Nonlinearity, Nonstationarity, and Thick Tails: How They Interact to Generate Persistency in Memory[1]

J. Isaac Miller
and
Joon Y. Park
Department of Economics
Rice University

Abstract

In this paper, we consider nonlinear transformations of random walks driven by thick-tailed innovations with infinite means or variances. In particular, we show how nonlinearity, nonstationarity, and thick tails interact to generate persistency in memory, and we clearly demonstrate that this triad may generate a broad spectrum of persistency patterns. Time series generated by nonlinear transformations of random walks with thick-tailed innovations have asymptotic autocorrelations that decay very slowly as the number of lags increases or do not even decay at all and remain constant at all lags. Depending upon the type of transformation considered and how the model error is specified, they are given by random constants, deterministic functions which decay slowly at polynomial rates, or mixtures of the two. These autocorrelation patterns, along with other sample characteristics of the transformed time series, suggest the possibility that these three ingredients are involved in the data generating processes for many actual economic and financial time series data. We also discuss nonlinear regression asymptotics when the regressor is observable and an alternative regression technique when it is unobservable. We use our model to analyze two empirical applications: exchange rates governed by a target zone and electricity price spikes driven by capacity shortfalls.

First Draft: December 2003
This version: January 2005

Key words and phrases: persistency in memory, nonlinear transformations, random walks, thick tails, stable distributions, target zone exchange rate models, wholesale electricity prices

[1]We are grateful to Yoosoon Chang for many helpful discussions, and we would like to thank participants at the Rice University econometric workshop, Texas Camp Econometrics IX (Fort Worth), the 2004 North American Summer Meeting of the Econometric Society (Providence), and the 2004 Symposium of the Center for Computational Finance and Economic Systems (Houston) for helpful comments.

1. Introduction

In this paper, we consider nonlinear transformations of random walks driven by thick-tailed innovations with infinite variances and possibly infinite means. We show that this specification generates a wide spectrum of patterns of persistency in memory. The triad of nonstationarity, nonlinearity, and thick tails generates time series with asymptotic autocorrelations that decay very slowly as the number of lags increases or do not even decay at all and remain constant at all lags. Depending upon the type of transformation considered and how the model error is specified, they are given by random constants, deterministic functions which decay slowly at polynomial rates, or mixtures of the two. This combination therefore holds the potential to generate the persistent memory patterns that are present in many economic and financial time series data. It may also generate several other prominent properties of many observed time series, such as jumps in the sample paths, excessive volatility and skewness, and leptokurtosis.

The theory underlying our model depends crucially on the type of transformation functions involved. We therefore consider separately two types of functions for the underlying transformations: integrable and asymptotically homogeneous functions. These are the classes of functions introduced by Park and Phillips (1999, 2001) in their studies on nonlinear transformations of integrated time series. Our models with integrable transformations are referred to as *ITS models*, where ITS denotes "integrable transformation of a stable process". On the other hand, we refer to those belonging to the class of models employing asymptotically homogeneous transformations as *AHTS models*, where AHTS signifies "asymptotically homogeneous transformation of a stable process". These models yield very different time series characteristics, in terms of the asymptotics of the sample moments and differing rates of convergence of the parameters estimates from regression.

We establish various time series properties for ITS and AHTS models. ITS models yield time series that have characteristics similar to those of stationary long-memory processes. More precisely, the transformed processes have asymptotic autocorrelations decaying at a polynomial rate with the exact rate depending upon the thickness of the tails of the innovations driving the underlying random walks. We find that they generate autocorrelation patterns consistent with fractionally integrated $\mathrm{I}(d)$ processes with memory parameter d between $0 < d \leq 1/4$. When model error is present, it is also possible to get autocorrelations that have these patterns with additional Gaussian noise or that are determined by pure Gaussian noise at all lags. In contrast, AHTS models generate time series that have asymptotic autocorrelation functions that are constant and do not decay at all. The asymptotic autocorrelations of non-constant asymptotically homogeneous transformations of random walks are unity at all lags, just like those of untransformed random walks.

