Unit roots: A selective review of the contributions of Peter C. B. Phillips

Author: Zhijie Xiao

Post-print version of an article published in Econometric Theory (2012), http://journals.cambridge.org/action/displayJournal?jid=ECT.

Post-print version of an article published in Econometric Theory (2012).

Unit Roots: A Selective Review of the Contributions of Peter C. B. Phillips

Zhijie Xiao*

Boston College

June 28, 2011

Abstract

Peter C. B. Phillips has made fundamental contributions to unit root econometrics. This paper provides a selective review of Peter's contribution to unit roots, with a focus on unit root asymptotics, unit root tests, and testing for stationarity against the unit root alternative. The discussion puts a relatively heavier weight on Peter's most recent work.

1 lntroduction

In 1998, Phoebus J. Dhrymes published a book "Time series, unit roots and cointegration". He dedicated this book to Peter C.B. Phillips, "whose work on integrated processes infused clarity and depth into the subject" (Dhrymes, 1998). This is unusual but is very indicative about Peter's contribution to unit root econometrics.

Nonstationarity is an important empirical feature in many macroeconomic and financial time series. In the last 20 years, hundreds of economic time series have been examined by unit root tests. Nowadays, unit root testing is a common exercise in macroeconomic time series applications. Peter

*I thank Peter Phillips, Bruce Hansen, Joon Park, Rustam Ibragimov, Ted Juhl, an anonymous referee, and participants at the 2008 New Zealand Econometrics Group Meeting for their very helpful comments. Address correspondence: Department of Economics, Boston College, Chestnut Hill, MA 02467, USA. Tel: 617-552-1709. Email: xiaoz@bc.edu.

Phillips' contribution to unit root time series analysis is fundamental, and the amount of his work on this topic is enormous. Peter has made important contributions in almost every sub-field in unit root analysis, and is still very active in research in this field. For these reasons, the current paper only provides a relatively selective review of Peter's contribution on unit root econometrics. In particular, I will focus on selected papers about (1) Unit root asymptotics; (2) Unit root tests; and (3) Testing stationarity against the unit root alternatives.

The first part on unit root asymptotics focuses on three papers: Phillips (1987a), "Time series regression with a unit root," *Econometrica*; Phillips and Solo (1992), "Asymptotics for linear processes," *Annals of Statistics*; and Ibragimov and Phillips (2008), "Regression Asymptotics using Martingale Convergence Methods," *Econometric Theory*. Phillips (1987a) is a classic paper in this field and has influenced much of the research on unit roots in the last 20 years. Phillips and Solo (1992) provides a powerful approach for time series asymptotic analysis. This approach is now widely used in econometrics. The recent paper by Ibragimov and Phillips (2008) develops a new and conceptually simple method that has great generality and a wide range of applicability.

The second part of this paper reviews the unit root tests. The semiparametric unit root tests proposed in Phillips (1987a) and Phillips and Perron (1988) have numerous applications in econometrics; the approach of nonparametric correction for serial correlation and endogeneity suggested by Phillips (1987a) and Phillips and Hansen (1990) is now an important method in econometrics; the local asymptotics of Phillips (1987b, 1988) provide a statistical foundation for local power analysis in unit root tests; Peter's work on nonlinear and nonparametric unit root models stimulates numerous extensions and applications along these directions.

The third part of this paper considers testing stationarity against the unit root alternative - the KPSS test of Kwiatkowski, Phillips, Schmidt, and Shin (1992). Testing the null hypothesis of stationarity is subtle. Additional regularity conditions are necessary to guarantee a consistent test against the unit root alternative. The KPSS test has had a huge impact on subsequent research on this topic.

There are lots of other important contributions of Peter on unit roots that are not included in this paper.

Notation in this paper is standard, we use the symbol "$\Rightarrow$" to signify weak convergence of the associated probability measures, continuous stochastic process such as the Brownian motion $B(r)$ on [0,1] are usually written simply as B and integrals like $\int$ are understood to be Lebesgue integrals over the interval [0,1], the measure of integration " dr " being omitted for simplicity. In addition, a standard Brownian motion is usually denoted by $W(r)$, and $BM(\omega^2)$ signifies a Brownian motion with variance ω^2.

2 Unit Root Asymptotics

2.1 Unit root asymptotics via representations as polynomials in sample moments

Peter has made fundamental contributions to unit root asymptotics. Phillips (1987a) and Phillips (1986) pioneered the use of functional limit theory in econometrics and showed for the first time how to use this theory in regression problems with nonstationary data. In some ways, this was the most immediate and long-lasting contribution in the field. It has been quickly adopted by econometricians and there were many papers using this method and Brownian functionals even in general interest economics journals.

Phillips (1987a) plays a crucial role in unit root econometrics. In this paper he derived, under mixing conditions on the residual process, the limiting distribution of the unit root autoregression estimator in stochastic process (Brownian motion) representations on function spaces, and proposed semiparametric unit root tests that use a nonparametric treatment for weak correlation and endogeneity. In particular, functional limit theory for multilinear forms of weakly dependent mixing random variables is derived by using their representations as polynomials in sample moments (via summation by parts arguments) and then using standard weak convergence results for sums of weakly dependent sequences. This classic work gives invariance principles for partial sums, sample variances and sample moments, and convergence to stochastic integrals. Much of these results and analysis have been extensively used in econometrics in the last 20 years, and the methods are part of the general econometric toolkit.

If we consider the following autoregressive process:

$$y_t = \alpha y_{t-1} + u_t, \tag{1}$$

where y_0 is any random variable with finite variance, and u_t is a weakly dependent process such that

1. $\mathrm{E}(u_t) = 0$;

2. $\sup_t \mathrm{E}|u_t|^\beta < \infty$, $\beta > 2$;

3. The long-run variance of u_t, $\omega^2 = \lim n^{-1} E\left(\sum_{t=1}^n u_t\right)^2$, and the variance of u_t, $\sigma_u^2 = \lim n^{-1} \sum_{t=1}^n E u_t^2$, exist and are positive;

4. u_t is strong mixing with $\sum \alpha_m^{1-2/\beta} < \infty$, where α_m is the mixing coefficient.

The ordinary least square estimator of α is given by

$$\widehat{\alpha} = \frac{\sum_{t=2}^n y_{t-1} y_t}{\sum_{t=2}^n y_{t-1}^2}. \tag{2}$$

Notice that

$$\widehat{\alpha} = \alpha + \frac{\sum_{t=2}^n y_{t-1} u_t}{\sum_{t=2}^n y_{t-1}^2}, \tag{3}$$

the asymptotic behavior of $\widehat{\alpha}$ is determined by limit theorems for the following bilinear forms:

$$\sum_{t=2}^n y_{t-1}^2, \text{ and } \sum_{t=2}^n y_{t-1} u_t.$$

When $\alpha = 1$,

$$\frac{1}{\sqrt{n}} y_{[nr]} = \frac{1}{\sqrt{n}} y_0 + \frac{1}{\sqrt{n}} \sum_{j=1}^{[nr]} u_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{[nr]} u_j + o_p(1),$$

if we denote $Y_n(r) = n^{-1/2} \sum_{j=1}^{[nr]} u_j$, then $Y_n(r) \Rightarrow B(r) = BM(\omega^2)$ by the invariance principle for mixing processes. The limiting behavior of the denominator of (3), $\sum_{t=1}^n y_{t-1}^2$, can then be immediately derived from the invariance principle and the continuous mapping theorem:

$$\begin{aligned}
\frac{1}{n^2} \sum_{t=2}^n y_{t-1}^2 &= \frac{1}{n} \sum_{t=2}^n \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{t-1} u_j \right)^2 + o_p(1) \\
&= \frac{1}{n} \sum_{t=2}^n \left[\int_{(t-1)/n}^{t/n} Y_n(r)^2 dr \right] + o_p(1) \\
&= \int_0^1 Y_n(r)^2 dr + o_p(1) \\
&\Rightarrow \int_0^1 B(r)^2 dr.
\end{aligned}$$

The numerator of (3), $\sum_{t=2}^{n} y_{t-1}u_t$, is the sample covariance between I(1) and I(0) components. Phillips (1987a) analyzed this term using its representation as a polynomial in sample moments via summation by parts. In particular, noticing that

$$\frac{1}{n}\sum\nolimits_{t=2}^{n} y_{t-1}u_t = \frac{1}{n}\sum\nolimits_{t=2}^{n}\left(\sum\nolimits_{j=1}^{t-1} u_j\right)u_t + o_p(1),$$

and

$$\begin{aligned}\frac{1}{n}\sum\nolimits_{t=2}^{n}\left(\sum\nolimits_{j=1}^{t-1} u_j\right)u_t &= \frac{1}{2}\left[\frac{1}{n}\left(\sum\nolimits_{j=1}^{n} u_j\right)^2 - \frac{1}{n}\sum\nolimits_{j=1}^{n} u_j^2\right] \\ &\Rightarrow \frac{1}{2}\left[B(1)^2 - \sigma_u^2\right] \\ &= \frac{1}{2}\left[B(1)^2 - \omega^2\right] + \frac{1}{2}\left[\omega^2 - \sigma_u^2\right] \\ &= \int_0^1 B(r)dB(r) + \lambda,\end{aligned}$$

where $\lambda = \frac{1}{2}\left[\omega^2 - \sigma_u^2\right]$ is the one-sided long-run variance of u_t and the last equality comes from Ito's Lemma. Consequently,

$$n\left(\widehat{\alpha} - 1\right) = \frac{\frac{1}{n}\sum_{t=2}^{n} y_{t-1}u_t}{\frac{1}{n^2}\sum_{t=2}^{n} y_{t-1}^2} \Rightarrow \frac{\omega^2\int_0^1 W(r)dW(r) + \lambda}{\omega^2\int_0^1 W(r)^2 dr}. \tag{4}$$

A detailed analysis of the above asymptotics is given by Phillips (1987a), and has been widely used in the literature. Also see White (1958), Lai and Wei (1982), Solo (1984) and Chan and Wei (1988) for related studies on this topic.

Recently, Ibragimov and Phillips (2008) show that the above results are a natural outcome of convergence of a sequence of martingales to a continuous martingale. This is another important contribution to unit root asymptotics, and is applicable to a wide range of econometric models. We illustrate the basics of this approach in the following section.

2.2 Unit root asymptotics using martingale convergence methods

Ibragimov and Phillips (2008) develop a new and conceptually simple method based on martingale convergence for obtaining weak convergence of partial sums and multilinear forms in independent random variables and linear processes to stochastic integrals. The martingale convergence method has great generality and a wide range of applicability. It can be used in developing weak convergence of different types of multilinear forms, giving invariance principles for partial sums, sample variances and sample covariances; convergence to stochastic integrals; and asymptotics for general functionals of partial sums. The approach reduces all asymptotics to the weak convergence of (semi)martingales, i.e. convergence of a sequence of (semi)martingales to a continuous (semi)martingale. In particular, under appropriate assumptions (including identification of the limit), semimartingale convergence can be derived based on convergence of its predictable characteristics. Using the martingale convergence method, a unified treatment of the asymptotics for stationary, near unit root, unit root, local to unit root, and explosive autoregressions can be constructed.

To study unit root asymptotic distribution of the AR estimator using martingale convergence theory, Ibragimov and Phillips (2008) consider the following recursive OLS estimator

$$\widehat{\alpha}_r = \frac{\sum_{t=1}^{[nr]} y_{t-1} y_t}{\sum_{t=1}^{[nr]} y_{t-1}^2}, \text{ where } r \in (0,1].$$

Let us start with the simplest case with $u_t \equiv \mathrm{iid}(0,\sigma^2)$ in (1). In the presence of a unit root,

$$n\left(\widehat{\alpha}_r - 1\right) = \frac{n^{-1}\sum_{t=1}^{[nr]} y_{t-1}u_t}{n^{-2}\sum_{t=1}^{[nr]} y_{t-1}^2}. \tag{5}$$

First, by Skorohod embedding as in Park and Phillips (1999), Ibragimov and Phillips (2008) construct a probability space supporting $S_t = \sum_{j=1}^{t} u_j$, Brownian motion $B(r)$, and an increasing sequence of nonnegative stopping times $(T_t)_{t\geq 0}$ such that

$$\frac{u_t}{\sqrt{n}} =_d B\left(\frac{T_t}{n}\right) - B\left(\frac{T_{t-1}}{n}\right),$$

so that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{t} u_j =_d B\left(\frac{T_t}{n}\right),$$

and

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{k} y_{t-1}u_t &= \sum_{t=1}^{k}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{t-1} u_j\right)\frac{u_t}{\sqrt{n}} + o_p(1)\\ &= \sum_{t=1}^{k} B\left(\frac{T_{t-1}}{n}\right)\left\{B\left(\frac{T_t}{n}\right) - B\left(\frac{T_{t-1}}{n}\right)\right\} + o_p(1).\end{aligned}$$

We next construct the following sequence of processes:

$$X_n(r) = \sum_{t=1}^{[nr]} B\left(\frac{T_{t-1}}{n}\right)\left\{B\left(\frac{T_t}{n}\right) - B\left(\frac{T_{t-1}}{n}\right)\right\} + B\left(\frac{T_{[nr]}}{n}\right)\left\{B\left(r\right) - B\left(\frac{T_{[nr]}}{n}\right)\right\}, \tag{6}$$

which is a *martingale* and is *continuous in* r. By Skorohod embedding, we have the following martingale representation

$$
\begin{aligned}
& \sum_{t=1}^{[nr]} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{t-1} u_j \right) \frac{u_t}{\sqrt{n}} \\
= & \sum_{t=1}^{[nr]} B\left(\frac{T_{t-1}}{n}\right) \left\{ B\left(\frac{T_t}{n}\right) - B\left(\frac{T_{t-1}}{n}\right) \right\} \\
= & {}_d X_n \left(\frac{T_{[nr]}}{n} \right),
\end{aligned}
$$

thus

$$
\frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1} u_t = \sum_{t=1}^{[nr]} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{t-1} u_j \right) \frac{u_t}{\sqrt{n}} + o_p(1) = X_n \left(\frac{T_{[nr]}}{n} \right) + o_p(1).
$$

Using the above device, Ibragimov and Phillips (2008) show that the limiting behavior of $n^{-1} \sum_{t=1}^{[nr]} y_{t-1} u_t$ can be obtained from convergence of a sequence of martingales $X_n\left(\frac{T_{[nr]}}{n}\right)$ to the following continuous martingale:

$$
X(r) = \int_0^r B(s) dB(s). \tag{7}
$$

Weak Convergence of (Semi)martingales. Using a very general convergence result for (semi)martingales obtained by Jacod and Shiryaev (2003), and overcoming some technical problems in the existing literature, Ibragimov and Phillips (2008) show that the study of weak convergence for the (semi)martingales is reduced to the study of convergence of their *predictable characteristics*.

