Nonstationary Nonlinearity:
An Outlook for New Opportunities [1]

Joon Y. Park
Department of Economics
Rice University

Abstract

In this paper, we look for new opportunities that can be exploited using some of the recent developments on the theory of nonlinear models with integrated time series. Heuristic introductions on the basic tools and asymptotics are followed by the opportunities in three different directions: in data generation, in mean and in volatility. In the direction of data generation, we investigate the nonlinear transformations of random walks. It is shown in particular that they can generate stationary long memory as well as bounded nonstationarity and leptokurticity, which we commonly observe in many of economic and financial data. We then discuss how the nonlinear mean relationships between integrated processes can be appropriately formulated, interpreted and estimated within the regression framework. Both the nonlinear least squares regression and the nonparametric kernel regression are considered. Such formulations allow us to explore the nonlinear and nonparametric cointegration, which may be used in modelling the nonlinear and nonparametric longrun relationships among various economic and financial time series. Finally, a stochastic volatility model with the conditional variance specified as a nonlnear function of a random walk is examined. Established are various time series properties of the model, which are shown to be largely consistent with the observed characteristics of many time series data.

First Draft: July 2001
This version: March 2003

Key words and phrases: integrated processes, nonstationarity, nonlinear transformations, nonlinear and nonparametric regressions, conditional heteroskedasticity.

1. Introduction

It has long been recognized that many of economic and financial time series data exhibit nonstationarity that can be reasonably well modelled as integrated processes. Integrated processes have stochastic trends, which allow us to build their relationships using the notion of cointegration. As is well known, cointegration refers to the presence of a linear relationship among multiple integrated processes that holds up to stationary and mean-reverting residuals, and it has thus been widely used to describe various longrun economic equilibria. While the concept of cointegration has been received enthusiastically by many applied researchers, its practical implementation appears to be not entirely successful. In particular, we have not witnessed many meaningful empirical findings on the subject, in spite of the numerous attempts that have been made by many practitioners for the past two decades. Most of them seem to have found nothing beyond our common sense.

Both integration and cointegration, respectively as means of modelling observed individual time series and describing relationships among them, often appear to be too restrictive to be much useful in practical applications. For instance, there are many time series observed in reality that are bounded, yet locally nonstationary, and behave like integrated processes. Clearly, such time series cannot be effectively modelled as integrated processes that should necessarily be unbounded. It is also apparent for an obvious reason that any time series taking only nonnegative values cannot be generated by an integrated process. Moreover, in many cases it seems extremely unrealistic to presume that a simple linear relationship persists among integrated processes over a long period of time, which is implied by the presence of cointegration. Undoubtedly, there are many economic problems and phenomena that the concepts of integration and cointegration, both of which are very linear, appear to be too simple to effectively deal with.

This paper considers the nonlinear models with integrated time series. The relevant theory has been developed in a series of recent papers by Park and Phillips (1998, 1999, 2000, 2001, 2003), Chang, Park and Phillips (2001), Park (2002, 2003), Phillips and Park (1997) and several other papers that are still in progress. We concentrate on the positive side of the theory, and seek new research opportunities, rather than pitfalls, in nonstationary nonlinearity. An outlook for the new opportunities is made into three different directions: in data generation, in mean and in volatility. Presented are various possibilities that can come out of nonstationary nonlinearity in data generations, in regressions and in models for conditional variance. They are given in a unified framework. We also introduce the basic tools to effectively deal with nonstationary nonlinearity, and subsequently to develop some fundamental nonstationary nonlinear asymptotics.

First, we consider the nonstationary nonlinearity in data generation. Here we look for the possibility that integrated time series undergo a nonlinear transformation prior to our final observation. The transformation may come as a result of policy or market intervention as in the case of the exchange rate regime with a target zone, or simply from taking variable transformations such as logs and/or ratios. To investigate the effect of a nonlinear transformation, we examine the time series properties of the transformed integrated processes. Quite a few interesting results are obtained here. It is shown, for instance, that the nonlinear transformation of an integrated time series may yield bounded nonstationarity

and/or long memory. Leptokurticity and/or thick tails can also be generated by the non-stationary nonlinearity in data generation. Moreover, we demonstrate that the nonlinear transformation may or may not preserve the unit root depending upon the transformation function.

Second, the nonlinear relationships between integrated time series are discussed. We consider both the parametric nonlinear least squares regression and the nonparametric kernel regression, which are useful to make inference on the nonlinear and nonparametric cointegration models. As shown in Park and Phillips (2001), the usual nonlinear least squares yields consistent estimate for the nonlinear cointegrating regression. Also, we may use the standard kernel method to consistently estimate a cointegrating relationship nonparametrically. The nonparametric method, however, reduces the convergence rate dramatically, espectially when the regression function is increasing at infinity. We suggest as a practical solution to this problem the partial parametric approach, which in the first step models and estimates the dominent component parametrically. The residual component can be estimated nonparametrically in the second step.

Third, we look at a stochastic volatility model with conditional variance given by a nonlinear function of a random walk. The model, referred to as *nonstationary nonlinear heteroskedasticity*, was introduced and investigated earlier by Park (2002). Here we clearly demonstrate that the nonstationary nonlinearity can also play an important role in modelling volatility. In particular, it is shown that the nonstationary nonlinearity in volatility may generate samples with volatility clustering and leptokurticity, which are observed commonly for many economic and financial time series data. There appear to be many potential examples for such time series. For instance, the volatility of a stock return may well be positively related with the level of interest rate and/or transactions quantity, which are believed by many to be integrated.

All our subsequent discussions are heuristic. This is to focus effectively on the main theme that we want to deliver in this paper: new research opportunities from nonstationary nonlinearity. The assumptions made here are far from being necessary to obtain the stated results. Also, the results are not proved rigorously in the paper. The readers are referred to other existing literature for the minimal conditions and rigorous proofs. The standard notations are used without specific references. In particular, $=_d$ signifies the equivalence in distribution, $\rightarrow_d$ denotes the convergence in distribution. Likewise, we use $\rightarrow_{a.s}$ and $\rightarrow_p$, respectively, to signify almost sure convergence and convergence in probability. The stochastic order symbols o_p and O_p are also used, and **MN** stands for mixed normal distribution.

The rest of the paper is organized as follows. Section 2 introduces the technical preliminaries that are necessary to understand the subsequent development of our theory. Basic tools to deal with nonstationary nonlinearity are introduced, and some fundamental nonstationary nonlinear asymptotics are also developed. In Sections 3 to 5, various new opportunities from nonstationary nonlinearity are proposed. Section 3 lays out the statistical properties of nonlinear transformations of integrated processes. The nonlinear relationships for nonstationary processes are considered in Section 4. Here we develop their asymptotics, and discuss how they can be properly formulated and effectively estimated. In Section 5, we introduce nonstationary nonlinear heteroskedasticity as a volatility model, and compare

it with other competing models. Section 6 concludes the paper.

2. Theoretical Background

We consider an integrated time series (x_t), which is given by

$$x_t = x_{t-1} + v_t \qquad (1)$$

where (v_t) is an independent and identically distributed sequence of random variables with mean zero and variance ω^2. It is assumed that (v_t) has continous type distribution. Some of our subsequent results require the existence of higher moments for (v_t), i.e., $\mathbf{E}|v_t|^r < \infty$ for some $r > 2$. The assumption made here is far from being necessary for most of our subsequent results. Many of the results derived in the paper are applicable for more general integrated processes driven by linear processes or mixing sequences. In those cases, we may obtain similar results using higher moment conditions on innovations sequences and/or stronger assumptions on nonlinear transformation functions. As we move along, we will indicate what types of extensions are possible and where to find more general results.

Our primary objective in this paper is to investigate the probabilistic and statistical properties of an integrated process, such as (x_t) introduced in (1), under a transformation. The transformation is presumed, though not restricted, to be nonlinear. In general, the asymptotics for integrated time series under the nonlinear transformation depend crucially on the type of the transformation involved. This was clearly demonstrated by Park and Phillips (1999). They indeed introduced three different classes of transformation functions – integrable functions, asymptotically homogeneous functions and exponential functions – to characterize the asymptotics for the nonlinear transformations of integrated processes. Of the three classes, we consider only the first two classes. The integrated processes have stochastic trends, and their exponential transformations appear to be too explosive to be useful for practical applications.

Obviously, the transformations like $T(x) = 1\{a \leq x \leq b\}$ with some constants a and b, and $T(x) = e^{-x^2}$ are integrable. Asymptotically homogeneous functions are roughly the functions that behave asymptotically as homogeneous functions. In short, the transformation T is asymptotically homogeneous, if we have

$$T(\lambda x) \approx \nu(\lambda)S(x) \qquad (2)$$

for all large λ and for all x on any bounded interval. We will call ν and S, respectively, the *asymptotic order* and the *limit homogeneous function* of T. The class of asymptotically homogeneous transformations includes a wide variety of functions. It includes, for instance, constant functions, all distribution function-like functions, logarithmic functions and all functions that behave asymptotically as polynomials.

The transformation given by $T(x) = |x|^k$ is homogeneous, and is certainly asymptotically homogeneous with asymptotic order $\nu(\lambda) = \lambda^k$ and limit homogeneous function $S(x) = |x|^k$. The logarithmic transformation $T(x) = \log|x|$ is also asymptotically homogeneous with asymptotic order $\nu(\lambda) = \log\lambda$ and limit homogeneous function $S(x) = 1$. Note that

$\log|\lambda x| = \log \lambda + \log|x| \approx \log \lambda$ for all large λ and for all x on any bounded interval. The logistic transformation $T(x) = e^x/(1+e^x)$ can be decomposed as $T(x) = S(x)+R(x)$, where $S(x) = 1\{x \geq 0\}$ is homogeneous and $R(x) = 1\{x < 0\}e^x/(1 + e^x) - 1\{x \geq 0\}1/(1 + e^x)$ is integrable. However, it follows that $S(\lambda x) = S(x)$ for all λ and for all x, and $R(\lambda x) \to 0$ as $\lambda \to \infty$ for all x on any bounded interval. We thus have $T(\lambda x) \approx 1\{x \geq 0\}$ for all large λ and for all x on a bounded interval. Consequently, it has asymptotic order $\nu(\lambda) = 1$ and limit homogeneous function $S(x) = 1\{x \geq 0\}$. The same result applies to any distribution function-like transformation.