We study other time series properties of these models, as well. In particular, we derive asymptotics for the sample variance, skewness, and kurtosis. Calculating such statistics for a time series implicitly assumes that the series is stationary, because these statistics are meant to characterize the underlying distribution. These are spurious statistics when applied to a nonstationary time series, but they still carry meaningful information that allows one to distinguish between a stationary series and one that may have a data generating process described by our model. In terms of sample moments, an ITS process behaves like a stationary time series if observed with error. If observed without error, however, it has vanishing sample variance but diverging sample skewness and kurtosis. The sample moments of an AHTS process do not depend upon whether or not it is observed with error. In both cases, the sample variance diverges, and the sample skewness and kurtosis are random in the limit.

The explanatory variable in our models may or may not be observed. If it is observable, then the transformation function may be properly specified and can be consistently estimated by the usual nonlinear least squares method. Here we extend the theories developed by Park and Phillips (2001) for nonlinear regressions with integrated processes to our models driven by thick-tailed random walks. We find that all of the results in Park and Phillips (2001) apply to our models, with different rates of convergence. If, on the other hand, the explanatory variable is not observable, we suggest that it may be estimated together with the transformation function using the extended Kalman filter. Although we do not develop a rigorous theory to justify this approach, the method seems to work reasonably well in extracting the unobserved explanatory variable and estimating the transformation function. We evaluate the performance of the extended Kalman filter by simulations.

As illustrative examples of empirical applications of our models, we consider two models: exchange rates governed by a target zone and electricity price spikes driven by capacity shortfalls. The target zone exchange rate model is an example of an AHTS model with an unobserved explanatory variable. For the actual application, we analyze DEM/FRF exchange rates and extract what is believed to be the fundamental driving the exchange rate. The model for electricity prices is an example of an ITS model with capacity utilization as the observed explanatory variable. Price is specified as an integrable function of a measure of excess capacity, and the model is estimated by standard nonlinear least squares. The fitted model appears to be quite reasonable and it generates time series patterns similar to those of the observed prices.

The remainder of the paper is structured as follows. Section 2 describes the general model. We formalize the concept of thick tails by introducing the class of α-stable distributions, which may have undefined moments. Section 3 defines the transformations we employ in our analysis and derives sample statistics for series generated by ITS and AHTS models. Section 4 discusses regression using ITS and AHTS models. Regression asymptotics are presented for the case in which (x_t) are observable, and we discuss using the extended Kalman filter to estimate the model parameters when (x_t) are not observable. Section 5 presents the two empirical applications, with results from Monte Carlo simulations for the specific functional forms employed there, as well as empirical findings based on our model. Section 6 concludes. Appendix A contains useful lemmas and their proofs, and Appendix B contains proofs of the main results of our theoretical analysis.

2. The Model and Preliminaries

Let (x_t) be a time series generated as

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

where (v_t) is a sequence of random variables, the densities of which have thick tails, as will be specified in more detail below. We consider the time series (y_t), the conditional mean of which is defined as a nonlinear transformation of (x_t) with the transformation function F on $\mathbb{R}$. More specifically, we let

$$y_t = F(x_t) + \varepsilon_t, \tag{2}$$

where the series (ε_t) is assumed to be a martingale difference sequence (an MDS) with respect to a filtration $(\mathcal{F}_t)$ to which (x_{t+1}) are adapted, and $\mathbf{E}\,|\varepsilon_t|^p < \infty$ for some $p \geq 6$. We further assume that (v_t) and (ε_t) are independent, or equivalently that (x_t) are strictly exogenous. This assumption may be relaxed for many of our results, but it is especially convenient when dealing with regression asymptotics.

Let $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2$. We consider two plausible alternative modeling assumptions in this analysis:

$$\sigma_\varepsilon^2 > 0 \tag{3}$$

and

$$\sigma_\varepsilon^2 = 0. \tag{4}$$

The former amounts to including modeling error. In this case, (y_t) are observable only with noise. In the latter case, (y_t) are directly observable, and model error is omitted. In both cases, we have

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

if $\mathbf{E}\left[y_t|\mathcal{F}_{t-1}\right]$ is well-defined. As we explain below, this is not well-defined if the tails of the distribution of the innovations are too thick. Otherwise, the time series (y_t) specified by this model has the conditional mean given as a function of a random walk driven by innovations having thick tails. Our model thus has three ingredients that are commonly observed in many economic and financial time series: nonlinearity, nonstationarity, and thick tails.