In general, let $(X(r), r \geq 0)$ be a continuous d-dimensional semimartingale on a probability space $(\Omega, \mathcal{F}, P)$ with respect to a filtration $\mathbb{F} = (\mathcal{F}_r, r \geq 0)$ of sub σ-fields of $\mathcal{F}$, then $X(r)$ can be decomposed into the summation of an initialization component $X(0)$; a martingale component $M(r)$; and a predictable component $A(r)$. More specifically, let $X(r) = \left(X^1(r), \cdots, X^d(r)\right)$, then $X(r)$ admits the following unique decomposition

$$
X^j(r) = X^j(0) + M^j(r) + A^j(r)
$$

where $X^j(0)$ are finite valued $\mathcal{F}_0$-measurable random variables; $M^j(r)$ are continuous local $\mathbb{F}$-martingales with $M^j(0) = 0$; and $A^j(r)$ are continuous $\mathbb{F}$-adapted finite variation process.

If we denote (the quadratic variation process) $C(r) = \left(C^{ij}(r), 1 \le i, j \le d\right)$, where $C^{ij}(r) = [C^i, C^j](r)$ is the unique continuous process for which $C^iC^j - C^{ij}$ is a martingale, then the process

$$A(r) = \left(A^1(r), \cdots, A^d(r)\right)$$

is called the first predictable characteristic of X, and $C(r)$ is called the second predictable characteristic of X.

Given continuous locally square integrable semimartingales $X_n = (X_n(r), r \ge 0)$ and $X = (X(r), r \ge 0)$, let their predictable characteristics be $(A_n(r), C_n(r))$ and $(A(r), C(r))$, and initializations be $X_n(0)$ and $X(0)$, Jacod and Shiryaev (2003) provide sufficient condition for weak convergence of semimartingales $X_n \xrightarrow{d} X$. In particular, convergence of a sequence of martingales ($X_n \xrightarrow{d} X$) holds if

1. Their predictable characteristics (A_n, C_n) and the initial distributions ($X_n(0)$) tend to those of the limit martingale (i.e. A, C, and $X(0)$);
2. The predictable characteristics of the limit process grow in a regular way;
3. The process X is the only continuous martingale with characteristics A and C and the initial distribution $X(0)$.

Intuitively, conditions 2 and 1 guarantee tightness of the sequence and the uniqueness condition (Condition 3) guarantees identification of the limit. For a rigorous and detailed treatment of the martingale convergence analysis, see Jacod and Shiryaev (2003) and Ibragimov and Phillips (2008).

Going back to the recursive OLS estimator (5), let X_n and X be defined as (6) and (7), if we consider the quadratic variation processes[1] of X_n and X, denoted by $[X_n]_r$ and $[X]_r$ (existence and uniqueness of the quadratic variation processes hold by the Doob-Meyer decomposition), notice that the first predictable characteristics (A_n and A) corresponding to (X_n and X) are simply 0, and conditions 2 and 3 above are satisfied for the process X, Ibragimov and Phillips (2008) (also see Jacod and Shiryaev (2003)) show that

$$X_n(r) \xrightarrow{d} X(r)$$

[1] If $M(r)$ is a martingale, and its quadratic variation process is denoted by $[M]_r$, then $M(r)^2 - [M]_r$ is a martingale.

if the corresponding quadratic variation processes (the second predictable characteristics) converge in probability:

$$C_n(r) \xrightarrow{p} C(r) \ ,$$

i.e.

$$[X_n]_r \xrightarrow{p} [X]_r \, .$$

This is true since

$$\sup_{t \le n} \left| \frac{T_t}{n} - \frac{t}{n} \right| \xrightarrow{a.s.} 0,$$

by Park and Phillips (1999) and Phillips and Ploberger (1996), and

$$\left[B\left(\frac{T_{[nr]}}{n}\right) \left\{ B\left(r\right) - B\left(\frac{T_{[nr]}}{n}\right) \right\} \right]_r = B\left(\frac{T_{[nr]}}{n}\right)^2 \left\{ r - \frac{T_{[nr]}}{n} \right\} \sigma^2 \xrightarrow{p} 0,$$

thus

$$\begin{aligned} [X_n]_r &= \sum_{t=1}^{[nr]} B\left(\frac{T_{t-1}}{n}\right)^2 \left\{ \frac{T_t}{n} - \frac{T_{t-1}}{n} \right\} \sigma^2 + B\left(\frac{T_{[nr]}}{n}\right)^2 \left\{ r - \frac{T_{[nr]}}{n} \right\} \sigma^2 \\ &\approx \sum_{t=1}^{[nr]} B\left(\frac{T_{t-1}}{n}\right)^2 \left\{ \frac{T_t}{n} - \frac{T_{t-1}}{n} \right\} \sigma^2 \\ &\approx \sigma^2 \sum_{t=1}^{[nr]} B\left(\frac{s}{n}\right)^2 \frac{1}{n} \\ &\xrightarrow{p} \sigma^2 \int_0^r B\left(s\right)^2 ds = [X]_r \, . \end{aligned}$$

Consequently, $n\left(\widehat{\alpha} - 1\right) \Rightarrow \left[\int_0^1 B(r)^2 dr\right]^{-1} \left[\int_0^1 B(r) dB(r)\right]$ as a result of martingale convergence.

The above martingale convergence analysis can be conveniently extended to the case where u_t is weakly dependent. Ibragimov and Phillips (2008) provide an analysis for linear processes using martingale convergence arguments and the linear process method of Phillips and Solo (1992).

The linear process asymptotic theory of Phillips and Solo (1992) provides an extremely convenient and powerful approach for time series asymptotic analysis. The basic idea of this method is to use *simple limiting theory* and the *algebra of the linear operator* to extract limiting theory for linear processes. This approach is based on martingale transformations of linear processes. Such transformations reduce complicated asymptotics for dependent linear processes to those of i.i.d., i.ni.d., or martingale difference sequences, and deliver an ingenious approach to develop strong laws, central limit theorems, functional laws, and laws of iterated logarithms for time series. The approach applies not only to the linear process itself but also to products and multivariate processes, and it can be further used in panel data and frequency domain problems.

We assume that

$$u_t = C(L)\varepsilon_t, \tag{8}$$

where ε_t is an i.i.d. sequence with $\mathrm{Var}(\varepsilon_t^2) = \sigma_\varepsilon^2$, and

$$C(L) = \sum\nolimits_{j=0}^{\infty} c_j L^j,\ \sum\nolimits_{j=0}^{\infty} j|c_j| < \infty,\ C(1) \neq 0, \tag{9}$$

where L is the lag operator for which $Ly_t = y_{t-1}$. The initial condition in (1) is set at $t = 0$, and again we assume that y_0 may be a constant or a random variable with finite variance. The second and third conditions of (9) ensure that u_t is *covariance stationary* and has positive *spectral density* at the origin - see Phillips and Solo (1992) for discussions on various alternative settings for ε_t and different summability conditions. Under the summability condition in (9), Phillips and Solo (1992) validate the following expansion of the operator $C(L)$

$$C(L) = C(1) + \widetilde{C}(L)(L-1), \tag{10}$$

where $\widetilde{C}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j$ and $\widetilde{c}_j = \sum_{s=j+1}^{\infty} c_s$. This expansion gives rise to an explicit martingale difference decomposition of u_t

$$u_t = C(1)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t, \quad \text{with } \widetilde{\varepsilon}_t = \widetilde{C}(L)\varepsilon_t. \tag{11}$$

This decomposition is sometimes called the martingale decomposition in the probability literature (see, e.g., Hall and Heyde, 1980) because, if $\{\varepsilon_t\}$ is a martingale difference sequence, the first term of (11) is a martingale difference and the partial sums $\sum_{s=1}^{t} u_s$ correspondingly have the leading martingale term $C(1)\sum_{s=1}^{t} \varepsilon_s$. The expansion is also called the BN decomposition due to the work of Beveridge and Nelson (1981) on decomposing aggregated economic data into long run and short run components. If y_t is generated by (1) with $\alpha = 1$, then

$$y_t = C(1)\sum_{s=1}^{t} \varepsilon_s + \widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_t + y_0,$$

where $C(1)\sum_{s=1}^{t}\varepsilon_s$ and $\widetilde{\varepsilon}_0-\widetilde{\varepsilon}_t+y_0^s$ are the long run and short run components of y_t, respectively. Phillips and Solo (1992) show that

$$\frac{1}{\sqrt{n}}y_{[nr]}=C(1)\frac{1}{\sqrt{n}}\sum_{s=1}^{[nr]}\varepsilon_s+o_p(1)\Rightarrow C(1)B_\varepsilon(r)=B(r)$$

where $B_\varepsilon(r)=$BM(σ_ε^2).

Ibragimov and Phillips (2008) analyzed the limiting behavior of $n^{-1}\sum_{t=2}^{n}y_{t-1}u_t$ under assumptions (8) and (9). First, by the previous analysis in this section using the martingale convergence method and Skorohod embedding, we have

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t\Rightarrow\int_0^r B_\varepsilon(s)dB_\varepsilon(s)=\sigma_\varepsilon^2\int_0^r W(s)dW(s).$$

Let $\lambda=\sum_{j=1}^{\infty}\mathrm{E}(u_0u_j)$, and apply the BN decomposition (11) twice, we have

$$\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t=C(1)^2\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t+C(1)\sum_{t=2}^{[nr]}\left(\widetilde{\varepsilon}_0-\widetilde{\varepsilon}_{t-1}\right)\varepsilon_t+\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)\left(\widetilde{\varepsilon}_{t-1}-\widetilde{\varepsilon}_t\right),$$

thus

$$\begin{aligned}&\left|\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t-r\lambda-C(1)^2\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t\right|\\ =&\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\left(\widetilde{\varepsilon}_0-\widetilde{\varepsilon}_{t-1}\right)\varepsilon_t+\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)\left(\widetilde{\varepsilon}_{t-1}-\widetilde{\varepsilon}_t\right)-r\lambda\right|\\ \leq&\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_0\varepsilon_t\right|+\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_{t-1}\varepsilon_t\right|+\left|\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)\left(\widetilde{\varepsilon}_{t-1}-\widetilde{\varepsilon}_t\right)-r\lambda\right|.\end{aligned}$$

Notice that

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)\left(\widetilde{\varepsilon}_{t-1}-\widetilde{\varepsilon}_t\right)=-\frac{1}{n}\sum_{t=1}^{[nr]}u_t\widetilde{\varepsilon}_{[nr]}+\frac{1}{n}\sum_{t=1}^{[nr]}u_t\widetilde{\varepsilon}_t,$$

thus

$$\begin{aligned}&\left|\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t-r\lambda-C(1)^2\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t\right|\\ \leq&\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_0\varepsilon_t\right|+\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_{t-1}\varepsilon_t\right|+\left|\left(\frac{1}{n}\sum_{t=1}^{[nr]}u_t\right)\widetilde{\varepsilon}_{[nr]}\right|+\left|\frac{1}{n}\sum_{t=1}^{[nr]}u_t\widetilde{\varepsilon}_t-r\lambda\right|.\end{aligned}$$

The BN decomposition can be used again to analyze the term $n^{-1}\sum_{t=1}^{[nr]} u_t\widetilde{\varepsilon}_t - r\lambda$. Notice that $\widetilde{\varepsilon}_t = \widetilde{C}(L)\varepsilon_t$, therefore we have

$$
\begin{aligned}
u_t\widetilde{\varepsilon}_t &= \left(C(1)\varepsilon_t + \widetilde{C}(L)\varepsilon_{t-1} - \widetilde{C}(L)\varepsilon_t\right)\widetilde{C}(L)\varepsilon_t \\
&= h_0(L)\varepsilon_t^2 + \sum_{s=1}^{\infty} h_s(L)\varepsilon_t\varepsilon_{t-s},
\end{aligned}
$$

where $h_0(L) = \sum_{k=0}^{\infty} c_k\widetilde{c}_k L^k$, $h_s(L) = \sum_{k=0}^{\infty}\left(c_{k+s}\widetilde{c}_k + c_k\widetilde{c}_{k+s}\right)L^k$. Applying the BN decomposition again to $h_s(L)$ $(s = 0, 1, \cdots)$,

$$
h_s(L) = h_s(1) + \widetilde{h}_s(L)(L-1),
$$

we have

$$
u_t\widetilde{\varepsilon}_t = h_0(1)\varepsilon_t^2 + (L-1)\left[\widetilde{h}_0(L)\varepsilon_t^2\right] + \varepsilon_t\left(\sum_{s=1}^{\infty} h_s(1)\varepsilon_{t-s}\right) + (L-1)\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{t-s}\right)\varepsilon_t,
$$

consequently

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{[nr]} u_t\widetilde{\varepsilon}_t \\
=\;& h_0(1)\frac{1}{n}\sum_{t=1}^{[nr]}\varepsilon_t^2 + \frac{1}{n}\widetilde{h}_0(L)\varepsilon_0^2 - \frac{1}{n}\widetilde{h}_0(L)\varepsilon_{[nr]}^2 \\
&+\frac{1}{n}\sum_{t=1}^{[nr]}\varepsilon_t\left(\sum_{s=1}^{\infty} h_s(1)\varepsilon_{t-s}\right) + \frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{-s}\right)\varepsilon_0 - \frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{[nr]-s}\right)\varepsilon_{[nr]}.
\end{aligned}
$$

Thus, for all $N \in \mathbf{N} = \{1, 2, \cdots\}$, and all $r \in [0, N]$,

$$
\begin{aligned}
&\sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1} u_i\right)u_t - r\lambda - C(1)^2\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t\right| \\
\le\;& \sup_{0\le r\le N}\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_0\varepsilon_t\right| + \sup_{0\le r\le N}\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_{t-1}\varepsilon_t\right| + \sup_{0\le r\le N}\left|\left(\frac{1}{n}\sum_{t=1}^{[nr]} u_t\right)\widetilde{\varepsilon}_{[nr]}\right| \\
&+\sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=1}^{[nr]}\left[h_0(1)\varepsilon_t^2 - \lambda\right]\right| + \left|\frac{1}{n}\widetilde{h}_0(L)\varepsilon_0^2\right| + \sup_{0\le r\le N}\left|\frac{1}{n}\widetilde{h}_0(L)\varepsilon_{[nr]}^2\right| \\
&+\sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=1}^{[nr]}\varepsilon_t\left(\sum_{s=1}^{\infty} h_s(1)\varepsilon_{t-s}\right)\right| + \left|\frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{-s}\right)\varepsilon_0\right| \\
&+\sup_{0\le r\le N}\left|\frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{[nr]-s}\right)\varepsilon_{[nr]}\right|
\end{aligned}
$$

