A small sample correction of the Dickey-Fuller test

Søren Johansen
Department of Applied Mathematics and Statistics
University of Copenhagen
March 2004

Abstract

The purpose of this note is to investigate the small sample correction derived in Johansen (2002) for the test for cointegrating rank in the special case of the Dickey-Fuller test. We find an explicit form for the correction and investigate the relevance by simulations. It is seen that for small samples, the Dickey Fuller test has size distortions, and that the correction factor helps conducting reliable inference. For parameters close to the $I(2)$ boundary the size distortion can be serious. The simulation experiments indicate that when the correction factor is less than 1.14 the corrected test gives a rejection probability close to 5%.

JEL Classification: C32
Keywords: Small sample correction, likelihood ratio test, test for unit root, Dickey-Fuller test

1 Introduction and main result

We consider the likelihood ratio test for the hypothesis $\mathcal{M}_k^0 : \pi = \beta = 0$, in the autoregressive model with $k$ lags for $x_t$

$$
\mathcal{M}_k : \Delta x_t = \pi x_{t-1} + \beta t^d + \sum_{i=1}^{k-1} \gamma_i \Delta x_{t-i} + \sum_{i=0}^{d-1} \beta_i t^i + \epsilon_t,
$$

(1)

where $\epsilon_t$ are i.i.d. $N(0, \sigma^2)$, see Dickey and Fuller (1981). Note that under the null hypothesis $x_t$ becomes nonstationary, and that it has a trend of order $d$, both under the null hypothesis and the alternative hypothesis of nonstationarity. We choose this version of the unit root test because the distribution of the test statistic under the null hypothesis does not depend on the parameters of the deterministic terms.

If $\mathcal{M}_k^0$ holds, so that $\pi = \beta = 0$ and the process $\Delta x_t$ is stationary, the asymptotic distribution of $-2\log LR$, under the assumption of i.i.d. errors with mean zero and
finite variance, is given by

\[-2 \log LR(\pi = \beta = 0) \xrightarrow{w} \int_0^1 dB' \left( \int_0^1 F F' d\mu \right)^{-1} \int_0^1 F(dB),\]

where \(B\) is a standard Brownian motion on \([0, 1]\) and

\[F(u) = \left( \begin{array}{c} B(u) \\ u^d \\ 1, \ldots, u^{d-1} \end{array} \right),\]

that is, the Brownian motion \(B(u)\) and the trend \(u^d\) corrected for \((1, \ldots, u^{d-1})\) on the unit interval.

The asymptotic distribution is tabulated by simulation since it contains no parameters, but the finite sample distribution depends on \(T\) and the parameters \(\gamma = (\gamma_1, \ldots, \gamma_{k-1})\) under the null hypothesis, but not on \(\beta_0, \ldots, \beta_{d-1}\) and \(\sigma^2\). For \(T \to \infty\) the dependence on \(\gamma\) disappears, but not uniformly in \(\gamma\). If \(\gamma\) is close to the boundary where \(\Delta x_t\) becomes nonstationary, the approximation can be rather poor, as we shall demonstrate by simulations.

In Johansen (2002) an analytic approximation to the expectation of the likelihood ratio test for cointegrating rank is derived. A special case of that is of course the above Dickey-Fuller test, and the purpose of this note is to see what the formulae look like in this univariate case, and by simulations see if the approximation helps making reliable inference. The idea is to use the approximation of the expectation to derive a correction factor to the likelihood ratio test. It turns out that this correction factor depends on the parameters only through the quantity

\[
\frac{\sum_{i=1}^{k-1} i \gamma_i}{1 - \sum_{i=1}^{k-1} \gamma_i},
\]

see Theorem 1, and the main contribution here is the reduction of the general expression for the correction factor found in Johansen (2002) to an expression involving only (2).

**Theorem 1** Under the assumption that \(x_t\) is an I(1) process given by model (1) with \(\pi = \beta = 0\), the correction factor for the likelihood ratio test for \(\pi = \beta = 0\) is given by

\[
a_T(d)\left(1 + \frac{1}{T}[(k - 1 + 2 \sum_{i=1}^{k-1} m(d) + \frac{1}{2}((-1)^{k-1} - 1)g(d))]\right) .
\]

The functions \(g(d), m(d), \) and \(a_T(d) = 1 + a_1(d)T^{-1} + a_2(d)T^{-2},\) are given in Table 1 for \(d = 0, 1, 2,\) and for the model without deterministic terms, \(d = \ast.\)

We conclude this section by discussing briefly the Bartlett correction and the interpretation of the correction factor. The proofs are given in the Appendix.
<table>
<thead>
<tr>
<th>$d$</th>
<th>$g(d)$</th>
<th>$m(d)$</th>
<th>$sd~m(d)$</th>
<th>$a_1(d)$</th>
<th>$a_2(d)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>-0.5601</td>
<td>0.545</td>
<td>0.064</td>
<td>-0.0233</td>
<td>1.1518</td>
</tr>
<tr>
<td>0</td>
<td>-0.3262</td>
<td>0.748</td>
<td>0.035</td>
<td>0.2123</td>
<td>1.4259</td>
</tr>
<tr>
<td>1</td>
<td>-0.6209</td>
<td>0.860</td>
<td>0.044</td>
<td>0.1212</td>
<td>4.0538</td>
</tr>
<tr>
<td>2</td>
<td>-0.7304</td>
<td>0.932</td>
<td>0.049</td>
<td>0.0788</td>
<td>7.4031</td>
</tr>
</tbody>
</table>

Table 1: The values for $g(d)$, $d = *, 0, 1$ are taken from Nielsen (1997a). The coefficients $a_1(d), a_2(d), g(2)$ are found by 2.5 mill. simulations of some moments of random walk. The results for $d = *$ correspond to the model without deterministic terms. The number of simulations for determination of $m(*), m(0), m(1), m(2)$ is (11, 13.5, 4.7, 4) mill.

