On detection of unit roots generalizing the classic Dickey-Fuller approach

A. Steland

Ruhr-Universität Bochum

Fakultät für Mathematik

Building NA 3/71

D-44780 Bochum, Germany

February 18, 2005

Abstract

If we are given a time series of economic data, a basic question is whether the series is stationary or a random walk, i.e., has a unit root. Whereas the problem to test the unit root null hypothesis against the alternative of stationarity is well studied in the context of classic hypothesis testing in the sense of Neyman, sequential and monitoring approaches have not been studied in detail yet. We consider stopping rules based on a sequential version of the well known Dickey-Fuller test statistics in a setting, where the asymptotic distribution theory becomes a nice and simple application of weak convergence of Ito integrals. More sophisticated extensions studied elsewhere are outlined. Finally, we present a couple of simulations.

1 Introduction

Non-stationarity is a serious concern of many time series. A possible departure from the stationarity assumptions are trends. However, one has to distinguish between deterministic trends and stochastic trends, whereas the latter means that the process has mean zero, but is a random walk, whose trajectories often exhibit a trend-like behavior. Even if deterministic trends can be excluded, the problem to decide whether the process is stationary or a random walk is crucial from both a theoretical and practical point of view. As a practical problem it arises, e.g., in econometrics when analyzing log returns of assets, macroeconomic series as the GDP, or equilibrium errors of a known cointegration relationship. The question whether a time series is stationary or a random walk is of considerable importance for a correct interpretation of a time series, and is also crucial to choose a valid method when analyzing the series to detect or trends which usually assume stationarity as in Steland (2004), Pawlak et al. (2004), Husková (1999), Husková and Slabý (2001), among others. It is also interesting from a more theoretical point of view, since standard statistics as averages have different convergence rates and different limiting distributions in the random walk case. Various approaches to test the unit root null hypothesis against stationarity have been studied in the statistics and econometrics literature, in contrast to the monitoring perspective, where we aim at detecting stationarity as soon as possible. For a review see Stock (1994). In this article we consider some simple stopping times (control charts) to detect stationarity, being motivated by least squares, for which the asymptotic theory can be easily based on a known result on weak convergence of stochastic Ito integrals. A more sophisticated procedure (Steland, 2005b) requiring other techniques is briefly outlined. For a nonparametric approach based on the KPSS test (Kwiatkowski et al., 1982) we refer to Steland (2005a).

Let us assume that we are given a time series $\{Y_t\}$ with

$$Y_t = \rho Y_{t-1} + \epsilon_t, \qquad t \geq 1, \qquad Y_0 = 0, \tag{1}$$

where $\rho \in (-1, 1]$ is a deterministic but unknown parameter, and $\{\epsilon_t\}$ is a mean-zero sequence of i.i.d. error terms (innovations). If $|\rho| < 1$, $\{Y_t\}$ is stationary, whereas for $\rho = 1$, the differences ΔY_t form a stationary process. To estimate ρ one may use the ordinary least squares (OLS) estimator, which is given by

$$\widehat{\rho}_T = \sum_{t=1}^{T} Y_t Y_{t-1} \ / \ \sum_{t=1}^{T} Y_{t-1}^2$$

As well known, for $|\rho| < 1$, $\sqrt{T}(\widehat{\rho}_T - \rho)$ is asymptotically normal. To construct a test we need the distribution under the null hypothesis $H_0 : \rho = 1$, which has been studied by White (1958), Dickey and Fuller (1979), and many others. It is known that

$$D_T = T(\widehat{\rho}_T - 1) \xrightarrow{d} (1/2)(B(1)^2 - 1)/\int_0^1 B(r)^2\, dr,$$

as $T \to \infty$. Here B denotes standard Brownian motion. The result shows that $\widehat{\rho}_T$ has a different convergence rate in the random walk case, and a non-standard asymmetric limit distribution.