Noticing that $\lambda = h_0(1)\sigma_\varepsilon^2$, and the following sequences: $\{\widetilde{\varepsilon}_0\varepsilon_t\}$, $\{\widetilde{\varepsilon}_{t-1}\varepsilon_t\}$, $\left\{h_0(1)\varepsilon_t^2 - \lambda\right\}$, $\{\varepsilon_t\left(\sum_{s=1}^{\infty} h_s(1)\varepsilon_{t-s}\right.$ are all martingale-differences sequences with finite variances, Ibragimov and Phillips (2008) show that

$$\sup_{0\le r\le N}\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_0\varepsilon_t\right| \xrightarrow{p} 0,\ \sup_{0\le r\le N}\left|C(1)\frac{1}{n}\sum_{t=2}^{[nr]}\widetilde{\varepsilon}_{t-1}\varepsilon_t\right| \xrightarrow{p} 0,$$
$$\sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=1}^{[nr]}\left[h_0(1)\varepsilon_t^2 - \lambda\right]\right| \xrightarrow{p} 0,\ \sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=1}^{[nr]}\varepsilon_t\left(\sum_{s=1}^{\infty}h_s(1)\varepsilon_{t-s}\right)\right| \xrightarrow{p} 0.$$

In addition

$$\sup_{0\le r\le N}\left|\left(\frac{1}{n}\sum_{t=1}^{[nr]}u_t\right)\widetilde{\varepsilon}_{[nr]}\right| \xrightarrow{p} 0,\ \left|\frac{1}{n}\widetilde{h}_0(L)\varepsilon_0^2\right| \xrightarrow{p} 0,\ \sup_{0\le r\le N}\left|\frac{1}{n}\widetilde{h}_0(L)\varepsilon_{[nr]}^2\right| \xrightarrow{p} 0,$$
$$\left|\frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{-s}\right)\varepsilon_0\right| \xrightarrow{p} 0,\ \sup_{0\le r\le N}\left|\frac{1}{n}\left(\sum_{s=1}^{\infty}\widetilde{h}_s(L)\varepsilon_{[nr]-s}\right)\varepsilon_{[nr]}\right| \xrightarrow{p} 0,$$

under the summability condition and appropriate moment conditions (Ibragimov and Phillips (2008)), thus

$$\sup_{0\le r\le N}\left|\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t - r\lambda - C(1)^2\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t\right| \xrightarrow{p} 0,$$

and

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t \Rightarrow r\lambda + C(1)^2\int_0^r B_\varepsilon(s)dB_\varepsilon(s) = r\lambda + C(1)^2\sigma_\varepsilon^2\int_0^r W(s)dW(s).$$

Notice that, by continuous mapping:

$$\frac{1}{n^2}\sum_{t=2}^{n}y_{t-1}^2 = \frac{1}{n}\sum_{t=2}^{n}\left(\frac{y_{t-1}}{\sqrt{n}}\right)^2 = \int_0^1\left[\frac{1}{\sqrt{n}}y_{[nr]}\right]^2 dr \Rightarrow C(1)^2\int_0^1 B_\varepsilon^2(r)dr,$$

thus limiting distribution of $\hat{a}$ in the general case with weakly correlated u_t can be obtained as a result of martingale convergence:

$$n(\hat{a}-1) = \frac{n^{-1}\sum_{t=1}^{n}y_{t-1}u_t}{n^{-2}\sum_{t=1}^{n}y_{t-1}^2} \Rightarrow \frac{\lambda + \omega^2\int_0^1 W(s)dW(s)}{\omega^2\int_0^1 W^2(r)dr} \tag{12}$$

where $\omega^2 = C(1)^2\sigma_\varepsilon^2$. Similarly, the limiting distribution of the t-ratio statistic is given by

$$t_{\hat{a}} \Rightarrow \frac{\lambda + \omega^2\int_0^1 W(s)dW(s)}{\sigma_u\left[\omega^2\int_0^1 W^2(r)dr\right]^{1/2}}, \tag{13}$$

where $\sigma_u^2 = \mathrm{Var}(u_t)$.

2.3 General Autoregression Asymptotics

Peter's contribution to autoregression asymptotic theory goes much beyond the exact unit root models. Phillips (1987b, 1988) studied asymptotic theory for autoregressions with a root that is local to unity; Phillips, Moon and Xiao (2001), Giraitis and Phillips (2004), Phillips and Magdalinos (2004, 2005) studied asymptotic theory for autoregressions with moderate deviations from unity; and Ibragimov and Phillips (2008) studied a wide range of autoregressive models.

The asymptotic analysis described in the previous section embeds $n^{-1}\sum_{t=1}^{[nr]} y_{t-1}u_t$ in its continuous version and thus convergence results are obtained based on weak convergence of continuous martingales. Ibragimov and Phillips (2008) also show how to directly treat the discrete time martingales and semimartingales using their third predictable characteristic as the predictable measure of jumps. Such extensions provide a unified formulation for limiting theory of stationary, unit root, local to unit root models, and autoregressions with moderate deviations from unit root, as well as explosive autoregressions.

For this purpose it is useful to introduce the *compensator* of $[X]_r$ - the conditional quadratic variation process $\langle X\rangle_r$ (Phillips, 2005). In particular, the compensator of $[X]_r$ compensates $[X]_r$ to make it into a martingale, i.e.

$$[X]_r - \langle X\rangle_r = \text{martingale}.$$

If we consider autoregression (1), with $u_t = \text{i.i.d}(0, \sigma^2)$, and α may be:

1. $|\alpha| < 1$: Stationary
2. $\alpha = 1 + c/n^b$, $0 < b < 1$ and $c < 0$: Moderate deviations from unit root
3. $\alpha = 1 + c/n$: Local to unit root
4. $\alpha = 1$: Unit root
5. $\alpha > 1$: Explosive autoregression

Ibragimov and Phillips (2008) consider the following standardization on the recursive OLS estimator $\widehat{\alpha}_r$:

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2}\right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{\sum_{t=1}^{[nr]} y_{t-1} u_t}{\left(\sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2\right)^{1/2}},$$

which, as will be more clear later, can be represented in martingale form as a ratio $X_n(r)/\left(\langle X_n\rangle_r\right)^{1/2}$, and the limit theory can be delivered by martingale convergence in the form

$$\frac{X_n(r)}{\left(\langle X_n\rangle_r\right)^{1/2}} \xrightarrow{d} \frac{X(r)}{\left(\langle X\rangle_r\right)^{1/2}},$$

where $X(r)$ is the corresponding limiting martingale process.

Stationary case: When $|\alpha| < 1$, let

$$X_n\left(r\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} y_{t-1} u_t, \text{ and } X\left(r\right) = \sigma \int_0^r dB_y(s) = \sigma B_y(r),$$

where $B_y(r) = BM(\sigma_y^2)$, and $\sigma_y^2 = \sigma^2/(1-\alpha^2)$, then

$$[X_n]_r = \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1}^2 u_t^2, \;\; \langle X_n\rangle_r = \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2$$

and, notice that $X\left(r\right)$ is a continuous martingale,

$$[X]_r = \langle X\rangle_r = \sigma^2 \sigma_y^2 r \;,$$

thus

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2}\right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{X_n\left(r\right)}{\left(\langle X_n\rangle_r\right)^{1/2}} \xrightarrow{D} \frac{X\left(r\right)}{\left(\langle X\rangle_r\right)^{1/2}} = \frac{\sigma B_y(r)}{\left(\sigma^2 \sigma_y^2 r\right)^{1/2}} = N(0,1).$$

Notice that if we consider

$$\widehat{\sigma}_r^2 = \frac{1}{[nr]} \sum_{t=1}^{[nr]} \widehat{u}_t^2 \to \sigma^2, \text{ where } \widehat{u}_t = y_t - \widehat{\alpha}_r y_{t-1},$$

then

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\widehat{\sigma}_r^2}\right)^{1/2} (\widehat{\alpha}_r - \alpha) \xrightarrow{D} \frac{X\left(r\right)}{\left(\langle X\rangle_r\right)^{1/2}} = N(0,1).$$

Moderate deviation from unity: If $\alpha = 1 + c/n^b$, for some $b \in (0,1)$ and $c < 0$, notice that by Phillips (1987b)

$$\frac{1}{\sqrt{n^b}} y_{[n^b r]} \xrightarrow{P} \int_0^r e^{c(r-s)} dB(s) = J_c(r),$$

let

$$X_n(r) = \frac{1}{n^{(1+b)/2}} \sum_{t=1}^{[nr]} y_{t-1} u_t, \text{ and } X(r) = \sqrt{\frac{\sigma^4}{-2c}} W(r) = N\left(0, \frac{\sigma^4}{-2c} r\right),$$

then

$$\langle X_n \rangle_r = \frac{1}{n^{1+b}} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2, \text{ and } \langle X \rangle_r = \frac{\sigma^4}{-2c} r.$$

Thus,

$$\begin{aligned} X_n(r) &= \frac{1}{n^{(1+b)/2}} \sum_{t=1}^{[nr]} y_{t-1} u_t = \frac{1}{n^{(1-b)/2}} \sum_{\tau=1}^{n^{1-b}} \sum_{t=n^b(\tau-1)}^{n^b \tau} \left(\frac{y_{t-1}}{n^{b/2}}\right) \left(\frac{u_t}{n^{b/2}}\right) \\ &= \frac{1}{n^{(1-b)/2}} \sum_{\tau=1}^{n^{1-b}} \int_{\tau-1}^{\tau} J_c(s) dB(s) + o_p(1). \end{aligned}$$

and $\xi_\tau = \int_{\tau-1}^{\tau} J_c(s) dB(s)$ is a martingale difference sequence with variance

$$\mathrm{E}\left(\xi_\tau^2\right) = \mathrm{E}\left(\sigma^2 \int_{\tau-1}^{\tau} J_c^2(s) ds\right) = \frac{\sigma^4}{-2c}\left(1 - e^{2c\tau}\right) = \frac{\sigma^4}{-2c} + O\left(e^{2c\tau}\right), \ c < 0,$$

thus,

$$X_n(r) = \frac{1}{n^{(1-b)/2}} \sum_{\tau=1}^{n^{1-b} r} \int_{\tau-1}^{\tau} J_c(s) dB(s) + o_p(1) \xrightarrow{P} \sqrt{\frac{\sigma^4}{-2c}} W(r) = X(r)$$

and

$$\begin{aligned} \langle X_n \rangle_r &= \frac{1}{n^{1+b}} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2 = \frac{1}{n^{1-b}} \sum_{\tau=1}^{n^{1-b} r} \frac{1}{n^b} \sum_{t=n^b(\tau-1)}^{n^b \tau} \left(\frac{y_{t-1}}{n^{b/2}}\right)^2 \sigma^2 \\ &= \frac{\sigma^2}{n^{1-b}} \sum_{\tau=1}^{n^{1-b} r} \int_{\tau-1}^{\tau} J_c^2(s) ds + o_p(1) \\ &\to \sigma^2 \lim_n \frac{1}{n^{1-b}} \sum_{\tau=1}^{n^{1-b} r} \mathrm{E}\left(\int_{\tau-1}^{\tau} J_c^2(s) ds\right) \\ &= \frac{\sigma^4}{-2c} r \\ &= \langle X \rangle_r \end{aligned}$$

thus,

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2}\right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{X_n(r)}{\left(\langle X_n \rangle_r\right)^{1/2}} \xrightarrow{D} \frac{X(r)}{\left(\langle X \rangle_r\right)^{1/2}} = N(0,1).$$

Local to unit root: In the case with $\alpha = 1 + c/n$, notice that $n^{-1/2} y_{[nr]} \xrightarrow{P} \int_0^r e^{c(r-s)} dB(s) = J_c(r)$ (Phillips 1987b), let

$$X_n(r) = \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1} u_t, \text{ and } X(r) = \int_0^r J_c(s) dB(s),$$

then

$$[X_n]_r = \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 u_t^2, \ \langle X_n \rangle_r = \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2$$

and

$$[X]_r = \langle X \rangle_r = \sigma^2 \int_0^r J_c(s)^2 ds,$$

thus, the martingale convergence theorem delivers the following result:

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2} \right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{X_n(r)}{\left(\langle X_n \rangle_r \right)^{1/2}} \xrightarrow{D} \frac{X(r)}{\left(\langle X \rangle_r \right)^{1/2}} = \frac{\int_0^r J_c(s) dB(s)}{\left(\sigma^2 \int_0^r J_c(s)^2 ds \right)^{1/2}} = \frac{\int_0^r W_c(s) dW(s)}{\left(\int_0^r W_c(s)^2 ds \right)^{1/2}},$$

where

$$W_c(r) = \int_0^r e^{c(r-s)} dW(s).$$

Unit root case: When $\alpha = 1$, let

$$X_n(r) = \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1} u_t, \text{ and } X(r) = \int_0^r B(s) dB(s),$$

then,

$$[X_n]_r = \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 u_t^2, \ \langle X_n \rangle_r = \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2$$

and

$$[X]_r = \langle X \rangle_r = \sigma^2 \int_0^r B(s)^2 ds,$$

by martingale convergence we obtain

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2} \right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{X_n(r)}{\left(\langle X_n \rangle_r \right)^{1/2}} \xrightarrow{D} \frac{X(r)}{\left(\langle X \rangle_r \right)^{1/2}} = \frac{\int_0^r B(s) dB(s)}{\left(\sigma^2 \int_0^r B(s)^2 ds \right)^{1/2}} = \frac{\int_0^r W(s) dW(s)}{\left(\int_0^r W(s)^2 ds \right)^{1/2}}.$$

Explosive case: If $\alpha > 1$, let

$$X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1} u_t,$$

then,

$$\langle X_n \rangle_r = \frac{1}{n^{2[nr]}} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma^2.$$

Notice that

$$\frac{y_t}{\alpha^t} = \frac{u_t}{\alpha^t} + \frac{u_{t-1}}{\alpha^{t-1}} + \cdots + \frac{u_1}{\alpha}$$

is a martingale, and

$$\sup_t E\left(\frac{y_t}{\alpha^t}\right)^2 = \sum_{s=1}^{\infty} \frac{\sigma^2}{\alpha^{2s}} = \frac{\sigma^2}{\alpha^2 - 1} < \infty,$$

by the martingale convergence theorem, as $t \to \infty$, $\alpha^{-t} y_t \overset{a.s.}{\to} Y_\alpha = y_0 + \sum_{s=1}^{\infty} \alpha^{-s} u_s$,

$$X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1} u_t = \sum_{t=1}^{[nr]} \frac{y_{t-1}}{\alpha^{t-1}} \frac{u_t}{\alpha^{[nr]-(t-1)}} \overset{a.s.}{\to} Y_\alpha U_\alpha = X(r),$$

with $U_\alpha = \sum_{s=1}^{\infty} \alpha^{-s} u'_s$ where u'_s is an i.i.d. sequence that is distributionally equivalent to u_s. The limit $X(r)$ is the product of two independent random variables Y_α and U_α, and

$$\langle X \rangle_r = Y_\alpha^2 \sum_{s=1}^{\infty} \alpha^{-2s} \sigma^2 = \frac{\sigma^2}{\alpha^2 - 1} Y_\alpha^2,$$

thus we obtain

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma^2}\right)^{1/2} (\widehat{\alpha}_r - \alpha) = \frac{X_n(r)}{(\langle X_n \rangle_r)^{1/2}} \overset{D}{\to} \frac{X(r)}{(\langle X \rangle_r)^{1/2}} = \frac{Y_\alpha U_\alpha}{\left(\frac{\sigma^2}{\alpha^2-1} Y_\alpha^2\right)^{1/2}} = \operatorname{sign}(Y_\alpha)\left(\frac{\sigma^2}{\alpha^2 - 1}\right) U_\alpha.$$

If $y_0 = 0$, and $u_t =$ i.i.d. $\mathrm{N}(0, \sigma^2)$, then Y_α and U_α are two independent $\mathrm{N}\left(0, \frac{\sigma^2}{\alpha^2-1}\right)$ random variables and the above limit becomes N(0,1).