1.1 The Bartlett correction

Bartlett (1937) suggested finding the expectation of the likelihood ratio test statistic and use that, or an approximation of it, to correct the likelihood ratio statistic to have (approximately) the same mean as the limit distribution. He proved that in certain cases not only the mean was better approximated, but it also held for all higher moments. In order to exploit this idea we therefore want an approximation to

$$E_{\gamma,T}[-2 \log LR\{\pi = \beta = 0\}] = E_{\gamma,T}[-2 \log LR\{\mathcal{M}_k^0|\mathcal{M}_k\}],$$

which is a function of $\gamma$ and $T$ under the assumption of Gaussian errors. In Johansen (2002) we derived an analytic approximation of the ratio

$$\frac{E_{\gamma,T}[-2 \log LR\{\mathcal{M}_k^0|\mathcal{M}_k\}]}{E_T[-2 \log LR\{\mathcal{M}_1^0|\mathcal{M}_1\}]} = 1 + \frac{b(\gamma)}{T} + \ldots,$$

where $\mathcal{M}_1^0$ corresponds to $\pi = \beta = 0$ in the model

$$\mathcal{M}_1 : \Delta x_t = \pi x_{t-1} + \beta t^d + \sum_{i=0}^{d-1} \beta_i t^i + \varepsilon_t. \quad (5)$$

When $\pi = \beta = 0$, the expectation

$$f_T(d) = E_T[-2 \log LR\{\mathcal{M}_1^0|\mathcal{M}_1\}]$$

only depends on $d$ and $T$, and can therefore be tabulated by simulation. We define $f(d) = \lim_{T \to \infty} f_T(d)$, and approximate $f_T(d)$ by

$$f(d)(1 + a_1(d)T^{-1} + a_2(d)T^{-2}) = f(d)a_T(d).$$

From Nielsen (1997a) we have the values calculated from analytic expressions

$$f(*) = 1.1416, \quad f(0) = 4.0560, \quad f(1) = 6.3207, \quad (6)$$
and the remaining coefficients $f(2)$ and $a_i(d)$ are determined by regression of the simulated values of $f_T(d)$ on $(1, T^{-1}, T^{-2})$.

We therefore suggest to use the test statistic

$$\frac{f(d) - 2 \log \text{LR}}{f_T(d)(1 + T^{-1}b(\hat{\gamma}))} = \frac{-2 \log \text{LR}}{a_T(d)(1 + T^{-1}b(\hat{\gamma}))},$$

or equivalently use the usual likelihood ratio test statistic and correct the asymptotic quantiles by the factor $a_T(d)(1 + T^{-1}b(\hat{\gamma}))$.

For the Dickey-Fuller test we cannot expect that the correction factor improves the approximation by an order of magnitude in an expansion, because, as Jensen and Wood (1997) show for the model without deterministics and one lag, this does not hold for the Dickey-Fuller test. However, the correction idea seems to work in practice as investigated by Johansen (2002) and Nielsen (1997b, 2004).

### 1.2 The correction factor

We next want to give an interpretation of the factor (2), which depends on the parameters, and therefore study the univariate autoregressive process $x_t$ under the null hypothesis by means of the stacked process

$$y_t = (\Delta x_t, \Delta x_{t-1}, \ldots, \Delta x_{t-k+2})'$$

corrected for its mean. The process $y_t$ is a stationary AR(1) process of dimension $k - 1$, which satisfies the equations

$$y_t = Py_{t-1} + Q \varepsilon_t,$$

where

$$P = \begin{pmatrix} \gamma_1 & \cdots & \gamma_{k-2} & \gamma_{k-1} \\ 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, Q = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (7)$$

The process is given by

$$y_t = \sum_{m=0}^{\infty} P^m Q \varepsilon_{t-m},$$

with variance $\Sigma = \text{Var}(y_t)$ and autocovariance function

$$\Sigma = \sigma^2 \sum_{m=0}^{\infty} P^m Q \gamma^T P^{m'}, \quad \gamma(h) = \text{Cov}(y_t, y_{t+h}) = \Sigma^{hp}.$$

We define the long-run coefficient

$$\psi = \sum_{m=0}^{\infty} P^m Q = (I_{k-1} - P)^{-1} Q,$$ 

(9)
and the long-run variance
\[ \Sigma_{long} = \lim_{T \to \infty} T^{-1} \text{Var} \left( \sum_{t=1}^{T} y_t \right) = \sigma^2 (I_{k-1} - P)^{-1} Q Q' (I_{k-1} - P')^{-1} = \frac{\sigma^2}{\phi(1)^2} \Sigma', \]
where \( \Sigma = (1, \ldots, 1)' \), and \( \phi(z) = 1 - \sum_{i=1}^{k-1} \gamma_i z^i \), so that \( (I_{k-1} - P) \Sigma = \phi(1) \Sigma \).

It follows from the general expression for the correction factor, see Section 4.2, that the main parameter dependence is through
\[ tr\{\Sigma^{-1} \Sigma_{long}\} = \frac{\sigma^2 \Sigma^{-1} \Sigma}{\phi(1)^2}. \tag{10} \]

Another interpretation of this quantity is
\[ as.\text{Var}(\log \hat{\phi}(1)) = \frac{as.\text{Var}(\hat{\phi}(1))}{\phi(1)^2} = \frac{\sigma^2 \Sigma^{-1} \Sigma}{\phi(1)^2}, \]
because the asymptotic variance of \( T^{1/2} (\hat{\gamma}_1 - \gamma_1, \ldots, \hat{\gamma}_{k-1} - \gamma_{k-1}) \) is \( \sigma^2 \Sigma^{-1} \).