2.1 Statistical Tools

The classical Donsker's theorem applies to (v_t). Therefore, if we let $[nr]$ be the integral part of nr for $r \in [0, 1]$ and define

$$V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t \qquad (3)$$

then it follows that

$$V_n \to_d V \qquad (4)$$

where V is the Brownian motion with variance ω^2 on the unit interval $[0, 1]$. The invariance principle (4) holds under the conditions that are much weaker than those we impose here. We may obtain it for, e.g., linear processes driven by martingale difference sequences satisfying some mild summability and moment conditions. See Hall and Heyde (1981) and Phillips and Solo (1992) for technical details.

Our subsequent asymptotic theory is presented using various functionals of the limit Brownian motion V introduced above in (4). It also involves the local time L of the Brownian motion V. The Brownian local time L is a stochastic process with two parameters, t and s, say, which can be defined as

$$L(t, s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t 1\{|V(r) - s| < \epsilon\} \, dr \qquad (5)$$

Roughly, it may be interpreted as the time, measured as an instantaneous rate, spent by V in an immediate neighborhood of s up to time t.[2] The reader is referred to Chung and Williams (1990) for an elementary introduction to the concept of local time.

The Brownian local time is known to have a version continuous both in t and s, so we may assume L is given as such. For each t, $L(t, \cdot)$ has a compact support a.s. This should be obvious, since Brownian motion has continuous sample path a.s. Some of our asymptotics, in particular, include the Brownian local time at $t = 1$ and $s = 0$, i.e., $L(1, 0)$. It is well known that $L(1, 0)$ has the same distribution as $|V(1)|$, i.e., the modulus of the normal variate with mean zero and variance ω^2. This follows immediately from the fact that the local time of standard Brownian motion has distribution given by the truncated standard

---

[2]The local time is defined for a broader class of semi-martingales. Also, it is more customary to measure the time in the unit of the quadratic variation of the underlying process. Our local time L in (5) may thus be defined more conventionally using $d[V](r) = \omega^2 dr$ in place of dr. However, it is more convenient to use our definition (5) for the subsequent asymptotics on nonstationary nonlinearity.
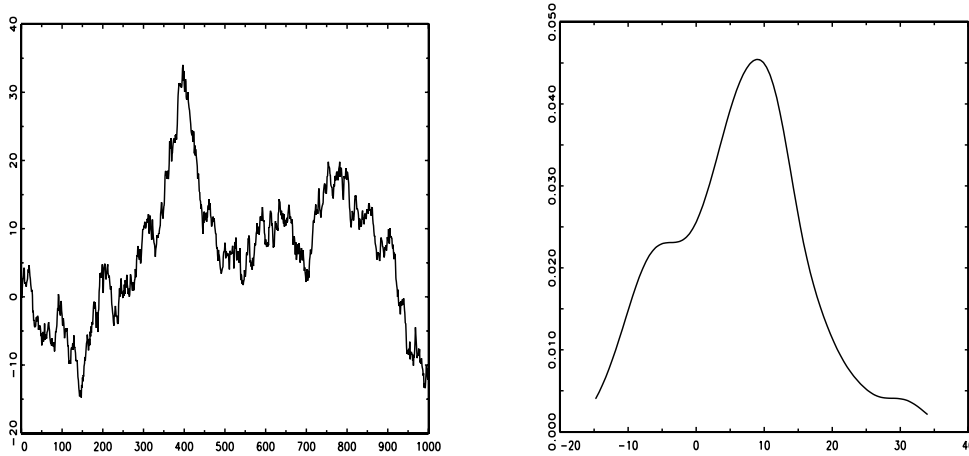
Figure 1: Sample Path and Local Time of Brownian Motion

normal supported on the positive half of **R**. See, e.g., Revuz and Yor (1994, Proposition 3.7, p100 and Theorem 2.3, p230). Note that the local time L of V at time and spatial parameters t and s has the same value with $1/\omega$ times of the local time of W at t and s/ω. This can be easily deduced from our defition in (5).

The local time yields the formula

$$\int_0^t T(V(r))\,dr = \int_{-\infty}^{\infty} T(s)L(t,s)\,ds \qquad (6)$$

for any T locally integrable. The relationship (6) is commonly referred to as the *occupation times formula*. For each fixed t, it allows us to evaluate the integral of any locally integrable transformation of a Brownian motion with respect to time as the integral of the function itself weighted by the local time. In this sense, we may legitimately interpret the local time, as a function of its spatial parameter, as the spatial density of the values that the underlying Brownian motion takes. The interpretation of the local time as the spatial density is illustrated in Figure 1. There we present a simulated Brownian sample path with its estimated local time, which shows the relative frequencies for the realized values of the underlying Brownian motion.

2.2 Basic Asymptotics

Now we are ready to derive some basic asymptotics for the nonlinear models with integrated time series. To develop the covariance asymptotics, as well as the mean asymptotics, we introduce (u_t) in addition to (x_t) defined in (1). In what follows, we assume that $(u_t, \mathcal{F}_t)$ is a martingale difference sequence, where $(\mathcal{F}_t)$ is a filtration such that (x_t) is adapted to $(\mathcal{F}_{t-1})$. We let

$$\sigma^2 = \mathbf{E}(u_t^2 | \mathcal{F}_{t-1})$$

and $\mathbf{E}(|u_t|^r | \mathcal{F}_{t-1}) < \infty$ a.s. for some $r > 2$. Moreover, we define a partial sum process U_n on $[0,1]$ from (u_t), similarly as V_n in (3) constructed from (v_t), and denote by U the limit

Brownian motion of U_n. Clearly, U is a Brownian motion with variance σ^2, and allowed to be dependent with the limit Brownian motion V of V_n. We may indeed permit U and V to be perfectly correlated and identical.

We let $T_n : \mathbf{R} \to \mathbf{R}$ and consider the asymptotic behaviors of

$$\sum_{t=1}^{n} T_n(x_t) \quad \text{and} \quad \sum_{t=1}^{n} T_n(x_t)\, u_t$$

under appropriate normalizations. Under suitable conditions, we may expect

$$\frac{1}{n} \sum_{t=1}^{n} T_n(x_t) = \int_0^1 T_n(\sqrt{n}V_n(r))\, dr \approx \int_0^1 T_n(\sqrt{n}V(r))\, dr \tag{7}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T_n(x_t)\, u_t = \int_0^1 T_n(\sqrt{n}V_n(r))\, dU_n(r) \approx \int_0^1 T_n(\sqrt{n}V(r))\, dU(r) \tag{8}$$

for large n.

If, for instance, T_n is given by

$$T_n(x) = T\left(\frac{x}{\sqrt{n}}\right)$$

for some locally integrable T, then both the mean and covariance asymptotics follow immediately from (7) and (8). It is indeed easy to see that (7) and (8) imply in this case

$$\frac{1}{n} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) \to_d \int_0^1 T(V(r))\, dr \tag{9}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T\left(\frac{x_t}{\sqrt{n}}\right) u_t \to_d \int_0^1 T(V(r))\, dU(r) \tag{10}$$

as $n \to \infty$.

If, on the other hand, we let $T_n = T$ for all n, the asymptotics become crucially dependent upon the type of function T. If T is asymptotically homogeneous and given as in (2), then the relevant asymptotics follow somewhat similarly as in (9) and (10). We only need different normalizations. Indeed, since

$$T(x_t) \approx \nu(\sqrt{n})S\left(\frac{x_t}{\sqrt{n}}\right)$$

in this case, we may readily deduce that

$$\frac{1}{n}\nu^{-1}(\sqrt{n}) \sum_{t=1}^{n} T(x_t) \to_d \int_0^1 S(V(r))\, dr \tag{11}$$

$$\frac{1}{\sqrt{n}}\nu^{-1}(\sqrt{n}) \sum_{t=1}^{n} T(x_t)\, u_t \to_d \int_0^1 S(V(r))\, dU(r) \tag{12}$$

as $n \to \infty$.

The asymptotics for $T_n = T$ with an integrable T are more intriguing. To obtain the mean asymptotics in this case, we need to note that

$$
\begin{aligned}
\sqrt{n} \int_0^1 T(\sqrt{n}V(r))\,dr &= \sqrt{n} \int_{-\infty}^{\infty} T(\sqrt{n}s)L(1,s)\,ds \\
&= \int_{-\infty}^{\infty} T(s)L\left(1, \frac{s}{\sqrt{n}}\right)\,ds \\
&\to_{a.s.} L(1,0) \int_{-\infty}^{\infty} T(s)\,ds \qquad (13)
\end{aligned}
$$

as $n \to \infty$. This follows by the successive applications of occupation times formula, change-of-variables and dominated convergence. Consequently, we may obtain the mean asymptotics

$$
\frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) \to_d L(1,0) \int_{-\infty}^{\infty} T(s)\,ds \qquad (14)
$$

as $n \to \infty$, due to (7) and (13).

The covariance asymptotics are more difficult to obtain. Due to the representation theorem of continuous martingale by time-changed Brownian motion [see, e.g., Revuz and Yor (1994, Theorem 1.6, p173)], we may regard for each n a continuous martingale

$$
M_n(t) = \sqrt[4]{n} \int_0^t T(\sqrt{n}V(r))\,dU(r)
$$

as Brownian motion, say, W_n evaluated at time

$$
[M_n](t) = \sigma^2 \sqrt{n} \int_0^t T^2(\sqrt{n}V(r))\,dr
$$

where $[M_n]$ is the quadratic variation of M_n. That is, $M_n(t) = W_n([M_n](t))$, where W_n is often referred to as Dambis-Dubins-Schwarz (DDS) Brownian motion. It is obvious that we may write

$$
M_n(t) =_d W([M_n](t)) \text{ a.s.}
$$

for all n.