.

The martingale convergence approach unifies the limit theory for stationary, unit root and explosive autoregressions. The above analysis immediately delivers the following results:

$$\sqrt{n}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} N(0,1-\alpha^2), \qquad \text{if } |\alpha|<1;$$

$$n^{(1+b)/2}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} N(0,-2c), \qquad \text{if } \alpha = 1 + c/n^b;$$

$$n\,(\widehat{\alpha}-\alpha) \xrightarrow{d} \int_0^1 W(s)dW(s)/\int_0^1 W(s)^2 ds, \qquad \text{if } \alpha = 1;$$

$$n\,(\widehat{\alpha}-\alpha) \xrightarrow{d} \int_0^1 W_c(s)dW(s)/\int_0^1 W_c(s)^2 ds, \qquad \text{if } \alpha = 1 + c/n;$$

$$\frac{\alpha^n}{(\alpha^2-1)}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} \frac{U_\alpha}{Y_\alpha}, \text{ where } Y_\alpha \text{ and } U_\alpha \text{ are independent}, \qquad \text{if } \alpha > 1.$$

This approach has great generality and can be used in a wide range of applications.

3 Unit Root Tests

Unit root testing is probably the most active area in unit root econometrics. A large amount of research effort on unit root studies is devoted to the development of unit root tests. A widely used method in unit root testing is the semiparametric approach proposed by Phillips (1987a). The semiparametric procedures proposed in this paper not only provide important unit root tests, but also provide a different direction of dealing with the serial correlation and endogeneity problem. Nowadays, the semiparametric unit root tests are widely used in time series applications, and the method of nonparametric correction for serial correlation and endogeneity has become an important technique in econometrics.

For a simple random walk with i.i.d. residuals u_t, the limiting distributions of the OLS estimator of α given by (2) and its t-ratio statistic $t_{\widehat{\alpha}}$ are free of nuisance parameters:

$$n(\widehat{\alpha}-1) \Rightarrow \left[\int_0^1 W(r)dW(r)\right]\left[\int_0^1 W(r)^2\right]^{-1}, \tag{14}$$

$$t_{\widehat{\alpha}} \Rightarrow \left[\int_0^1 W(r)dW(r)\right]\left[\int_0^1 W(r)^2\right]^{-1/2}. \tag{15}$$

When the residual process u_t is weakly dependent, analysis in Section 2 indicates that the limit distributions of $\widehat{\alpha}$ and $t_{\widehat{\alpha}}$ have additional bias terms due to the presence of serial correlation, and thus these two statistics can not be directly used in testing for a unit root.

Phillips (1987a) proposed a semiparametric approach to adjust $\widehat{\alpha}$ and $t_{\widehat{\alpha}}$ based on nonparametric estimates of the nuisance parameters ω and λ to account for the serial correlation. This approach is said to be semi-parametric since its treatment of the regression coefficient α is parametric but it deals with the correlation in the stationary residual u_t nonparametrically.

Notice that ω^2 and λ in (12) and (13) are nuisance parameters and may be consistently estimated by nonparametric kernel methods (e.g., see, Phillips (1987a), Andrews (1991)), let $\hat{\omega}^2$ and $\hat{\lambda}$ be such estimates, the semiparametric unit root tests can be constructed as follows:

$$Z_\alpha = n(\hat{\alpha} - 1) - \hat{\lambda}\left(n^{-2}\sum\nolimits_{t=2}^{n} y_{t-1}^2\right)^{-1} \Rightarrow \left[\int_0^1 W dW\right]\left[\int_0^1 W^2\right]^{-1},$$

$$Z_t = \hat{\sigma}_u \hat{\omega}^{-1} t_\alpha - \hat{\lambda}\left\{\hat{\omega}\left(n^{-2}\sum\nolimits_{t=2}^{n} y_{t-1}^2\right)^{1/2}\right\}^{-1} \Rightarrow \left[\int_0^1 W dW\right]\left[\int_0^1 W^2\right]^{-1/2}.$$

The limiting null distributions of the above semiparametric statistics Z_α and Z_t are the same as those given in (14) and (15).

The weak correlation in u_t can also be handled in a parametric way as Dickey and Fuller (1979, 1981) and Said and Dickey (1984). The parametric approach assumes that the autocorrelation in the stationary residual process u_t can be captured by a parametric model. In particular, if u_t is an AR(p) process, the original model (1) can be re-written as

$$\Delta y_t = a y_{t-1} + \sum\nolimits_{j=1}^{p} \varphi_j \Delta y_{t-i} + \varepsilon_t, \tag{16}$$

where ε_t are i.i.d. with $\mathrm{Var}(\varepsilon_t^2) = \sigma_\varepsilon^2$, and the unit root hypothesis corresponds to the hypothesis $a = 0$, which can be tested by means of the regression coefficient $\widehat{a}$ or its t-ratio statistic $t_{\hat{a}}$, which have the same limiting distributions as those given in (14) and (15). For more general time series processes with a valid AR(∞) representation (and fast enough convergence rate in the coefficients), we may still construct unit root tests based on autoregression (16) with $p \to \infty$ as sample size $n \to \infty$. These tests are called the augmented Dickey–Fuller (ADF) tests. The t-ratio statistic corresponding to $\widehat{a}$ can be directly used in unit root testing. A rescaled regression coefficient $\widehat{a}$ can also be used in testing for a unit root.

If we denote the long-run variance of u_t and ε_t as ω^2 and σ^2 and let $\widehat{\omega}^2$ and $\widehat{\sigma}^2$ be their consistent estimates (which can be obtained from regression (16)), the ADF coefficient test can be constructed as

$$ADF_{\alpha} = (\widehat{\omega}/\widehat{\sigma})n\widehat{a}.$$

We denote the t-ratio statistic corresponding to $\widehat{a}$ as ADF_t, under the null hypothesis of unit root, if $p \to \infty$ at a rate less than $n^{1/3}$,

$$ADF_{\alpha} \Longrightarrow \frac{\int W dW}{\int W^2},\ ADF_t \Longrightarrow \frac{\int W dW}{\left(\int W^2\right)^{1/2}},\ \text{as } n \to \infty.$$

Said and Dickey (1984) prove the validity of the ADF t-ratio test (ADF_t) in general ARMA processes of unknown order, and Xiao and Phillips (1998) studied the ADF coefficient test (ADF_{α}).

To study power properties of the unit root tests, it is useful to look at their behavior under the local alternative hypothesis $H_c : \alpha = 1 + c/n$. Phillips (1987b) shows that under the local alternative H_c, y_t behaves asymptotically like a linear diffusion: $n^{-1/2}y_{[nr]} \Rightarrow J_c(r) = \omega W_c(r)$, and the limit distributions of the unit root tests are functionals of $W_c(r)$:

$$Z_{\alpha},\ ADF_{\alpha} \Rightarrow c + \left[\int_0^1 W_c(r)dW(r)\right]\left[\int_0^1 W_c(r)^2\right]^{-1},$$

and

$$Z_t,\ ADF_t \Rightarrow c\left[\int W_c(r)dr\right]^{1/2} + \left[\int W_c(r)^2 dr\right]^{-1/2}\int W_c(r)dW(r).$$

The corrections for residual serial correlation in u_t do not lead to loss in asymptotic power.

The local asymptotic theory may be used to construct asymptotic power envelopes for unit root tests. In the simplest framework where the model is a Gaussian AR(1) with unit error variance, the Neyman–Pearson lemma can be used to construct the most powerful test of a unit root against a simple point alternative. Such a test is point optimal for a unit root at the specific point alternative that is selected. King (1988) provides a general discussion of such point optimal invariant tests, and Dufour and King (1991) developed the family of exact most powerful invariant tests. Also see Elliott, Rothenberg, and Stock (1996) for related studies.

Under the alternative that $|\alpha| < 1$, Z_{α}, $ADF_{\alpha} = O_p(n)$, and Z_t, $ADF_t = O_p(n^{1/2})$ as $n \to \infty$.

One assumption typically imposed in the literature is that the initialization y_0 has no impact on the limit theory under the unit root hypothesis. This is true if y_0 is a constant, or any random variable with finite variance, as we assumed in the beginning. The limit results hold even if the innovations

u_{-j} go back into the distant past, as long as y_0 is small relative to the sample size n, i.e. $y_0 = o_p(\sqrt{n})$. If $y_0 = O_p(\sqrt{n})$, the initialization will then affect the limiting distribution and the above limiting results will change. For example, if we consider a distant past initialization such that

$$\frac{1}{\sqrt{n}} y_0 \Rightarrow B_0(r_0) = \omega_0 W_0(r_0)$$

for some Brownian motion $B_0(\cdot) = BM(\omega_0^2)$, then

$$\frac{1}{\sqrt{n}} y_{[nr]} = \frac{1}{\sqrt{n}} y_0 + C(1) \frac{1}{\sqrt{n}} \sum_{s=1}^{[nr]} \varepsilon_s \Rightarrow B_0(r_0) + B(r),$$

and the limiting distributions of $\widehat{\alpha}$ and its t-ratio $t_{\widehat{\alpha}}$ will change correspondingly. In practice, the initial condition is not observable or estimable, thus it is hard to directly use information about the initial condition. However, the initial condition uncertainty does affect the selection of, say, a more efficient detrending procedure (see later discussions on trending time series), and researchers want to take into account of this uncertainty when selecting a detrending method. For additional discussions related to initial observations, see, e.g. Phillips (1987a), Phillips and Lee (1996), Canjels and Watson (1997), Muller and Elliott (2003), Andrews and Guggenberger (2008), Ayat and Burridge (2000), Harvey, Leybourne and Taylor (2008).

Another important condition is the condition (9) imposed on the linear process coefficients, which ensures that u_t is covariance stationary and has positive spectral density at the origin, thereby ensuring that the unit root in y_t does not cancel. In practice, if the root of the error process is close to the unit circle, unit root tests have distorted sizes. Campbell and Perron (1991) and Blough (1992) indicated that this property may affect unit root tests. Perron and Ng (1996) analyzed the local asymptotic properties and proposed some useful modifications to the semiparametric unit root tests. Faust (1996) studied this issue and concludes that the two classes of processes: $I(1)$ sequences for which $C(1) \neq 0$ and $I(0)$ sequences for which $C(1) = 0$, are nearly observationally equivalent. This near observational equivalence may be prevented by introducing a smoothness requirement on spectrum of $\widetilde{\varepsilon}_t$, which can be accomplished by appropriate summability condition on the linear process coefficients. It turns out that a strengthening of the summability condition used in Phillips and Solo (1992) to validate the BN decomposition is sufficient to rule out the pathology of $I(0)$ sequences with near $I(1)$ behavior. See related discussions and additional literature in Section 4 on bandwidth selection in KPSS test.

3.1 Unit Root Tests In The Presence of a Deterministic Trend

In practice, many observed time series display a tendency of growth. This characteristic is especially evident in time series that represent aggregate economic behavior like GDP and industrial production. To capture such secular movements in the series, deterministic trending functions (usually time polynomials) are commonly included in the above mentioned regressions.

Consider an observed time series y_t that can be written as the sum of a deterministic trend d_t and a stochastic component y_t^s :

$$y_t = d_t + y_t^s,\ t = 1, ..., n, \tag{17}$$

where $d_t = \gamma' x_t$, x_t is a k-vector of deterministic trends and $\gamma = (\gamma_0, ..., \gamma_p)'$ is a vector of trend coefficient. If the trend is assumed to be a time polynomial, $x_t = (1, t, ..., t^p)'$. In practice, the leading cases of the deterministic component are: (i) a constant $x_t = 1$; (ii) a linear time trend $x_t = (1, t)'$. To develop an asymptotic theory it is assumed that there exists a scaling matrix D_n and a limiting trend function $X(r)$ such that $D_n^{-1} x_{[nr]} \to X(r)$ as $n \to \infty$ uniformly in $r \in [0, 1]$. For example, if the trend is a p-degree time polynomial, then $D_n = \operatorname{diag}(1, n, ..., n^p)$ and $X(r) = (1, r, ..., r^p)'$.

