Testing for unit root against stationarity

using the likelihood ratio test

Nikolay Angelov† and Rolf Larsson‡

August 17, 2005

Abstract

In a first order autoregressive model with drift, we derive the likelihood ratio test for a unit root against the stationary alternative. We also derive the test in a state space model with trend. Finite sample and asymptotic critical values are obtained by Monte Carlo simulations. A simulation study investigates the power performance of the likelihood ratio test.

Keywords: unit root; LR test; stationary alternative.

1 Introduction

During the last decades, unit root testing in autoregressive models has been given extensive attention in the literature, see e.g. Dickey and Fuller (1981) and Phillips and Xiao (1998). However, focus is seldom on the case when only stationary models are allowed under the alternative. (In the following, we will refer to this as the constrained case.) For example, Dickey and Fuller derive the likelihood ratio test of a unit root versus the alternative of no unit root, which contains both

†Department of Economics, Uppsala University. E-mail: Nikolay.Angelov@nek.uu.se
‡Department of Information Science, Division of Statistics, Uppsala University. E-mail: Rolf.Larsson@dis.uu.se
stationary and explosive cases. This is an F type test, which fails to discriminate between the two cases under the alternative. An attempt to solve this problem is made by Abadir and Distaso (2003), who propose a general class of Wald type statistics where the direction of one of the alternatives is incorporated. A test which is applicable to our situation (but in a state space model) arises as a special case.

In the present paper, we derive the likelihood ratio test for a unit root against stationarity in autoregressive models of order one with a drift or a linear trend. This extends the work of Dickey and Fuller. For the sake of completeness, we re-derive some of their results. After a short presentation of our models in section 2, these results are collected in section 3. Section 4 extends the discussion to the constrained case. In section 5, we discuss the state space model of Abadir and Distaso. Monte Carlo simulation of critical values and a power study are presented in section 6. Finally, some concluding thoughts are given in section 7. Omitted proofs are given in the appendix.

2 Models and parametrisations

Consider the model

\[ x_t = \alpha + \rho x_{t-1} + \epsilon_t, \]  

where we have observations at \( t = 1, 2, \ldots, T \), \( x_0 \) is fixed and the \( \epsilon_t \) are independent \( N(0, \sigma^2) \). We want to test \( H_0: (\alpha, \rho) = (0, 1) \). We may test against either the unconstrained alternative \( H_1: \{ \alpha \neq 0, \rho < 1 \} \). (Observe that \( \alpha \) is not restricted under the alternatives.) We may also include a time trend in our model as follows:

\[ x_t = \alpha + \beta t + \rho x_{t-1} + \epsilon_t, \]  

where we test \( H_0: (\alpha, \beta, \rho) = (0, 0, 1) \) against \( H_1: \{ \alpha \neq 0, \beta \neq 0, \rho < 1 \} \). Alternatively, we may test \( H_0': (\beta, \rho) = (0, 1) \) against \( H_1': \{ \beta \neq 0, \rho < 1 \} \).
Both models may be reformulated as

\[ x_t - \mu_t = \rho (x_{t-1} - \mu_t) + \varepsilon_t, \]  \tag{3}\]

where we get (1) if \( \mu_t = \mu \), i.e. \( (1 - \rho) \mu = \alpha \). With \( \mu_t = \mu_0 + \mu_1 t \), we obtain (2).

Here, \( (1 - \rho) (\mu_0, \mu_1)' = (\alpha, \beta)' \).

3 The unconstrained case

Under the model with constant and no trend, the likelihood ratio test statistic and its asymptotic distribution are given in the following theorem.

**Theorem 1** Assume that (1) holds. Let

\[
Y \equiv \begin{pmatrix} \Delta x_1 & \cdots & \Delta x_T \end{pmatrix},
\]

\[
X_1 \equiv \begin{pmatrix} 1 & x_0 \\ \vdots & \vdots \\ 1 & x_{T-1} \end{pmatrix}.
\]

Then,

1. The maximum Likelihood Estimator (MLE) of \( \rho \) is given by

\[
\hat{\rho} = 1 + \begin{pmatrix} 0 & 1 \end{pmatrix} X_1 (X'_1 X_1)^{-1} X'_1 Y.
\]

2. The likelihood ratio test, \( Q_{1T} \) say, of \( H_0 : (\alpha, \rho) = (0, 1) \) against \( H_1 : \lnot H_0 \) satisfies

\[
-2 \log Q_{1T} = -T \log \left( 1 - \frac{Y'X_1 (X'_1 X_1)^{-1} X'_1 Y}{Y'Y} \right).
\]

3. Under \( H_0 \) and as \( T \to \infty \),

\[
-2 \log Q_{1T} \overset{d}{\to} Z_1.
\]
where
\[
Z_1 \equiv \frac{W(1)^2 \int_0^1 W(t)^2 \, dt - 2W(1) \int_0^1 W(t) \, dt \int_0^1 W(t) \, dW(t) + \left( \int_0^1 W(t) \, dW(t) \right)^2}{\int_0^1 W(t)^2 \, dt - \left( \int_0^1 W(t) \, dt \right)^2},
\]

\(W(t)\) being a standard Wiener process.

**Proof.** See the appendix. ■

The corresponding result for the model with trend is the following.