Now we obtain the limit distribution of $M_n(1)$ using its representation as the DDS Brownian motion. We have exactly as in (13)

$$
[M_n](1) \to_{a.s.} \sigma^2 L(1,0) \int_{-\infty}^{\infty} T^2(s)\,ds
$$

and, as a result,

$$
M_n(1) \to_d W\left(\sigma^2 L(1,0) \int_{-\infty}^{\infty} T^2(s)\,ds\right)
$$

as $n \to \infty$. Moreover, if we denote by $[M_n, V]$ the quadratic covariation of M_n and V, then

$$
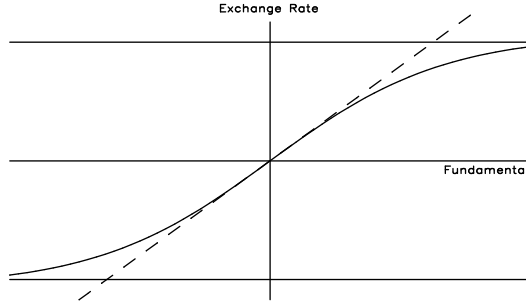[M_n, V](t) = \sqrt[4]{n} \int_0^t T(\sqrt{n}V(r))\,d[U,V](r) \to_{a.s.} 0
$$

Figure 2: Target Zone Exchange Rate Model

for all $t \geq 0$. Therefore, M_n is asymptotically independent of V, and hence, of L which is the local time of V. This, however, implies in turn that L is independent of W, since W is the distributional equivalence of the DDS Brownian motion of M_n. Consequently, it follows that

$$W\left(\sigma^2 L(1,0) \int_{-\infty}^{\infty} T^2(s)\,ds\right) =_d \mathbf{MN}\left(0, \sigma^2 L(1,0) \int_{-\infty}^{\infty} T^2(s)\,ds\right)$$

[see, e.g., Revuz and Yor (1994, Chapter XIII) for a more detailed derivation]. The covariance asymptotics for $T_n = T$ with integrable T now follow immediately and given by

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^{n} T(x_t)\,u_t \to_d \mathbf{MN}\left(0, \sigma^2 L(1,0) \int_{-\infty}^{\infty} T^2(s)\,ds\right) \tag{15}$$

as $n \to \infty$, dut to (8).

The asymptotics for more general T_n can be obtained in a completely analogous manner, once the approximations in (7) and (8) are validated. The reader is referred to Akonom (1993), Borodin and Ibragimov (1995) and Park (2003) for the precise conditions and the exact orders of approximation errors. The asymptotics derived here hold for much more general time series than are assumed here. The asymptotics for integrable transformation in (14) and (15) hold for general integrated processes driven by linear processes with iid innovations satisfying some mild regularity conditions. The asymptotics in (9) and (10) for normalized transformations, and those in (11) and (12) for asymptotically homogeneous transformations are obtainable under even much weaker conditions. For these results, the invariance principle (4) is virtually all that is required. See Park (2003) for the exact conditions and technical details.

3. Nonstationary Nonlinearity in Data Generation

In this section, we look at the time series (y_t) given by

$$y_t = F(x_t)$$

and consider its time series properties. Such a time series (y_t) may be generated in reality due to the presence of institutional restriction and/or barrier or policy intervention, in which
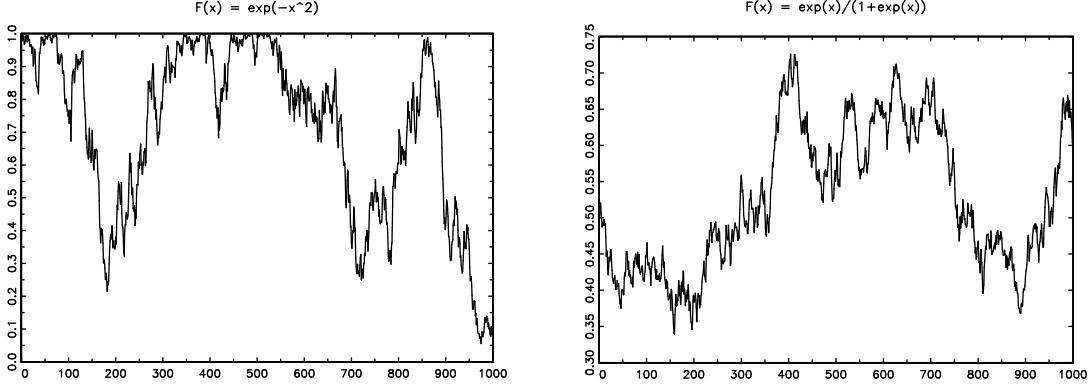
Figure 3: Simulated Sample Paths

case (x_t) would represent the corresponding economic fundamentals. A prime example for (y_t) would be the observed exchange rates in the presence of target zone. In this example, F is given by a bounded and S-shaped nonlinear function, which is shown in Figure 2. The resulting time series (y_t) may exhibit very different characteristics, depending upon the type of function F. If F is integrable, then (y_t) would look more like a stationary process. If F is asymptotically homogeneous, then (y_t) would behave like an integrated process having a unit root. The time series (y_t) itself of course becomes an integrated process when F is the identity. In Figure 3, the sample paths of the transformed time series are simulated for $F(x) = e^{-x^2}$ and $F(x) = e^x/(1 + e^x)$.

To investigate the statistical properties of the transformed series (y_t), we define the sample autocorrelation for (y_t) by

$$R_{nk} = \frac{\displaystyle\sum_{t=k+1}^{n}(y_t - \bar{y}_n)(y_{t-k} - \bar{y}_n)}{\displaystyle\sum_{t=1}^{n}(y_t - \bar{y}_n)^2}$$

where $\bar{y}_n$ denotes the sample mean of (y_t). We also look at the sample variance of (y_t) given by

$$S_n^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

and the sample kurtosis

$$K_n^4 = \frac{\dfrac{1}{n}\displaystyle\sum_{t=1}^{n}(y_t - \bar{y})^4}{\left(\dfrac{1}{n}\displaystyle\sum_{t=1}^{n}(y_t - \bar{y})^2\right)^2}$$

which will be analyzed subsequently for both integrable and asymptotically homogeneous transformations.

3.1 Integrable Transformations

We assume that F is bounded and integrable. Then it follows from (14) that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F^p(x_t) \to_d L(1,0) \int_{-\infty}^{\infty} F^p(x)\,dx \tag{16}$$

for $p = 1, 2, \ldots$, as $n \to \infty$. Moreover, if we denote by D_k the density of $(x_t - x_{t-k})$ that is assumed to be independent of (x_{t-k}) for $k = 1, 2, \ldots$, then we have $\mathbf{E}(F(x_t)|\mathcal{F}_{t-k}) = G(x_{t-k})$ with an integrable function G given by

$$G(x) = \int_{-\infty}^{\infty} F(x+y) D_k(y)\,dy \tag{17}$$

We may therefore deduce under suitable regularity conditions that

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} F(x_t)F(x_{t-k}) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (GF)(x_{t-k}) + o_p(1) \\
&\to_d L(1,0) \int_{-\infty}^{\infty} (GF)(x)\,dx \\
&= L(1,0) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x)F(x+y) D_k(y)\,dx\,dy
\end{aligned}
\tag{18}
$$

as $n \to \infty$.

We may now easily deduce from (16) and (18) that

$$R_{nk} \to_p R_k$$

where

$$R_k = \frac{\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x)F(x+y) D_k(y)\,dx\,dy}{\displaystyle\int_{-\infty}^{\infty} F^2(x)\,dx}$$

for $k = 1, 2, \ldots$, as $n \to \infty$. For the integrable transformation, the sample autocorrelation R_{nk} therefore converges in probability to a nonrandom limit. Of course, the limit R_k may be regarded as its asymptotic autocorrelation function.

We have

$$
\begin{aligned}
\int_{-\infty}^{\infty} G^2(x)\,dx &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} F(x+y) D_k(y)dy \right)^2 dx \\
&< \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, D_k(y) F^2(x+y) = \int_{-\infty}^{\infty} F^2(x)\,dx
\end{aligned}
$$

from which, together with the Cauchy-Schwarz inequality

$$\left| \int_{-\infty}^{\infty} (GF)(x)\,dx \right| < \left(\int_{-\infty}^{\infty} G^2(x)\,dx \right)^{1/2} \left(\int_{-\infty}^{\infty} F^2(x)\,dx \right)^{1/2}$$

it follows that $R_k < 1$ for all $k = 1, 2, \ldots$. This implies, in particular, that the first order autoregression run with the integral transformation of a random walk would asymptotically yield the coefficient strictly less than unity. Hence, the unit root disappears under integrable transformations. The actual values of the asymptotic autocorrelation function R_k are determined by the distribution of (v_t), as well as the transformation function.

It is important to investigate the behavior of the asymptotic autocorrelation R_k as a function of k. This is to find the degree of persistency for the integral transformation of a random walk. If (v_t) are normally distributed, then $x_t - x_{t-k} =_d k^{1/2} v_t$. We therefore have $D_k(x) = k^{-1/2} D(k^{-1/2} x)$, where D is the density of (v_t). Consequently, we may deduce that

$$k^{1/2} R_k = \frac{\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) F(x+y) D(k^{-1/2} y) \, dx \, dy}{\displaystyle\int_{-\infty}^{\infty} F^2(x) \, dx}$$

$$\to D(0) \frac{\left(\displaystyle\int_{-\infty}^{\infty} F(x) \, dx\right)^2}{\displaystyle\int_{-\infty}^{\infty} F^2(x) \, dx}$$

as $k \to \infty$, due to dominated convergence. The transformation function $F(x) = e^{-x^2}$ actually yields the asymptotic autocorrelation function $R_k = 1/\sqrt{k+1}$ for $k = 1, 2, \ldots$.