The stochastic component y_t^s is usually modeled as

$$y_t^s = \alpha y_{t-1}^s + u_t,\ t = 1, ..., n, \tag{18}$$

where α is the largest autoregressive root, and the disturbances u_t follow a general linear process as described by (8). To take into account the deterministic trend, we may consider the following regression model

$$y_t = \beta' x_t + \alpha y_{t-1} + u_t. \tag{19}$$

The large sample theory for the coefficient estimator $\widehat{\alpha}$ in (19) and its regression t-ratio statistic $t_{\widehat{\alpha}}$ can be derived by partitioned regression using the asymptotics given earlier in this paper:

$$n(\hat{a} - 1) \Rightarrow \left[\omega^2 \textstyle\int_0^1 W_X(r) dW(r) + \lambda\right]\left[\omega^2 \textstyle\int_0^1 W_X^2(r) dr\right]^{-1},$$

and

$$t_{\hat{\alpha}} \Rightarrow \sigma_u^{-1} \left[\omega^2 \textstyle\int_0^1 W_X(r) dW(r) + \lambda\right]\left[\omega^2 \textstyle\int_0^1 W_X^2(r) dr\right]^{-1/2},$$

where σ_u^2, ω^2, and λ are the same quantities as before and $W_X(r)$ is detrended Brownian motion defined by the $L_2[0,1]$ Hilbert space projection of $W(r)$ onto the space orthogonal to the span of $X(r)$, viz.,

$$W_X(r) = W(r) - \left[\int_0^1 WX\right]\left[\int_0^1 XX\right]^{-1} X(r),$$

and $W(r)$ is standard Brownian motion. Again, let $\hat{\omega}^2$ and $\hat{\lambda}$ be consistent estimates of ω^2 and λ, and let $y_{X,t}$ be the residual from a regression of y_t on x_t, the following semiparametric tests can be formed to test the unit root hypothesis:

$$Z_\alpha = n(\hat{\alpha}-1) - \hat{\lambda}\left(n^{-2}\sum\nolimits_{t=2}^{n} y_{X,t-1}^2\right)^{-1} \Rightarrow \left[\int_0^1 W_X dW\right]\left[\int_0^1 W_X^2\right]^{-1}, \tag{20}$$

$$Z_t = \hat{\sigma}_u\hat{\omega}^{-1}t_\alpha - \hat{\lambda}\left\{\hat{\omega}\left(n^{-2}\sum\nolimits_{t=2}^{n} y_{X,t-1}^2\right)^{1/2}\right\}^{-1} \Rightarrow \left[\int_0^1 W_X dW\right]\left[\int_0^1 W_X^2\right]^{-1/2}. \tag{21}$$

Phillips and Perron (1988) developed the above semiparametric tests for the leading cases where $x_t = 1$ and $x_t = (1,t)'$. These tests are also called the Phillips-Perron tests in the literature and are widely used in econometric applications. Ouliaris et al. (1989), and Park and Sung (1994) also gave various extensions of the original semiparametric tests.

We may also introduce a deterministic trend into the ADF regression (16)

$$\Delta y_t = a y_{t-1} + \sum\nolimits_{j=1}^{k-1} \varphi_j \Delta y_{t-i} + \beta' x_t + \varepsilon_t,$$

again the unit root hypothesis can be tested based on the regression coefficient $\widehat{a}$ or its t-ratio statistic $t_{\hat{a}}$, which have the same limiting distributions as those given in (20) and (21).

The limit distributions of the detrended unit root tests depend on limiting trend functions. Numerical tabulations for the leading cases can be found in Fuller (1976/1996) and many time series textbooks. Ouliaris and Phillips (1994) provide critical values for the case of polynomial trends. These limit distributions are asymmetric and have long left tails. In the case of the Z_α test, for instance, we reject the null hypothesis of a unit root at the 5% level if $Z_\alpha < cv(Z_\alpha; 5\%)$, the 5% critical value of the test.

3.2 Unit Root Tests Based on Efficient Detrending

The discriminatory power in unit root tests between models with a root at unity and a root close to unity is generally low, and is reduced further by detrending the data. A careful comparison between (17) and (19) reveals that there are surplus trend variables in the regression equation (19). The inclusion of redundant trend variables in regression (19) ensures that the maximum trend degrees in (19) and (17) are the same, and thus the limiting result is invariant to the parameters in the trend function.

However, the redundant trend variables also lead to some inefficiency in unit root testing. A large amount of research effort has been devoted to improving the sampling performance of detrended unit root tests, including Sargan and Bhargava (1983), Shimidt and Phillips (1992), Stock (1995), and Elliot et al. (1996), Xiao (2001b).

Sargan and Bhargava (1983) suggested using the von Neumann (VN) ratio (the ratio of the sample variances of the differences and the levels of a time series) for testing the Gaussian random walk hypothesis, and Bhargava (1986) extends it to the case of a time trend. Using nonparametric estimates of the nuisance parameter ω^2 to rescale the VN ratio, a unit root test against the stationary alternative can be obtained. Stock (1995) does this for the case where there is a linear trend. Using a different approach and working with polynomial trends, Schmidt and Phillips (1992) show that for a Gaussian likelihood the Lagrange multiplier (LM) principle leads to a VN test. For the model (17) and (18), under the null hypothesis and after differences are taken, we get

$$\Delta y_t = \Delta d_t + \Delta y_t^s. \tag{22}$$

This equation is trend stationary, so that the trend function can be efficiently estimated by an OLS regression by the Grenander–Rosenblatt theorem (Grenander & Rosenblatt, 1957). Let $\Delta \hat{y}_t^s = \Delta y_t - \Delta \hat{d}_t$ be the residuals from the above detrending regression and let $\hat{y}_t^s = \sum_{s=2}^{t} \Delta \hat{y}_t^s$ be the associated estimate of y_t^s, the rescaled von Neumann ratio is given by

$$R_{VN} = \frac{\hat{\omega}^2}{\hat{\sigma}^2} \frac{n^{-1} \sum_{t=2}^{n} (\Delta \hat{y}_t^s)^2}{n^{-2} \sum_{t=1}^{n} (\hat{y}_t^s)^2} \Rightarrow \left[\int_0^1 \hat{V}_X^2 \right]^{-1}, \tag{23}$$

The limit process $\hat{V}_X(r)$ is a detrended generalized *Brownian bridge*, whose precise form depends on the deterministic trend d_t.

Another way to improve the power of unit root tests is to perform the detrending regression in a way that is efficient under the alternative hypothesis as well, an idea that was suggested in Elliot et al. (1996). To obtain large sample approximations, we consider an local alternative $\alpha = 1 + n^{-1}c$ for some fixed $c = \bar{c}$. In this case, in order to efficiently estimate the trend coefficient under the alternative hypothesis, we should use quasi-differencing in the construction of the detrending regression. Define the quasi-difference (QD) operator $\Delta_{\overline{c}}$ as $\Delta_{\bar{c}} y_t = (1 - L - n^{-1}\bar{c}L)y_t = \Delta y_t - n^{-1}\bar{c}y_{t-1}$, take quasi-differences of (17) and run the detrending regression

$$\Delta_{\bar{c}} y_t = \tilde{\gamma}' \Delta_{\bar{c}} x_t + \Delta_{\bar{c}} \tilde{y}_t^s. \tag{24}$$

Using the fitted coefficients $\tilde{\gamma}$ from the above regression, the levels data are detrended according to $\tilde{y}_t = y_t - \tilde{\gamma}' x_t$. The QD detrended data $\tilde{y}_t$ may be used in the construction of modified semiparametric unit root tests and ADF tests. Phillips and Lee (1996) show that such a regression leads to estimates of the trend coefficients that are asymptotically more efficient than an OLS regression in levels.

The QD detrending procedure involves the choice of the prespecified local parameter $\bar{c}$ that is used in the quasi-differencing, and the limit theory of unit root tests depends on $\bar{c}$. Elliot et al (1996) use a default choice of $\bar{c}$ to be the value for which local asymptotic power is 50%. Juhl and Xiao (2003) discussed the choice of local parameter in related models. If the power function of the test is $\Pi(c, \bar{c})$, where c is the true value of the local parameter under the alternative and $\bar{c}$ is the parameter used in quasi-differencing, then the power envelope is given by $\Pi(c, c)$. Juhl and Xiao (2003) suggest choosing $\bar{c}$ by

$$\min_{\bar{c}} \int_0^\infty \left[\Pi(c, c) - \Pi(c, \overline{c})\right] dc. \tag{25}$$

Notice that the power envelope is above the power function $\Pi(c, \overline{c})$, the difference between $\Pi(c, c)$ and $\Pi(c, \overline{c})$ is always non-negative. Juhl and Xiao (2003) derive power functions and envelopes of unit root tests using covariates as developed in Hansen (1995) and unit root tests based on nonnormal quasi-likelihood functions as in Rothenberg and Stock (1997). The power functions are calculated by integrating the characteristic function as in Tanaka (1996).

3.3 Unit Root Tests Against Trends With Structural Breaks

Since Nelson and Plosser (1982), many macroeconomic time series are found to contain a unit root when standard unit root tests are applied. These findings are challenged by Perron (1989) who argues that in the presence of a structural break, the standard unit root tests are biased towards the unit root hypothesis.

Trend dummies can be used to capture structural breaks in trend. If we consider the vector of polynomial trend $x_t = (1, t, \cdots, t^p)^\top$, and the vector of trend dummy given by

$$x_{tb} = \begin{cases} (0, 0, \cdots, 0)^\top, & t \in \{1, ..., b\} \\ (1, t-b, \cdots, (t-b)^p)^\top, & t \in \{b+1, ..., n\} \end{cases},$$

to account for the presence of a structural change in the trend at the data point $t = b+1$, and assume that $0 < \mu = \lim_{n\to\infty}(b/n) < 1$, ADF type tests for the presence of a unit root in models allowing for a broken trend against the alternative hypothesis of stationarity about a broken deterministic trend can be constructed based on the coefficient estimator $\widehat{a}$ and its t-ratio statistic in the following modified ADF regression

$$\Delta y_t = a y_{t-1} + \sum_{j=1}^{k-1} \varphi_j \Delta y_{t-i} + \beta' x_t + \gamma' x_{tb} + \varepsilon_t. \tag{26}$$

The limit theory for these trend break ADF statistics is similar to the traditional ADF tests, for example, the limiting distribution of trend break ADF_t test is given by

$$ADF_{tb} \Rightarrow \left[\int_0^1 W_{X_\mu} dW\right]\left[\int_0^1 W_{X_\mu}^2\right]^{-1/2},$$

where $W_{X_\mu}(r) = W(r) - \left[\int_0^1 W X_\mu\right]\left[\int_0^1 X_\mu X_\mu'\right]^{-1} X_\mu(r)$, and $X_\mu(r)$ is the limit trend function that depends on the break point μ. Similar extensions to trend breaks are possible for the other unit root tests such as the semiparametric tests. Perron (1989) considered linear trends with a single exogenous break point. Critical values of the above limiting test statistic are naturally further out in the tail than those of the corresponding conventional unit root test. As a result, it is harder to reject the null hypothesis of a unit root when the break point is introduced.

Notice that the limit theory in this case depends on the limiting trend functions that are dependent on the break point μ, in order to construct unit root tests that allow for such breaking trends, it is necessary to specify the break point b. Christiano (1992), Zivot & Andrews (1992) and Banerjee et al. (1992) argue that the assumption of a known break date is subject to the criticism of "data mining". An alternative approach is to treat the break point(s) as endogenous, and take into account this fact in the construction of the unit root tests. In this case, alternative trend break unit root tests have been suggested that endogenize the break point by choosing the value of b that gives the least favorable view of the unit root hypothesis. If $ADF_t(b)$ denotes the ADF statistic given by the t-ratio for a in the ADF regression (26), then the endogenous trend break ADF_t statistic can be constructed as

$$ADF_t(\hat{b}) = \min_{\underline{b} \leq b \leq \bar{b}} ADF_t(b), \text{ where } \underline{b} = [n\underline{\mu}],\ \bar{b} = [n\overline{\mu}] \text{ and } 0 < \underline{\mu} < \overline{\mu} < 1, \tag{27}$$

and $[\cdot]$ signifies the integer part of its argument. The limit theory for this trend break ADF statistic is given by

$$ADF_t(\hat{b}) \Rightarrow \inf_{\mu \in [\underline{\mu}, \overline{\mu}]} \left[\int_0^1 W_{X_\mu} dW \right] \left[\int_0^1 W_{X_\mu}^2 \right]^{-1/2}. \tag{28}$$

In the presence of endogenous breaks, critical values of the limiting test statistic (28) are even further out in the tail than those of the exogenous trend break tests, and in practice it is harder to reject the null hypothesis of a unit root when the break point is considered to be endogenous. Asymptotic and finite sample critical values for the endogenized trend break ADF_t unit root test are given in Zivot and Andrews (1992).

The trend break tests can be naturally extended to allow for multiple break points in the sample and in the limit process without affecting the general theory. See, inter alia, Lumsdaine and Papell (1997), Bai and Perron (1998), Hansen (2001), Lee and Strazicich (2003), Kapetanois (2005).

Structural break models attribute the persistency in time series data to the effect of structural changes at particular times in the sample period (so that it can be parameterized by dummy variables). Unit root models have persistent shocks throughout the entire history of the process. Such a difference makes it possible to test for a unit root model against stationary time series with a finite number of structural breaks. Simulations studies indicate that the introduction of trend break functions leads to further reductions in the power of unit root tests and to substantial finite sample size distortion in the tests. Sample trajectories of a random walk are often similar to those of a process that is stationary about a broken trend for some particular breakpoints. Phillips (1998, 2002) explores these issues in a systematic way.

Phillips (1998) shows that Brownian motion can be represented as an infinite linear random combination of deterministic functions of time. He argues that carefully chosen trend stationary models can always be expected to provide reasonable representations of given random walk data, but such models are certain to fail in post sample projections as the post sample data drifts away from the final trend line. In particular, for a Brownian motion $B(r)$, we have the following L_2 -representation (the Loeve-Karhunen expansion):

$$B(r) = \omega\sqrt{2}\sum_{k=1}^{\infty}\frac{\sin\left[\left(k-1/2\right)\pi r\right]}{(k-1/2)\pi}\xi_k = \omega\sum_{k=1}^{\infty}\varphi_k(r)\xi_k, \tag{29}$$

where the components ξ_k are independently and identically distributed (iid) as $N(0,1)$ and the functions $\varphi_k(r)$ form an orthogonal set in $L_2[0,1]$. Consequently, empirical regressions of y_t on $\varphi_{K,t} = \left(\varphi_1\left(\frac{t}{T}\right), ..., \varphi_K\left(\frac{t}{T}\right)\right)'$ accurately reproduce in the limit the first K terms of the expansion (29). Further, when $K \to \infty$ and $K/T \to 0$ as $T \to \infty$, such regressions succeed in reproducing the entire representation (29). It follows that these deterministic functions are capable of successfully representing a nonstationary time series like y_t in the limit as $T \to \infty$. Such regressions on deterministic functions then become an alternate way of modelling a nonstationary time series. Consequently, one might mistakenly "reject" a unit root model in favor of a trend "alternative" when in fact the alternative model is nothing but an alternative representation of the unit root process itself.

Phillips (2002) shows that unit root tests which involve deterministic functions will inevitably lead to the rejection of the unit root hypothesis when $K, T \to \infty$, and in this sense the conventional critical values used in unit root tests are invalid asymptotically when the competing deterministic functions that appear in the maintained hypothesis provide an alternative mechanism of modelling the nonstationarity, as in cases like polynomial trends and trend break polynomials they will.