**Theorem 2** Assume that (2) holds. Let
\[
X_2 \equiv \begin{pmatrix}
1 & 1 & x_0 \\
\vdots & \vdots & \vdots \\
1 & T & x_{T-1}
\end{pmatrix}.
\]

Then,

1. The MLE of \(\rho\) is given by
\[
\hat{\rho} = 1 + \begin{pmatrix}
0 & 0 & 1
\end{pmatrix} X_2 (X_2'X_2)^{-1} X_2' Y.
\]

2. The likelihood ratio test, \(Q_{2T}\) say, of \(H_0: (\alpha, \beta, \rho) = (0, 0, 1)\) against \(H_1: \nabla H_0\) satisfies
\[
-2 \log Q_{2T} = -T \log \left( 1 - \frac{Y'X_2 (X_2'X_2)^{-1} X_2' Y}{Y'Y} \right).
\]

3. Under \(H_0\) and as \(T \to \infty\),
\[
-2 \log Q_{2T} \xrightarrow{d} Z_2 \equiv \frac{D_{\infty}}{C_{\infty}},
\]
where
\[
C_\infty \equiv -4 \left( \int_0^1 W(t) \, dt \right)^2 + 12 \int_0^1 W(t) \, dt \int_0^1 tW(t) \, dt - 12 \left( \int_0^1 tW(t) \, dt \right)^2 + \int_0^1 W(t)^2 \, dt
\]

and
\[
D_\infty \equiv -12 \left\{ W(1) \int_0^1 tW(t) \, dt - W(1) \int_0^1 W(t) \, dt + \left( \int_0^1 W(t) \, dt \right)^2 \right\} - 4 \left\{ -6 \int_0^1 W(t) \, dt \int_0^1 tW(t) \, dt + 3 \left( \int_0^1 W(t) \, dt \right)^2 \right\}
- W(1) \int_0^1 W(t) \, dt + 3W(1) \int_0^1 tW(t) \, dt \right\} \int_0^1 W(t) \, dW(t)
- 4 \left\{ -3 \left( \int_0^1 W(t) \, dt \right)^2 + 3W(1) \int_0^1 W(t) \, dt - W(1)^2 \right\} \int_0^1 W(t)^2 \, dt
+ \left( \int_0^1 W(t) \, dW(t) \right)^2.
\]

4. The likelihood ratio test, \( Q_{3T} \) say, of \( H'_0: (\beta, \rho) = (0, 1) \) against \( H'_1: |H'_0 \)
satisfies
\[
-2 \log Q_{3T} = -T \log \left( 1 - \frac{Y' \left( X_2 (X_2'X_2)^{-1} X_2' - T^{-1}X_0X_0' \right) Y}{Y' (I_T - T^{-1}X_0X_0') Y} \right),
\]
where
\[
X_0 = (1, \ldots, 1)'.
\]

5. Under \( H'_0 \) and as \( T \to \infty \),
\[
-2 \log Q_{3T} \xrightarrow{d} Z_3 \equiv \frac{D_\infty}{C_\infty} - W(1)^2.
\]
4 The constrained case

Consider the model formulation in (3). It follows that the log likelihood, maximized over \( \sigma^2 \) and \( \mu_t \), is

\[
    l(\hat{\mu}_t, \hat{\sigma}^2, \rho) = c - \frac{T}{2} \log \sum_{t=1}^{T} \{(x_t - \hat{\mu}_t) - \rho (x_{t-1} - \hat{\mu}_t)\}^2,
\]

where \( c \) is a constant. Moreover, differentiating w.r.t. \( \rho \), we obtain

\[
    \frac{\partial}{\partial \rho} l(\hat{\mu}_t, \hat{\sigma}^2, \rho) = T \sum_{t=1}^{T} \frac{(x_t - \hat{\mu}_t) ((x_t - \hat{\mu}_t) - \rho (x_{t-1} - \hat{\mu}_t)) \{ (x_t - \hat{\mu}_t) - \rho (x_{t-1} - \hat{\mu}_t) \}^2}{\sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_t)^2}, \tag{4}
\]

which, in the unconstrained case, yields the ML estimator

\[
    \rho^* = \frac{\sum_{t=1}^{T} (x_t - \hat{\mu}_t) (x_{t-1} - \hat{\mu}_t)}{\sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_t)^2}. \tag{5}
\]

In the constrained case, however, the r.h.s. of (4) is positive for all \( \rho \leq 1 \) if \( \sum_{t=1}^{T} (x_t - \hat{\mu}_t) (x_{t-1} - \hat{\mu}_t) > \sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_t)^2 \). This means that if the r.h.s. of (5) is larger than one, the likelihood is maximized for \( \rho = 1 \). Conclusively, under the constraint \( \rho \leq 1 \),

\[
    \hat{\rho} = \min \left\{ \frac{\sum_{t=1}^{T} (x_t - \hat{\mu}_t) (x_{t-1} - \hat{\mu}_t)}{\sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_t)^2}, 1 \right\}.
\]

If \( \hat{\rho} = 1 \), it is easily seen that the likelihood attains the same maximum under \( H_0 \) and \( H_1 \), and so, the likelihood ratio is 1 in this case. Hence, the likelihood ratio test rejects \( H_0 \) for small values of \( Q \) (large values of \(-2 \log Q\)) in connection with \( \hat{\rho} < 1 \). In other words, the critical value of the likelihood ratio test may be found by Monte Carlo simulation by repeating the following two steps a large number of times:

1. Generate the \( \varepsilon_t \) and calculate \( \rho^* \) and \(-2 \log Q\) as in the unconstrained case.
2. If \( \rho^* < 1 \), save the value of \(-2 \log Q\), otherwise put \(-2 \log Q = 0\).
The so obtained vector of simulated statistics is an estimator of the empirical distribution function of $-2 \log Q$. This procedure may be performed for any sample size $T$, and increasing $T$ gives convergence to the asymptotic distribution.

An alternative way of estimating the empirical distribution function of $-2 \log Q$ is to use the asymptotic distribution directly. For the case with no trend, the distribution of $Z_1$ from theorem 1 can be simulated under the constraint $\rho \leq 1$. In terms of sample moments (see the appendix for a definition of $\lambda$), the estimator of $\rho = 1 + \delta$ is given in theorem 1, and it follows that the condition for $\hat{\rho} < 1$ can be expressed as

$$T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta x_t < T^{-1} \lambda_1 (x_T - x_0).$$

Asymptotically and under $H_0$, (8), (10), and (11) yield that this inequality boils down to

$$\frac{1}{2} \left\{ W(1)^2 - 1 \right\} < W(1) \int_0^1 W(t)dt.$$

Consequently, the empirical asymptotic distribution of $-2 \log Q$ under $\rho \leq 1$ can be estimated in a similar manner as above. The integrals involving the Wiener process can be approximated using pseudo-random normally distributed vectors of large size. One such vector gives one observation from the empirical distribution of $Z_1$; if the constraint is satisfied, the observation is saved, otherwise $Z_1$ is set to zero. For a large $T$, this conditional distribution converges to the conditional distribution described previously.

