1) \cup R_{\mathsf{R}}^{-}(\gamma_1)$ and of $S_{\mathsf{R}^2}(\gamma_2)$ will asymptotically be the same. The sets in the complex plane $R_{\mathsf{C}}^{\pm}(\gamma)$ are simply given by $S_{\mathsf{C}}(\gamma)$ and its mirror image in the real axis.

The finite sample performances of Theorem 3 relative to Theorem 2 was investigated by simulation. First, the last column of Table 1 shows a marked improvement in the finite sample performance from using Theorem 3. Secondly, Figure 1 illustrates the sampling regions for the same experiments $A - F$ In contrast to the δ-method Theorem 2 can capture the non-symmetric variation on the real axis and the non-ellipsoid variation in the complex plane. The δ-method fares particularly poorly in experiments B, C, E and F where there is a considerable probability that $\widehat{D}$ has the wrong sign.

3 Higher order autoregression

The distribution theory for higher order autoregressions is much more complicated than for the second order case. In the third order case not only the vertical location of graph of the polynomial can change, but also the shape of the function can change from having two turning points to having none. For this reason the δ-method is no longer helpful in establishing an asymptotic theory for the discriminant. Moreover, the relevant convergence rates are even slower than in the second order case.

3.1 The discriminant in the third order case

The discriminant can be defined in two ways, see Basu, Pollack and Roy (2003, §4.2), either in terms of the characteristic roots or in terms of the coefficients of the polynomial through the determinant of the socalled Sylvester matrix. For the third order characteristic polynomial

$$p(z) = z^3 - \beta_1 z^2 + \beta_2 z + \beta_3 \tag{3.1}$$

the discriminant is therefore given by

$$\begin{aligned} D &= \left(\lambda_1^2 - \lambda_2^2\right)\left(\lambda_1^2 - \lambda_3^2\right)\left(\lambda_2^2 - \lambda_3^2\right) \\ &= \beta_1^2\beta_2^2 + 4\beta_2^3 + 4\beta_1^3\beta_3 - 27\beta_3^2 - 18\beta_1\beta_2\beta_3 \end{aligned} \tag{3.2}$$

The determinant has the property that $D > 0$ if all roots are real, $D = 0$ if there are double roots, and $D < 0$ if one root is real.

As in the two dimensional case it would be of interest to derive an asymptotic theory for the discriminant from the asymptotic distribution of the autoregressive coefficients using the δ-method. The next result, which is proved in the Appendix, shows that the relevant Jacobian is zero if there are triple roots and consequently a second order δ-method is needed.