From a practical point of view, both the unit root models and stationary time series models with structural breaks are useful alternative models to capture the empirical features in macroeconomic time series. In practice, important external shocks such as institutional events may have substantial impact on the dynamics of a time series. The number of such shocks is usually small and it is often of interest to identify the break points endogenously. If we allow for a large number of breaks in a sample, the distinction between a unit root series and a stationary time series with broken trends is less clear in empirical time series analysis.

3.4 Nonparametric Unit Root Tests

It is possible to construct nonparametric tests for a unit root, such tests having relative advantages against nonlinear/nonparametric alternatives. Peter has made a fundamental contribution in developing nonparametric and nonlinear methods for unit root models.

Unit root tests constructed in the frequency domain have some nonparametric features. Choi and Phillips (1993) consider a frequency domain regression of the AR model and constructed a unit root test based on the frequency domain estimates of α and its t-ratio statistic. An important feature of the frequency domain estimator is that no serial correlation correction is needed. The limiting distribution of the AR estimator based on a first order autoregression of the Fourier transformation is free of nuisance parameters.

Another direction to develop nonparametric tests for a unit root is to consider the class of processes that are fractionally integrated:

$$(1-L)^d y_t^s = u_t, \tag{30}$$

where d may be fractional and the operator $(1-L)^d$ is defined by the formal binomial expansion

$$(1-L)^d = 1 + \sum_{j=1}^{\infty} \frac{(-d)_j}{j!} L^j, \quad (a)_j = (-a)(-a+1), ..., (-a+j-1) \tag{31}$$

whose convergence properties depend on the value of d. Within this family it is possible to test for 'unit root' nonstationarity by estimating d and test the null hypothesis $d = 1$ against the alternative $d < 1$. Shimotsu and Phillips (2005) propose an exact local Whittle estimator $\widehat{d}$ for d that is consistent for all d and show that $\sqrt{n}(\widehat{d} - d) \Rightarrow N(0, 1/4)$. This estimator can then be used to test the unit root hypothesis $\mathrm{H}_0 : d = 1$ against alternatives such as $\mathrm{H}_1 : d < 1$.

Peter also made important contributions to asymptotic analysis of nonlinear or nonparametric transformation on unit root processes. Park and Phillips (1998), Wang and Phillips (2006) studied asymptotic theory for nonstationary density estimation and kernel regression with unit root processes.

Consider the following first order autoregression model

$$y_t = m(y_{t-1}) + u_t, \ t = 1, ..., n, \tag{32}$$

where u_t are i.i.d. $(0, \sigma^2)$ random variables satisfying appropriate regularity assumptions, the unit root hypothesis corresponds to $m(y_{t-1}) = y_{t-1}$. If we estimate $m(x)$ using the following nonparametric kernel smoother:

$$\widehat{m}(x) = \frac{\sum_t K_h(x - y_{t-1}) y_t}{\sum_t K_h(x - y_{t-1})}$$

where $K_h(u) = h^{-1}K(u/h)$, $K(\cdot)$ is a kernel function and h is the bandwidth parameter, under the unit root hypothesis and regularity assumptions, Park and Phillips (1998), Wang and Phillips (2006) show that $\widehat{m}(x)$ is a consistent estimator, but the convergence rate is slower than that in stationary nonparametric autoregression. In particular,

$$\left[h \sum_t K_h(x - y_{t-1})\right]^{1/2} (\widehat{m}(x) - m(x)) \Rightarrow N\left(0, \sigma^2 \int K(u)^2 du\right).$$

Nonparametric tests for the null of a unit root against nonlinear or nonparametric alternatives may be developed based on the asymptotic analysis of Phillips and Park (1998), Wang and Phillips (2006). If we rewrite (32) as

$$\Delta y_t = f(y_{t-1}) + u_t,$$

the unit root hypothesis corresponds to H_0: $f(y_{t-1}) = 0$. Nonparametric unit root tests can then be constructed based on kernel estimates of $f(y_{t-1})$. For example, notice that $\mathrm{E}(\Delta y_t f(y_{t-1})) = E\left(f(y_{t-1})^2\right) > 0$ if $f(y_{t-1}) \neq 0$, replacing the expectation by sample average and $f(y_{t-1})$ by its kernel estimate, we may construct the following density-weighted nonparametric statistic

$$U_n = \sum_t \sum_{s \neq t} K_h(y_{s-1} - y_{t-1}) \Delta y_s \Delta y_t$$

and, under the unit root hypothesis

$$n^{-3/4}\sqrt{h}U_n \Rightarrow N\left(0, \sigma^6 \int_0^1 K(u)^2 du \int_0^1 L(\tau, W(\tau)) d\tau\right)$$

where $L(\tau, v) = \lim_{r \to 0^+} \int_0^t 1(|W(s) - v| \leq r) \, ds$ is a local time. Let

$$V_n = n^{-3/2} h \sum_{t=2}^{T} \sum_{s=1}^{t-1} K_h^2(y_{s-1} - y_{t-1}) (\Delta y_s)^2 (\Delta y_t)^2$$

then

$$n^{-3/4}\sqrt{h}V_n^{-1/2}U_n \Rightarrow N(0,1),$$

providing a unit root test for the simple AR(1) model. Also see Juhl (2008) for a similar test in the predictive regression model.

3.5 Peter's Shortcut and Robust Inference on Unit Roots

Phillips (1991) and Phillips (1995a) propose a "shortcut" to asymptotic analysis of estimators with nonsmooth criterion function. The approach treats nonsmooth objective criterion functions as generalized functions and uses generalized Taylor series expansions to represent their local behavior. This approach provides a convenient way to construct unit root tests based on robust regressions such as Least Absolute Deviation (LAD) regression.

To illustrate Peter's shortcut, we may consider, say, the following ADF regression

$$\Delta y_t = \mu + \rho y_{t-1} + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + u_t \tag{33}$$

where u_t =i.i.d.$(0,\sigma^2)$, with CDF $F(\cdot)$. If we consider a quantile regression estimation of the above model:

$$\min \sum_t \rho_\tau(\Delta y_t - \mu - \rho y_{t-1} - \sum_{j=1}^{p} \beta_j \Delta y_{t-j}) \tag{34}$$

the asymptotic behavior of the unit root quantile regression can be analyzed using Peter's shortcut, and unit root tests can be constructed based on such estimators. Let

$$\theta = (\mu, \rho, \beta_1, \cdots, \beta_p)^\top,\ x_t = (1, y_{t-1}, \Delta y_{t-1}, \cdots, \Delta y_{t-p})^\top,\ D_n = diag(\sqrt{n}, n, \sqrt{n}, \cdots, \sqrt{n})$$

then the unit root quantile regression may be written as

$$\widehat{\theta}(\tau) = \arg\min_\theta \sum_t \rho_\tau(\Delta y_t - \theta' x_t).$$

Denote

$$\eta = \eta(\tau) = D_n(\theta - \theta(\tau)), \widehat{\eta}(\tau) = D_n\left(\widehat{\theta}(\tau) - \theta(\tau)\right),$$

$$\theta(\tau) = (\mu + F^{-1}(\tau), \rho, \beta_1, \cdots, \beta_p)^\top,$$

and define

$$Z_n(\eta) = \sum_t \left\{ \rho_\tau(u_{t\tau} - D_n^{-1} x_t^\top \eta) - \rho_\tau(u_{t\tau}) \right\}$$

then

$$\widehat{\eta} = \arg\min Z_n(\eta).$$

Denote $w_t = \Delta y_t$, and $\psi_\tau(u) = \tau - I(u < 0)$, under regularity assumptions :

$$n^{-1/2} \sum_{t=1}^{[nr]} (w_t, \psi_\tau(u_{t\tau}))^\top \Rightarrow (B_w(r), B_\psi^\tau(r))^\top = BM(0, \underline{\Sigma}(\tau))$$

where

$$\underline{\Sigma}(\tau) = \begin{bmatrix} \sigma_w^2 & \sigma_{w\psi}(\tau) \\ \sigma_{w\psi}(\tau) & \sigma_\psi^2(\tau) \end{bmatrix}$$

Following Phillips (1991, 1995a), under regularity conditions, notice that $\rho_\tau(u)$ can be treated as a generalized function with a smooth regular sequence

$$\rho_{\tau m}(u) = \int_{-\infty}^{\infty} \rho_\tau(v) S\left[m(v-u)\right] m e^{-v^2/m^2} dv$$

where $S(\cdot)$ is a smudge function whose role in $\rho_{\tau m}(u)$ is to smudge out $\rho_\tau(v)$ when v is outside the interval $(u - m^{-1}, u + m^{-1})$ (see Phillips (1995a) for more discussions on smudge function and related literature), then $Z_n(\eta)$ is a generalized process defined by the following regular sequence of processes

$$Z_{nm}(\eta) = \sum_t \left\{ \rho_{\tau m}(u_{t\tau} - D_n^{-1} x_t^\top \eta) - \rho_{\tau m}(u_{t\tau}) \right\}.$$

Notice that $\rho_\tau(u)$ has first derivative everywhere except $u = 0$ and

$$\dot{\rho}_\tau(u) = \tau - \frac{1}{2} + \frac{1}{2}\mathrm{sign}(u)$$

has a regular sequence $\dot{\rho}_{\tau m}(\cdot) = \tau - \frac{1}{2} + \frac{1}{2}\mathrm{sign}_m(\cdot)$, where

$$\mathrm{sign}_m(u) = \int_{-\infty}^{\infty} \mathrm{sign}(v) S\left[m(v-u)\right] m e^{-v^2/m^2} dv.$$

In addition, $\rho_\tau(u)$ has second order derivative everywhere except $u = 0$ and

$$\ddot{\rho}_\tau(u) = \frac{1}{2}\frac{d\mathrm{sign}(u)}{du} = \delta(u) = \text{ Dirac delta function}$$

has a regular sequence $\ddot{\rho}_{\tau m}(\cdot) = (m/\pi)^{1/2}e^{-mu^2} = \delta_m(u)$. By Taylor expansion of $Z_{nm}(\eta)$ around $\eta = 0$, we have

$$
\begin{aligned}
Z_{nm}(\eta) &= -\sum_t \dot{\rho}_{\tau m}(u_{t\tau})D_n^{-1}x_t^\top \eta \\
&\quad +\frac{1}{2}\eta^\top \left[\sum_t \ddot{\rho}_{\tau m}(u_{t\tau} - \lambda D_n^{-1}x_t^\top \eta)D_n^{-1}x_t x_t^\top D_n^{-1}\right]\eta \qquad (35)
\end{aligned}
$$

where $\dot{\rho}_{\tau m}(\cdot)$ and $\ddot{\rho}_{\tau m}(\cdot)$ are first and second order derivatives of $\rho_{\tau m}(\cdot)$ and $\lambda \in (0,1)$.

Following a similar argument as Example 3.5 in Phillips (1995a), it can be shown that

$$\sum_t \dot{\rho}_{\tau m}(u_{t\tau})D_n^{-1}x_t^\top \eta \Rightarrow \xi_m^\top \eta$$

where

$$\xi_m \Rightarrow \xi = \begin{bmatrix} \int \overline{B}_y(r)dB_\psi^\tau(r) \\ \mathrm{N}(0, \tau(1-\tau)\Sigma_p) \end{bmatrix}, \text{ as } m \to \infty.$$

For the second term on right hand side of (35), notice that the regular sequence $\delta_m(\cdot)$ is differentiable and has bounded derivative,

$$\left|\ddot{\rho}_{\tau m}(u_{t\tau} - \lambda D_n^{-1}x_t^\top \eta) - \ddot{\rho}_{\tau m}(u_{t\tau})\right| \le C_m \left|D_n^{-1}x_t^\top \eta\right|$$

thus

$$
\begin{aligned}
&\left|\sum_t \left[\ddot{\rho}_{\tau m}(u_{t\tau} - \lambda D_n^{-1}x_t^\top \eta) - \ddot{\rho}_{\tau m}(u_{t\tau})\right]\eta^\top D_n^{-1}x_t x_t^\top D_n^{-1}\eta\right| \\
\le\ & C_m \sum_t \left|D_n^{-1}x_t^\top \eta\right| \left|\eta^\top D_n^{-1}x_t x_t^\top D_n^{-1}\eta^\top\right| \\
\to\ & 0, \text{ uniformly over } \eta \text{ in compact sets.}
\end{aligned}
$$

In addition,

$$\sum_t \ddot{\rho}_{\tau m}(u_{t\tau})D_n^{-1}x_t x_t^\top D_n^{-1} \Rightarrow \left(\mathrm{E}\left[\ddot{\rho}_{\tau m}(u_{t\tau})\right]\right)\Pi$$

where

$$\Pi = \begin{bmatrix} \int \overline{B}_y(r)\overline{B}_y(r)^\top dr & 0 \\ 0 & \Sigma_p \end{bmatrix}.$$

Notice that $\mathrm{E}[\ddot{\rho}_{\tau m}(u_{t\tau})]$ is an ordinary function as long as density at $F^{-1}(\tau)$ exists and continues, and as $m \to \infty$,

$$\mathrm{E}\left[\ddot{\rho}_{\tau m}(u_{t\tau})\right] \to f(F^{-1}(\tau)) \ .$$

Then

$$Z_{nm}(\eta) \Rightarrow Z_{\cdot m}(\eta) = -\xi_m^\top \eta + \left(\mathrm{E}\left[\ddot{\rho}_{\tau m}(u_{t\tau})\right]\right) \frac{1}{2} \eta^\top \Pi \eta,$$

and as $m \to \infty$, over η on a compact set,

$$Z_{\cdot m}(\eta) \Rightarrow Z(\eta) = -\xi^\top \eta + f(F^{-1}(\tau)) \frac{1}{2} \eta^\top \Pi \eta,$$

Notice that (1) $Z_{\cdot m}(\eta)$ is convex; (2) $Z_{\cdot m}(\eta) \Rightarrow Z(\eta)$; (3) $Z(\eta)$ has an unique minimum at

$$\frac{1}{f(F^{-1}(\tau))} \Pi^{-1} \xi = \arg\min \left[-\xi^\top \eta + f(F^{-1}(\tau)) \frac{1}{2} \eta^\top \Pi \eta \right],$$

we have

$$D_n \left(\widehat{\theta}(\tau) - \theta(\tau) \right) \Rightarrow \frac{1}{f(F^{-1}(\tau))} \left[\begin{array}{c} \left[\int \overline{B}_w \overline{B}_w{}^\top \right]^{-1} \int \overline{B}_w dB_\psi^\tau \\ \mathrm{N}\big(0, \tau(1-\tau)\Sigma_p^{-1}\big) \end{array} \right]$$

where $\overline{B}_w(r) = [1, B_w(r)]^\top$, $\mathrm{N}\big(0, \tau(1-\tau)\Sigma_p^{-1}\big)$ is a p-dimensional normal variate with covariance matrix $\tau(1-\tau)\Sigma_p^{-1}$, and is independent with $\int_0^1 \overline{B}_w dB_\psi^\tau$. In particular

$$n\widehat{\rho}(\tau) \Rightarrow \frac{1}{f(F^{-1}(\tau))} \left[\int_0^1 \underline{B}_w^2 \right]^{-1} \int_0^1 \underline{B}_w dB_\psi^\tau,$$

where $\underline{B}_w(r) = B_w(r) - \int_0^1 B_w$ is a demeaned Brownian Motion. Similarly, the corresponding t-ratio statistic has the following limiting distribution

$$t_n(\tau) \Rightarrow \frac{1}{\sqrt{\tau(1-\tau)}} \left[\int_0^1 \underline{B}_w^2 \right]^{-1/2} \int_0^1 \underline{B}_w dB_\psi^\tau. \tag{36}$$

For related literature, also see Knight (1989, 1998), Pollard (1991).