We require some technical assumptions about (v_t). Throughout the paper, we assume that (v_t) are iid and have regularly varying tail probabilities, i.e.,

$$\mathbf{P}\{|v_t| > x\} = x^{-\alpha}\ell(x) \tag{5}$$

with $\alpha > 0$ and ℓ a slowly varying function at infinity. Moreover, we let the tail balancing condition hold, i.e.,

$$\frac{\mathbf{P}\{v_t > x\}}{\mathbf{P}\{|v_t| > x\}} \to p, \quad \frac{\mathbf{P}\{v_t < -x\}}{\mathbf{P}\{|v_t| > x\}} \to q \tag{6}$$

as $x \to \infty$, $0 \leq p, q \leq 1$, and $p + q = 1$. The conditions in (5) and (6) are essential for our subsequent theoretical developments. However, the iid assumption of (v_t) can be relaxed at the cost of more involved exposition, as explained below.

As we show in what follows, the standardized sum of (v_t) satisfying the conditions introduced in (5) and (6) converges to what is known as a *stable distribution*. Formally, a random variable v is said to have a stable distribution $S_\alpha(\sigma, \beta, \mu)$, for $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, and μ real, if it has the characteristic function $\varphi(s)$ given by

$$\log \varphi(s) = i\mu s - \sigma^\alpha |s|^\alpha \left(1 - i\beta\varpi(s,\alpha)\right),$$

where

$$\varpi(s,\alpha) = \begin{cases} \operatorname{sgn}(s)\tan(\pi\alpha/2), & \alpha \neq 1 \\ -(2/\pi)\operatorname{sgn}(s)\log|s|, & \alpha = 1 \end{cases}$$

and $\operatorname{sgn}(s)$ is the usual sign function taking values -1, 0, and 1 respectively for $s < 0$, $s = 0$, and $s > 0$. See Samorodnitsky and Taqqu (1994, pg. 5) for the characteristic function of the stable distribution given above.[2] The parameters μ, σ and β are called the shift, scale, and skewness parameters, respectively. The densities of stable distributions are not known in closed form with a few exceptions, notably Gaussian ($\alpha = 2$) and Cauchy ($\alpha = 1$ and $\beta = 0$). For $0 < \alpha < 2$, (v_t) have infinite variance, and for $0 < \alpha \leq 1$, they have an infinite mean, as well.

Central Limit Theory

We now introduce the central limit theory for the sum of iid sequences with thick tails. In particular, the required normalization and centering that play important roles in our subsequent theory will be discussed in detail. We first assume $0 < \alpha < 2$. The case of $\alpha = 2$ will be considered later. Define numerical sequences (a_n) and (b_n) by

$$n\mathbf{P}\{|v_t| > a_n x\} \to x^{-\alpha}$$

as $n \to \infty$, and

$$b_n = \mathbf{E} v_t 1\{|v_t| \leq a_n\}.$$

Then it follows that

$$a_n^{-1}\sum_{i=1}^{n}(v_t - b_n) \to_d S_\alpha(\sigma, \beta, 0), \tag{7}$$

where

$$\sigma^\alpha = \begin{cases} \Gamma(1-\alpha)\cos(\pi\alpha/2), & \alpha \neq 1 \\ \pi/2, & \alpha = 1 \end{cases}$$

and $\beta = 2p - 1$. This is well-known. See, e.g., Feller (1971, Theorem 3, pg. 580). According to our definition of (a_n), we have $C(2-\alpha)/\alpha = 1$ in his formula.[3]

It is well-known that we may set

$$a_n = n^{1/\alpha}\ell(n), \tag{8}$$

[2] The characteristic function of stable distribution given in Borodin and Ibragimov (1995) is in error, and has the term $1 + i\beta\varpi(s,\alpha)$ instead of $1 - i\beta\varpi(s,\alpha)$ as we have here.