5 The state space model

Abadir and Distaso (2003) consider the specification
\[ x_1 = \alpha - \beta \frac{T}{2} + \varepsilon_1, \]
\[ \Delta x_t = \delta x_{t-1} - \alpha \delta + \beta (\delta + 1) - \beta \delta (t - 1 - \frac{T}{2}) + \varepsilon_t, \quad (6) \]

where \( \delta \equiv \rho - 1, \ t = 2, 3, \ldots, T \). This model is not equivalent to (2), since it is nonlinear in the parameters. It is hard to derive the likelihood ratio test explicitly, but as a help when calculating it numerically, at first assume \( \delta \) fixed and rewrite (6) as
\[ \tilde{Y} = \tilde{X}B + \varepsilon, \]
where \( \varepsilon \) is as above and
\[
\begin{align*}
\tilde{Y} &\equiv \left( \Delta x_1 - \delta x_0 \quad \cdots \quad \Delta x_T - \delta x_{T-1} \right)^\prime, \\
\tilde{X} &\equiv \begin{pmatrix}
\delta & 1 - \delta (-1 - \frac{T}{2}) \\
\vdots & \vdots \\
\delta & 1 - \delta (T - 2 - \frac{T}{2})
\end{pmatrix}, \\
B &\equiv \begin{pmatrix}
-\alpha & \beta
\end{pmatrix}^\prime.
\end{align*}
\]

Hence, defining
\[ \tilde{\sigma}^2(\delta) \equiv T^{-\frac{1}{2}} \tilde{Y}^\prime \left( I_3 - \tilde{X} \left( \tilde{X}^\prime \tilde{X} \right)^{-1} \tilde{X}^\prime \right) \tilde{Y}, \]
the log likelihood, maximized w.r.t. \( \alpha, \beta \) and \( \sigma^2 \) is
\[ l_1 (\delta) \equiv -\frac{T}{2} \{ \log (2\pi) + \log \tilde{\sigma}^2 (\delta) + 1 \}. \]

This is a function of \( \delta \), which may be maximized numerically to yield the maximum log likelihood \( l_1 (\delta) \) under \( H_1 \). This is equivalent to minimizing \( \tilde{\sigma}^2 (\delta) \). The
maximum log likelihood under $H_0$ is

$$l_0 \equiv \frac{T}{2} \{ \log (2\pi) + \log \hat{\sigma}_0^2 + 1 \},$$

and so, in the unconstrained case,

$$-2 \log Q = -T \log \left\{ \frac{\hat{\sigma}_0^2}{\hat{\sigma}_0^2} \right\}.$$ 

If $l_1(\hat{\delta})$ is monotone increasing in $\delta$ when $\hat{\delta} > 0$, we may go on as outlined in the previous section to calculate the likelihood ratio test and to simulate its critical values. Unlike for the specification in (3), it is difficult to prove this conjecture analytically for the state space model. Nevertheless, it can be checked numerically, which we will return to in the next section.

<table>
<thead>
<tr>
<th>$T$</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>0.6122</td>
<td>0.6081</td>
<td>0.5946</td>
<td>0.5829</td>
<td>0.5978</td>
</tr>
<tr>
<td>2.5%</td>
<td>0.8056</td>
<td>0.7959</td>
<td>0.7895</td>
<td>0.7823</td>
<td>0.7898</td>
</tr>
<tr>
<td>5%</td>
<td>1.0279</td>
<td>1.0170</td>
<td>1.0211</td>
<td>1.0044</td>
<td>1.0083</td>
</tr>
<tr>
<td>10%</td>
<td>1.3536</td>
<td>1.3591</td>
<td>1.3534</td>
<td>1.3408</td>
<td>1.3408</td>
</tr>
<tr>
<td>90%</td>
<td>7.6364</td>
<td>7.5869</td>
<td>7.5618</td>
<td>7.5435</td>
<td>7.5341</td>
</tr>
<tr>
<td>97.5%</td>
<td>10.8995</td>
<td>10.7748</td>
<td>10.8062</td>
<td>10.7218</td>
<td>10.6643</td>
</tr>
<tr>
<td>99%</td>
<td>12.9047</td>
<td>12.7982</td>
<td>12.8697</td>
<td>12.7718</td>
<td>12.6644</td>
</tr>
</tbody>
</table>

Table 1: Simulated percentiles of the distribution of $-2 \log Q$, the case without trend; 100,000 replications.

### 6 Simulated distributions and a power study

The critical values are computed using simulated unit-root series of various sizes. For each series of size $T$, we compute both the unrestricted and the restricted likelihood-ratio statistics from section 4, and by doing so $R$ times, we get their simulated distributions. We set $R = 100,000$, and the data-generating process is
\[
x_0 = 0
\]
\[
x_t = x_{t-1} + \epsilon_t,
\]
where \(\{\epsilon_t\}_{t=1}^T\) is a sequence of pseudo-random i.i.d. \(N(0, 1)\) variables. (All numerical results in the paper are generated using Ox version 3.40.) The drift parameter is set to zero since it does not affect the estimated distribution under the null.

The critical values reported in Table 1 are the percentiles from the simulated distribution of \(-2 \log Q\). For ease of comparison, the same seed was used for simulating both sets of critical values. As apparent from the table, the critical values differ mostly in the left-hand tail of the distribution, while the right-hand tail (which is the relevant one for performing the LR-test) does not change a lot when the condition \(\hat{\rho} < 1\) is applied.

Tables 2 and 3 show that the same applies when we allow for a deterministic trend in the model (see (2)). In fact, only a small share of ML-estimates in the simulation are larger than one: about 3.5 percent when testing in (1), 0.5 percent when testing \(H_0\) in (2), and 0.4 percent when testing \(H_0'\) in (2).