The above asymptotic results can be immediately used to construct unit root tests based on a quantile regression. Like the conventional augmented Dickey-Fuller (ADF) t-ratio test, we may consider the corresponding t-ratio statistic $t_n(\tau)$. Following Phillips and Hansen (1990) and Phillips (1995b) we have the following decomposition:

$$\int_0^1 \underline{B}_w dB_\psi^\tau = \int \underline{B}_w dB_{\psi.w}^\tau + \lambda_{\omega\psi}(\tau) \int \underline{B}_w dB_w,$$

where $\lambda_{\omega\psi}(\tau) = \sigma_{w\psi}(\tau)/\sigma_w^2$ and $B_{\psi.w}^\tau$ is a Brownian motion with variance $\sigma_{\psi.w}^2(\tau) = \sigma_\psi^2(\tau) - \sigma_{w\psi}^2(\tau)/\sigma_w^2$, and is independent of $\underline{B}_w$. The limiting distribution of $t_n(\tau)$ can therefore be decomposed as

$$\delta \left(\int_0^1 \underline{W}_1{}^2 \right)^{-1/2} \int_0^1 \underline{W}_1 dW_1 + \sqrt{1-\delta^2} N(0,1), \tag{37}$$

where

$$\delta = \frac{\sigma_{w\psi}(\tau)}{\sigma_w \sigma_\psi(\tau)} = \frac{\sigma_{w\psi}(\tau)}{\sigma_w \sqrt{\tau(1-\tau)}}.$$

Thus the limiting distribution of $t_n(\tau)$ is a mixture of the well-known Dickey-Fuller distribution and a standard normal distribution (which is independent with the DF distribution), with the weight δ being simply the long-run correlation coefficient between $\{w_t\}$ and $\{\psi_\tau(u_{t\tau})\}$. The limiting distribution (37) is the same as that of the covariate-augmented Dickey-Fuller (CADF) test of Hansen (1995).

Koenker and Xiao (2004) studied unit root tests based on the above quantile regression. In the special case $\tau = 0.5$, this gives the unit root tests based on LAD regression. Herce (1996) studied semiparametric unit root tests based on LAD regressions. Hasan and Koenker (1997) developed rank test of unit roots based on quantile regression.

4 Testing Stationarity Against the Unit Root Alternative

Evidence has been accumulated that many economic and financial time series contain a unit root. Kwiatkowski, Phillips, Schmidt, and Shin (1992) argued that most standard testing procedures consider the null hypothesis of a unit root, which ensures that it is accepted unless there is strong evidence against it. Therefore, it is of considerable interest to propose tests of the null hypothesis that the time series is trend stationary.

Kwiatkowski, Phillips, Schmidt, and Shin (1992) propose a LM test for the null of trend stationarity against the alternative of a unit root. The original test was derived based on a *component* model where the observed time series y_t can be decomposed into a deterministic trend $\beta' x_t$, a (unit root) stochastic trend y_t^s, and a stationary residual v_t:

$$y_t = \beta' x_t + y_t^s + v_t, \quad y_t^s = y_{t-1}^s + u_t. \tag{38}$$

The null hypothesis that y_t is trend stationary then corresponds to the case $\sigma_u^2 = \mathrm{var}(u_t) = 0$. Under Gaussian assumptions and iid error conditions, a LM test for the null of trend stationarity against the alternative of a unit root can be constructed from testing H_0: $\sigma_u^2 = 0$ against the one-sided alternative H_1: $\sigma_u^2 > 0$. Let $\widehat{e}_t$ be the residuals from the regression of y_t on the deterministic trend x_t and $\widehat{\sigma}_v^2 = n^{-1} \sum \widehat{e}_t^2$, then the LM statistic can be constructed as follows:

$$LM = \frac{1}{n^2 \widehat{\sigma}_v^2} \sum_t S_t^2,$$

where $S_t = \sum_{j=1}^t \widehat{e}_j$. Under the null hypothesis of trend stationarity, $LM \Rightarrow \int_0^1 V_X^2$, where

$$V_X(r) = W(r) - \left[\int_0^r X'\right] \left[\int_0^1 XX'\right]^{-1} \left[\int_0^1 X dW\right]$$

is a generalized Brownian bridge process.

In the case that v_t is a general stationary residual with long-run variance ω_v^2, consistent test for stationarity can still be obtained by replacing the variance estimator $\widehat{\sigma}_v^2$ by a long run variance estimator $\widehat{\omega}_v^2$:

$$KPSS = \frac{1}{n^2 \widehat{\omega}_v^2} \sum S_t^2. \tag{39}$$

The KPSS test is widely used in time series applications. Intuitively, if y_t is a stationary time series around a deterministic function, it fluctuates around the deterministic trend function and cannot grow indefinitely. However, an unstable (unit root) or explosive process has unbounded variance and grows over long period of time. As a result, the fluctuation of a unit root or explosive process is much larger than that of a stationary process. Xiao (1999) proposes testing trend stationarity by looking at the fluctuation in the detrended time series. If we look at the fluctuation from the partial sum process of $\widehat{e}_t$:

$$U_n(r) = \frac{1}{\widehat{\omega}_v \sqrt{n}} \sum_{t=1}^{[nr]} \widehat{e}_t, \tag{40}$$

and use a continuous functional $h(\cdot)$ (say, the Kolmogoroff-Smirnoff type or Cramer-von Mises type measures) to measure the fluctuation, we can construct $h(U_n(r))$ as a test statistic for trend stationarity against the unit root alternative. If we choose $h(\cdot)$ as the Cramer-von-Mises measure,

$$h(U_n(r)) = \int_0^1 U_n(r)^2 dr,$$

we obtain the KPSS test.

The KPSS test can be applied and extended along various directions: In particular, the KPSS test can be extended to different models, different estimation methods, and alternative measurement of fluctuation. Leybourne and McCabe (1994) suggested a similar test for stationarity which differs from the test of Kwiatkowski et al. (1992) in its treatment of autocorrelation and applies when the null hypothesis is an AR(k) process.

The problem is particularly delicate and interesting in the multivariate case, where several time series may have nonstationary characteristics and we are interested in testing the cointegration relationship among these I(1) variables. To test the null hypothesis of cointegration[2], using a component representation again and applying the KPSS test to the residual process of a cointegrating regression, Shin (1994) proposed a residual-based test. Xiao and Phillips (1998) show that the conventional CUSUM test for structural change can be applied to cointegrating regression residuals, leading to a consistent residual-based test for the null hypothesis of cointegration. In this residual based cointegration test, nonparametric corrections are used to remove nuisance parameters associated with serial correlation and endogeneity and the test is constructed based on the fully modified (Phillips and Hansen 1991) residuals. Again, like the univariate unit root tests and stationarity tests, the limiting distributions of these cointegration tests are nonstandard, depending on not only the deterministic trend function, but also the dimension of regressors. Consequently, critical values are dependent on

[2] Another important contribution of Peter is on testing the null of non-cointegration against the alternative of cointegration. Unit root tests can be used to test the null of non-cointegration using residual based approaches. The tests are used in the same way as standard unit root tests and have the same null hypothesis, but the data are the residuals from a least squares cointegrating regression, and the alternative hypothesis (of cointegration) is now the main hypothesis of interest (Engle and Granger, 1987; Phillips and Ouliaris, 1990). The limit theory for these residual based tests was developed in Phillips and Ouliaris (1990).

the trend function and dimension parameter and have to be generated by Monte Carlo experiments.

Extensive Monte Carlo experiments have been conducted to evaluate the finite sample performance of the aforementioned stationarity and cointegration tests. A well-documented fact is that the sampling performance of these stationarity tests is sensitive to the bandwidth selection in the long-run variance estimation.

In constructing the KPSS statistic given by (39), ω_v^2 is usually estimated by the nonparametric kernel method and has the following form

$$\widehat{\omega}_v^2 = \sum_{h=-q}^{q} k\left(\frac{h}{q}\right)\widehat{\gamma}_{vv}(h), \tag{41}$$

where $\widehat{\gamma}_{vv}(h)$ is the h-th order sample autocovariance of v_t, $k(\cdot)$ is a kernel function, and q is the bandwidth (lag truncation) parameter satisfying the property that $q \to \infty$ and $q/n \to 0$ as the sample size $n \to \infty$. The KPSS test is consistent because it diverges to ∞ at rate n/q under the unit root alternative.

A popular bandwidth choice in long-run variance estimation is the data-dependent plug-in bandwidth

$$q = \mu_k \widehat{\delta}(f,k) n^{1/(2p+1)}, \tag{42}$$

where μ_k is a constant associated with the kernel function, $\delta(f,k)$ is a function of the unknown spectral density and is estimated using a plug-in (usually AR(1) plug-in) method, and p is the characteristic exponent of the kernel. This bandwidth choice has been studied by Andrews (1991) in the context of estimation of a covariance matrix for stationary time series. Such a bandwidth has the advantage that it partially adapts the serial correlation in the underlying time series through the data-dependent component $\widehat{\delta}(f,k)$. However, such a bandwidth choice is not appropriate in distinguishing between I(0) and I(1) processes.

To test the I(0) hypothesis against an I(1) alternative, the bandwidth q should be large enough to capture the short-range dependence under the null. On the other hand, it should not be too large that it also captures dependence under the unit root alternatives. Unfortunately, the data-dependent bandwidth (42) not only captures the short-range dependence under the null, but also captures the dependence under the alternatives. This is reflected on the value of the plug-in component $\widehat{\delta}(f,k)$ in

(42). As the temporal dependence increases, the value of the data-dependent $\widehat{\delta}(f,k)$ becomes larger and larger. For example, if we use the popular AR(1) plug-in, the magnitude of the bandwidth value is determined by the first order autoregression coefficient $\widehat{\rho} = (\sum_t X_{t-1}^2)^{-1}(\sum_t X_{t-1}X_t)$, and

$$q = \mu_k \widehat{\delta} n^{1/3},\ \widehat{\delta} = \left[\frac{2\widehat{\rho}}{1-\widehat{\rho}^2}\right]^{2/3}, \text{ for first order kernels,} \tag{43}$$

and

$$q = \mu_k \widehat{\delta} n^{1/5},\ \widehat{\delta} = \left[\frac{2\widehat{\rho}}{(1-\widehat{\rho})^2}\right]^{2/5}, \text{ for second order kernels.} \tag{44}$$

Under the unit root alternative, $\widehat{\rho} - 1 = O_p(n^{-1})$. Thus $q = O_p(n)$, and $n/q \nrightarrow \infty$, consequently the KPSS test based on the data-dependent bandwidth will be inconsistent.

Bandwidths as fixed functions of n such as $[8(n/100)^{1/3}]$ or $[12(n/100)^{1/4}]$ provide consistent tests. For many fixed functions of n, they generally have reasonable performance when the short-term memory is strong, but have poor performance if the short-range dependence is weak. The data-dependent bandwidth (42) is able to pick up the weak dependence in data under the null and generally gives better size properties than other choices, but it also captures the strong dependency under the unit root alternative and reduces the power to inconsistency. Xiao (1998) suggests a partially data-dependent bandwidth choice: the data-dependent plug-in bandwidth $\mu_k \widehat{\delta(f,k)} n^{1/(2q+1)}$ coupled with an upper bound. The upper bound is a fixed increasing function of the sample size n. Thus, the suggested bandwidth is partially data-dependent and has the following form:

$$M^* = \min\{\mu_k \widehat{\delta(f,k)} n^{1/(2q+1)}, B(n)\},$$

where $B(n)$ is an upper bound function, say $[8(n/100)^{1/4}]$. When the serial correlation is weak, $\mu_k \widehat{\delta(f,k)} n^{1/(2q+1)}$ generally has a smaller value than $B(n)$ and M^* is determined by the data-dependent formula $\mu_k \widehat{\delta(f,k)} n^{1/(2q+1)}$ and gives better size than fixed bandwidth or fixed functions of n. Under the alternative hypothesis, $\widehat{\delta(f,k)}$ is generally very large. In this case the upper bound function $B(n)$ prevents M^* from being too big and thus retains reasonable power. Similar problems exist in other types of stationarity tests, residual based tests for the null of cointegration, and testing for stationarity against long memory alternatives. See Xiao (2003), Xiao and Phillips (2002), Sul, Phillips and Choi (2005), Harris, Leybourne, and McCabe (2007) for related discussions on this issue.

The power problem and the associated bandwidth selection problem in KPSS test are natural consequences that the unit root ($\alpha = 1$) is on the boundary of the stationary range $|\alpha| < 1$, and are again related to the observational equivalence issue mentioned in Section 3. The KPSS test and other stationarity tests (against the unit root alternative) are consistent tests under regularity conditions that ensure invariance principles to hold and long-run variance be consistently estimated. The regularity assumptions are typically used in the literature and are sufficient, but not necessary, for, say, the invariance principles. Without sufficient restrictions on the model, it is impossible to consistently discriminate between I(0) and I(1) processes. For example, Pötscher (2002) shows that the minimax risk for estimating the value of the long-run variance is infinite. See Faust (1996) and Müller (2008) for additional related discussion on this issue.

5 Conclusion

Peter has made the most important contributions to the development of unit root econometrics. Unit root theory plays a major role in modern time series econometrics and weak convergence methods and function space asymptotics have opened up the econometric analysis of nonstationary regression models. This paper is only a very selective review of Peter's contribution on unit roots. Lots of other important topics, including nonlinear methods in unit root models, Bayesian inference in unit root models, etc, are not discussed in this paper. In addition, Peter has a lot of ongoing research on unit roots and we look forward to seeing this work in the literature.