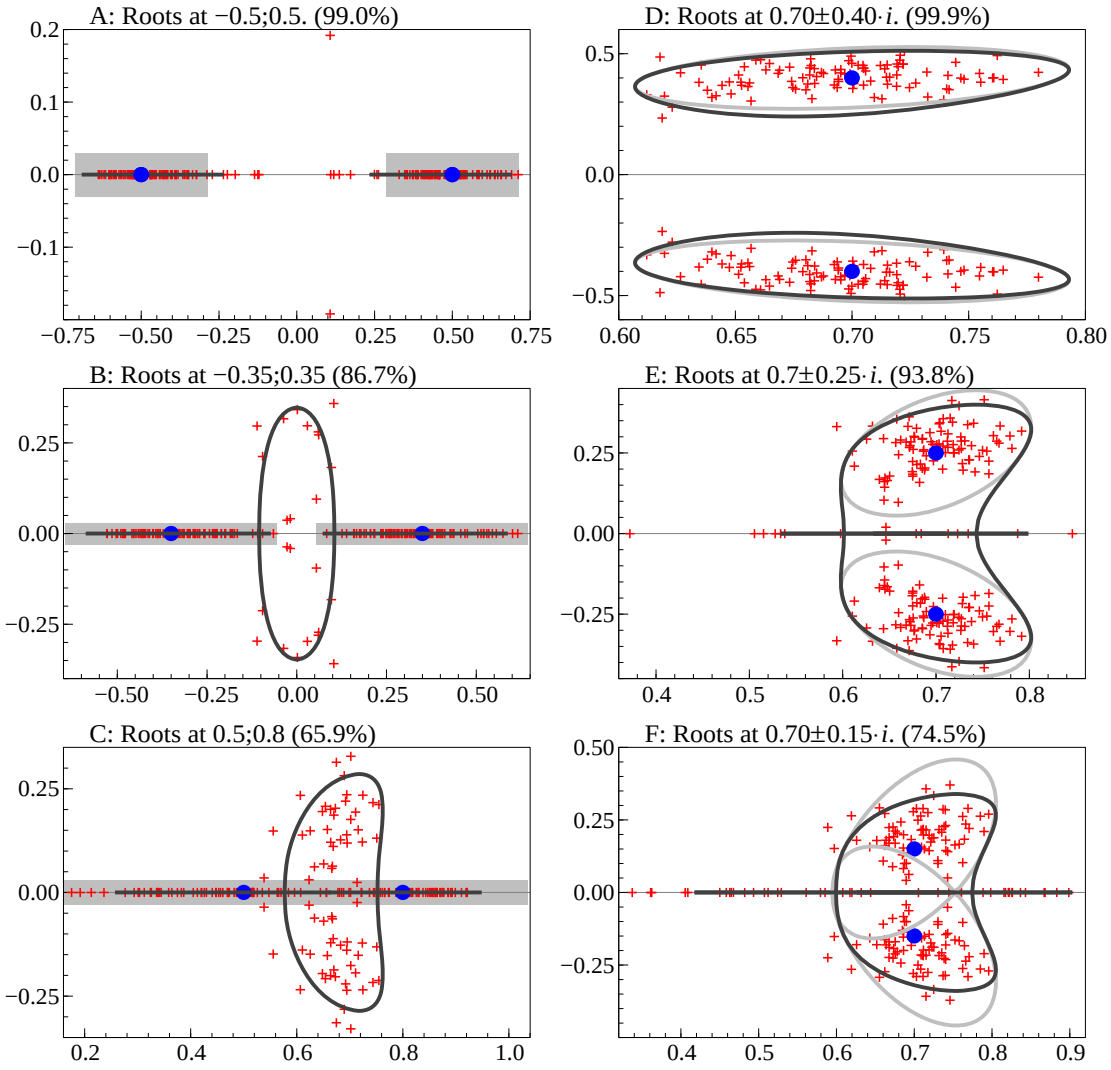

Figure 1: Sampling regions of nominal 95% level for same experiments as in Table 1. Sample length $T = 100$ observations and 100 replications. The grey shaded lines are based on Theorem 2. The bold black lines are based on Theorems 3 and Remark 4. The estimated probabilities of $\text{sign}(\hat{D}) = \text{sign}(D)$ in 10^5 replications are reported in parentheses.

	β_3	-0.32	0	0.384	-0.192	-0.192
	Roots	0.8	0.8	0.8	-0.4	-0.6
		0.8	0.8	0.8	0.6	0.4
		-0.5	0	0.6	0.8	0.8
Asymptotic	$T = 10^2$	0.50	0.50	0.50	0.60	0.82
	$T = 10^3$	0.50	0.50	0.50	0.80	1.00
	$T = 10^4$	0.50	0.50	0.50	1.00	1.00
Simulated	$T = 10^2$	0.46	0.37	0.07	0.54	0.72
	$T = 10^3$	0.48	0.48	0.10	0.77	0.99
	$T = 10^4$	0.49	0.50	0.26	0.99	1.00

Table 3: Probability that $\widehat{D}$ is positive. Upper panel: Asymptotic approximation based on δ-method with numerical derivatives. Lower panel: simulated probability based on 5000 repetitions, and zero initial values. The parameter β_3 is the product of the characteristic roots. When $|\beta_3| < T^{-1/2}2$ lag length tests would typically point to a shorter lag length than 3.