When applying the results we need the process \( \Delta x_t \) to be stationary, that is, \( \sum_{i=1}^{k-1} \gamma_i < 1 \), and it is seen that
\[ \frac{\phi(1)^2}{\sigma^2 \Sigma^{-1} \Sigma} = \frac{1 - \sum_{i=1}^{k-1} \gamma_i}{\sigma^2 \Sigma^{-1} \Sigma} \]
measures the deviation from the unit root using the variance of the estimate.

When the distance is small, we get a large correction factor but then we would probably accept another unit root and the results derived here would not be valid.

We also want to give an interpretation in terms of the eigenvalues \( \rho_1, \ldots, \rho_k \) of the matrix \( P \), see (7). From \( \phi(z) = \prod_{i=1}^{k-1} (1 - z \rho_i) \), and \( \log \phi(1) = \sum_{i=1}^{k-1} \log(1 - \rho_i) \), we find
\[ as.\text{Var}(\log \hat{\phi}(1)) = \sum_{i,j} \frac{1}{1 - \rho_i} as.\text{Var}(\hat{\rho}_i, \hat{\rho}_j) \frac{1}{1 - \rho_j} \]
\[ = \left\{ \frac{1}{1 - \rho_i} \right\}' \left\{ \frac{1}{1 - \rho_i \rho_j} \right\}^{-1} \left\{ \frac{1}{1 - \rho_j} \right\}, \]
if the eigenvalues are distinct and real, see Johansen (2003). It follows from (14) that \( tr\{\Sigma^{-1} \Sigma_{long}\} = \sigma^2 \Sigma^{-1} \Sigma/\phi(1)^2 \) and from (19) in Corollary 3, that it equals
\[ \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i} = k - 1 - 2 \frac{d \log \phi(z)}{dz} \bigg|_{z=1} = k - 1 - 2 \frac{\sum_{i=1}^{k-1} i \gamma_i}{1 - \sum_{i=1}^{k-1} \gamma_i}.
\]
Thus the main parameter dependence is summarized in the ratio of the long-run to the short-run variance, or simply in the asymptotic variance of \( \log(1 - \sum_{i=1}^{k-1} \gamma_i) \), or as the inverse distance to the unit root, and finally as the ratio \( (\sum_{i=1}^{k-1} i \gamma_i)/(1 - \sum_{i=1}^{k-1} \gamma_i) \).

It is seen that when the model for \( \Delta x_t \) has a near unit root, \( \sum_{i=1}^{k-1} \gamma_i \sim 1 \), the correction factor becomes very large and, as will be seen from the simulation experiments, this corresponds to the situation where it is difficult to make reliable inference, because the asymptotic test may have serious size distortions.
2 A simulation experiment

In order to illustrate the results we consider the simulation of the rejection probability for the likelihood ratio test for the hypothesis \( \pi = \beta = 0 \) in the model

\[
\Delta x_t = \pi x_{t-1} + \beta t + \sum_{i=1}^{k-1} \gamma_i \Delta X_{t-i} + \beta_0 + \varepsilon_t. \tag{11}
\]

Under the null hypothesis, the distribution depends on the parameters \( \gamma_1, \ldots, \gamma_{k-1} \) and \( T \). We choose \( k = 2, 3, \) and \( T = 10, 15 \) and \( k = 7 \) and \( T = 25, 50 \). We tabulate the rejection probabilities by simulation for some values of \( \gamma_1 \) and \( \gamma_2 \), and for \( k = 7 \) we choose \( \gamma_3 = \ldots = \gamma_6 = 0 \).

In all experiments we assume that the lag length is known, and we give three rejection probabilities. First the one we get by using the asymptotic tables, \( p \), next the one we get by using the correction factor for the true parameter value, \( p_{corr} \), and finally the one we get by estimating the parameter in every simulation, and then use the estimated correction factor, \( p_{est}^{corr} \).

For \( k = 3, 7 \), the parameter values \( \gamma_1 + \gamma_2 = 1 \) and \( 0 < \gamma_1 < 2 \) correspond to \( x_t \) being \( I(2) \), and for the point \( \gamma_1 = 2, \gamma_2 = -1 \), the process is \( I(3) \). For \( k = 2 \), we get an \( I(2) \) process when \( \gamma_1 = 1 \). The results, given in Table 2, are based on 10,000 simulations.

As an example consider the experiment with \( k = 3 \), \( T = 10 \), and \( \gamma_2 = 0 \), where the rejection probability of a nominal 5\% test ranges from 13\% to 23\% close to the \( I(2) \) boundary given by \( \gamma_1 = 0 \). Introducing the correction with known parameters inserted, brings these values down to 8\% – 6\% for \( \gamma_1 \leq 0.3 \), but close to the \( I(2) \) boundary the correction is too large due to the factor \( 1/(1 - \gamma_1) \). Strangely enough, by inserting the estimated parameter in the correction for each simulation, we get value from 8\% – 12\% for the whole range. This is due to the bias in the estimation of \( \gamma_1 \). Obviously, increasing the sample size to \( T = 15 \) improves matters, so that the range 8\% – 18\% can be reduced to 5\% – 8\% for the corrected test.

If one should extract a rule of thumb from these experiments, it would be that when the correction factor is less than 1.14 the corrected test improves the accuracy of the Dickey-Fuller test. Note that the factor \( a_T(d) \), for \( T = 10 \), contributes at most 8\% (for \( d = 2 \)) to the correction factor. The rest is due to the parameter dependence.