6 Reference

Andrews and Guggenberger, 2008, "Asymptotics for Stationary Very Nearly Unit Root Processes," Journal of Time Series Analysis (2008), 29(1): 203-210.

Andrews, D.W.K., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, **59**, 817–858.

Ayat, L. and P. Burridge (2000), Unit root tests in the presence of uncertainty about the nonstochastic trend, Journal of Econometrics, 95, 71-90.

Bai, J., and P. Perron (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," Econometrica, vol. 66(1), pages 47-78.

Banerjee, A., R.L. Lumsdaine and J.H. Stock, 1990, "Recursive and sequential tests for a unit root: Theory and international evidence," manuscript, Harvard University, Cambridge, MA.

Beveridge, S. and R. Nelson, 1981, "A new approach to decomposition of time series in permanent and transitory components with particular attention to measurement of the 'business cycle'," *Journal of Monetary Economics*, **7**, 151–174.

Bhargava, A., 1986, "On the theory of testing for unit roots in observed time series," *Review of Economic Studies*, **53**, 369-384.

Blough, S.R., 1992, "The relationship between power and level for generic unit root tests in finite samples," *Journal of Applied Econometrics*, **7**, 295–308.

Campbell, J.Y. and Perron, P. 1991. Pitfalls and opportunities: what macroeconomists should know about unit roots (with discussion). In NBER Macroeconomics Annual 1991, ed. O.J. Blanchard and S. Fischer. Cambridge, MA: MIT Press.

Canjels, N. And M. Watson, 1997, "Estimating deterministic trends in the presence of serially correlated errors," *Review of Economics and Statistics, 79,* 184-200.

Chan, N.H. and C.Z Wei, 1987, "Asymptotic inference for nearly nonstationary AR(1) processes," *Annals of Statistics*, **15**, 1050–1063.

Chan N. H. & C. Z. Wei, 1988, "Limiting distributions of least squares estimates of unstable autoregressive processes," *Annals of Statistics*, **16**, 367–401.

Choi, I., and P.C.B. Phillips, 1993, Testing for a unit root by frequency domain regression, Journal of Econometrics, 59, 263-286.

Christiano, L.J. 1992. Searching for a break in GNP. Journal of Business & Economic Statistics 10, 237–50.

Dickey, D.A. and W.A. Fuller, 1979, "Distribution of estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, D.A. & W.A. Fuller, 1981, "Likelihood ratio tests for autoregressive time series with a unit root," *Econometrica*, **49**, 1057–1072.

Dhrymes, P., Time series, unit roots and cointegration, 1998.

Dufour, J. and M. L. King, 1991, "Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary AR(1) errors," *Journal of Econometrics*, **47**, 115–143.

Elliot, G., T. J. Rothenberg and J. H. Stock, 1996, "Efficient tests for an autoregressive unit root," *Econometrca*, **64**, 813–836.

Engle, R.F. and C.W.J. Granger, 1987, "Cointegration and error correction: representation, estimation and testing," *Econometrica*, **55**, 251–276.

Faust, J., 1996. Near observational equivalence and theoretical size problems with unit root tests. Econometric Theory, 12, 724-731.

Giraitis and P.C.B. Phillips, 2006, "Uniform Limit Theory for Stationary Autoregression" (with Liudas Giraitis), Journal of Time Series Analysis, 27(1): 51-60.

Grenander, U. and M. Rosenblatt, 1957, *Statistical Analysis of Stationary Time Series.* New York: John Wiley.

Hall, P. and C.C. Heyde, 1980, *Martingale Limit Theory and its Applications.* New York: Academic Press.

Hansen, B., 1995, Rethinking the Univariate Approach to Unit Root Tests: How to Use Covariates to Increase Power, Econometric Theory, 11, 1148-1171.

Hansen, B., 2000, Testing for structural change in conditional models," Journal of Econometrics, 97, 93-115.

Hansen, B. 2001, The New Econometrics of Structural Change: Dating Changes in U.S. Labor Productivity." Journal of Economic Perspectives, 15, 117-128.

Hansen, B., and Mehmet Caner, 2001, Threshold autoregression with a unit root," with . Econometrica, 69, 1555-1596.

Harris, Leybourne, and McCabe, 2007, Modified KPSS tests for near integration, Econometric Theory, 23, 355-363.

Harvey, Leybourne and Taylor, 2008, Unit root testing in practice: dealing with uncertainty over the trend and initial conditions, *Econometric Theory (forthcoming).*

Hasan, M.N. and R.W. Koenker, 1997, "Robust rank tests of the unit root hypothesis," *Econometrica, 65, 133-161.*

Hercé, M., 1996, Asymptotic Theory od LAD estimation in a unit root process with finite variance errors, Econometric Theory, 12, 129-153.

Ibragimov, R., and P.C.B. Phillips, 2008, Regression Asymptotics using Martingale Convergence Methods, *Econometric Theory*, **24**(4): 888-947.

Jacod, J. and A.N. Shiryaev, 2003, Limit Theorems for Stochastic Processes, Springer-Verlag, New York.

Juhl, T. 2008, A nonparametric test of the predictive regression model, mimeo.

Juhl, T., and Z. Xiao, 2003, Power Functions and Envelopes for Unit Root Tests, Econometric Theory, V.19, 240-253.

Kapetanois, G., 2005, Unit root testing against the alternative hypothesis up to m structural breaks, Journal of Time Series Analysis.

Knight, K., 1989, Limit Theory for Autoregressive-parameter estimates in an infinite-variance random walk, The Canadian Journal of Statistics, V. 17, 261-278.

Knight, K., 1998, Asymptotics for L1 regression estimates under general conditions, Annals of Statistics.

King M. L., 1988, "Towards a theory of point optimal testing," *Econometric Reviews*, **6**, 169–218.

Koenker, R. and Z. Xiao, 2004, "Unit Root Quantile Autoregression," *Journal of American Statistical Association*, .

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, "Testing the null hypothesis of stationarity 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.

Lai, T.L. and C.Z. Wei, 1982. "Least squares estimation in stochastic regresssion models with applications to identification and control of dynamic systems," *Annals of Statistics*, **10**, 154–166.

Lee, J. and MC Strazicich**,** Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks, *Review of Economics and Statistics,* Vol. 85, No. 4, Pages 1082-1089.

Leybourne S.J. and B.P.M. McCabe, 1994, "A consistent test for a unit root," *Journal of Business and Economic Statistics*, **12**, 157–166.

Lumsdaine, R.L., and D.H. Papell, 1997, "Multiple trend breaks and the unit root hypothesis," Review of Economics and Statistics.

Müller, U.K., 2008. The impossibility of consistent discrimination between I(0) and I(1) processes, Econometric Theory 24 (2008), 616 – 630.

Müller, U.K., and G. Elliot, 2003, Tests for Unit Roots and the Initial Condition, Econometrica 71 (2003), 1269 – 1286.

Nelson, C.R. and C. Plosser, 1982, "Trends and random walks in macro-economic time series: some evidence and implications," *Journal of Monetary Economics*, **10**, 139–162.

Ng, S. and P. Perron, 1995, "Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag", *Journal of the American Statistical Association*, **90**, 268–281.

Ouliaris, S., J.Y. Park and P.C.B. Phillips, 1989, "Testing for a unit root in the presence of a maintained trend." In B. Raj (ed.), *Advances in Econometrics and Modelling.* Amsterdam: Kluwer, pp. 7–28.

Ouliaris, S. and P.C.B. Phillips, 1990, "Asymptotic Properties of Residual Based Tests for Cointegration", Econometrica, Vol. 58, No. 1, January 1990: 165–193.

Ouliaris, S. and P.C.B. Phillips, 1994, *Coint 2.0.* Maple Valley, WA: Aptech Systems.

Park, J.Y. and J. Sung, 1994, "Testing for unit roots in models with structural change," *Econometric Theory*, **10**, 917–936.

Park, J. and P.C.B. Phillips, 1998, "Nonstationary Density Estimation and Kernel Autoregression", CFDP 1181, Yale University.

Park, J. and P.C.B. Phillips, 1999, "Asymptotics for Nonlinear Transformations of Integrated Time Series", Econometric Theory, Vol. 15, No. 3, June 1999: 269–298.

Perron, P., 1989, "The great crash, the oil price shock and the unit root hypothesis," *Econometrica*, **57**, 1361–1401.

Perron, P. and S. Ng., 1996, "Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," Review of Economic Studies 63 (1996), 435-463.

Phillips, P.C.B., 1986, "Understanding spurious regressions in econometrics," *Journal of Econometrics*, Volume 33, Issue 3, December 1986, Pages 311-340.

Phillips, P.C.B., 1987a, "Time series regression with a unit root," *Econometrica*, 55, 277–302.

Phillips, P.C.B., 1987b, "Towards a unified asymptotic theory of autoregression." *Biometrika*, 74, 535–547.

Phillips, P.C.B., 1988, "Regression Theory for Near-Integrated Time Series," Econometrica, Vol. 56, No. 5, September 1988: 1021–1044.

Phillips, P.C.B., 1991, A Shortcut to LAD Estimator Asymptotics, *Econometric Theory*, **7**.

Phillips, P.C.B., 1995a, Robust Nonstationary Regression, *Econometric Theory*, Vol. 12, No. 5, December 1995: 912–951.

Phillips, P.C.B., 1995b, "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, Vol. 63, No. 5, September 1995: 1023–1078.

Phillips, P.C.B., 1996, "Econometric model determination," *Econometrica*, **64**, 763–812.

Phillips, P.C.B., 1998, "New tools for understanding spurious regression," *Econometrica*, Vol. 66, No. 6, November, 1998: 1299–1326.

Phillips, P.C.B., 2002, "The invalidity of critical values of tests for unit roots against trend alternatives," Journal of Econometrics..

Phillips, P.C.B., 2005, "Some Themes in Modern Econometrics," Lecture Notes, Singapore Management University.

Phillips, P.C.B. and B.E. Hansen, 1990, Statistical Inference in Instrumental Variable Regressions with I(1) Processes", Review of Economic Studies, Vol. 57, 99–125.

Phillips, P.C.B. and C.C. Lee, 1996, "Efficiency gains from quasi-differencing under nonstationarity." In P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan*, Springer–Verlag: New York.

Phillips, P.C.B., Tassos Magdalinos, 2007, "Limit Theory for Moderate Deviations from a Unit Root", Journal of Econometrics, 136:115-130

Phillips, P.C.B., Tassos Magdalinos, 2006, "Limit Theory for Moderate Deivations from Unity under Weak Dependence", in G.D.A. Phillips and E. Tzavalis, eds., The Refinement of Econometric Estimation and Test Procedures: finite Sample and Asymptotic Analysis. Cambridge: Cambridge University Press, forthcoming.

Phillips, P.C.B., R.H. Moon, and Z. Xiao, 2001, “"How to Estimate Autoregressive Roots Near Unity" *Econometric Theory*, Vol. 17 No. 1, February, 2001: 29–69.

Phillips, P.C.B. and S. Ouliaris, 1990, “Asymptotic properties of residual based tests for cointegration,” *Econometrica*, **58**, 165–194.

Phillips, P.C.B. and P. Perron, 1988, “Testing for unit roots in time series regression,” *Biometrika*, **75**, 335–346.

Phillips, P.C.B. and W. Ploberger, 1994, “Posterior odds testing for a unit root with data-based model selection,” *Econometric Theory*, **10**, 774–808.

Phillips, P.C.B. and W. Ploberger, 1996, “An asymptotic theory of Bayesian inference for time series,” *Econometrica, 64, 381-412.*

Phillips, P.C.B. and V. Solo, 1992, “Asymptotics for linear processes,” *Annals of Statistics*, **20**, 971–1001.

Pollard, D., 1991, Asymptotics for Least Absolute Deviation Regression Estimators, Econometric Theory, 7, 186-199.

Pötscher, B. M., 2002, Lower risk bounds and properties of confidence sets for ill-posted estimation problems with applications to spectral density and persistency estimation, unit roots and estimation of long memory parameters., Econometrica, 1035 -1068.

Rothenberg, T. and J. Stock, 1997, “Inference in a nearly integrated autoregressive model with nonnormal innovations”, *Journal of Econometrics* 80, 269-286.

Said, S.E. and D.A. Dickey, 1984, “Testing for unit roots in autoregressive-moving average models of unknown order,” *Biometrika*, **71**,599–608.

Sargan, J. D. and A. Bhargava, 1983, "Testing residuals from least squares regression for being generated by the Gaussian random walk," *Econometrica*, **51**, 153–174.

Schmidt, P. and P.C.B. Phillips, 1992, "Testing for a unit root in the presence of deterministic trends," *Oxford Bulletin of Economics and Statistics*, **54**, 257–288.

Shimotsu, K. and P.C.B. Phillips, Exact Local Whittle Estimation of Fractional Integration", The Annals of Statistics (2005), 33(4): 1890-1933.

Stock, J.H., 1995, "Unit roots, structural breaks and trends." In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics,* Vol. 4, 2739-2841, Amsterdam: North–Holland.

Sul, D., P.C.B. Phillips, C. Choi, "Prewhitening Bias in HAC Estimation", Oxford Bulletin of Economics and Statistics (2005), 67(4): 517-546.

Wang, Q., and P.C.B. Phillips, 2006, Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression" CFDP 1594.

White, J.S., 1958, "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Math. Statist.*, **29**, 1188–1197.

Xiao, Z., 1998, A Note on Bandwidth Selection in Hypothesis Tests with Integrated Time Series, mimeo, University of Illinois.

Xiao, Z., 2001a, Testing the Null hypothesis of Stationarity against an Autoregressive Unit Root Alternative, Journal of Time Series Analysis, Vol. 22, No. 1, 87-105.

Xiao, Z., 2001b, Likelihood Based Inference in Trending Time Series with a Root Near Unity, Econometric Theory 17, 1082-1112.

Xiao, Z., 2003, Bandwidth Selection in Testing for Long Range Dependence, 2003, Economics Letters, Vol. 78, No. 1, 33-39.

Xiao, Z. and P.C.B. Phillips, 1998, "An ADF coefficient test for ARMA models with unknown orders," *The Econometrics Journal*, 27-43. .

Xiao, Zhijie, and P.C.B. Phillips, 2002, A CUSUM test for Cointegration Using Regression Residuals, Journal of Econometrics 108, 43-61.

Zivot, E. and D.W.K. Andrews, 1992, "Further evidence on the great crash, the oil price shock, and the unit root hypothesis," *Journal of Business and Economic Statistics*, **10**, 251–270.