In this article we are interested in the construction of detection procedures to detect the stationarity alternative as soon as possible. In practice such a monitoring scheme is often applied until a certain time horizon T. After T observations corresponding to a certain time interval, a conclusion should be made in any case. Let us define the stopping rule

$$S_T = \inf\{1 \leq t \leq T : D_t < c\}$$

for some fixed control limit c. If $S_T < T$, we reject the unit root hypothesis after S_T observations in favor of stationarity. If $S_T = T$, we accept the unit root hypothesis as a plausible model for the time series. One reasonable approach to specify the control limit c is to ensure that the above detection procedure has a controlled type I error of size α, when interpreted as a classic hypothesis test, i.e., we reject the unit root null hypothesis if $S_T < T$. However, one could also choose the control limit to ensure that the average run length defined as

$$\mathrm{ARL}(S_T) = E_0(S_T),$$

where E_0 indicates that the expectation is calculated under H_0, is sufficiently large, i.e., $\mathrm{ARL}(S_T) \geq \xi$ for some prespecified so-called in-control average run length ξ. Having in mind that the distributions of stopping times are typically skewed, it would also be reasonable to take the median instead of the mean. For these reasons we provide asymptotic results which provide assertions about the asymptotic distribution and are not restricted to certain functionals of it.

The key to derive such limit theorems is to the following simple observation. Note that S_T can be written as $T \inf\{s \in (0,1) : DF_T(s) < c\}$, where

$$DF_T(s) = \lfloor Ts \rfloor (\widehat{\rho}_{\lfloor Ts \rfloor} - 1) = \frac{\lfloor Ts \rfloor^{-1} \sum_{t=1}^{\lfloor Ts \rfloor} Y_t \Delta Y_t}{\lfloor Ts \rfloor^{-2} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1}^2}, \qquad s \in [0,1].$$

Here and in the sequel $\lfloor z \rfloor$ is the greatest integer less or equal to $z \in \mathbf{R}$. Using that representation in terms of a inf functional of a stochastic process allows to obtain the desired asymptotic results about the distribution of S_T.

The organization of the paper is as follows. Section 2 provides a functional central limit theorem for the process underlying the detection procedure S_T. A more sophisticated procedure studied in detail in a separate article is briefly outlined in Section 3. Simulation resuls proving the applicability of the procedure and studying its statistical behavior in some respects are given in the last section.

2 A functional central limit theorem

In this section we provide an elegant proof of the asymptotic distribution of DF_T and S_T/T using a theorem about weak convergence of Ito integrals. The general and more involved case dealing with a general weighting function is studied in Steland (2005b), where a detailed treatment of the asymptotic theory under various stochastic models is given.

Noting that all processes defined in the previous section are elements of the function space $D[0,1]$, we show weak convergence with respect to the Skorokhod topology, denoted by $\Rightarrow$. Recall that for a semimartingale X and a predictable process H the Ito integral is defined as the process $t \mapsto \int_0^t H(s)\,dX(s)$, $t \in [0,1]$, and abbreviated as $\int H\,dX$. We consider the canonical filtration $\{\sigma(Y_1,\dots,Y_t)\}$ associated to the time series $\{Y_t\}$. The following theorem is taken from Kurtz and Protter (2004, Sec. 7). For an introduction to Ito integrals we also refer to Oksendahl (1992).

Theorem 1 *Suppose X_n is a semimartingale for each n, and H_n is predictable for each n. If $(H_n, X_n) \Rightarrow (H, X)$ in the space $D_{\mathbf{R}^2}[0,1]$ equipped with the Skorokhod topology, and $\sup_n Var(X_n) < \infty$, then, as $n \to \infty$,*

$$\left(H_n, X_n, \int H_n\,dX_n\right) \Rightarrow \left(H, X, \int H\,dX\right).$$

The following result is a nice application of Theorem 1.

Theorem 2 (Functional Central Limit Theorem) *Assume $\rho = 1$. We have*

$$DF_T(s) \Rightarrow (2s)^{-1}(B(s)^2 - s) / \int_0^s B(r)^2\,dr,$$

as $T \to \infty$.