It is interesting to note that the autocorrelation function of the integrable transformation of a random walk decays, as for the stationary time series. The decaying rate, however, is much slower, and given by

$$R_k = c \, k^{-1/2}$$

for some constant $c > 0$. The integral transformations of random walks would thus be expected to show considerable persistency in memory. The rate of decay for their autocorrelations is indeed comparable to that of the stationary long memory processes. For the I(d) process with $0 < d < 1/2$, the autocorrelation function is given by $R_k = c \, k^{2d-1}$ for some constant $c > 0$ [see, e.g., Beran (1994, p63)]. The autocorrelation function of the integrable transformation of a random walk has therefore the same rate of decay as that of the I(d) process with $d = 1/4$. Stationary long memory can thus be generated by nonstationary nonlinearity.

The asymptotics of the sample variance and kurtosis for the integrable transformations of random walks can also be easily obtained. Indeed, it follows immediately from (16) and (18) that

$$\sqrt{n} S_n^2 \to_d L(1,0) \int_{-\infty}^{\infty} F^2(x) \, dx$$

$$\frac{1}{\sqrt{n}} K_n^4 \to_d \frac{\displaystyle\int_{-\infty}^{\infty} F^4(x) \, dx}{L(1,0) \left(\displaystyle\int_{-\infty}^{\infty} F^2(x) \, dx\right)^2}$$

as $n \to \infty$.

The sample variance of the integral transformation of a random walk is thus expected to decrease at the rate of $n^{-1/2}$. Roughly, this implies that the second-order variations of the integral transformations of random walks are increasing, but not as fast as the sample size. On the other hand, the sample kurtosis of the integral transformation of a random walk increases at an $n^{1/2}$ rate as the sample size gets large. We may thus expect the sample kurtosis to be large for the integral transformations of random walks. The leptokurticity, along with the aforementioned long memory property, is frequently observed in many macroeconomic and financial time series.

All of our results previously derived in this subsection apply to the transformations that can be defined as the constant shifts of integrable functions, as well as the integrable functions themselves. This is so, since the statistics, sample autocorrelation, sample variance and sample kurtosis considered here, are all invariant with respect to the constant shift. More precisely, all the previous results hold also for the function F, which can be written as

$$F(x) = c + G(x) \tag{19}$$

where c is some constant, and G is an integrable function that has been considered thus far. The function F in (19) is, of course, not integrable. It is indeed asymptotically homogeneous with the unit asymptotic order and the constant limit homogeneous function. The analysis for general asymptotically homogeneous functions with nonconstant limit homogeneous functions will follow in the next subsection.

3.2 Asymptotically Homogeneous Transformations

We now let F be an asymptotically homogeneous function with asymptotic order κ and limit homogeneous function H. Under very mild regularity conditions that are satisfied by virtually all asymptotically homogeneous functions used in practical applications, F^p would also be asymptotically homogeneous with asymptotic order κ^p and limit homogeneous function H^p, i.e.,

$$F^p(\lambda x) \approx \kappa^p(\lambda) H^p(x)$$

for $p = 1, 2 \ldots$. This will be assumed in what follows. Then it follows immediately from (11) that

$$\frac{1}{n} \kappa^{-p}(\sqrt{n}) \sum_{t=1}^{n} F^p(x_t) \to_d \int_0^1 H^p(V(r)) \, dr \tag{20}$$

as $n \to \infty$.

Moreover, if we let G be defined as in (17), then we may expect

$$
\begin{aligned}
G(\lambda x) &= \int_{-\infty}^{\infty} F(\lambda x + y) D_k(y) \, dy \\
&\approx \kappa(\lambda) \int_{-\infty}^{\infty} H\left(x + \frac{y}{\lambda}\right) D_k(y) \, dy \\
&\approx \kappa(\lambda) H(x)
\end{aligned}
$$

for all large λ and for all x on any bounded interval. That is, G is asymptotically homogeneous with asymptotic order κ and limit homogeneous function H, exactly as for F. As a result, we have

$$(GF)(\lambda x) \approx \kappa^2(\lambda)H^2(x)$$

for all large λ and for all x on any bounded interval. We may therefore establish that

$$\frac{1}{n}\kappa^{-2}(\sqrt{n})\sum_{t=1}^{n} F(x_t)F(x_{t-k}) = \frac{1}{n}\kappa^{-2}(\sqrt{n})\sum_{t=1}^{n}(GF)(x_{t-k}) + o_p(1)$$

$$\to_d \int_0^1 H^2(V(r))\,dr \tag{21}$$

as $n \to \infty$, under suitable technical assumptions.

Here we assume that the limit homogeneous function H of F is non-constant. The case where H is a constant yields quite different results and has already been considered. The asymptotically homogeneous function will be called *nontrivial*, if it has non-constant limit homogeneous function. Otherwise, it will be called *trivial*. For the nontrivial asymptotically homogeneous function F, it only requires some simple algebra to deduce from (20) and (21) that

$$R_{nk} \to_p 1$$

as $n \to \infty$, for every $k = 1, 2, \ldots$. The asymptotic autocorrelations for the nontrivial asymptotically homogeneous transformation of a random walk are therefore unity at all lags, just like those of untransformed random walks. In particular, the asymptotic first-order autocorrelation remains to be unity. The nontrivial asymptotically homogeneous transformation would thus preserve the unit root, i.e., the transformed time series continue to have unit roots if the nontrivial asymptotically homogeneous function is used. This is in sharp constrast with the integrable or trivial asymptotically homogeneous functions, for which the unit root disappears upon transformations.

Once again, we let F be a nontrivial asymptotically homogeneous function with asymptotic order κ and limit homogeneous function H. It is straightforward to derive

$$\kappa^{-2}(\sqrt{n})S_n^2 \to_d \int_0^1 \left(H(V) - \int_0^1 H(V)\right)$$

$$K_n^4 \to_d \frac{\int_0^1 \left(H(V) - \int_0^1 H(V)\right)^2}{\int_0^1 \left(H(V) - \int_0^1 H(V)\right)^4}$$

as $n \to \infty$, using the results in (20) and (21). Note that we use the shorthand notation for the integrals to simplify the exposition. This convention will be followed also in the subsequent presentation of our results, whenever they become lengthy and require simplifications in expressions.

The sample variance of the nontrivial asymptotically homogeneous transformation of a random walk may diverge or converge, depending upon its asymptotic order. If $\kappa(\lambda) \to 0$ as

$\lambda \to \infty$, as in the case of the power function with a negative power, then $S_n^2 \to_p 0$ as $n \to \infty$. On the other hand, $S_n^2 \to_p \infty$ as $n \to \infty$ for all power functions with positive powers and any other functions with diverging asymptotes including the logarithmic function, since for them $\kappa(\lambda) \to \infty$ as $\lambda \to \infty$. For the bounded asymptotically homogeneous functions such as the logistic function or any other distribution function-like functions, we have $\kappa(\lambda) = 1$, and the sample variance would be stochastically bounded, though it remains to be random in the limit. Indeed, the sample variance in this case has a well-defined limit distribution that is represented by a functional of the limit homogeneous transformation of Brownian motion.

In contrast, the sample kurtosis has random limit in all cases. As long as the underlying asymptotically homogeneous transformation is nontrivial, the sample kurtosis has the limit distribution that is given as a functional of the limit homogeneous transformation of Brownian motion. The sample kurtosis, just as the sample variance in the case of the asymptotically homogeneous transformations that are nontrivial and bounded, is expected to be given randomly, with distribution depending upon the limit homogeneous function of the underlying transformation. The limit distributions for the sample variance and sample kurtosis can easily be simulated for any asymptotically homogeneous transformation, if its limit homogeneous function is given.

4. Nonstationary Nonlinearity in Mean

In this section, we statistically analyze the nonlinear relationships between nonstationary time series. We consider the regression model

$$y_t = F(x_t) + u_t \tag{22}$$

where (x_t) is an integrated regressor as specified earlier in (1), and (u_t) is a stationary error. We assume the regression function F to be either integrable or asymptotically homogeneous. The regressand (y_t) can therefore have characteristics of both stationary and integrated time series, depending upon F. Our results in the previous section imply that the process (y_t) would behave like a stationary time series if F is integrable, whereas it would look more like an integrated time series if F is asymptotically homogeneous.

Assume that (u_t) satisfies the conditions introduced in Section 2.2. Then we have in particular

$$\mathbf{E}(y_t|\mathcal{F}_{t-1}) = F(x_t)$$

as in the usual regression model. However, our regression involves integrated time series, and would thus have different interpretations. If the regression function F is asymptotically homogeneous, the regression model specifies a nonlinear cointegrating relationship, i.e., a longrun nonlinear equilibrium relationship, between (y_t) and (x_t). On the other hand, the integrable regression function F would remove the stochastic trend in (x_t), and therefore, the model specifies a usual stationary relationship between (y_t) and $(F(x_t))$. We may leave the regression function F completely unrestricted, or further parametrize it using some appropriate family of functions. We consider both cases in the paper, which will be analyzed below in sequel.

4.1 Nonlinear Regressions with Integrated Processes

The model (22) becomes, and may be analyzed as such, a nonlinear regression, if we further specify F as
$$F(x) = G(x, \theta_0)$$
where $G(\cdot, \theta)$ with $\theta \in \Theta$ represents a parametric family of functions. The function G is assumed to be known, while the true parameter value θ_0 is regarded as being unknown. The family of functions should, of course, be chosen so that the regresion is *balanced*, i.e., we should choose an integrable family if the regressand (y_t) behaves like a stationary process. If the regressand is closer to an integrated process, an asymptotically homogeneous family should be more appropriate. The linear cointegrating regression can be considered as a special case with $G(x, \theta) = x\theta$.