Theorem 4 *The discriminant of the polynomial* $p(z) = z^3 - \beta_1 z^2 + \beta_2 z + \beta_3$ *satisfies*

$$\frac{\partial D}{\partial \beta} = 0 \qquad \Longleftrightarrow \qquad p(z) = (z - \lambda)^3 \quad \textit{for some } \lambda \in \mathbf{R}.$$

The fact that a δ-method of second order rather than first order is needed in parts of parameter space implies that the asymptotic theory delivered by the δ-method is not going to be uniform in the parameter space. Table 3 reports the results from a simulation study of this effect. The probability $\mathsf{P}(D > 0)$ is found for different sample lengths and various combinations of roots. The probability was found partly by simulation and partly using the δ-method with numerical derivatives. It is seen that in some situations the sample size has to be enormous before the first order δ-method gives a good approximation.

A consequence of Theorem 4 is that if it is of interest to test the hypothesis $D = 0$, one would have to contemplate the hypothesis of a triple root. The triple root hypothesis can be formulated in terms of the autoregressive parameters using the first subdiscriminant, see Basu, Pollack and Roy (2003, Proposition 4.29). For the third order polynomial (3.1) this is $sD_1 = 2(3\beta_2 + \beta_1^2)$, which is zero if and only if $p(z)$ has a triple root. The Jacobian of sD_1 with respect to β is always non-zero, so reliable inference can be made using the δ-method. Mimicking the procedure of Pantula (1989) for determining the number of unit roots in an autoregression, the hypothesis $D = 0$ can be tested by first testing $sD_1 = 0$ and then testing $D = 0$ using the δ-method for both tests.

3.2 Convergence rates

For autoregressions of order k the relevant convergence rates are as slow as $T^{-1/2k}$. To explain this slow convergence it is useful to keep in mind the perturbation theory for the algebraic eigenvalue problem, in which the coefficients of the polynomials are known but errors arise in the numerical analysis, see Wilkinson (1965). To keep the problem tractable consider simply a regression of a series X_t on its k-th lag, $X_t = \beta X_{t-k} + \varepsilon_t$, in which the population value is $\beta = 0$. That model has a characteristic polynomial of the type $p(z) = z^k - \beta$. The least squares estimator for β has an asymptotic standard error given by $\mathsf{se}(\widehat{\beta}) = T^{-1/2}$. The population characteristic equation of $z^k = 0$ is then perturbed as $z^k = \widehat{\beta}$. The perturbed equation has k distinct roots at

$$\widehat{\lambda}_j = \exp\left(\frac{j2\pi}{k}i\right)\widehat{\beta}^{1/k} = \mathrm{O_P}\{T^{-1/(2k)}\} \quad \text{for} \quad j = 1, \ldots, k. \tag{3.3}$$

This results in roots at the polar arguments $k^{-1}2j\pi$ when $\widehat{\beta} > 0$ and arguments $k^{-1}(2j-1)\pi$ when $\widehat{\beta} < 0$. The distribution of $\widehat{\lambda}_j$ will then have $2k$ rays and the modulus of the roots will be $T^{-1/(2k)}$. This explains the observation of Granger and Jeon (2006) that over-fitting can give spurious characteristic roots.

The effect from over-fitting is illustrated in Figure 2 A and B for 10^2 and 10^3 observations, respectively. The true roots are located at the origin, corresponding to generating independent, identically distributed variables, while the estimated model is a fourth order autoregression. Most of the estimated roots are located in a torus around the origin, with marked concentration on $2k = 8$ symmetric rays from the origin as predicted. The radius of the torus shrinks slowly consistent with a rate of $T^{-1/8}$ in that $T^{-1/8}$ takes the values 0.56, and 0.42 for T taking values 10^2 and 10^3, respectively. The proportion of real estimated roots is constant (26%) with the sample length in line with the situation of double roots in the second order case.

An example of a cyclical process is given in Figure 2 C. Here, the four population roots are complex and located at $(-0.6 \pm 0.3i, 0.6 \pm 0.3i)$. For $T = 100$ there is a non-negligible probability of real estimated roots. Non-cyclical processes can generate similar patterns, and Figure 2 D shows the estimated roots when the population roots are simple, real, and located at $(\pm 0.6, \pm 0.8)$. For these examples the product of the roots are $\beta_4 = 0.2025$ and $\beta_4 = 0.2304$, which just about satisfy $|\beta_4| > T^{-1/2}2$ so that a test for lag length would tend to point towards a lag length of 4. In other words, it is rather difficult to place four roots inside the unit circle that are so far from each other that the δ-method is reliable and at the same time avoid the over-fitting problem.