When we apply the constrained LR-test on a specific series, we first calculate \(-2 \log Q\) based on the constrained ML-estimator of \(\rho\), and then compare that
Table 3: Simulated percentiles of the distribution of $-2 \log Q$, deterministic trend allowed under both $H_0$ and $H_1$; 100,000 replications.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1.6223</td>
<td>1.5991</td>
<td>1.5462</td>
<td>1.5346</td>
<td>1.5309</td>
</tr>
<tr>
<td>2.5%</td>
<td>1.9746</td>
<td>1.9412</td>
<td>1.9020</td>
<td>1.8852</td>
<td>1.8901</td>
</tr>
<tr>
<td>5%</td>
<td>2.3382</td>
<td>2.3006</td>
<td>2.2747</td>
<td>2.2709</td>
<td>2.2710</td>
</tr>
<tr>
<td>10%</td>
<td>2.8603</td>
<td>2.8313</td>
<td>2.8053</td>
<td>2.7942</td>
<td>2.8105</td>
</tr>
<tr>
<td>90%</td>
<td>10.7560</td>
<td>10.6593</td>
<td>10.6966</td>
<td>10.6881</td>
<td>10.6558</td>
</tr>
<tr>
<td>95%</td>
<td>12.6343</td>
<td>12.5136</td>
<td>12.5644</td>
<td>12.5321</td>
<td>12.5144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1.3781</td>
<td>1.4060</td>
<td>1.3853</td>
<td>1.3907</td>
<td>1.3988</td>
</tr>
<tr>
<td>2.5%</td>
<td>1.8568</td>
<td>1.8434</td>
<td>1.8225</td>
<td>1.8096</td>
<td>1.8206</td>
</tr>
<tr>
<td>5%</td>
<td>2.2649</td>
<td>2.2442</td>
<td>2.2210</td>
<td>2.2171</td>
<td>2.2247</td>
</tr>
<tr>
<td>10%</td>
<td>2.8131</td>
<td>2.7914</td>
<td>2.7675</td>
<td>2.7638</td>
<td>2.7788</td>
</tr>
<tr>
<td>90%</td>
<td>10.7475</td>
<td>10.6540</td>
<td>10.6909</td>
<td>10.6852</td>
<td>10.6530</td>
</tr>
<tr>
<td>95%</td>
<td>12.6270</td>
<td>12.5067</td>
<td>12.5594</td>
<td>12.5299</td>
<td>12.5130</td>
</tr>
</tbody>
</table>

The statistic with the relevant critical value. If the sample statistic is larger than the critical value, we reject the null hypothesis of a unit root. But since the right-hand tails of both the constrained and the unconstrained distributions do not differ a lot, neither do the critical values.

Moreover, if $-2 \log Q$ computed from a certain series is based on the constrained ML-estimator of $\rho$, the test will not reject $H_0$ for $\hat{\rho} > 1$. But for the cases when $\hat{\rho} < 1$, the constrained test will be almost equivalent to the unconstrained one, given the small differences between the critical values. Thus, for $\hat{\rho} < 1$ the constrained LR-test will only give very minor power gains compared to the unconstrained one, if any at all.

We estimate the powers of the various tests using a Monte Carlo simulation performed in the standard manner: For given $\alpha, \beta, \rho$, and $T$, we construct 100,000 series according to model (1) or (2). For each draw and set of parameter values, we compute the test statistic $-2 \log Q$ (with the constraint $\hat{\rho} < 1$), thus acquiring a simulated distribution. Finally, we calculate the portion of the distribution which is above the relevant 5% critical value.

Table 4 contains the estimated power function of the LR-test for the case without trend. As expected, for $\rho < 1$ the power is decreasing in $\rho$. The power increases with $\alpha$. Within parentheses is the power of the test involving the $\Phi_1$ statistic from Dickey and Fuller (1981), which has $H_1 : \{\rho \neq 1\}$ and is equivalent.
Table 4: Estimated power function of the test at 5% significance level, constrained case without trend. The data-generating process is \( x_t = \alpha + \rho x_{t-1} + \epsilon_t \), \( H_0 : \{ \alpha = 0, \rho = 1 \} \), and \( H_1 : \{ \alpha \neq 0, \rho < 1 \} \). Calculated power of two-sided test involving the \( \Phi_1 \)-statistic from Dickey and Fuller (1981) within parentheses. The power function was simulated using 100,000 replications.

Table 5 shows the estimated power of the test where we allow for a deterministic time trend under the alternative. The overall power is lower than in the case without trend, and it is increasing in \( \alpha \). Interestingly, in some intervals, the power is not falling in \( \rho \). For instance when \( \beta = 0, \alpha = 0.5, T = 100 \), and when \( \rho \) falls from 0.99 to 0.95, the estimated power falls from 0.67 to 0.10, and then increases for lower values of \( \rho \). This result is shared with the test involving the \( \Phi_2 \) statistic from Dickey and Fuller (1981). We see that also here, the unconstrained case and the constrained one give the same estimated power when \( \rho < 1 \), and that the constrained one has no power in the explosive case. Thus, when \( \beta > 0 \) and...
\( \rho = 1 \), the constrained test is less likely to reject the null – a nice result, as already mentioned.

We get similar results when we allow for a time trend also under the null hypothesis, as shown in Table 6.

Finally, Table 7 contains the calculated power of the standard Dickey and Fuller one-sided \( \tau \)-test of \( H_0 : \rho = 1 \) against \( H_1 : \rho < 1 \) in (2). Comparing the results with the ones for the \( \Phi_3 \) test reveals that the the \( \tau \)-test is more powerful when \( \rho \leq 0.9 \), but less powerful for \( \rho > 0.9 \). The difference in favour of the constrained test increases with \( \beta \), and is considerable for \( \beta = 0.1 \).

Still one difference between the tests is that the \( \tau \)-test in the explosive case has lower power for small sample sizes. Asymptotically, none of the tests should have any power in the explosive case, which can be seen in both cases already when \( T = 100 \). But when \( T = 25 \), the \( \tau \)-test has much lower power.