The nonlinear regression
$$y_t = G(x_t, \theta_0) + u_t \tag{23}$$
can be estimated by the nonlinear least squares (NLLS), in which case the estimator $\hat{\theta}_n$ can be obtained by
$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{t=1}^{n} (y_t - G(x_t, \theta))^2$$

It is shown by Park and Phillips (2001) that the estimator $\hat{\theta}_n$ of θ in nonlinear regression (23) is asymptotically equivalent to the usual least squares (LS) estimator of θ in the linear regression
$$y_t = \dot{G}(x_t, \theta_0)'\theta + u_t \tag{24}$$
where $\dot{G}$ denotes the partial derivative of G with respect to θ. The asymptotic equivalence of (23) and (24) is well established and long known for the standard stationary regressions.

The asymptotic distribution of $\hat{\theta}_n$ can be readily obtained from the asymptotic equivalence of (23) and (24), using the basic nonstationary nonlinear asymptotics introduced earlier. Note that the aforementioned asymptotic equivalence implies

$$
\begin{aligned}
\hat{\theta}_n &\approx \left(\sum_{t=1}^{n} \dot{G}(x_t, \theta_0) \dot{G}(x_t, \theta_0)' \right)^{-1} \sum_{t=1}^{n} \dot{G}(x_t, \theta_0) y_t \\
&= \theta_0 + \left(\sum_{t=1}^{n} \dot{G}(x_t, \theta_0) \dot{G}(x_t, \theta_0)' \right)^{-1} \sum_{t=1}^{n} \dot{G}(x_t, \theta_0) u_t
\end{aligned}
\tag{25}
$$

for large n. The asymptotic distribution of $\hat{\theta}_n$ may therefore be derived from our mean and covariance asymptotics developed in Section 2.

If $\dot{G}(\cdot, \theta_0)$ is integrable, then we may deduce from (14) and (15) that

$$\sqrt[4]{n}(\hat{\theta}_n - \theta_0) \to_d \mathbf{MN} \left(0, \sigma^2 \left(L(1,0) \int_{-\infty}^{\infty} \dot{G}(s, \theta_0) \dot{G}(s, \theta_0)' ds \right)^{-1} \right) \tag{26}$$

as $n \to \infty$. Moreover, if $\dot{G}(\cdot, \theta_0)$ is asymptotically homogeneous with asymptotic order $\dot{\kappa}(\cdot)$ and limit homogeneous function $\dot{H}(\cdot, \theta_0)$, then we have due to (11) and (12)

$$\sqrt{n}\dot{\kappa}(\sqrt{n})(\hat{\theta}_n - \theta_0) \to_d \left(\int_0^1 \dot{H}(V, \theta_0)\dot{H}(V, \theta_0)' \right)^{-1} \int_0^1 \dot{H}(V, \theta_0)\, dU \qquad (27)$$

as $n \to \infty$.

The NLLS estimator $\hat{\theta}_n$ in both cases is consistent. The convergence rate, however, is different, depending upon the regression function. If $\dot{G}(\cdot, \theta_0)$ is integrable, $\hat{\theta}_n$ converges at the $\sqrt[4]{n}$ rate, which is of an order of magnitude slower than the standard linear stationary case. For asymptotically homogeneous $\dot{G}(\cdot, \theta_0)$, the actual convergence rate is dependent upon its asymptotic order $\dot{\kappa}$, and is given by $\sqrt{n}\dot{\kappa}(\sqrt{n})$. The rate becomes n for the linear cointegrating regression, since $\dot{\kappa}(\lambda) = \lambda$ in this case. The distribution of $\hat{\theta}_n$ is asymptotically mixed normal, when $\dot{G}(\cdot, \theta_0)$ is integrable. If it is asymptotically homogeneous, however, the limiting distribution is generally non-Gaussian. We have the Gaussianity only under the strictly exogeneity, i.e., the case where the regressors are uncorrelated with the regression errors in all leads and lags. This is unrealistic.

To obtain our results in (26) and (27), we need to require some obvious identifiability assumptions as well as technical regularity conditions. In particular, we should have $\int_{-\infty}^{\infty} \dot{G}(s, \theta_0)\dot{G}(s, \theta_0)'ds > 0$ for $\dot{G}(\cdot, \theta_0)$ integrable. Likewise, asymptotically homogeneous $\dot{G}(\cdot, \theta_0)$ must have limit homogeneous function $\dot{H}$ satisfying $\int_0^1 \dot{H}(V, \theta_0)\dot{H}(V, \theta_0)' > 0$ a.s. If these conditions fail, then our results here are obviously no long valid. This, however, does not imply that such identifiability assumption is absolutely necessary. It is often the case that the failure of the identifiability condition simply results in asymptotic multicollinearity, which can be avoided by an appropriate reparametrization.

For instance, if we let $\theta = (\alpha, \beta)'$, and let

$$G(x, \alpha, \beta) = \alpha + \beta \log |x|$$

Since $\log |x|$ is asymptotically homogeneous with asymptotic order $\log \lambda$ and limit homogeneous function 1, $\dot{H}$ becomes $(1, 1)'$ and the identifiability condition does not hold. We may nevertheless reformulate the regression as

$$y_t = \alpha + \beta \log |x_t| + u_t$$
$$= (\alpha + \beta \log \sqrt{n}) + \beta \log \frac{|x_t|}{\sqrt{n}} + u_t$$

to obtain the proper asymptotics. After some simple algebra, we may obtain

$$\sqrt{n} \begin{pmatrix} 1 & \log \sqrt{n} \\ 0 & 1 \end{pmatrix} (\hat{\theta}_n - \theta_0) \to_d \begin{pmatrix} 1 & \int_0^1 \log |V| \\ \int_0^1 \log |V| & \int_0^1 \log^2 |V| \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 dU \\ \int_0^1 dU \log |V| \end{pmatrix}$$

as $n \to \infty$. Similar reformulations are possible for other cases having asymptotic multicollinearities.

The detailed statistical theory of the NLLS estimator for the nonlinear regression with integrated time series is developed in Park and Phillips (2001) and Chang, Park and Phillips

(2001). The rigorous foundation for the formal asymptotics for the models with a single regressor is develped in the former. In the latter, the theory is extended for more realistic models having multiple integrated regressors that are additively nonlinear, as well as other stationary and deterministic regressors. Efficient estimation and chi-square testing procedures are also proposed there. Park and Phillips (2000) consider the binary choice model with integrated explanatory variables, and derive the asymptotic distributions of the maximum likelihood estimator. Chang and Park (1999) investigate the index-type model such as the neural network and the smooth transition regression, where the multiple integrated regressors are included in an index form. They develop the relevant statistical theory for the efficient estimators and the chi-square testing procedures.

The nonlinear regression with general non-additive regression function with multiple regressors yields the results that are drastically different from those in the single regressor or additively separable multiple regressors case. Roughly, this is because integrated processes are represented in the limit by Brownian motions, and a vector Brownian motion has the recurrence property that is very distinctive from that of an univariate Brownian motion. As can be shown using the asymptotic results obtained in Chang, Park and Phillips (2001), the NLLS estimator is consistent, yet converges very slowly at a logarithmic rate. This is so, as long as we have more than two regressors that are integrated. We thus have very severe *curse of dimensionality* in this case. The curse of dimensionality, however, does not get exacerbated along with the increase in the number of integrated regressors. At any rate, the general nonlinear regression with multiple integrated regressors does not seem to be of reasonable choice in practical applications.

4.2 Nonparametric Regressions with Integrated Processes

The nonparametric estimation of the regression function F in (22) is also possible. To estimate F nonparametrically, it is customary to use the Nadaraya-Watson estimator defined as

$$\hat{F}_n(x) = \frac{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right) y_t}{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)}$$

where K is the kernel function and h_n is the bandwidth parameter. We assume in particular that $\int_{-\infty}^{\infty} K(s)\,ds = 1$, $\int_{-\infty}^{\infty} sK(s)\,ds = 0$ and $\int_{-\infty}^{\infty} s^2 K(s)\,ds \neq 0$, i.e., K is a second-order kernel, and that $h_n \to 0$ as $n \to \infty$. We let F be twice continuously differentiable.

We decompose

$$\hat{F}_n(x) = \hat{F}_n^A(x) + \hat{F}_n^B(x)$$

where

$$\hat{F}_n^A(x) = \frac{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right) F(x_t)}{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)}, \quad \hat{F}_n^B(x) = \frac{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right) u_t}{\displaystyle\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)}$$

The two components $\hat{F}_n^A(x)$ and $\hat{F}_n^B(x)$ represent, respectively, the conditional mean and variance parts of $\hat{F}_n(x)$. In our context, the decomposition may be viewed as the discrete analogue of the representation of a continuous semimartingale as the sum of a bounded variation process and a continuous martingale.

Under suitable regularity conditions, we have

$$
\frac{1}{\sqrt{nh_n^2}} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right) = \frac{\sqrt{n}}{h_n} \int_0^1 K\left(\frac{\sqrt{n}V_n(r) - x}{h_n}\right) dr
$$

$$
= \frac{\sqrt{n}}{h_n} \int_0^1 K\left(\frac{\sqrt{n}V(r) - x}{h_n}\right) dr + o_p(1)
$$

and may deduce as in (13) that

$$
\int_0^1 K\left(\frac{\sqrt{n}V(r) - x}{h_n}\right) dr \to_{a.s.} L(1, 0)
$$

as $n \to \infty$. Consequently, we have

$$
\frac{1}{\sqrt{nh_n^2}} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right) \to_d L(1, 0) \tag{28}
$$

as $n \to \infty$.