3 Acknowledgement

The author would like to thank Henrik Hansen, David Hendry, Bent Nielsen, Peter Boswijk and the participants in the ESF-EMM network for useful references and discussions. The author is grateful for support from the Danish Social Sciences Research Council.
<table>
<thead>
<tr>
<th>$k = 2$</th>
<th>$T = 10$</th>
<th>$T = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>$p$</td>
<td>$p_{corr}$</td>
</tr>
<tr>
<td>-0.9</td>
<td>0.082</td>
<td>0.050</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.097</td>
<td>0.051</td>
</tr>
<tr>
<td>0.3</td>
<td>0.125</td>
<td>0.049</td>
</tr>
<tr>
<td>0.9</td>
<td>0.178</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k = 3$</th>
<th>$T = 10$</th>
<th>$T = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
<td>$p$</td>
</tr>
<tr>
<td>-0.9</td>
<td>0.0</td>
<td>0.132</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.0</td>
<td>0.129</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.172</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0</td>
<td>0.233</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.9</td>
<td>0.130</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.3</td>
<td>0.130</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.165</td>
</tr>
<tr>
<td>0.0</td>
<td>0.9</td>
<td>0.205</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$k = 7$</th>
<th>$T = 25$</th>
<th>$T = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
<td>$p$</td>
</tr>
<tr>
<td>-0.9</td>
<td>0.0</td>
<td>0.117</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.0</td>
<td>0.130</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.150</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0</td>
<td>0.238</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.9</td>
<td>0.104</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.3</td>
<td>0.125</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.155</td>
</tr>
<tr>
<td>0.0</td>
<td>0.9</td>
<td>0.225</td>
</tr>
</tbody>
</table>

Table 2: For the likelihood ratio test of $\pi = \beta = 0$ in model (10) we give the rejection probability of a nominal 5% test based upon 10,000 simulations, $p$. We also give the corrected rejection probabilities with known parameter, $p_{corr}$, and corrected rejection probabilities with estimated parameters, $p_{cor}^{est}$. For $k = 7$, we take $\gamma_3 = \ldots = \gamma_6 = 0$. The lag length is assumed known.
4 Appendix

In this Appendix we first prove some identities, which we then apply to reduce the general expression in the formula for the correction factor.

4.1 Some matrix identities

We study the general expression for the correction factor using an eigenvalue decomposition of the matrix $P$, and we assume in the following that the eigenvalues are all of multiplicity one. This is no loss of generality as the results we prove are continuous functions of $P$. Let therefore $\rho$ be an eigenvalue of $P$ of multiplicity one. The left eigenvector is

$$\tau'_\rho = (1, \rho - \gamma_1, \rho^2 - \rho \gamma_1 - \gamma_2, \ldots, \rho^{k-2} - \rho^{k-3} \gamma_1 - \ldots - \gamma_{k-2})$$

and the right eigenvector is

$$\kappa'_\rho = (\rho^{k-2}, \rho^{k-3}, \ldots, 1)' .$$

We define the vector

$$v_\eta = (1, \eta, \ldots, \eta^{k-2})' ,$$

and find the relation

$$(I_{k-1} - \eta P)v_\eta = \phi(\eta)Q . \tag{12}$$

Multiplying by $\tau'_\rho$ we find

$$\tau'_\rho v_\eta = \frac{\phi(\eta)}{1 - \rho \eta}, \rho \eta \neq 1. \tag{13}$$

We collect the left eigenvectors in the matrix

$$T = (\tau_1, \tau_2, \ldots, \tau_{k-1}) ,$$

corresponding to the eigenvalues $\rho_1, \ldots, \rho_{k-1}$, and find that the variance can be simplified because

$$\tau'_\rho \Sigma \tau_\eta = \sigma^2 \sum_{m=0}^{\infty} \rho^m \tau'_\rho Q Q' \tau_\eta \eta^m = \frac{\sigma^2}{1 - \rho \eta} .$$

We introduce the coefficient

$$c(\eta) = \sigma^2 v'_1 \Sigma^{-1} v_\eta = \frac{\sigma^2 v'_1 T (T' \Sigma T)^{-1} T' v_\eta}{\phi(1) \phi(\eta)} \tag{14}$$

$$= \left\{ \frac{1}{1 - \rho_i} \right\}' \left\{ \frac{1}{1 - \rho_i \rho_j} \right\}^{-1} \left\{ \frac{1}{1 - \eta \rho_i} \right\} ,$$

where the last result follows from (13). For $\eta = 1$ we get the coefficient in (10).
We find the coefficient \( c(\eta) \) in Corollary 3 expressed in terms of the eigenvalues of \( P \), but the coefficient \( c(1) \) can also be found from a result of Wise (1955), where an explicit expression for \( \Sigma^{-1} \) is given in terms of \( \gamma \). For \( \gamma_0 = -1 \) we have

\[
\Sigma^{ij} = \sum_{m=0}^{i-1} \gamma_m \gamma_{m+j-i} - \sum_{m=k-j}^{k-1+i-j} \gamma_m \gamma_{m+j-i}, \quad i \leq j,
\]

see Galbraith and Galbraith (1974, formula (16) page 70), or Shaman and Stine (1988), where the result \( c(1) = k - 1 - 2d \log \phi(z)/dz|_{z=1} \) is given.

We next prove some identities involving the distinct complex numbers \( \rho_1, \ldots, \rho_l \).