Before continuing, a note on starting values is in place. As noted in Elder and Kennedy (2001), the power of the \( \Phi_1 \)-test is unaffected by \( \alpha \), as long as the starting value is the same as the unconditional mean of the process, \( \alpha/(1 - \rho) \).

In our simulations, we set \( x_0 = 0 \) irrespective of \( \alpha \). This means that the power variation over different values of \( \alpha \) depends on the difference between \( x_0 \) and \( \alpha/(1 - \rho) \), rather than on the value of \( \alpha \) per se. Furthermore, when reporting the power for \( \alpha = 0 \) with starting value \( x_0 = 0 \), we also account for all values of \( \alpha \) when \( x_0 = \alpha/(1 - \rho) \).

Moving to the state space model, we first need to explore our conjecture about monotonicity of the log likelihood. Since we cannot prove the conjecture analytically, we perform a simple numerical investigation. For \( T = 25, 50, \) or 100, \( \rho = 1 \), and \( \alpha = \beta = 0 \), we create 20,000 series according to (6). For each series, we numerically find \( \hat{\delta} \), and in all cases where \( \delta > 0 \) (approximately 0.4 percent of the series in our simulation), we evaluate the variance function at 100 points on the interval \( \delta \in [-0.5, 0] \). We find that the log likelihood function is monotonically increasing in \( \delta \).

Although this is not a proof of our conjecture, we find the results above con-
Table 5: Estimated power function of the test at 5% significance level, constrained case with trend. The DGP is $x_t = \alpha + \beta t + \rho x_{t-1} + \varepsilon_t$, $H_0 : \{\alpha = 0, \beta = 0, \rho = 1\}$, and $H_1 : \{\alpha \neq 0, \beta \neq 0, \rho < 1\}$. Calculated power of two-sided test involving the $\Phi_2$-statistic from Dickey and Fuller (1981) within parentheses. The power function was simulated using 100,000 replications.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$ = 0</th>
<th>$\beta$ = 0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T$</td>
<td>25</td>
</tr>
<tr>
<td>0.6</td>
<td>0.09</td>
<td>0.46</td>
</tr>
<tr>
<td>(0.09)</td>
<td>(0.45)</td>
<td>(0.99)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>(0.05)</td>
<td>(0.22)</td>
<td>(0.86)</td>
</tr>
<tr>
<td>0.8</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.09)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.9</td>
<td>0.03</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>0.95</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>0.99</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>1.02</td>
<td>0.08</td>
<td>0.16</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.16)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>1.05</td>
<td>0.20</td>
<td>0.33</td>
</tr>
<tr>
<td>(0.20)</td>
<td>(0.58)</td>
<td>(0.95)</td>
</tr>
</tbody>
</table>

$\alpha$ and $\beta$ not reported.
with deterministic trend under both hypotheses. The DGP is

\[ \Phi = 0 \{ \beta = 0 \}, \quad \beta = 0.1 \]

Table 6: Estimated power function of the test at 5% significance level, constrained case involving the \( \Phi \) function was simulated using 100,000 replications.

\[
\begin{array}{cccc|cccc|cccc}
\alpha & 0 & 0.5 & 1 & 0 & 0.5 & 1 & 0 & 0.5 & 1 \\
T & 25 & 50 & 100 & 25 & 50 & 100 & 25 & 50 & 100 \\
0.6 & 0.15 & 0.61 & 1.00 & 0.17 & 0.64 & 1.00 & 0.22 & 0.71 & 1.00 \\
0.7 & 0.10 & 0.34 & 0.95 & 0.11 & 0.37 & 0.94 & 0.16 & 0.48 & 0.97 \\
0.8 & 0.06 & 0.15 & 0.55 & 0.08 & 0.18 & 0.61 & 0.12 & 0.29 & 0.73 \\
\rho & 0.9 & 0.05 & 0.06 & 0.14 & 0.06 & 0.09 & 0.20 & 0.09 & 0.20 & 0.44 \\
0.95 & 0.05 & 0.05 & 0.07 & 0.05 & 0.07 & 0.11 & 0.07 & 0.16 & 0.40 \\
0.99 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.06 & 0.05 & 0.06 & 0.14 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1.02 & 0.05 & 0.05 & 0.06 & 0.05 & 0.09 & 0.08 & 0.06 & 0.34 & 0.00 \\
1.05 & 0.05 & 0.08 & 0.00 & 0.08 & 0.04 & 0.00 & 0.27 & 0.00 & 0.00 \\
\beta = 0.05 \\
0.6 & 0.16 & 0.61 & 1.00 & 0.17 & 0.64 & 1.00 & 0.22 & 0.71 & 1.00 \\
0.7 & 0.10 & 0.34 & 0.94 & 0.10 & 0.36 & 0.94 & 0.15 & 0.46 & 0.96 \\
0.8 & 0.07 & 0.15 & 0.55 & 0.07 & 0.17 & 0.58 & 0.10 & 0.24 & 0.70 \\
\rho & 0.9 & 0.06 & 0.08 & 0.18 & 0.05 & 0.07 & 0.15 & 0.06 & 0.10 & 0.23 \\
0.95 & 0.07 & 0.13 & 0.32 & 0.05 & 0.06 & 0.09 & 0.05 & 0.05 & 0.07 \\
0.99 & 0.10 & 0.65 & 0.94 & 0.04 & 0.55 & 0.93 & 0.08 & 0.43 & 0.91 \\
1 & 0.12 & 0.72 & 0.61 & 0.11 & 0.72 & 0.61 & 0.12 & 0.72 & 0.60 \\
1.02 & 0.17 & 0.24 & 0.00 & 0.25 & 0.16 & 0.00 & 0.34 & 0.10 & 0.00 \\
1.05 & 0.32 & 0.00 & 0.00 & 0.39 & 0.00 & 0.00 & 0.44 & 0.10 & 0.00 \\
\beta = 0.1 \\
0.6 & 0.16 & 0.61 & 1.00 & 0.17 & 0.64 & 1.00 & 0.21 & 0.71 & 1.00 \\
0.7 & 0.10 & 0.34 & 0.93 & 0.10 & 0.36 & 0.94 & 0.14 & 0.44 & 0.96 \\
0.8 & 0.07 & 0.16 & 0.57 & 0.07 & 0.16 & 0.56 & 0.07 & 0.15 & 0.66 \\
\rho & 0.9 & 0.08 & 0.14 & 0.31 & 0.05 & 0.08 & 0.16 & 0.05 & 0.07 & 0.15 \\
0.95 & 0.15 & 0.61 & 0.98 & 0.09 & 0.28 & 0.72 & 0.06 & 0.11 & 0.24 \\
0.99 & 0.37 & 0.84 & 0.99 & 0.33 & 0.83 & 0.99 & 0.29 & 0.84 & 0.99 \\
1 & 0.45 & 0.64 & 0.55 & 0.45 & 0.64 & 0.55 & 0.45 & 0.64 & 0.55 \\
1.02 & 0.56 & 0.02 & 0.00 & 0.58 & 0.01 & 0.00 & 0.58 & 0.01 & 0.00 \\
1.05 & 0.25 & 0.00 & 0.00 & 0.15 & 0.00 & 0.00 & 0.09 & 0.00 & 0.00 \\
\end{array}
\]