We also have

$$
\frac{1}{\sqrt{nh_n^2}} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right) F(x_t) = \frac{\sqrt{n}}{h_n} \int_0^1 K\left(\frac{\sqrt{n}V_n(r) - x}{h_n}\right) F(\sqrt{n}V_n(r)) \, dr
$$

$$
= \frac{\sqrt{n}}{h_n} \int_0^1 K\left(\frac{\sqrt{n}V(r) - x}{h_n}\right) F(\sqrt{n}V(r)) \, dr + o_p(1) \tag{29}
$$

and may derive similarly as in (13) that

$$
\frac{\sqrt{n}}{h_n} \int_0^1 K\left(\frac{\sqrt{n}V(r) - x}{h_n}\right) F(\sqrt{n}V(r)) \, dr
$$

$$
= \frac{\sqrt{n}}{h_n} \int_{-\infty}^{\infty} K\left(\frac{\sqrt{n}s - x}{h_n}\right) F(\sqrt{n}s) L(1, s) \, ds
$$

$$
= \int_{-\infty}^{\infty} K(s) F(x + h_n s) L\left(1, \frac{1}{\sqrt{n}}(x + h_n s)\right) ds
$$

$$
= L(1, 0) \left(F(x) + \frac{1}{2} h_n^2 F''(x) \int_{-\infty}^{\infty} s^2 K(s) \, ds\right) + o(h_n^2) \text{ a.s.}
$$

for large n. It follows that

$$
\frac{1}{\sqrt{nh_n^2}} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right) F(x_t) =_d L(1, 0) \left(F(x) + \frac{1}{2} h_n^2 F''(x) \int_{-\infty}^{\infty} s^2 K(s) \, ds\right) + o_p(h_n^2) \tag{30}
$$

as $n \to \infty$, as long as the approximation error in (29) is small enough. We now have from (28) and (30) that

$$\hat{F}_n^A(x) = F(x) + \frac{1}{2}h_n^2 F''(x) \int_{-\infty}^{\infty} s^2 K(s)\,ds + o_p(h_n^2)$$

for large n.

On the other hand, since

$$\frac{1}{\sqrt{nh_n^2}}\sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h_n}\right) = \frac{\sqrt{n}}{h_n}\int_0^1 K^2\left(\frac{\sqrt{n}V_n(r) - x}{h_n}\right)dr$$

$$= \frac{\sqrt{n}}{h_n}\int_0^1 K^2\left(\frac{\sqrt{n}V(r) - x}{h_n}\right)dr + o_p(1)$$

and we may show as in (13)

$$\int_0^1 K^2\left(\frac{\sqrt{n}V(r) - x}{h_n}\right)dr \to_{a.s.} L(1,0)\int_{-\infty}^{\infty} K^2(s)\,ds$$

as $n \to \infty$, it follows that

$$\frac{1}{\sqrt{nh_n^2}}\sum_{t=1}^{n} K^2\left(\frac{x_t - x}{h_n}\right) \to_d L(1,0)\int_{-\infty}^{\infty} K^2(s)\,ds$$

as $n \to \infty$. We may therefore show as in (15) that

$$\frac{1}{\sqrt[4]{nh_n^2}}\sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right)u_t \to_d \mathbf{MN}\left(0, L(1,0)\int_{-\infty}^{\infty} K^2(s)\,ds\right) \tag{31}$$

as $n \to \infty$. Consequently, it follows from (28) and (31) that

$$\sqrt[4]{nh_n^2}\,\hat{F}_n^B(x) \to_d \mathbf{MN}\left(0, L(1,0)^{-1}\int_{-\infty}^{\infty} K^2(s)\,ds\right) \tag{32}$$

as $n \to \infty$.

We now have from (30) and (32) that the optimal bandwidth is given by

$$h_n = n^{-1/10} \tag{33}$$

which balances off the bias and variance terms. The rate here suggests a bandwidth wider than the usual stationary case, for which the optimal bandwidth is given by $h_n = n^{-1/5}$. This is due to that an integrated regressor has a stochastic trend and observed more scarcely around any fixed spatial point. We therefore need a wider bandwidth to have enough number of observations to balance off the bias and variance terms. The bandwidth optimal for the stationary regressor would make the variance term dominate the bias term in our case.

With the choice of the optimal bandwidth given in (33), we have

$$\hat{F}_n(x) = F(x) + O_p(n^{-1/5})$$

Table 1: Rates of Convergence

		Parametric Approach	Nonparametric Approach
Stationary Regressor		$n^{1/2}$	$n^{2/5} \sim n^{1/2}$
Nonstationary Regressor	H-regular I-regular	$n^{1/2}\dot{\kappa}(n^{1/2})$ $n^{1/4}$	$n^{1/5} \sim n^{1/4}$

for each $x \in \mathbf{R}$. The Nadaraya-Watson estimator for the regression function thus converges at rate $n^{1/5}$, when the regressor is an integrated time series. We may obtain faster convergence rates, if we use a higher-order kernel, i.e., the kernel function K such that $\int_{-\infty}^{\infty} s^k K(s)\, ds = 0$ for some integer $k \geq 2$. It is possible, at least theoretically, that we may make the convergence rate arbitrarily close to $n^{1/4}$ by using higher-order kernels. Note that this convergence rate applies to any regression function F, in particular, for both asymptotically homogeneous and integrable functions. This is in sharp contrast with the nonlinear regression, for which the convergence rate is dependent upon the type of regression function.

It is interesting to compare the convergence rates of the parametric and nonparametric regressions for the regression with an integrated regressor. The comparison is made explicitly in Table 1, and contrasted with the standard stationary regression. The rates for the nonparametric approach depend upon the choice of kernels, and therefore, are given by the ranges. In general, the nonparametric approach yields slower rates of convergence for both statioary and nonstationary regressions. This might be considered as a price to be paid for using unrestrictive and flexible specification for the regression function. The reduction in the convergence rates can however be recovered for stationary regression if we rely on higher-order kernels, which reduce the order of bias term. This is also true for nonstationary regression if the regression function is integrable, though the convergence rates are overall smaller for both parametric and nonparametric approaches.

For the nonstationary regression with asymptotically homogeneous regression function, a quite different picture emerges. In this case, nonparatric approach yields a drastic reduction in convergence rates that cannot be recovered simply by the use of higher-order kernels. For instance, the convergence rate in the parametric linear cointegrating regression is given by n, in comparison with the maximal rate $n^{1/4}$ for the nonparametric regression. This differing convergence rates come from the presence of stochastic trend in the regressor. For the parametric approach, the stochastic trend amplifies the signal if the regression function is increasing, and we may say that the convergence rate is given by the behavior of the function at infinity. This is not so for the nonparametric approach, which essentially estimates the regression function locally at each given point. The nonparametric method thus would not utilize the effect of magnifying signals made by an integrated regressor for the regression function increasing at infinity.

The usual nonparametric approach is thus not very attractive for the estimation of non-linear relationship between integrated time series. An obviously better method is given by the *partial parametric* approach. Suppose that the regression function F can be decomposed as $F = H + G$, where H is homogeneous and G is integrable. We may then specify H parametrically as

$$H(x,\theta) = \alpha|x|^\beta \quad \text{or} \quad \alpha_1|x|^{\beta_1}1\{x \geq 0\} + \alpha_2|x|^{\beta_2}1\{x < 0\}$$

with parameter $\theta = (\alpha, \beta)'$ or $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2)'$, while we leave G unrestricted and estimate it nonparametrically.

The partial parametric model can be estimated consistently in two steps. In the first step, we run the parametric NLLS regression

$$y_t = H(x_t, \theta_0) + w_t$$

where (w_t) is the regression error newly defined by $w_t = u_t + G(x_t)$. This first step NLLS regression yields consistent estimator $\hat{\theta}_n$, say, for θ_0. It is indeed not difficult to show that the NLLS estimator $\hat{\theta}_n$ of θ here has the same limiting distribution as in the regression having error (u_t). The asymptotic behavior of the NLLS estimator $\hat{\theta}_n$ of θ in this regression is thus given by our earlier results. In the second step, we estimate G nonparametrically from the regression

$$y_t - H(x_t, \theta_0) = G(x_t) + u_t$$

The limit theory is not affected by using $H(x_t, \hat{\theta}_n)$ in place of $H(x_t, \theta_0)$ in the second-step regression. The asymptotics for the second-step regression are therefore identical to our theories for nonparametric regression given above.

5. Nonstationary Nonlinearity in Volatility

The nonstationary nonlinearity can also be used in modelling volatilities. If we let

$$y_t = \sigma_t \varepsilon_t \tag{34}$$

where (ε_t) is a sequence of independent and identically distributed random variables with mean zero and unit variance, and (σ_t) signifies the conditional variance at time t given information up to time $t-1$, which is represented by the filtration $(\mathcal{F}_{t-1})$. We thus have $\mathbf{E}(y_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$. Furthermore, we let

$$\sigma_t^2 = F(x_t) \tag{35}$$

where (x_t) is an integrated time series as we specified earlier, which is assumed to be $(\mathcal{F}_{t-1})$-measurable and independent of (ε_t).

The model given by (34) and (35) can be regarded as a stochastic volatility model, where the conditional heterogeneity is generated by a nonlinear function of an integrated process. It is investigated by Park (2002), and referred to as *nonstationary nonlinear heteroskedasticity* (NNH). The function F in (35), which should necessarily be nonlinear

to ensure nonnegativity of the conditional variance, is called the heterogeneity generating function (HGF). If (x_t) in (35) is stationary, then it is said that we have *stationary nonlinear heteroskedasticity*. Instead of modelling the conditional variance σ_t in (34) explcitly as a function of some explanatory variable, the autoregressive conditional heteroskedasticity (ARCH) or the generalized autoregressive conditional heteroskedasticity (GARCH) model assumes that it is given by the past values of the squared process (y_t^2).

It appears that there are many potential examples, for which volatilities may reasonably be specified as functions of integrated processes. For instance, the volatility of the return from a stock may well be given as a function of interest rate or transaction quantity, which are considered by many to be well approximated by integrated processes. We may likewise specify the volatility in the nominal interest rate differentials as a function of inflation, which is also commonly believed to have a unit root. It may also seem reasonable to model the volatility of the spread between forward and spot rates as a function of the level of the spot rate, as is demonstrated in Park (2002).