It is interesting to compare the estimated roots for the cyclical and non-cyclical case. In a small sample, $T = 100$, the visual appearances of the graphs are broadly similar and the graphs contain very noisy information about the true characteristic

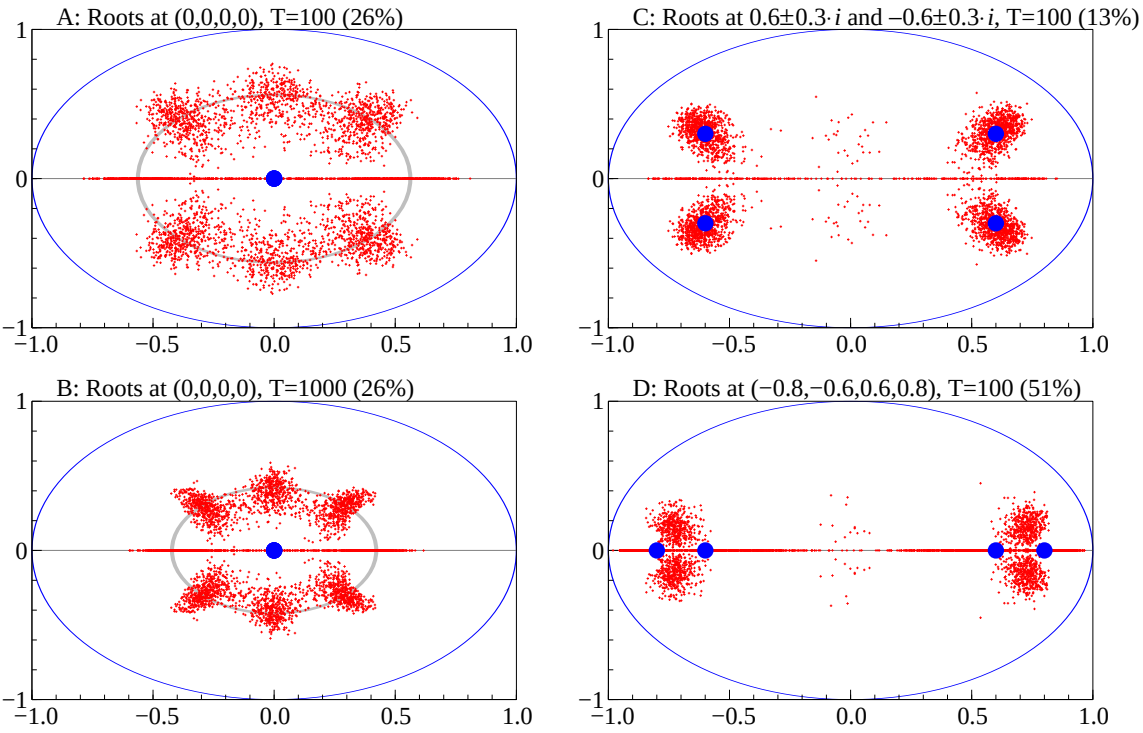

Figure 2: Each graph shows the true characteristic roots of a fourth order autoregression and the estimated roots in 10^3 replications. Proportion of real estimated roots in parentheses. The gray circles in A and B correspond to a modulus of $T^{-1/(2k)}$ consistent with overfitting.

roots. This suggests that the estimated roots should be interpreted with great care in finite samples. Large estimated complex roots with an apparently strong cyclical or seasonal behaviour, could arise just as well from an over-fitted model, a model with real roots, and from a model with complex roots. In practice the over-fitted models could be avoided to a large extent by determining the lag length carefully.