**Lemma 2** Let the \( \rho \neq 0, \rho_1, \ldots, \rho_l \) be distinct and less than one in absolute value, then the identities hold

\[
\sum_{i=1}^{l} \frac{1 + \rho_i}{1 - \rho_i \rho_m} \prod_{j \neq i} \frac{1 - \rho_i \rho_j}{\rho_i - \rho_j} = \frac{1}{1 - \rho_m}, \tag{15}
\]

\[
\sum_{i=1}^{l} \frac{1 + \rho_i}{1 - \rho_i \rho_m} \prod_{j \neq i} \frac{1 - \rho_i \rho_j}{\rho_i - \rho_j} = \sum_{i=1}^{l} \frac{1 + \rho_i}{\rho_i - \rho_m} \prod_{j=1}^{i} \frac{1 - \rho_i \rho_j}{1 - \rho_i \rho_j}, \tag{16}
\]

**Proof.** We first prove (15). We consider the left hand side as a function of \( z = \rho_m \) and find for \( z \neq 1, \rho_1, \ldots, \rho_l \)

\[
f(z) = \frac{1}{1 - z} \frac{\prod_{j \neq m} (1 - z \rho_j)}{\prod_{j \neq m} (z - \rho_j)} + \sum_{i \neq m} \frac{1 + \rho_i}{\rho_i - z} \frac{\prod_{j \neq i} (1 - \rho_i \rho_j)}{\prod_{j \neq i} (\rho_i - \rho_j)}.
\]

The rational function \( f(z) \) has poles of order one at the points \( z = 1, \) and \( z = \rho_i, i \neq m, \) and the value zero at \( \infty \). Hence it has the expansion

\[
f(z) = \frac{b_m}{1 - z} + \sum_{i \neq m} b_i \frac{1}{z - \rho_i}.
\]

We find the coefficients from \( b_m = (1 - z)f(z)|_{z=1} = 1 \), and \( b_i = (z - \rho_i)f(z)|_{z=\rho_i}, \) which gives

\[
b_i = \frac{1 - \rho_i \rho_m \prod_{j \neq m,i} (1 - \rho_i \rho_j)}{1 - \rho_i \prod_{j \neq m} (\rho_i - \rho_j)} - (1 + \rho_i) \frac{\prod_{j \neq i} (1 - \rho_i \rho_j)}{\prod_{j \neq i} (\rho_i - \rho_j)} = 0,
\]

so that \( f(z) = \frac{1}{1 - z}, \) and \( f(\rho_m) = \frac{1}{1 - \rho_m}, \) which proves (15).

The next result (16) is proved by induction in a similar way. We define

\[
S_m = \sum_{i=1}^{m} \frac{1 + \rho_i \prod_{j \neq i} (1 - \rho_i \rho_j)}{1 - \rho \rho_i \prod_{j \neq i} (\rho_i - \rho_j)}, \quad m = 1, \ldots, l,
\]
and replace \( \rho_i \) by \( z \) and define the rational function \( f(z) \) for \( z \neq \rho^{-1}, \rho_1, \ldots, \rho_{l-1} \)

\[
f(z) = \frac{1 + z}{1 - \rho z} \prod_{j=1}^{l-1}(1 - z \rho_j) \sum_{i=1}^{l-1} \frac{1 + \rho_i}{1 - \rho \rho_i} \frac{1 - \rho_i z}{1 - \rho_i} \prod_{j \neq i}^{l-1}(1 - \rho_i \rho_j) \prod_{j \neq i}^{l-1}(\rho_i - \rho_j).
\]

The function \( f \) has poles at the points \( z = \rho^{-1}, \rho_1, \ldots, \rho_{l-1} \), and a value \( b_0 \) at infinity. It therefore has the representation

\[
f(z) = b_0 + \frac{b_l}{1 - \rho z} + \sum_{i=1}^{l-1} \frac{b_i}{\rho_i - z}.
\]

We first show that \( b_i = (z - \rho_i)f(z)\big|_{z=\rho_i} = 0 \) for \( i < l \):

\[
b_i = \frac{1 + \rho_i}{1 - \rho \rho_i} \prod_{j \neq i}^{l-1}(1 - \rho_j \rho_i) \prod_{j \neq i}^{l-1}(\rho_i - \rho_j) - \frac{1 + \rho_i}{1 - \rho \rho_i} \prod_{j \neq i}^{l-1}(1 - \rho_i \rho_j) = 0
\]

Next we find \( b_l = (1 - \rho z)f(z)\big|_{z=\rho^{-1}} \):

\[
b_l = (1 + \rho^{-1}) \prod_{j=1}^{l-1}(1 - \rho^{-1} \rho_j) = \frac{1 + \rho}{\rho} \prod_{j=1}^{l-1}(\rho_j - \rho) - \prod_{j=1}^{l-1}(1 - \rho \rho_j)
\]

Finally we determine \( b_0 \) by setting \( z = -1 \) and find

\[
f(-1) = \sum_{i=1}^{l-1} \frac{1 + \rho_i}{1 - \rho \rho_i} \prod_{j \neq i}^{l-1}(1 - \rho_j \rho_i) = S_{l-1} = b_0 + \frac{1}{\rho} \prod_{j=1}^{l-1}(\rho_j - \rho)
\]

so that

\[
f(z) = b_0 + \frac{b_l}{1 - z \rho} = S_{l-1} + \frac{1 + z}{1 - z \rho} \prod_{j=1}^{l-1}(\rho - \rho_j)
\]

For \( z = \rho_i \) we find

\[
f(\rho_i) = S_i = S_{l-1} + \frac{1 + \rho_i}{1 - \rho \rho_i} \prod_{j=1}^{l-1}(\rho_j - \rho) = S_{l-1} + \frac{1 + \rho_i}{\rho - \rho_i} \prod_{j=1}^{l-1}(1 - \rho \rho_j)
\]

Continuing like this we find

\[
f(\rho_i) = \sum_{i=1}^{l} \frac{1 + \rho_i}{\rho - \rho_i} \prod_{j=1}^{l}(\rho_j - \rho) = \sum_{i=1}^{l} \frac{1 + \rho_i}{\rho - \rho_i} \prod_{j=1}^{l}(1 - \rho \rho_j)
\]

and proves (16).