Proof. *For a proof, define $Z_T(r) = T^{-1/2} Y_{\lfloor Ts \rfloor}$, $s \in [0,1]$, and note that Z_T is a L_2-martingale and therefore a semimartingale. Thus, for each T the numerator of DF_T can be represented via an Ito integral,*

$$\lfloor Ts \rfloor^{-1} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1}(Y_t - Y_{t-1}) = \lfloor Ts \rfloor^{-1} T \int_0^s Z_T(r)\,dZ_T(r).$$

More generally, since under the assumptions of the theorem, $Z_T \Rightarrow \sigma B(r)$*, as* $T \to \infty$*, we have joint weak convergence of integrand and integrator, and since* Z_T *is a local martingale with* $\sup_T \mathit{Var} Z_T = \sup_T(\lfloor Ts \rfloor / T)\, \mathit{Var}(\epsilon_1) < \infty$*, we can apply Theorem 1 to obtain*

$$\lfloor Ts \rfloor^{-1} T \int_0^s Z_T(r)\, dZ_T(r) \Rightarrow \eta^2 s^{-1} \int_0^s B\, dB = \eta^2 (2s)^{-1} (B^2(s) - s).$$

Since any linear combination of $\lfloor T\circ \rfloor T^{-1} \int_0^\circ Z_T(r)\, dZ_T(r)$ *and* $\int_0^\circ Z_T^2(r)\, dr$ *is a functional of* Z_T*, it converges weakly to the associated linear combination of* $\eta^2 \int B\, dB$ *and* $\eta^2 \int_0^\circ B(r)\, dr$*, yielding the assertion of the theorem.* ○

Due to a.s. continuity of the limit process, the following corollary can be proved using arguments given in greater detail in Steland (2005b).

Corollary 1 *If* $\rho = 1$*, we have for* $T \to \infty$*,*

$$S_T/T \to \inf\{s \in (0,1] : (2s)^{-1}(B(s)^2 - s) / \int_0^s B(r)^2\, dr < c\}$$

Remark 1 *Let us briefly discuss the benefits from these limit theorems. Firstly, they show that asymptotically the distribution of the detection procedure* S_T *is a functional of standard Brownian motion, at least if the model assumptions are satisfied. Second, and for applications more important is that the established representations of the limit distributions can be used to obtain study the shape of the asymptotic distribution and to obtain asymptotic critical values. By simulating trajectories of the Brownian motion, which is rather simple, and approximating the integrals by appropriate sums, one can simulate trajectories of the limit processes. These can be then used to simulate the distribution of* S_T *and to obtain critical values.*

It is interesting to look how the Dickey-Fuller test relates to the optimal test for simple hypotheses. If $\epsilon_t \sim N(0,\sigma^2)$, $\sigma^2 > 0$, the critical region of the Neyman-Pearson test of $H_0 : \rho = 1$ against $H_1 : \rho = \rho'$, $\rho' < \rho$, can be written as

$$(T(\rho' - 1))^2 T^{-2} \sum_{t=1}^{T} Y_{t-1}^2 - 2T(\rho' - 1)T^{-1} \sum_{t=1}^{T} Y_{t-1}\Delta Y_t < k,$$

for a constant k, see Stock (1994). This fact shows that there is no uniformly most powerful test against the composite alternative $H_1 : \rho < 1$. However, this fact motivates to consider linear combinations of the two statistics defining the optimal critical region. Let

$$DF_T^*(s) = w_1 \lfloor Ts \rfloor^{-2} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1}^2 + w_2 \lfloor Ts \rfloor^{-1} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1}\Delta Y_t, \qquad s \in [0,1],$$

for two constants w_1, w_2, and the corresponding stopping time

$$S_T^* = \inf\{s \in (0,1] : DF_T^*(s) < c\}.$$

Noting that DF_T^* is a linear combination of two processes that have already been dealt with in Theorem 2, one can verify the following result.