4 Conclusion

Estimated characteristic roots have been shown to give rather noisy information about their population equivalents. The δ-method often gives unreliable approximations to the distribution of the characteristic roots. Thus, it appears that it is typically difficult to construct reliable inferential procedures directly based on the roots, whereas

reliable inferences can more easily be drawn from autoregressive coefficients. Examples are lag length determination through the last autoregressive coefficient vis-a-vis the product of the roots; unit root analysis through the sum of the autoregressive coefficients vis-a-vis unit roots, which is standard; and determination of the number of real roots and the multiplicity of roots through the discriminant and subresultants.

5 Appendix: Proof of Theorem 4

The discriminant given by (3.2) has partial derivatives

$$\frac{\partial D}{\partial \beta} = \begin{pmatrix} 2\beta_1\beta_2^2 - 12\beta_1^2\beta_3 - 18\beta_2\beta_3 \\ 2\beta_1^2\beta_2 + 12\beta_2^2 - 18\beta_1\beta_3 \\ -4\beta_1^3 - 54\beta_3 - 18\beta_1\beta_2 \end{pmatrix}.$$

Expressing the polynomial in terms of its roots

$$\begin{aligned} p(z) &= (z-\gamma)(z-\lambda-\alpha)(z-\lambda+\alpha) \\ &= z^3 - (\gamma+2\lambda)z^2 + \left(2\gamma\lambda + \lambda^2 - \alpha^2\right)z - \gamma\left(\lambda^2-\alpha^2\right), \end{aligned}$$

where $\gamma, \lambda, \alpha^2 \in \mathbf{R}^3$ shows that the partial derivatives satisfy

$$\frac{\partial D}{\partial \beta} = 4(\lambda-\gamma)\begin{pmatrix} (\lambda-\gamma)^2\lambda^2 - 2\lambda^2 + 4\lambda\gamma + 3\gamma^2\alpha^2 \\ (\lambda-\gamma)^2\lambda - (4\lambda+5\gamma)\alpha^2 \\ (\lambda-\gamma)^2 - 9\alpha^2 \end{pmatrix} + 4\alpha^4\begin{pmatrix} 5\gamma+\lambda \\ 3 \\ 0 \end{pmatrix}$$

Four different special cases need to be considered depending on $\lambda-\gamma$ and α.

1. Triple roots: $\lambda-\gamma=\alpha=0$. $\partial D/\partial\beta = 0$.

2. $\lambda-\gamma=0$, but $\alpha\neq 0$. Then $\partial D/\partial\beta_2 = 12\alpha^4 \neq 0$.

3. $\lambda-\gamma\neq 0$, but $(\lambda-\gamma)^2 - 9\alpha^2 \neq 0$. Then $\partial D/\partial\beta_3 \neq 0$.

4. $\lambda-\gamma\neq 0$, but $(\lambda-\gamma)^2 - 9\alpha^2 = 0$ so that $\lambda-\gamma = 3\alpha \neq 0$. Then $\partial D/\partial\beta_2 = 228\alpha^4 \neq 0$.

6 References

Anderson, T.W. (1971) *The statistical analysis of time series.* New York: John Wiley & sons.

Basu, S., Pollack, R., and Roy, M.-F. (2003) *Algorithms in Real Algebraic Geometry.* Berlin: Springer-Verlag.

Brockwell, P.J., and Davis, R.A. (1996) *Introduction to time series and forecasting.* New York: Springer.

Doornik, J.A. (2002) *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

Granger, C.W.J. and Jeon, Y. (2006) Dynamics of Model Overfitting Measured in Terms of Autoregressive Roots. *Journal of Time Series Analysis*, 27(3), 347-365.

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

Johansen, S. (2003) The asymptotic variance of the estimated roots in a cointegrated vector autoregressive model. *Journal of Time Series Analysis*, 24, 663-678.

Pantula, S.G. (1989) Testing for unit roots in time series data. *Econometric Theory,* 5, 256-271.

Priestley, H.A. (1990) *Introduction to complex analysis.* Revised edition. Oxford: Oxford University Press.

Wilkinson, J.H. (1965) *The algebraic eigenvalue problem*. Oxford: Oxford University Press.

Wymer, C.R. (1972) Econometric estimation of stochastic differential equation systems. *Econometrica*, 40, 565-577.