Next we apply Lemma 2 for the eigenvalues \( \rho_i \) and \( l = k - 1 \) to calculate the coefficient \( c(\eta) \), see (14).
Corollary 3 It follows from Lemma 2 that

\[
c(\eta) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{\eta - \rho_i} \prod_{j=1}^{i-1}(\eta - \rho_j), \eta \neq \rho_i^{-1} 
\]

(17)

\[
c(-1) = \frac{1}{2}(1 - (-1)^{k-1}),
\]

(18)

\[
c(1) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i}
\]

(19)

\[
c(\rho_m) = \frac{1}{1 - \rho_m}.
\]

(20)

**Proof.** The identities (15) and (16) involve the coefficients

\[
a_i = (1 + \rho_i) \frac{\prod_{j \neq i}(1 - \rho_i \rho_j)}{\prod_{j \neq i}(\rho_i - \rho_j)}, i = 1, \ldots, k - 1
\]

and (15) just states that

\[
\left\{ \frac{1}{1 - \rho_i \rho_m} \right\} \{a_i\} = \left\{ \frac{1}{1 - \rho_m} \right\},
\]

and hence that

\[
\left\{ \frac{1}{1 - \rho_i \rho_m} \right\}^{-1} \left\{ \frac{1}{1 - \rho_m} \right\} = \{a_i\}.
\]

This proves (17):

\[
c(\eta) = \left\{ \frac{1}{1 - \rho_j} \right\} \left\{ \frac{1}{1 - \rho_i \rho_j} \right\}^{-1} \left\{ \frac{1}{1 - \eta \rho_j} \right\} = \sum_{j=1}^{k-1} a_j \frac{1}{1 - \eta \rho_j}
\]

\[
= \sum_{j=1}^{k-1} \frac{1 + \rho_j}{1 - \eta \rho_j} \prod_{j \neq j}^{j-1}(1 - \rho_i \rho_j) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i} \prod_{j=1}^{i-1}(\eta - \rho_j)
\]

For \(\eta = -1\) in this expression, we find (18):

\[
c(-1) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i} \prod_{j=1}^{i-1}(1 - \rho_j) = -\sum_{i=1}^{k-1} (-1)^i = \frac{1}{2}(1 - (-1)^{k-1}).
\]

and for \(\eta = 1\), we find (19):

\[
c(1) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i} \prod_{j=1}^{i-1}(1 - \rho_j) = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i}
\]

Finally in order to prove (20). We observe that

\[
c(\rho_m) = \left\{ \frac{1}{1 - \rho_j} \right\} \left\{ \frac{1}{1 - \rho_i \rho_j} \right\}^{-1} \left\{ \frac{1}{1 - \eta \rho_j} \right\} |_{\eta = \rho_m} = \frac{1}{1 - \rho_m},
\]

because \(\left\{ \frac{1}{1 - \rho_i \rho_j} \right\}^{-1} \left\{ \frac{1}{1 - \eta \rho_j} \right\} |_{\eta = \rho_m}\) is the m'th unit vector. ■
4.2 Proof of Theorem 1

In Johansen (2002) the correction factor was given for an $n-$dimensional process with highest order deterministic term $f^{m+1}$ by the expression $a(T, n, n_d)(1 + T^{-1}b(\theta))$, where

$$b(\theta) = c_1(1 + h(n, n_d)) + (nc_2 + 2(c_3 + dc_1))g(n, n_d)/n^2$$

(21)

and

$$a(T, n, n_d) = 1 + a_1(n_d)(n/T) + a_2(n_d)(n/T)^2 + a_3(n_d)(n/T)^3 + b(n_d)/T.$$  

The functions $a(T, n, n_d), h(n, n_d)$ and $g(n, n_d)$ are determined by simulation of expectations of functions of random walk, whereas $c_i$ are analytic functions of the parameters of the model. In the present case $n = 1, n_d = d$, and we decided to perform new simulations to determine the coefficients $f_T(d), a_1(d), a_2(d), h(1, d)$, and $g(1, d)$.

4.2.1 The coefficients $f_T(d), a_1(d), a_2(d)$

We simulate the model (1) with $k = 1$. In this case $f_T(d) = E_T[-2 \log LR\{M_0|\mathcal{M}_1\}]$ is a function of $T$ and $d$ only and for $T = 10, 25, 50, 100, 300, 500, 700, 1000, 2000, 3000$ we run 2.5 mill. simulations and fit a second order polynomial in $T^{-1}$, using the intercept $f(\ast), f(0), f(1)$ given in (6). We thereby find the coefficients $(f(2), a_1(d), a_2(d))$ in the approximation

$$f_T(d) \sim f(d)(1 + a_1(d)T^{-1} + a_2(d)T^{-2}) = f(d)a_T(d).$$

The values for $a_1(d)$ and $a_2(d)$ are given in Table 1. The values of $a_1(d)$ for $d = \ast$ and 0 are also found analytically in Nielsen (1997a), but fitting the two coefficients by regression seems to describe the functions better.