Theorem 3 *If $\rho = 1$, we have*

$$DF_T^*(s) \Rightarrow w_1 s^{-2} \int_0^s B(r)^2\, dr + w_2 s^{-1} \int_0^s B(r)\, dB(r)$$

and

$$S_T^* \xrightarrow{d} \inf\{s \in (0,1] : w_1 s^{-2} \int_0^s B(r)^2\, dr + w_2 s^{-1} \int_0^s B(r)\, dB(r) < c\},$$

as $T \to \infty$.

However, to apply the procedure S_T^* one has to choose the weights w_1 and w_2. In the last section we examine this issue by simulations.

3 An outline of kernel-weighted processes

Note that the procedures of the previous section cumulate observations. If the time series is a random walk at the beginning, and changes to stationarity at a certain change-point q, then the stopping time S may react slowly, since the terms corresponding to time points before q can dominate the control statistic. To circumvent that problem, one considers a kernel-weighted sequential Dickey-Fuller type process

$$D_T(s) = \frac{\lfloor Ts \rfloor^{-1} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1} \Delta Y_t K((\lfloor Ts \rfloor - t)/h)}{\lfloor Ts \rfloor^{-2} \sum_{t=1}^{\lfloor Ts \rfloor} Y_{t-1}^2}, \qquad s \in [0, 1].$$

and the associated stopping time

$$S_T = \inf\{s \in (0, 1] : D_T(s) < c\}. \tag{2}$$

$K : \mathbf{R} \to \mathbf{R}$ is a kernel function to downweight summands in the numerator with large distances, $|t - \lfloor Ts \rfloor|$, to the current time $\lfloor Ts \rfloor$. K is assumed to have the following properties: (i) $K \geq 0$, $\int K(x)\, dx = 1$, $\|K\|_\infty < \infty$, (ii) K is of class C^2 with $\|K''\|_\infty < \infty$, and (iii) K is of bounded variation. $h = h_T$ is a sequence of bandwidths $h_T \geq 0$ such that

$$\lim T/h_T = \zeta \in (1, \infty).$$

Notice that if K has support $[-1, 1]$, the detection rule S_T uses exactly h past observations.

In the previous section we assumed that the innovation process, $\{\epsilon_t\}$, is i.i.d. That assumption is often too restrictive for applications. What happens if the innovations are correlated? In this case Theorem 1 does no longer apply, and it turns out that in this case the limiting distributions of the processes defined in Section 1 depend on the nuisance parameter $\vartheta = \sigma/\eta$, where

$$\sigma^2 = E(\epsilon_1^2), \qquad \eta^2 = r(0)(1 + 2\sum_{k=1}^{\infty} r(k)).$$

Here $r(k) = E(\epsilon_1 \epsilon_{1+k})/E(\epsilon_1^2)$. In Steland (2005b) it is shown that under weak regularity conditions on the moments of $\{\epsilon_t\}$ the limit process of $D_T(s)$ is given by

$$(s/2)\{K(0)B(s)^2 + \zeta \int_0^s B(r)^2 K'(\zeta(s-r))dr - \vartheta^2 \int_0^s K(\zeta(s-r))dr\} / \int_0^s B(r)^2 dr$$

This means, the control limit c in (2) ensuring an asymptotic level α test becomes a function of ϑ. Using at each time point t a nonparametric estimator of ϑ using only past and current data yields an estimated control limit $c(\widehat{\vartheta}_t)$. This means, we consider the stopping time

$$\widehat{S}_T = \inf\{s \in (0,1) : D_T(s) < c(\widehat{\vartheta}_s)\}.$$

Simulations given in Steland (2005b) show that this procedure behaves considerably better than S_T, if $\vartheta \neq 1$.

4 Simulations

In this section we briefly present some simulations to analyze the detection procedure given by the stopping rule S_T. The computer programs were developed under a Linux system using the statistics software R. To speed up calculations, a shared library of C routines was developed using the GNU C-compiler.