[3] The sign $\mp$ in the formula is in error and should be corrected to $\pm$.

where ℓ is slowly varying at infinity. Moreover, we may let

$$b_n = \begin{cases} 0, & 0 < \alpha < 1 \\ \mathbf{E}\left(\sin\left(a_n^{-1} v_t\right)\right), & \alpha = 1 \\ \mathbf{E}\left(v_t\right), & 1 < \alpha < 2 \end{cases} .$$

Note that if $\alpha = 1$ and v_t has a symmetric distribution, then $b_n = 0$ for all n. If condition (5) holds for large $x > 0$ with $\ell(x) = c$ for some constant $c > 0$, then we have

$$a_n = c^{1/\alpha} n^{1/\alpha} \tag{9}$$

as one may easily check.

If (7) holds with (8), then we say that the law of (v_t) belongs to the *domain of attraction of a stable law*. If (7) holds with (9), then it is said to belong to the *domain of normal attraction of a stable law*. Any stable law itself belongs to the domain of normal attraction of a stable law. If (v_t) are iid $S_\alpha(\sigma, \beta, \mu)$, then (5) indeed holds with $\ell(x) = c$, where $c > 0$ is given by

$$c = \begin{cases} \sigma^\alpha / (\Gamma(1-\alpha)\cos(\pi\alpha/2)), & \alpha \neq 1 \\ 2\sigma^\alpha/\pi, & \alpha = 1 \end{cases} .$$

See Brockwell and Davis (1987, pg. 480). Therefore, the conditions we introduced earlier in (5) and (6) are necessary and sufficient in order that the underlying distribution of (v_t) belongs to the domain of attraction of a stable law.

Now we let $\alpha = 2$. In this case, the limit theorem in (7) holds under somewhat weaker conditions than those we require previously, with $b_n = \mathbf{E}(v_t)$ for all n. It is indeed shown in Ibragimov and Linnik (1971, Theorem 2.6.2, pg. 79), e.g., that the condition in (5) alone is sufficient to have (7) with (a_n) specified in (8). Moreover, it is also well-known that (7) holds with (a_n) in (9) and with $\alpha = 2$, if and only if (v_t) have finite variance. See, e.g., Ibragimov and Linnik (1971, Theorem 2.6.6, pg. 92). Similarly as above, we say that the law of (v_t) belongs to the domain of attraction of a normal law if (7) holds with (8). If we have (7) with (9), then the law of (v_t) is said to belong to the domain of normal attraction of a normal law.

From now on, we assume that (v_t) are properly centered. For $1 < \alpha \leq 2$, centering simply requires demeaning or assuming zero mean. For $\alpha = 1$, the proper centering can be difficult and more involved unless we assume that the underlying distribution is symmetric. No centering is necessary for the case of $0 < \alpha < 1$. The limiting distribution has the zero shift parameter, i.e., $\mu = 0$ if (v_t) are centered. Furthermore, we let the adjustment for scales be done apriorily so that the normalized sum of (v_t) converges in distribution to a stable distribution with unit scale parameter, i.e., $\sigma = 1$. The scale of the limit distribution only has a trivial effect on our subsequent results, since the rescaling of (v_t) amounts to merely redefining the transformation function F by a constant multiplication of its argument. The skewness parameter β is not restricted to zero, so we allow for an asymmetric limit distribution for (v_t). Finally, the normalizing sequence (a_n) will be assumed to be given by (8) or (9), depending upon whether the distribution of (v_t) belongs the domain of attraction or of normal attraction of a stable law.