4.2.2 The coefficients $c_i$

We find the coefficients $c_i$ from Johansen (2002, page 1937):

$$c_1 = \text{tr}\{(I_{n_y} - P)^{-1}Q\Omega\Psi^{-1}\Psi^{-1}\Omega'(I_{n_y} - P)^{-1}\Sigma^{-1}\},$$
$$c_2 = \text{tr}\{(I_{n_y} - (I_{n_y} - P)^{-1}Q\Omega'Q'(I_{n_y} - P)'\Sigma^{-1}\},$$
$$c_3 = \text{tr}\{\psi'\Sigma^{-1}\sum_{m=0}^{\infty} \psi_m\text{tr}\{\Sigma^{-1}\gamma(m + 1)\} + \text{tr}\{\psi'\Sigma^{-1}\sum_{m=0}^{\infty} \gamma(m + 1)'\Sigma^{-1}\psi_m\}.}$$

In the present case $\alpha = 0, \alpha_\perp = 1, n_y = k - 1$, and $\Omega = \sigma^2$, so that from (12) we get $\psi = (I_{k-1} - P)^{-1}Q = \phi(1)^{-1}v_1$ and find from (19)

$$c_1 = \sigma^2\text{tr}\{(I_{k-1} - P)^{-1}Q'Q'(I_{k-1} - P)'\Sigma^{-1}\} = \frac{\sigma^2}{\phi(1)^{1/2}}v_1'\Sigma^{-1}v_1 = c(1),$$
$$c_2 = k - 1 - c_1.
The expression for \(c_3\) depends on the coefficient \(\gamma(m + 1)' = P^{m+1} \Sigma\) and \(\psi_m = P^m Q\). We find, using that \((I_{k-1} - \eta P)^{-1} Q = \phi(\eta)^{-1} v_n\), see (12),

\[
c_3 = \frac{1}{\phi(1)} v_1^t \Sigma^{-1} [\sum_{m=0}^\infty P^m Q \text{tr} \{P^{m+1}\} + \sum_{m=0}^\infty P^{2m+1} Q]
= \frac{1}{\phi(1)} v_1^t \Sigma^{-1} \sum_{i=1}^{k-1} \rho_i (I_{k-1} - \rho_i P)^{-1} Q + \frac{1}{2} (I_{k-1} - P)^{-1} (I_{k-1} + P)^{-1} Q)
= \frac{1}{\phi(1)} v_1^t \Sigma^{-1} \sum_{i=1}^{k-1} \rho_i (I_{k-1} - \rho_i P)^{-1} Q + \frac{1}{2} (I_{k-1} - P)^{-1} \left[ v_1 - \frac{1}{2 \phi(1)} v_1 - \frac{1}{2 \phi(-1)} v_{-1} \right]
= \sum_{i=1}^{k-1} \rho_i c(\rho_i) + \frac{1}{2} (c(1) - c(-1)).
\]

The first term can be expressed in terms of \(c_2\)

\[
2 \sum_{i=1}^{k-1} \rho_i c(\rho_i) = \sum_{i=1}^{k-1} \frac{2 \rho_i}{1 - \rho_i} = \sum_{i=1}^{k-1} \frac{1 + \rho_i}{1 - \rho_i} - (k - 1) = c(1) - (k - 1) = -c_2,
\]

see Corollary 3, so that the coefficient of \(g(d)\) becomes

\[
c_2 + 2(c_3 + dc_1) = c_2 - c_2 + c_1 - c(-1) + 2dc_1 = (1 + 2d)c_1 - c(-1).
\]

Therefore for \(n = 1\) we find from (21)

\[
b(\theta) = c_1(1 + h(1, d)) + (c_2 + 2(c_3 + dc_1)) g(d)
= c_1(1 + h(1, d)) + ((1 + 2d)c_1 - c(-1)) g(d)
= c_1(1 + h(1, d) + (1 + 2d)g(d)) - c(-1) g(d)
= c_1 m(d) - \frac{1}{2} (1 - (-1)^{k-1}) g(d),
\]

where we have defined

\[
m(d) = 1 + h(1, d) + (1 + 2d) g(d),
\]

where \(d = \ast\) is interpreted as \(0\).

### 4.2.3 The coefficients \(g(d)\)

The bias of the least squares estimator is also found analytically in Nielsen (1997a) for \(d = \ast, 0, 1\). These results can be used to find the function \(g(d) = g(1, d)\). In order to see this we need some notation. In the simulations we generate data from

\[
\mathcal{M}_1^0 : \Delta x_t = \sum_{i=0}^{d-1} \beta_t t^i + \varepsilon_t
\]

with \(\sigma^2 = 1\), and we define the notation

\[
d_t = (1, \ldots, t^{d-1})', \quad D_t = t^d, \quad S_t = \sum_{i=1}^t \varepsilon_i, \quad A_{t-1} = \left( \begin{array}{l} S_{t-1} \\ D_t \end{array} \right),
\]
that is, $S_{t-1}$ and $D_t$ corrected for $d_t$. We define the product moments

$$M_{aa} = \sum_{t=1}^{T} A_{t-1}A_{t-1}', \quad M_{ae} = \sum_{t=1}^{T} A_{t-1}\varepsilon_t, \quad M_{ae}^+ = \sum_{t=1}^{T} A_{t-1}\varepsilon_{t-1},$$

$$M_{\varepsilon\varepsilon} = \sum_{t=1}^{T} \varepsilon_t^2, M_{ad} = \sum_{t=1}^{T} A_{t-1}d_t', \quad M_{ed} = \sum_{t=1}^{T} \varepsilon_t d_t', \text{ etc.}$$

We use the notation

$$M_{aa,d} = M_{aa} - M_{ad}M_{dd}^{-1}M_{da}. \quad \text{In terms of these the definition of } g(d) \text{ is}$$

$$g(d) = \lim_{T \to \infty} E[M_{ea,d}M_{aa,d}^{-1}M_{ae,d}^+]/E[M_{ea,d}M_{aa,d}^{-1}M_{ae,d}].$$

We also define $\tilde{\pi}_d = M_{\varepsilon,S.D,D}/M_{SS,D,D}$, the least squares estimator of $\pi$ in model $M_1$.