When applying the detection procedure S_T, prerun data are required, i.e., one starts monitoring after, say, l, observations. The following table provides some simulated control limits (20,000 repetitions) under $H_0 : \rho = 1$ assuming $\vartheta = 1$ for various strategies given by (T, h, l), in particular for small values of T and h. Brownian motion was approximated by scaled partial sums with

T	h	l	1%	5%	10%
50	5	5	-3.73	-2.60	-2.13
50	5	10	-3.20	-2.18	-1.74
75	5	5	-3.79	-2.60	-2.13
75	5	10	-3.18	-2.19	-1.74
100	10	10	-4.12	-2.96	-2.35
100	10	20	-3.44	-2.33	-1.79
500	50	100	-3.62	-2.36	-1.78
500	50	150	-2.57	-1.68	-1.30
$T = \infty$			-2.97	-1.40	-0.86

Table 1: Simulated control limits for various design strategies (T, h, l).

250 terms. We see that the influence is also present for relatively large time horizons.

The following table provides the simulated type I error if $\alpha = 0.05$, power, and average run lengths, ES_T^*, for the procedure based on linear combinations motivated by the optimal test, using simulated critical values. Comparing the different weigthing schemes, it seems that the component $\sum_t Y_t \Delta Y_t$ contributing to the DF statistic, is preferable to detect $\rho < 1$. This motivates to study DF type detection procedures in greater detail, see Steland (2005b).

Many time series arising in economics and finance have fat tails. Thus, let us take a brief look at this stylized fact. To study the sensitivity of the Dickey-Fuller type detection rule with estimated nuisance parameters outlined in Section 3 we simulated time series of $N = 250$ observations satisfying model

w_1	w_2	ρ							
		1		0.98		0.95		1.1	
0.5	0.5	0.051	242.7	0.146	233.9	0.543	200.5	0.004	250
0.2	0.8	0.048	243.4	0.154	233.5	0.556	200.2	0.003	250
0.8	0.2	0.053	242.5	0.143	234.0	0.529	201.1	0.005	249.7

Table 2: Simulation results for the linear combination method. Table entries are empirical rejection rates (first column) and average run length (second column). Nominal significance level is 0.05.

ρ	df			
	2	3	10	∞
1	0.119	0.069	0.035	0.033
0.95	0.653	0.541	0.353	0.326

Table 3: Size and power for fat tailed error terms. Nominal significance level is 0.05.

(1) with $t(df)$-distributed error terms with asymptotic control limits ($h = 10$, thus $\zeta = 10$). The following empirical rejection rates under $H_0 : \rho = 1$ and the alternative $\rho = 0.95$ show a considerable sensitivity w.r.t. fat tails.

Acknowledgements

The support of the Deutsche Forschungsgemeinschaft, SFB 475 (Reduction of complexity in multivariate data structures) is gratefully acknowledged.

References

[1] Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimates for autoregressive time series with a unit root. *JASA*, **74**, 427-431.

[2] Husková, M. (1999). Gradual change versus abrupt change. *Journal of Statistical Planning and Inference*, **76**, 109-125.

[3] Hušková, M. and Slabý, A. (2001). Permutation tests for multiple changes. *Kybernetika*, **37**, 5, 605-622.

[4] Kurtz, and Protter (2004). *Weak Convergence of Stochastic Integrals and Differential Equations*, Lecture Notes.

[5] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159-178.

[6] Oksendal, B. (1992). *Stochastic Differential Equations.* Springer.

[7] Pawlak, M., Rafajłowicz, E., and Steland, A. (2004). Detecting jumps in time series - Nonparametric setting. *Journal of Nonparametric Statistics*, **16**, 329-347.

[8] Steland, A. (2004). Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives. *Metrika*, **60**, 229-249.

[9] Steland, A. (2005a). Stopping rules to detect unit roots and stationarity. *Submitted.*

[10] Steland, A. (2005b). Weighted Dickey-Fuller processes for detecting stationarity, *Preprint.*

[11] Stock, J.H. (1994). Unit roots, structural breaks and trends. In: *Handbook of Econometrics*, **4**, 2739-2841.