Using the asymptotics developed in Section 2, we may readily derive various statistical properties of the NNH model. Here we consider the limiting behavior of the sample autocorrelations for the squared process (y_t^2), which is defined as

$$
R_{nk}^2 = \frac{\displaystyle\sum_{t=k+1}^{n} (y_t^2 - \bar{y}_n^2)(y_{t-k}^2 - \bar{y}_n^2)}{\displaystyle\sum_{t=1}^{n}(y_t^2 - \bar{y}_n^2)^2}
$$

where $\bar{y}_n^2$ denotes the sample mean of (y_t^2). Moreover, we investigate the asymptotics for the sample kurtosis

$$
K_n^4 = \frac{\dfrac{1}{n}\displaystyle\sum_{t=1}^{n} y_t^4}{\left(\dfrac{1}{n}\displaystyle\sum_{t=1}^{n} y_t^2\right)^2}
$$

Other sample statistics, of course, can be analyzed similarly, though we do not report the details here.

There are two prominent characteristics that are revealed by many time series observations in economics and finance: volatility clustering and leptokurticity. For any volatility model to be useful in describing data having such properties, it should therefore be able to predict the persistency in the autocorrelations of the squared process and the large values of the sample kurtosis. The asymptotics for R_{nk}^2 and K_n^4 can be obtained similarly as in Section 3, since

$$
\begin{aligned}
y_t^p &= F^{p/2}(x_t)\,\varepsilon_t^p \\
&= (\mathbf{E}\varepsilon_t^p)\,F^{p/2}(x_t) + F^{p/2}(x_t)\,(\varepsilon_t^p - \mathbf{E}\varepsilon_t^p)
\end{aligned}
\tag{36}
$$

and, for all $k = 1, 2, \ldots$,

$$
y_t^2 y_{t-k}^2 = F(x_t)F(x_{t-k})\,\varepsilon_t^2\varepsilon_{t-k}^2
$$

$$= F(x_t)F(x_{t-k}) + F(x_t)F(x_{t-k})\left(\varepsilon_t^2 \varepsilon_{t-k}^2 - 1\right) \tag{37}$$

and their first terms dominate.

5.1 NNH with Integrable HGF

We first consider the NNH model with integrable HGF. Let the HGF F be bounded and integrable. Then we may easily deduce from (36) and our earlier result (16) in Section 3 that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t^p = (\mathbf{E}\varepsilon_t^p)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} F^{p/2}(x_t) + o_p(1)$$

$$\to_d (\mathbf{E}\varepsilon_t^p)\, L(1,0)\int_{-\infty}^{\infty} F^{p/2}(x)\,dx \tag{38}$$

as $n \to \infty$. Note that

$$\sum_{t=1}^{n} F^{p/2}(x_t)\,(\varepsilon_t^p - \mathbf{E}\varepsilon_t^p) = O_p(n^{1/4})$$

for all large n, due to (15).

Moreover, it follows from (37) and (18) in Section 3 that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t^2 y_{t-k}^2 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} F(x_t)F(x_{t-k}) + o_p(1)$$

$$\to_d L(1,0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D_k(y)\,dx\,dy \tag{39}$$

as $n \to \infty$. Similarly as above,

$$\sum_{t=1}^{n} F(x_t)F(x_{t-k})\left(\varepsilon_t^2 \varepsilon_{t-k}^2 - 1\right) = O_p(n^{1/4})$$

for all large n.

We may now readily deduce from (38) and (39) that

$$R_{nk}^2 \to_p R_k^2$$

where

$$R_k^2 = \frac{\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D_k(y)\,dx\,dy}{\displaystyle\mathbf{E}\varepsilon_t^4\int_{-\infty}^{\infty} F^2(x)\,dx}$$

As in Section 3, D_k, $k = 1, 2, \ldots$, denotes the density of $(x_t - x_{t-k})$. If the HGF F is integrable, the sample autocorrelation R_{nk}^2 of the squared process has a nonrandom probability limit R_k^2 for each $k = 1, 2, \ldots$, which we may regard as the asymptotic autocorrelation function of the squared process.

The asymptotic autocorrelation function R_k^2 of the squared process can be analyzed similarly as in Section 3. First, we may easily show that

$$R_k^2 < 1/\mathbf{E}\varepsilon_t^4$$

at all lags $k = 1, 2, \ldots$. Note that $\mathbf{E}\varepsilon_t^4 > 1$. The autocorrelations of the squared NNH process with integrable HGF are therefore expected to be strictly less than unity. In particular, if the distribution of (ε_t) has thick tails and large kurtosis, they should be small at all lags $k = 1, 2, \ldots$. Under Gaussianity, we have $\mathbf{E}\varepsilon_t^4 = 3$, and therefore, $R_k^2 < 1/3$ for all lags $k = 1, 2, \ldots$. Moreover, it follows as in Section 3 that

$$k^{1/2} R_k^2 \to D(0) \frac{\left(\int_{-\infty}^{\infty} F(x)\, dx \right)^2}{\mathbf{E}\varepsilon_t^4 \int_{-\infty}^{\infty} F^2(x)\, dx}$$

as $k \to \infty$. The autocorrelations of the squared NNH process with integrable HGF therefore decay slowly at the rate of $k^{-1/2}$, yielding persistency in memory, i.e., long memory. The NNH model may thus generate long memory in volatility, if an integrable HGF is used.

Now we consider the asymptotic behavior of the sample kurtosis for the NNH model with integrable HGF. Once again, it may easily be deduced from (36) that

$$\frac{1}{\sqrt{n}} K_n^4 \to_d \frac{\mathbf{E}\varepsilon_t^4 \int_{-\infty}^{\infty} F^2(x)\, dx}{L(1,0) \left(\int_{-\infty}^{\infty} F(x)\, dx \right)^2}$$

as $n \to \infty$. For the NNH model with integrable HGF, the sample kurtosis K_n^4 diverges as $n \to \infty$. It is therefore expected to have larger values as the sample size increases, which provides an explanation for the leptokurticity observed in many economic and financial data.

The asymptotics for the sample autocorrelations R_{nk}^2 of the squared process obtained here for the NNH model with integrable HGF are applicable also for the NNH model with trivial asymptotically homogeneous HGF. That is, if F is an asymptotically homogeneous function given by a constant shift of an integrable function as in (19), R_{nk}^2 asymptotically behaves exactly the same as in the case of an integrable function. This is an obvious consequence of the definition of R_{nk}^2, which is invariant with respect to the constant shift of (y_t^2). Contrarily, such invariance does not hold for our definition of the sample kurtosis K_n^4.[3] The results here for K_n^4 therefore do not extend to the NNH model with any asymptotically homogeneous HGF.

5.2 NNH with Asymptotically Homogeneous HGF

We now investigate the NNH model with asymptotically homogeneous HGF. If the HGF F is an asymptotically homogeneous function with asymptotic order κ and limit homogeneous

---

[3]Recall that we assume $\mathbf{E}y_t = 0$ is known to define the sample kurtosis K_n^4.

function H, then we may easily derive that

$$\frac{1}{n}\kappa^{-p/2}(\sqrt{n})\sum_{t=1}^{n} y_t^p = (\mathbf{E}\varepsilon_t^p)\frac{1}{n}\kappa^{-p/2}(\sqrt{n})\sum_{t=1}^{n} F^{p/2}(x_t) + o_p(1)$$

$$\to_d \int_0^1 H^{p/2}(V(r))\,dr \qquad (40)$$

as $n \to \infty$, from (36), our earlier result (20) in Section 3, and since

$$\kappa^{-p/2}(\sqrt{n})\sum_{t=1}^{n} F^{p/2}(x_t)\,(\varepsilon_t^p - \mathbf{E}\varepsilon_t^p) = O_p(n^{1/2})$$

for all large n, due to (12).

Moreover, it follows from (37), (21) in Section 3, and

$$\kappa^{-2}(\sqrt{n})\sum_{t=1}^{n} F(x_t)F(x_{t-k})\left(\varepsilon_t^2\varepsilon_{t-k}^2 - 1\right) = O_p(n^{1/2})$$

for all large n, which follows similarly as above, that

$$\frac{1}{n}\kappa^{-2}(\sqrt{n})\sum_{t=1}^{n} y_t^2 y_{t-k}^2 = \frac{1}{n}\kappa^{-2}(\sqrt{n})\sum_{t=1}^{n} F(x_t)F(x_{t-k}) + o_p(1)$$

$$\to_d \int_0^1 H^2(V(r))\,dr \qquad (41)$$

as $n \to \infty$.

We assume that the asymptotically homogeneous HGF is nontrivial, i.e., it has nonconstant limit homogeneous function H. Then we may easily deduce from (40) and (41) that

$$R_{nk}^2 \to_d R^2$$

where

$$R^2 = \frac{\displaystyle\int_0^1 H^2(V) - \left(\int_0^1 H(V)\right)^2}{\mathbf{E}\varepsilon_t^4 \displaystyle\int_0^1 H^2(V) - \left(\int_0^1 H(V)\right)^2}$$

as $n \to \infty$.

It is important to note that the limit R^2 of R_{nk}^2 is *random* and *independent of k*. The squared NNH process with nontrivial asymptotically homogeneous HGF therefore has a random asymptotic correlation function that is identical for all lags. It is thus expected that their sample autocorrelations are given randomly and independently of the lag order even when the sample size is fairly large. The distribution of the asymptotic correlation R^2, of course, depends upon the fourth moment of (ε_t) and the limit homogeneous function H of the HGF F. It is obtained through simulation in Park (2002) for the Gaussian NNH model with asymptotically homogeneous HGF having limit homogeneous function $H(x) = 1\{x \geq 0\}$ and $H(x) = |x|$. The distribution of R^2 may vary substantially across

different H's. For the models with $H(x) = 1\{x \geq 0\}$, R^2 is likely to take values that could be either very small or very large. In contrast, if $H(x) = |x|$, R^2 is most likely to have moderate values. For the former, the sample paths may show either little or heavy volatility clusterings. The latter in most cases generates samples with moderate volatility clusterings.