Table 6: Estimated power function of the test at 5% significance level, constrained case with deterministic trend under both hypotheses. The DGP is \( x_t = \alpha + \beta t + \rho x_{t-1} + \epsilon_t \). 

*H0 : \( \beta = 0, \rho = 1 \), and *H1 : \( \beta \neq 0, \rho < 1 \). Calculated power of two-sided test involving the \( \Phi \) statistic from Dickey and Fuller (1981) within parentheses. The power function was simulated using 100,000 replications.
Table 7: Estimated power function of the test at 5% significance level, ordinary one-sided DF test with trend. The data-generating process is \( x_t = \alpha + \beta t + \rho x_{t-1} + \epsilon_t \), \( H_0 : \{ \rho = 1 \} \), and \( H_1 : \{ \rho < 1 \} \). The power function was simulated using 100,000 replications.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( \rho )</th>
<th>( \beta = 0 )</th>
<th>( \beta = 0.05 )</th>
<th>( \beta = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>0.21 0.69 1.00</td>
<td>0.22 0.71 1.00</td>
<td>0.25 0.76 1.00</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
<td>0.13 0.41 0.96</td>
<td>0.14 0.44 0.97</td>
<td>0.16 0.51 0.98</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.08 0.19 0.64</td>
<td>0.09 0.21 0.68</td>
<td>0.10 0.28 0.78</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
<td>0.06 0.08 0.19</td>
<td>0.06 0.09 0.22</td>
<td>0.06 0.11 0.35</td>
</tr>
<tr>
<td>0.95</td>
<td>0.95</td>
<td>0.05 0.06 0.08</td>
<td>0.05 0.06 0.09</td>
<td>0.05 0.05 0.14</td>
</tr>
<tr>
<td>0.99</td>
<td>0.99</td>
<td>0.05 0.05 0.05</td>
<td>0.05 0.05 0.05</td>
<td>0.05 0.05 0.04</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.05 0.05 0.05</td>
<td>0.05 0.05 0.05</td>
<td>0.05 0.05 0.05</td>
</tr>
<tr>
<td>1.02</td>
<td>1.02</td>
<td>0.05 0.05 0.05</td>
<td>0.05 0.02 0.00</td>
<td>0.04 0.00 0.00</td>
</tr>
<tr>
<td>1.05</td>
<td>1.05</td>
<td>0.05 0.05 0.05</td>
<td>0.02 0.00 0.00</td>
<td>0.00 0.00 0.00</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>0.21 0.69 1.00</td>
<td>0.22 0.71 1.00</td>
<td>0.25 0.76 1.00</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
<td>0.13 0.41 0.96</td>
<td>0.14 0.43 0.96</td>
<td>0.16 0.50 0.98</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.08 0.20 0.64</td>
<td>0.09 0.20 0.66</td>
<td>0.10 0.25 0.75</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
<td>0.06 0.09 0.21</td>
<td>0.06 0.08 0.19</td>
<td>0.06 0.09 0.24</td>
</tr>
<tr>
<td>0.95</td>
<td>0.95</td>
<td>0.05 0.05 0.13</td>
<td>0.05 0.06 0.09</td>
<td>0.05 0.06 0.08</td>
</tr>
<tr>
<td>0.99</td>
<td>0.99</td>
<td>0.03 0.01 0.03</td>
<td>0.04 0.01 0.02</td>
<td>0.04 0.02 0.02</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.03 0.00 0.00</td>
<td>0.03 0.00 0.00</td>
<td>0.03 0.00 0.00</td>
</tr>
<tr>
<td>1.02</td>
<td>1.02</td>
<td>0.02 0.00 0.00</td>
<td>0.01 0.00 0.00</td>
<td>0.01 0.00 0.00</td>
</tr>
<tr>
<td>1.05</td>
<td>1.05</td>
<td>0.00 0.00 0.00</td>
<td>0.00 0.00 0.00</td>
<td>0.00 0.00 0.00</td>
</tr>
</tbody>
</table>
Table 8: Simulated percentiles of the distribution of $-2 \log Q$, based on numerical maximization of the likelihood; 100,000 replications.

Table 9: Estimated power function of the test at 5% significance level, constrained case. The DGP is according to (6), and within parentheses are values from Abadir and Distaso (2003, Table 3). The power function was simulated using 100,000 replications.

The data-generating process used in the leftmost block of the results – when $\beta = 0$ – is equivalent to the DGP used in the leftmost upper block of Table 5, and we can see that the estimated power of the test in this section is approximately the same as the previous one with trend. For $\beta > 0$, we see that the power function is not everywhere decreasing in $\rho$: for instance, when $\beta = 0.25$, the power decreases.
when $\rho$ falls from 0.99 to 0.95, and then increases for lower values of $\rho$. Remember that the power behaves similarly in Table 5.

Qualitatively, the estimated power of the LR-test behaves similarly to that of the test in Abadir and Distaso (2003), shown in parenthesis. But the latter is more powerful for each parameter combination.