In Nielsen (1997a) the asymptotic bias for $\tilde{\pi}_d, d = *, 0, 1$ when $\pi = \beta = 0$ is found to be

$$\lim_{T \to \infty} E(T\tilde{\pi}_*) = -1.7814, \lim_{T \to \infty} E(T\tilde{\pi}_0) = -5.3791, \lim_{T \to \infty} E(T\tilde{\pi}_1) = -10.2455. \quad (22)$$

The relation between this bias and the coefficient $g(d)$ is derived from the next Lemma

**Lemma 4** For all $d$ it holds that

$$M_{ea,d}M_{aa,d}^{-1}M_{ae,d}^+ - M_{ea,d}M_{aa,d}^{-1}M_{ae,d} - T\tilde{\pi} \overset{P}{\to} 0$$

**Proof.** We find from

$$M_{ea,d}M_{aa,d}^{-1}M_{ae,d}^+ = M_{d,d}M_{DD,d}^{-1/2}M_{De,d}^+ + M_{\varepsilon,S.D.d}M_{SS,D,d}^{-1}M_{S\varepsilon,D,d}^+$$

and a similar expression for $M_{sa,d}M_{aa,d}^{-1}M_{ae,d}$ that

$$M_{ea,d}M_{aa,d}^{-1}M_{ae,d}^+ - M_{ea,d}M_{aa,d}^{-1}M_{ae,d}^+ - T\tilde{\pi}_d$$

$$= M_{d,d}M_{DD,d}^{-1/2}(M_{De,d}^+ - M_{De,d}) + T\tilde{\pi}_d(T^{-1}(M_{S\varepsilon,D,d}^{-1} - M_{S\varepsilon,D,d}) - 1).$$

We see that

$$M_{\varepsilon,D,d}M_{DD,d}^{-1/2} \in O_P(1),$$

$$M_{DD,d}^{-1/2}(M_{De,d}^+ - M_{De,d}) = M_{DD,d}^{-1/2}\sum_{t=1}^{T} (D_t|d_t)(\varepsilon_{t-1} - \varepsilon_t) \in o_P(1),$$

so that the first term tends to zero. Next we use the notation $w_t = (d_t', D_t)'$ and consider

$$T^{-1}(M_{S\varepsilon,w}^+ - M_{S\varepsilon,w}) - 1$$

$$= (T^{-1}\sum_{t=1}^{T} S_{t-1}(\varepsilon_{t-1} - \varepsilon_t) - 1) - T^{-1}M_{Sw}M_{ww}^{-1/2}M_{ww}^{-1/2}\sum_{t=1}^{T} w_t(\varepsilon_{t-1} - \varepsilon_t).$$
The first term is
\[ T^{-1} \sum_{t=0}^{T-1} \varepsilon_t^2 - 1 - T^{-1} S_{T-1} \varepsilon_T \in o_P(1), \]
and the second term has the factors
\[ T^{-1} M_{sw} M_{ww}^{-1/2} \in o_P(1), \]
\[ M_{ww}^{-1/2} \sum_{t=1}^{T} w_t (\varepsilon_{t-1} - \varepsilon_t) \in o_P(1), \]
which completes the proof. □

We have proved convergence in probability, and we shall use that to suggest, as confirmed by the simulations, that we also have
\[ E[M_{ea,d} M_{aa,d}^{-1} M_{ae,d}^+] - E[M_{ea,d} M_{aa,d}^{-1} M_{ae,d}] - E[T \hat{\pi}_d] \to 0, \]
which together with
\[ E[M_{ea,d} M_{aa,d}^{-1} M_{ae,d}] \to f(d), \]
shows that
\[ g(d) = \lim_{T \to \infty} \frac{E[M_{ea,d} M_{aa,d}^{-1} M_{ae,d}^+]}{E[M_{ea,d} M_{aa,d}^{-1} M_{ae,d}]} = 1 + \lim_{T \to \infty} E[T \hat{\pi}_d]/f(d), \]
so that from (6) and (22) we find
\[ g(*) = -0.5601, \ g(0) = -0.3262, \ g(1) = -0.6209, \]
as given in Table 1.

4.2.4 The coefficients \( m(d) \)

The coefficients \( m(d) = 1 + h(1, d) + (1 + 2d) g(1, d) \) are more difficult to simulate accurately, as they contain the term
\[ \left( \sum_{t=1}^{T} \varepsilon_t \varepsilon_{t-1} \right) M_{ea,d} M_{aa,d}^{-1} M_{ae,d}, \]
which evidently has a variance proportional to \( T \), but many million simulations indicate that the mean is convergent to a non-zero limit. In Table 1 the uncertainty of the value of the simulated value \( m(d) \) is given as well.

5 References


Nielsen, B., 1997b. Bartlett correction of the unit root test in autoregressive models, Biometrika 84, 500–504.


Figure 1:

The rejection probabilities are plotted as a function of the parameter. At the top we have the uncorrected rejection probability, $p$, for a nominal 5% test. The rejection probability using the true parameter value, $p_{corr}$, decreases to zero for large parameter values, and the rejection probability, $p_{corr}^{est}$, using the estimated value in each simulation, is the middle curve.
Figure 2:

Figure 3: