Sequential Unit Root Test

Nagai, K. [1], K. Hitomi [2] and Y. Nishiyama [3]

[1] Department of Economics, Yokohama National University, Japan
[2] Department of Engineering, Kyoto Institute of Technology, Japan
[3] Institute of Economic Research, Kyoto University, Japan

Email: nishiyama@kier.kyoto-u.ac.jp

Keywords: *Unit root test, Sequential Analysis, Bessel processes*

EXTENDED ABSTRACT

Consider a scalar AR(1) process $x_t = \beta x_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ are iid disturbances. When $\beta = \pm 1$, the series is said to possess a unit root. Tests for the existence of unit roots in economic time series have been one of the main issues of interest in econometrics since the middle of 1980's. In practice, econometricians focus on testing the null hypothesis of $\beta = 1$ against the alternative of $|\beta| < 1$ since the explosive case $|\beta| > 1$ and negative unit root $\beta = -1$ are very unlikely in economic series.

The most standard unit root testing procedures must be Dicky-Fuller tests (DF tests hereafter) and its variants. We estimate β by the OLS and construct test statistics by suitably normalizing it. It is well known, however, that the DF test statistics do not have a good power property for small and medium sample sizes. Also, the limiting distribution is non-standard because the OLS estimator of β is not normally distributed, which is inconvenient in practice.

Lai and Siegmund (1983) (LS hereafter) and Shiryaev and Spokoiny (1997) respectively show that the OLS estimator of AR(1) coefficient, $\hat{\beta}$, is asymptotically normally distributed even if the true value of β equals to unity and it is greater than unity in fact under a sequential sampling scheme. This sequential procedure involves a stopping rule such that one stops sampling at time T_c if the generalized information $I_{T_c} \equiv \sigma^{-2} \sum_{t=1}^{T_c} x_{t-1}^2$, or its estimate in practice, exceeds a predetermined constant c.

The asymptotic normality of sequential OLS estimator implies that one can construct a test statistic for the unit root possessing a standard normal limit. It is obviously practically convenient compared with DF tests having non-normal limiting distribution.

In this sampling scheme, the number of observations T_c is also a random variable depending on the realization of the time series unlike the standard sampling case. We may like to know the statistical properties of the stopping time because it will be of some help to determine an appropriate value of c, and also because it may provide certain amount of additional information in unit root testing to that from $\hat{\beta}$. In this paper, we derive the joint asymptotic distribution of $\hat{\beta}$ and the stopping time suitably normalized. The marginal distribution of $\hat{\beta}$ is, of course, normal, while $T_c / \sqrt{c}$ is shown to have a non-standard distribution characterized by a functional of the Bessel process with the dimension 3/2 under the null. We also obtain the joint distribution under local alternatives where the marginal distribution of the stopping time is represented in terms of a Bessel process with drift. Using these asymptotic joint distributions, we can construct a likelihood ratio type test statistic for unit root. We find that the statistic does not depend on the stopping time, and thus the stopping time carries no additional information to the OLS estimate of the AR(1) coefficient in terms of testing for unit root.

The following section reviews DF test and sequential AR(1) parameter estimation, as well as the sequential unit root test based on LS including some simulation results. Section 2 presents the joint distribution of $\hat{\beta}$ and T_c under the null, while Section 3 provides that under a local alternative. Section 4 explains likelihood ratio type statistic and the sequential unit root test as well as the LAN property under the normal disturbances. Section 5 concludes.

1. INTRODUCTION

1.1. Dicky-Fuller Test

Suppose $\{x_i\}$ is generated from

$$x_t = \beta x_{t-1} + \varepsilon_t, \qquad (1)$$

where $\{\varepsilon_t\}$ are iid disturbances. When $\beta = 1$, the process is called a unit root process and its behaviour is very different from ones with $|\beta| < 1$. Some macroeconomic time series are said to have a unit root based on the results from DF test. We first briefly review the DF t-statistic. Given a sample $\{x_1, \cdots, x_T\}$, let the OLS estimator of β and an estimator of σ^2 be

$$\hat{\beta}_T = (\sum_{i=1}^{T} x_{i-1}^2)^{-1} \sum_{i=1}^{T} x_{i-1} x_i ,$$

$$\hat{\sigma}_T^2 = \frac{1}{T} \sum_{i=1}^{T} (x_i - \hat{\beta}_T x_{i-1})^2 .$$

Also denote $W(s)$ as a standard Brownian motion. Then, as $T \to \infty$, we have the following asymptotic results:

$$\frac{(\hat{\beta}_T - \beta)}{\hat{\sigma}_T \sqrt{\sum_{i=1}^{T} x_{i-1}^2}} \to^d \begin{cases} N(0,1) & if\ |\beta| < 1 \\ \dfrac{\int_0^1 W(s) dW(s)}{\sqrt{\int_0^1 W(s)^2 ds}} & if\ |\beta| = 1 \\ Cauchy\ \ otherwise. \end{cases}$$

Therefore to test the null of unit root against the alternative of stationarity, we use the t value $(\hat{\beta}_T - 1)/\{\hat{\sigma}_T \sqrt{\sum_{i=1}^{T} x_{i-1}^2}\}$ which converges to a functional of *W(s)* above under the null, while it explodes under the alternative. Table 1 shows the size and power of the test by simulation. The size seems to be acceptable, but the power is unsatisfactory for sample sizes for *T*=50~150.

Table 1. Rejection rate of DF t test (nominal size=5%).

	Beta=1.0 (size)	Beta=0.95 (power)
T=50	0.0453	0.1296
T=100	0.0460	0.3103
T=150	0.0473	0.5580

1.2. Sequential Estimation of AR(1) Parameter And the Stopping Rule

Sequential analysis was originally considered by Wald (1947). The idea is as follows. Suppose we can obtain one observation a day, say. We sample every day and when we accumulate "sufficient"' information, then we stop sampling and make a statistical decision (estimation or testing). How "sufficient"' is determined by the researchers through some user-determined parameter *c*, which controls the accuracy of the results. How we stop sampling is called the *stopping rule* and the time when we stop sampling is called the *stopping time*. Typically, we are better off if we can obtain conclusions earlier due to some cost of sampling or taking time. There exists a trade-off between accuracy and cost of sampling.

LS investigate the statistical properties of the sequential estimator of the AR(1) parameter in model (1). Formally, for a predetermined constant *c*, their stopping rule is defined as

$$T_c = \inf\{t > 0 \mid \sum_{i=1}^{t} x_{i-1}^2 / \hat{\sigma}_t^2 \geq c\}$$

where

$$\hat{\beta}_t = (\sum_{i=1}^{t} x_{i-1}^2)^{-1} \sum_{i=1}^{t} x_{i-1} x_i ,$$

$$\hat{\sigma}_t^2 = \frac{1}{t} \sum_{i=1}^{t} (x_i - \hat{\beta}_t x_{i-1})^2 .$$

We stop sampling when the estimated information $I_t = \sum_{i=1}^{t} x_{i-1}^2 / \hat{\sigma}_t^2$ exceeds a certain predetermined value *c*, which controls the accuracy of estimation through the "sample size" T_c. We write the stopping time as T_c to emphasize that it depends on the choice of c. c controls for the accuracy of the estimation in the sense that the variance of the estimator, $I_{T_c}^{-1}$, is guaranteed to be smaller than c^{-1}. There exists a trade-off between the accuracy of estimation and the cost of observations. If we set *c* large, T_c will tend to be also large by construction, which will yield a more accurate estimate. If we set *c* small, sampling will stop relatively earlier, but the accuracy will be lower. Note that T_c itself is a statistic depending on the observations.

Using this stopping time, we calculate sequential estimators by

$$\hat{\beta}_{T_c} = (\sum_{i=1}^{T_c} x_{i-1}^2)^{-1} \sum_{i=1}^{T_c} x_{i-1} x_i ,$$

$$\hat{\sigma}_{T_c}^2 = \frac{1}{T_c} \sum_{i=1}^{T_c} (x_i - \hat{\beta}_{T_c} x_{i-1})^2 .$$

LS prove the asymptotic normality of $\hat{\beta}_{T_c}$ in the case of $|\beta| \leq 1$:

$$\sqrt{I_{T_c}} (\hat{\beta}_{T_c} - \beta) \to^d N(0,1) .$$

Further, Shiryaev and Spokoiny (1997) obtain the same result in the explosive case of $|\beta| > 1$ under the assumption of normal disturbances. We can directly apply this result for unit root test which we call a sequential unit root test (SURT).

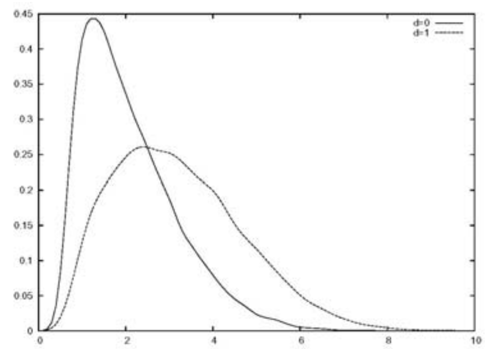
Table 2. Properties of SURT (size=5%)

***c*=600**	Beta=1.0 (size)	Beta=0.95 (power)
Rejection rate	0.0503	0.3376
E(Tc)	49.644	81.716
E(Beta)	0.999	0.950
Std(Tc)	25.302	35.894
Std(Beta)	0.0415	0.0419
***c*=2500**	Beta=1.0 (size)	Beta=0.95 (power)
Rejection rate	0.0516	0.7743
E(Tc)	100.181	251.129
E(Beta)	0.9996	0.9502
Std(Tc)	51.708	82.176
Std(Beta)	0.0210	0.0209

Table 2 shows Monte Carlo results of SURT for *c*=600 and 2500. *c* controls for the accuracy of inference through the stopping time. We set *c*=600 so that the average stopping time under the null is about 50, while *c*=2500 yields the average stopping time of 100 under the null. We will show a theoretical relationship between c and the average stopping time (sample size) later. Comparing the size results in Table 2 with those in Table 1, they are mostly satisfactory. In comparing the power, we need to be careful. The SURT procedure requires more sample sizes to stop sampling under the alternative, thus we cannot directly compare them with figures under the null. One point we can make is that DF test under the standard (fixed) sampling, we cannot conclude a unit root exists even if the null is not rejected from a sample of small or medium size (though it seems to prevail in economic literature). Under sequential sampling, however, researchers will be automatically forced to wait until a “sufficient” amount of information is accumulated both under the null and alternative hypotheses. The choice of c obviously becomes an important factor there. To choose c suitably, we need to study the statistical properties of T_c which also determines the power of the SURT procedure.

We lastly show the distribution of the stopping time from a Monte Carlo simulation. Figure 1 shows the density functions of $T_c / \sqrt{c}$ for $\beta = 1 - d/c$ with $d = 0, 1$ when $c = 10^8$. $d = 0$ and 1 respectively correspond to the cases of unit root and stationarity. The distributions are clearly quite different each other and the stationary case requires more sample size (or information) in making a statistical decision.

Figure 1. Densities of stopping time under the null (d=0) and the alternative (d=1)



2. JOINT DISTRIBUTION OF AR(1) PARAMETER ESTIMATOR AND THE STOPPING TIME UNDER THE NULL

The following theorem presents the joint distribution of $\hat{\beta}_{T_c}$ and T_c under the null.

THEOREM 1.

Suppose $\{\varepsilon_i\}$, $i = 1,2,\cdots$ is a stationary and ergodic martingale difference sequence with

$E(\varepsilon_1^2) = \sigma^2 < \infty$, and x_i are generated by (1) with $\beta = 1$ and $x_0 = 0$. Then, we have, as $c \to \infty$,

$$\left(\sqrt{c}(\hat{\beta}_{T_c} - 1), \frac{T_c}{\sqrt{c}} \right) \to^d \left(B_1, \int_0^1 \frac{1}{2X_s} ds \right)$$

where B_t is a standard Brownian motion on $[0, \infty)$ and X_t is a Bessel process satisfying

$$dX_t = \frac{\delta - 1}{2X_t} dt + dB_t$$

with the dimension $\delta = 3/2$ and the initial value $X_0 = 0$.

We note that $\hat{\beta}_{T_c}$ is asymptotically normally distributed, which is consistent with the result of Lai and Siegmund (1983), and that T_c is $O_p(\sqrt{c})$. Due to the results in Borodin and Salminen (2002, p.386), we know the analytical expression of the joint density:

$$P\left(B_1 \in dz, \int_0^1 (2X_s)^{-1} ds \in du \right)$$

$$= \frac{\sqrt{2(u+2z)}}{\pi} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-u-2z)^k \Gamma(1/2+k+l)}{k!\,l!\,\Gamma(1/2+k)}$$

$$\times exp\left[-\left\{ (1+2k+2l)u + z \right\}^2 / 4 \right] D_{3/2+k}\left((1+2k+2l)u + z \right)$$

where $D.(.)$ is the parabolic cylinder function defined by

$$D_p(z) = 2^{p/2} exp(-z^2/2)$$

$$\times \left(\frac{\sqrt{\pi}\, {}_1F_1\left(-\frac{p}{2}; \frac{1}{2}; \frac{z^2}{2} \right)}{\Gamma\left(\frac{1-p}{2} \right)} - \frac{\sqrt{2\pi}\, z\, {}_1F_1\left(-\frac{1-p}{2}; \frac{3}{2}; \frac{z^2}{2} \right)}{\Gamma\left(-\frac{p}{2} \right)} \right)$$

with the hypergeometric function

$${}_pF_q(\alpha_1, \cdots, \alpha_p; \beta_1, \cdots \beta_q; z) = \sum_{n=0}^{\infty} \frac{(\alpha_1)_n \cdots (\alpha_p)_n}{(\beta_1)_n \cdots (\beta_q)_n} \frac{z^n}{n!}$$

and

$$(\alpha)_n = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1).$$

Using this joint density, we can obtain the asymptotic moments of T_c, for example,

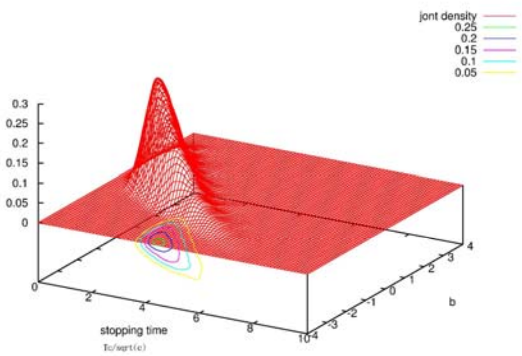
$$E\left(\frac{T_c}{\sqrt{c}} \right) \approx E\int_0^1 \frac{1}{2X_s} ds = \frac{2\sqrt{2}\Gamma(5/4)}{\Gamma(3/4)}$$

$$\approx 2.0921.$$

We set c=600 in the simulation presented in Table 2 such that the expected sample size is about 50, namely $E(T_c) \approx 2.0921\sqrt{c} = 50$, and similarly for c=2500.

Figure 2 shows the joint density function and its contour plot generated from a simulation. The two statistics are obviously highly correlated.

Figure 2. Joint density under the null



3. JOINT DISTRIBUTION UNDER LOCAL ALTERNATIVES

We consider the following local alternative;

$$H_0: \beta = 1 \text{ vs } H_\Delta: \beta_c = 1 - \frac{\Delta}{\sqrt{c}},$$

where Δ is a positive constant. We believe this is a natural local alternative setting as we consider the asymptotics of $c \to \infty$. The following theorem provides the joint density of AR(1) coefficient estimator and the stopping time under the above local alternatives.

THEOREM 2

Suppose the same conditions on ε's stated in Theorem 1 hold. Under the local alternative, we have,

(i) as $c \to \infty$,

$$\left(\sqrt{c}(\hat{\beta}_{T_c}-1), \frac{T_c}{\sqrt{c}}\right) \to^d \left(-\Delta + B_1, \int_0^1 \frac{1}{2X_s^{\Delta}} ds\right)$$

where X_t^{Δ} is the Bessel process with a drift solving the stochastic differential equation

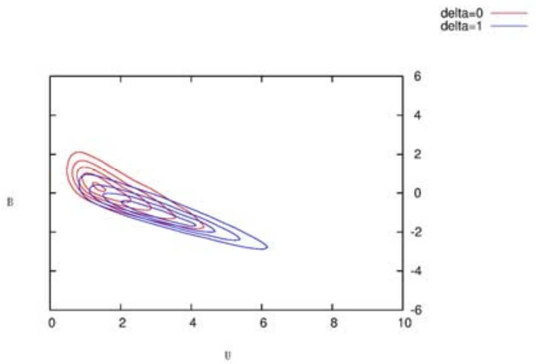
$$dX_t^{\Delta} = (-\Delta + \frac{\delta - 1}{2X_t^{\Delta}})dt + dB_t$$

with $\delta = 3/2$ and $X_0^{\Delta} = 0$.

(ii) Let $f^0(z,u)$ be the joint density under the null shown in Theorem 1, then that under the local alternatives is give by

$$f^{\Delta}(z,u) = \exp(-\Delta z - \frac{1}{2}\Delta^2) f^0(z,u).$$

Figure 3. Contours of joint densities under the null ($\Delta = 0$) and the alternative ($\Delta = 1$)



Obviously, $\Delta = 0$ reduces to Theorem 1. Figure 3 compares the contours of $f^0(z,u)$ and $f^{\Delta}(z,u)$.

4. LIKELIHOOD RATIO TYPE TEST AND LAN PROPERTY

4.1. Likelihood Ratio

Using Theorem 2 (ii), we can construct a likelihood ratio type test statistic based on the two statistics:

$$\log \frac{f^0(z,u)}{f^{\Delta}(z,u)} = \Delta z + \frac{1}{2}\Delta^2.$$

This indicates that the stopping time does not carry any additional information in testing the null of unit root to the information carried by the AR(1) coefficient estimate. This is a natural result in the case of normal disturbances, but it is also true for non-normal cases.

We note that this likelihood ratio is not exactly the likelihood ratio in the ordinary sense because it does not present the likelihood of the observations themselves, but only their functions, namely $\hat{\beta}_{T_c}$ and T_c. A test based on this likelihood ratio may be a reasonable approach especially when we do not know the distribution of the disturbances as we cannot write down the ordinary likelihood.

4.2. Local Asymptotic Normality

Suppose that the disturbances are normally independently distributed, $\varepsilon_i \sim iidN(0,\sigma^2)$. Then the log likelihood ratio of the observations is

$$\Lambda(x_1,\cdots,x_{T_c};\Delta/\sqrt{c})$$
$$= -\frac{\Delta}{\sqrt{c}\sigma^2}\sum_{i=1}^{T_c}(x_i - x_{i-1})x_{i-1} - \frac{\Delta^2}{c\sigma^2}\sum_{i=1}^{T_c}x_{i-1}^2.$$

The first term on the right is a martingale and asymptotically normally distributed, and the second term converges to a constant as $c \to \infty$ due to the definition of the stopping time. Therefore, it possesses the LAN property. The points are that we stop sampling when the summand of the second term hits c and that this quantity coincides with the quadratic variation of the first martingale term in the limit. Though we are not sure if the LAN property implies some optimality in making inferences in sequential sampling setup as in the standard sampling, it might be likely. We need further research on this.

5. CONCLUDING REMARKS

This paper considers testing for the existence of a unit root under the sequential sampling proposed by Lai and Siegmund (1983) and Shiryaev and Spokoiny (1997). We obtain joint distributions of AR(1) coefficient estimator and the stopping time both under the null and local alternatives. The null distribution of stopping time is characterized by a Bessel process with dimension 3/2, while the distribution under the local alternatives is represented in terms of the same Bessel process with a drift.

Though sequential sampling situation may not be very likely in most econometric time series except some cases where we need to make a, say, policy decision as soon as possible, the proposed sequential unit root test procedure may be a good alternative to the common DF test in terms of power. It is known that DF test does not have a sufficient power under small or medium sample size, but the SURT procedure automatically let econometricians wait until “sufficient” information is accumulated to make a statistical decision. It will be possible to apply this procedure to, for instance, the decision making of fund managers who may like to know if a series has a unit root or not, namely stable or not, as early as possible.

We provide a likelihood ratio type test statistic, which is shown to be independent of the stopping time in the first order asymptotics. It may become important in the second order. We also show that it has a LAN property when the disturbances are normally distributed.

There are some possibilities of extentions for future research. Firstly, we may need to compare the SURT with the sequential probability ratio test (SPRT) which is a standard testing procedure under sequential sampling. To the best of our knowledge, there has been considered no such test or its asymptotic theory in time series settings. Also, SPRT requires a specification in distribution, which we think may be too restrictive.

Secondly, we use the stopping time proposed in Lai and Siegmund (1983), which uses the generalized information. It is a suitable choice of stopping rule in the case of estimation since it coincides with the variance of the estimator, so that controlling this quantity means controlling the variance in fact. However, it may not be the best approach for the sake of testing since it may be more appropriate to control the accuracy of decision, or size and power. There is a possibility of pursuing different stopping rules for testing.

Thirdly, we treat the simplest case of scalar AR(1) without constant or drift terms. We may be able to relax these restrictions. We could extend the procedure to AR(p) processes, series with drift or trend, or long memory processes. Also, it may be more practically useful to consider tests for structural break or change point problems. Research toward these directions is currently under way.

6. REFERENCES

Borodin, A.N. and P. Salminen (2002), Handbook of Brownian Motion – Facts and Formulae, Secon edition, Birkhäuser.

Chang,Y. and J.Y. Park(2004), Taking a New Contour:A Novel View on Unit Root Test 1, mimeo.

Ferguson, T. S. (1967) Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York.

Lai, T.L. and D. Siegmund (1983), Fixed accuracy estimation of an autoregressive parameter,' *Annals of Statistics*, 11, 478-485.

Linetsky, V. (2004), The spectral representation of Bessel processes with constant drift: Applications in queueing and finance, *J. Appl. Probab*., 41(2), 327-344.

Revuz, D. and M. Yor (1994), Continuous Martingale and Brownian Motion, 2nd ed., Springer-Verlag, New York.

Shiryaev, A. N. and V.G. Spokoiny (1997), On Sequential Estimation of an Autoregressive Parameter, *Stochastics Stochastics Rep*., 60, 219-240.

Tanaka, K.(1996), Time series analysis, John Wiley and Sons.