7 Concluding remarks

We have presented a way to derive a LR unit-root test against a stationary alternative using the standard test with a double-sided alternative, first presented in Dickey and Fuller (1981), as a point of departure. Our results can be developed in at least two directions: further elaboration within the context of constrained unit-root testing, and likelihood-ratio testing in other situations where the alternative has to be constrained. The latter includes a wide range of possible econometric specifications, since we give examples of models that are either linear or non-linear in the parameters.

When it comes to further development of our proposed unit-root test, one possible way to go is to explore its properties when lagged values of $\Delta x_t$ are included, in the spirit of Dickey and Said (1984). Another potential path is to include structural breaks in the model, the way Perron (1989) does; it is likely that the latter offers a greater challenge.

Acknowledgements

We are grateful for comments from all seminar participants at the universities of Bergen and Uppsala, in particular Johan Lyhagen.
References


A Omitted proofs

A.1 Proof of theorem 1

It is standard to derive the likelihood ratio as

\[ Q_{1T} = \left( \frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_0^2} \right)^{T/2} = \left( 1 - \frac{1}{T} \frac{T (\tilde{\sigma}_0^2 - \tilde{\sigma}_1^2)}{\tilde{\sigma}_0^2} \right)^{T/2}, \]  

where \( \tilde{\sigma}_0^2 \) and \( \tilde{\sigma}_1^2 \) are the ML estimators of \( \sigma^2 \) under \( H_0 \) and \( H_1 \), respectively. Now, re-write the model as

\[ \Delta x_t = \alpha + \delta x_{t-1} + \varepsilon_t, \]

where \( \delta = \rho - 1 \), or in vector form,

\[ Y = X_1 A_1 + \varepsilon, \]

\[ A_1 \equiv \begin{pmatrix} \alpha & \delta \end{pmatrix}, \]

where \( Y \) and \( X_1 \) are as in the theorem. It follows that

\[ \tilde{\sigma}_1^2 = T^{-1} Y' \left( I_T - X_1 (X_1' X_1)^{-1} X_1' \right) Y, \]

\[ \hat{A}_1 = (X_1' X_1)^{-1} X_1' Y. \]

and the expression for \( \hat{\rho} \) readily follows. Here,

\[ X_1' X_1 = \begin{pmatrix} T & T \gamma \sum_{t=1}^T x_{t-1}^2 \end{pmatrix}, \]
where
\[ \mathbf{\tau}_- \equiv T^{-1} \sum_{t=1}^{T} x_{t-1}, \]
and
\[ X'_1 Y = \left( x_T - x_0 \sum_{t=1}^{T} x_{t-1} \Delta x_t \right)' \]
Moreover,
\[ \hat{\sigma}_0^2 = T^{-1} Y' Y = T^{-1} \sum_{t=1}^{T} \left( \Delta x_t \right)^2, \]
and so,
\[ T \left( \hat{\sigma}_0^2 - \hat{\sigma}_1^2 \right) = Y' X_1 \left( X'_1 X_1 \right)^{-1} X'_1 Y = \frac{D_1}{C_1}, \]
where
\[ C_1 \equiv \sum_{t=1}^{T} x_{t-1}^2 - T \mathbf{\tau}_-^2, \]
\[ D_1 \equiv T^{-1} (x_T - x_0)^2 \sum_{t=1}^{T} x_{t-1}^2 - 2 (x_T - x_0) \mathbf{\tau}_- \sum_{t=1}^{T} x_{t-1} \Delta x_t + \left( \sum_{t=1}^{T} x_{t-1} \Delta x_t \right)^2. \]

Now, from standard limit theory, as \( T \to \infty \), we have
\[ T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta x_t \xrightarrow{d} \int_0^1 W(t) dW(t) = \frac{W(1)^2 - 1}{2}, \]  
\[ T^{-2} \sum_{t=1}^{T} x_{t-1}^2 \xrightarrow{d} \int_0^1 W(t)^2 dt, \]
\[ T^{-1/2} \mathbf{\tau}_- \xrightarrow{d} \int_0^1 W(t) dt, \]
\[ T^{-1} (x_T - x_0)^2 \xrightarrow{d} W(1)^2, \]
where \( W(t) \) is a standard Wiener process, assuming without loss of generality that \( \sigma^2 = 1 \). Moreover, it is similarly seen that \( \hat{\sigma}_0^2 \) converges to 1 as \( T \to \infty \). Hence, \( T \left( \hat{\sigma}_0^2 - \hat{\sigma}_1^2 \right) / \hat{\sigma}_0^2 \) is \( O_P(1) \), and a Taylor expansion in (7) shows that \(-2 \log Q_{1T}\) has the same asymptotic distribution as \( T \left( \hat{\sigma}_0^2 - \hat{\sigma}_1^2 \right) / \hat{\sigma}_0^2 \), i.e.
\[ -2 \log Q_{1T} \xrightarrow{d} Z_1, \]
as \( T \to \infty \), where \( Z_1 \) is as given in the theorem.