We may also easily obtain from (40) that

$$K_n^4 \to_d \frac{\mathbf{E}\varepsilon_t^4 \int_0^1 H^2(V)}{\left(\int_0^1 H(V)\right)^2}$$

as $n \to \infty$. The sample kurtosis remains to be random in the limit for the NNH process with asymptotically homogeneous HGF. Its limit distribution depends upon the limit homogeneous function of the HGF. However, for all NNH processes with asymptotically homogeneous HGF's, the limit distribution of the sample kurtosis has support truncated on the left by the kurtosis $\mathbf{E}\varepsilon_t^4$ of the innovations (ε_t). To see this, note that

$$\left(\int_0^1 H(V)\right)^2 \leq \int_0^1 H^2(V) \text{ a.s.}$$

which holds for all H, due to Cauchy-Schwarz inequality. The inequality is strict unless H is a constant function. The leptokurticity is therefore naturally expected for the NNH processes with asymptotically homogeneous HGF's.

The limit distribution of the sample kurtosis for the NNH model with asymptotically homogeneous HGF is given in Park (2002) for the choices of the limit homogeneous function H, $H(x) = 1\{x \geq 0\}$ and $H(x) = |x|$. For the former, the limit density has the peak at the left boundary of its support. For the latter, the limit distribution for the latter is unimodal, though substantially skewed to the right. We may indeed obtain the limit distribution analytically for the former. As is well known, $\int_0^1 1\{V \geq 0\}$ has arcsine law with density $1/(\pi\sqrt{x(1-x)})$ on the unit interval $(0,1)$. The limit distribution in this case is therefore given by a constant multiple of the reciprocal of arcsine law, which has the density $1/(\pi x \sqrt{x-1})$ over the support $(1, \infty)$.

5.3 Comparisons with SNH and GARCH Models

It is illuminating to compare the properties of the NNH processes with those of the SNH and the GARCH models. In particular, the comparison is made with respect to their ability to explain the commonly observed patterns of the sample autocorrelations for the squared processes. In Table 2,[4] we present the sample autocorrelations of the squared returns for Dow-Jones Industrial Averages (DJ) and Standard & Poor 500 (SP500) stock indices at three different frequencies, i.e., daily, weekly and monthly.

[4]The data were obtained for the period of 1970.1.1 – 2003.3.3, and the returns were calculated as the first differences of logged stock price indices.

Table 2: Sample Autocorrelations of Squared Stock Returns

	Daily		Weekly		Monthly	
k	DJ	SP500	DJ	SP500	DJ	SP500
1	.0931	.1167	.1258	.2431	.0190	.0880
2	.1466	.1532	.0852	.1378	.0259	.0673
3	.0736	.0852	.0571	.1256	.0394	.0720
4	.0199	.0305	.0491	.1310	.0166	.0095
5	.1098	.1412	.0862	.0582	−.0069	−.0045
10	.0119	.0203	.0837	.0774	.0206	.0365
100	.0008	.0026	−.0110	.0094	−.0529	−.0705

Though we have different outcomes for different indices and frequencies, the sample autocorrelations have some common features. First, they are quite persistent. Sometimes, they do not decay at all. If they decay, they do so very slowly. Second, they are far below unity at all lags. Even at the first lag, they are usually pretty small, though they do not seem negligible. The patterns are quite consistent with those predicted by the NNH models. Recall that the autocorrelations for the NNH processes decay very slowly or are given randomly, depending upon whether the HGF is integrable (or trivial asymptotically homogeneous) or nontrivial asymptotically homogeneous.

Of the ARCH-GARCH class of models, we consider the simple, but most popular GARCH(1,1) model given by

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

where ω, α and β are parameters that are assumed to be nonnegative. For the usual GARCH model, we have $\alpha + \beta < 1$. If $\alpha + \beta = 1$, it becomes the so-called integrated GARCH (IGARCH).

For the usual GARCH model with $0 < \alpha + \beta < 1$, the theoretical autocorrelation of the squared process decreases at a geometric rate. Indeed, one may easily see that the sample autocorrelation of the squared process has probability limit given by

$$(\alpha + \beta)^{k-1} \frac{\alpha + \alpha\beta^2 + \beta^3}{1 - \alpha\beta}$$

which is just the k-th autocorrelation of the stationary process (y_t^2). When $\alpha + \beta = 1$, as is the case for IGARCH, the process (y_t^2) becomes an integrated process and R_{nk}^2 converges in probability to unity at all values of k. It is very clear that the observed patterns of the sample autocorrelations are consistent with neither GARCH nor IGARCH. They neither decay at geometric rates, nor are close to unity at any lag.

The SNH models are also inconsistent with the patterns that we observe in Table 2. Under SNH, the sample autocorrelations of the squared process are expected to decrease as

the number of lag increases. To see this more explicitly, we let (x_t) be generated as

$$x_t = \alpha x_{t-1} + v_t$$

with $|\alpha| < 1$. Also, denote by D and D_k respectively the densities of (x_t) and (x_t^k), where $x_t^k = v_t + \alpha v_{t-1} + \cdots + \alpha^{k-1} v_{t-k+1}$, and assume that $\mathbf{E}F^2(x_t) < \infty$ and $\mathbf{E}F^2(x_t^k) < \infty$ for all values of k. Then we have

$$R_{nk}^2 \to_p \frac{\mathbf{E}y_t^2 y_{t-k}^2 - \left(\mathbf{E}y_t^2\right)^2}{\mathbf{E}y_t^4 - \left(\mathbf{E}y_t^2\right)^2}$$

where $\mathbf{E}y_t^2 = \mathbf{E}F(x_t) = \int_{-\infty}^{\infty} F(x)D(x)\,dx$, $\mathbf{E}y_t^4 = \mathbf{E}\varepsilon_t^4 \mathbf{E}F^2(x_t) = \mathbf{E}\varepsilon_t^4 \int_{-\infty}^{\infty} F^2(x)D(x)\,dx$, and

$$
\begin{aligned}
\mathbf{E}\,y_t^2 y_{t-k}^2 &= \mathbf{E}\,F(x_t)F(x_{t-k}) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\alpha^k x + y)F(x)D(x)D_k(y)\,dx\,dy
\end{aligned}
$$

Let F be differentiable with derivative F', so that we have

$$F(\alpha^k x + y) \approx F(y) + \alpha^k x F'(y)$$

for large k. Then we may easily deduce under suitable technical conditions that for large k

$$R_{nk}^2 = c\,|\alpha|^k$$

with some constant $c > 0$. It is thus clear that the sample autocorrelations of the squared SNH processes decrease at a geometric rate as the order k of lags increases up to infinity. This is comparable to the usual GARCH process, and inconsistent with what we observe in Table 2.

6. Concluding Remark

In this paper, we present a unified theory for nonstationary nonlinearity, and lay out new research possibilities in several directions. One direction is omitted in our exposition: nonstationary nonlinearity in statistical inference. Nonstationary nonlinearity can also be used as testing instruments. Nonlinear transformations of integrated time series have some important statistical properties which can be exploited to develop tests with some desirable characteristics. Phillips, Park and Chang (1999) and Chang (2002) demonstrate this possibility very well. The asymptotic theory for unit root models is intrinsically non-Gaussian, rendering standard tests relying on Gaussian limiting distribution not applicable. The limiting Gaussianity of some nonlinear transformations of integrated time series, however, allows us to use the nonlinear IV method to test for a unit root using the normal table. Moreover, nonlinear transformations of integrated processes are asymptotically orthogonal, as long as they are not cointegrated. Chang (2001) uses this property to develop the panel unit root tests that are robust with respect to the cross-sectional dependency.

References

Akonom, J. (1993). Comportement asymptotique du temps d'occupation du processus des sommes partielles. *Annales de l'Institut Henri Poincaré*, 29, 57-81.

Beran, J. (1994). *Statistics for Long-Memory Processes.* New York: Chapman & Hall.

Borodin, A.N. and I.A. Ibragimov (1995). *Limit Theorems for Functionals of Random Walks*, Proceedings of the Steklov Institute of Mathematics, Vol. 195. American Mathematical Soceity.

Chang, Y. (2002). "Nonlinear IV unit root tests in panels with cross-sectional dependency," *Journal of Econometrics*, 110, 261-292.

Chang, Y. and J.Y. Park (1999). "Index models with integrated time series," forthcoming in *Journal of Econometrics*.

Chang, Y., J.Y. Park and P.C.B. Phillips (2001). "Nonlinear econometric models with cointegrated and deterministically trending regressors," *Econometrics Journal*, 4, 1-36.

Chung, K.L. and Williams, R.J. (1990). *Introduction to Stochastic Integration*, 2nd ed. Birkhäuser, Boston.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press: New York.

Park, J.Y. (2002). "Nonlinear nonstationary heteroskedasticity," *Journal of Econometrics*, 110, 383-415.

Park, J.Y. (2003). "Strong approximations for nonlinear transformations of integrated time series," Mimeograph, Department of Economics, Rice University.

Park, J.Y. and P.C.B. Phillips (1998). "Unit roots in nonlinear transformations of integrated time series," Mimeograph, Department of Economics, Rice University.

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, 15, 269-298.

Park, J.Y. and P.C.B. Phillips (2000). "Nonstationary binary choice," *Econometrica*, 68, 1249-1280.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear regressions with integrated time series," *Econometrica*, 69, 1452-1498.

Park, J.Y. and P.C.B. Phillips (2003). "Nonparametric cointegrating regression," Mimeograph, Department of Economics, Rice University.

Phillips, P.C.B. and J.Y. Park (1997). "Nonstationary density estimation and kernel autoregression," Mimeograph, Cowles Foundation for Research in Economics, Yale University.

Phillips, P.C.B., J.Y. Park and Y. Chang (1999). "Nonlinear instrumental variable estimation of an autoregression," forthcoming in *Journal of Econometrics*.

Phillips, P.C.B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics*, 20, 971-1001.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion*. New York: Springer–Verlag.