A.2 Proof of theorem 2

It is seen as above that the model with trend may be written as
\[ Y = X_2 A_2 + \varepsilon, \]
\[ A_2 \equiv \left( \alpha \ \beta \ \delta \right)' \]
which yields
\[ \hat{\sigma}_1^2 = T^{-1} Y' \left( I_T - X_2 (X_2' X_2)^{-1} X_2' \right) Y, \] (12)
\[ \hat{A}_2 = (X_2' X_2)^{-1} X_2' Y, \]
from which \( \hat{\rho} \) follows. Here, putting \( s_n \equiv \sum_{t=1}^T t^n \) for \( n = 1, 2, \ldots \) and \( \bar{s}_n \equiv \frac{1}{T} \sum_{t=1}^T t x_{t-1} \),
\[ X_2' X_2 = \begin{pmatrix} T & s_1 & T \bar{s}_- \\ s_1 & s_2 & T \bar{s}_- \\ T \bar{s}_- & T \bar{s}_- & \sum_{t=1}^{T-1} x_{t-1}^2 \end{pmatrix}, \]
and
\[ X_2' Y = \left( x_T - x_0 \ T (x_T - \bar{x}_-) \ \sum_{t=1}^T x_{t-1} \Delta x_t \right)'. \]
Moreover, \( \hat{\sigma}_0^2 \) is as before and
\[ T \left( \hat{\sigma}_0^2 - \hat{\sigma}_1^2 \right) = Y' X_2 (X_2' X_2)^{-1} X_2' Y, \]
which yields the expression for \( Q_{2T} \) as in the previous proof. Moreover, we have
\[ Y' X_2 (X_2' X_2)^{-1} X_2' Y = \frac{D_2}{C_2}, \]
where
\[ C_2 \equiv 12 \bar{s}_- \bar{s}_- + 12 T \bar{s}_- \bar{s}_- - 2 (\bar{s}_-)^2 - 12 (\bar{s}_-)^2 \]
\[ - 6 T (\bar{s}_-)^2 - 4 T^2 (\bar{s}_-)^2 - T^{-1} \sum_{t=1}^T x_{t-1}^2 + T \sum_{t=1}^T x_{t-1}^2 \]
\[ = -2 (T+1) (2T+1) (\bar{s}_-)^2 + 12 (T+1) \bar{s}_- \bar{s}_- - 12 (\bar{s}_-)^2 + \frac{T^2 - 1}{T} \sum_{t=1}^T x_{t-1}^2 \]
and
\[ D_2 = -12 T^{-1} D_{21} - 4 T^{-1} D_{22} \sum_{t=1}^T \Delta x_t x_{t-1} - 2 T^{-2} D_{23} \sum_{t=1}^T x_{t-1}^2 + \frac{T^2 - 1}{T} \left( \sum_{t=1}^T x_{t-1} \Delta x_t \right)^2, \]
with
\[ D_{21} \equiv T^2 \bar{s}_- - 2 T \bar{s}_- x_0 - 2 T^2 \bar{s}_- x_T + 2 T \bar{s}_- \bar{s}_- x_{t-1} \bar{s}_- x_T \]
\[ + \bar{s}_- (x_T - x_0)^2 - 2 T \bar{s}_- \bar{s}_- x_T (x_T - x_0) + T^2 \bar{s}_- x_T^2. \]
\[ D_{22} \equiv -6T \pi_+ \pi_- + 3T (T + 1) \pi_+^2 - (T + 1) (2T + 1) \pi_- x_0 + 3 (T + 1) \pi_- x_0 - (T^2 - 1) \pi_- x_T + 3 (T - 1) \pi_- x_T. \]

\[ D_{23} \equiv -6T^2 \pi_- + 6T (T + 1) \pi_- x_0 + 6T (T - 1) \pi_- x_T - 2 (T^2 - 1) x_0 x_T - (T + 1) (2T + 1) x_0^2 - (2T - 1) (T - 1) x_T^2. \]

Now, standard limit theory yields

\[ T^{-3/2} \pi_- \xrightarrow{d} \int_0^1 tW(t) \, dt, \]

which together with (8)-(11) implies

\[ T^{-3} C \xrightarrow{d} -4 \left( \int_0^1 W(t) \, dt \right)^2 + 12 \int_0^1 W(t) \, dt \int_0^1 tW(t) \, dt - 12 \left( \int_0^1 tW(t) \, dt \right)^2 + \int_0^1 W(t)^2 \, dt \equiv C_\infty \]

and

\[ T^{-3} D \xrightarrow{d} \]

\[ -12 \left\{ W(1) \int_0^1 tW(t) \, dt - W(1) \int_0^1 W(t) \, dt + \left( \int_0^1 W(t) \, dt \right)^2 \right\} \]

\[ -4 \left\{ -6 \int_0^1 W(t) \, dt \int_0^1 tW(t) \, dt + 3 \left( \int_0^1 W(t) \, dt \right)^2 \right\} \]

\[ -W(1) \int_0^1 W(t) \, dt + 3W(1) \int_0^1 tW(t) \, dt \right\} \int_0^1 W(t) \, dW(t) \]

\[ -4 \left\{ -3 \left( \int_0^1 W(t) \, dt \right)^2 + 3W(1) \int_0^1 W(t) \, dt - W(1)^2 \right\} \int_0^1 W(t)^2 \, dt \]

\[ + \left( \int_0^1 W(t) \, dW(t) \right)^2 \]

\[ \equiv D_\infty. \]

Hence, as in the proof of theorem 1,

\[ -2 \log Q_{2T} \xrightarrow{d} \frac{D_\infty}{C_\infty}. \]

Moreover, it is clear that \( \hat{\sigma}^2 \), the ML estimator of \( \sigma^2 \) under \( H_1' \), is given by (12). Under \( H_0' \), we have the model

\[ Y = X_0 \alpha + \varepsilon, \]
and so, in the usual manner the ML estimator of $\sigma^2$ is

$$\hat{\sigma}^2_0 = T^{-1} Y' \left( I_T - X_0 (X_0'X_0)^{-1} X_0' \right) Y = T^{-1} Y' \left( I_T - T^{-1} X_0 X_0' \right) Y,$$

implying

$$T \left( \hat{\sigma}^2_0 - \hat{\sigma}^2_1 \right) = Y' \left( X_2 (X_2'X_2)^{-1} X_2' - T^{-1} X_0 X_0' \right) Y.$$

Hence, the expression for $Q_{ST}$ follows as above. Moreover, simplifications yield

$$\hat{\sigma}^2_0 = T^{-1} \varepsilon' \varepsilon - T^{-2} (\varepsilon' X_0)^2,$$

which converges to $\sigma^2$ in probability as $T \to \infty$. Furthermore, we have

$$T \left( \hat{\sigma}^2_0 - \hat{\sigma}^2_1 \right) = \frac{D_2}{C_2} - T^{-1} (Y'X_0)^2,$$

where $C_2$ and $D_2$ are as in the previous proof and

$$T^{-1} (Y'X_0)^2 = T^{-1} (x_T - x_0)^2,$$

and the result follows via (11) and arguments as above.