Invariance Principle and Local Time

The central limit theorem in (7) is not sufficient to establish the limit theory for our model. To effectively deal with the nonstationarity in our models, we need a functional central limit theorem. Therefore, we construct a stochastic process V_n on $[0,1]$ by

$$V_n(r) = a_n^{-1} \sum_{t=1}^{[nr]} v_t,$$

where $[x]$ denotes the largest integer which does not exceed x, and invoke the functional central limit theorem as in e.g., Borodin and Ibragimov (1995, pg. 12, hereafter referred to as BI), which yields

$$V_n \to_d V, \tag{10}$$

where V is a standard α-stable Lévy motion on $[0,1]$. That is, $V_0 = 0$ a.s., V has independent increments, and $V_t - V_s$ has a $S_\alpha\left((t-s)^{1/\alpha}, \beta, 0\right)$ distribution for any $0 \le s < t$ and for some $0 < \alpha \le 2$ and $-1 \le \beta \le 1$, as introduced in Samorodnitsky and Taqqu (1994, pg. 113). The processes V_n and V take values in $D[0,1]$, the space of cadlag functions defined on $[0,1]$, and in (10) we have weak convergence probability measures in $D[0,1]$.

The nonlinearity in our models requires some additional tools. In particular, it is necessary to introduce the *local time* L of V. To do so, we first let the *sojourn time* of V in the subset A of $\mathbb{R}$ up to time $t > 0$ be given by

$$m(t, A) = \lambda\{s \in [0,t] | V(s) \in A\},$$

where λ is the usual Lebesgue measure on $\mathbb{R}$. Then the local time of L of V is defined by the Radon-Nikodym derivative of the sojourn time m with respect to λ, i.e.,

$$L(t,x) = \frac{dm}{d\lambda}(t,x).$$

Roughly, the local time L characterizes the portion of time the process V spends at x up to time t. As shown in BI (Theorem 4.1, pg. 18), standard Lévy motions have local times that are continuous with respect to both parameters, if $\alpha > 1$. For $0 < \alpha \le 1$, the local time does not exist.

Serial Correlation in the Innovations

It is possible to consider a more general process (x_t) driven by innovations that are correlated. In particular, we may set $x_t = x_{t-1} + u_t$, where

$$u_t = \sum_{k=0}^{\infty} c_k v_{t-k} \tag{11}$$

and

$$\sum_{k=0}^{\infty} |c_k|^\delta < \infty \tag{12}$$

for some $\delta \in (0, \alpha) \cap [0, 1]$. Under the summability condition in (12), the process (u_t) in (11) is well defined a.s., and if the underlying distribution of (v_t) belongs to the domain of normal attraction and (5) holds with $\ell(x) = c$, then

$$x^{\alpha}\mathbf{P}\{|u_t| > x\} \to c\left(\sum_{k=0}^{\infty}|c_k|^{\alpha}\right)$$

as $x \to \infty$. Therefore, condition (5) holds also for (u_t). Clearly, condition (6) can easily be satisfied with $p = q = 1/2$ if we assume that the underlying distribution of (v_t) is symmetric (and so is that of (u_t)). See for instance Brockwell and Davis (1987, Remarks 1 and 2, pg. 481).

All of our subsequent results hold, at least qualitatively, for (x_t) generated by the more general linear process (u_t) introduced in (11). Some are applicable without any modification. Others just need somewhat obvious modifications and some additional theoretical developments using the Beveridge-Nelson decomposition studied in Phillips and Solo (1992). This, however, will not be done in the present analysis, since it would simply add to expositional complexity without yielding any new features.

3. Time Series Properties of ITS and AHTS Models

In this section, we first introduce the function classes for the transformation F. We subsequently derive the asymptotics for some key sample statistics based on the time series (y_t) generated by ITS and AHTS models. These include the sample autocorrelation function, the sample variance, the sample skewness, and the sample kurtosis. We present asymptotics for ITS and AHTS models separately.

3.1. Classes of Transformation Functions

For the transformation function F in (2), we consider two classes of functions: *integrable* and *asymptotically homogeneous.*

Definition 3.1 (*Integrable Functions*). A transformation F in the class of integrable functions, satisfies

$$|F(x)| < c/(1 + |x|^p)$$

for some constants $c > 0$ and $p > 1$.

Definition 3.2 (*Asymptotically Homogeneous Functions*). A transformation F in the class of asymptotically homogeneous functions satisfies $F(\lambda x) = \nu(\lambda) H(x) + R(x, \lambda)$ for large λ, where H is locally integrable and R is such that

(a) $|R(x, \lambda)| \leq a(\lambda) P(x)$, where $\limsup_{\lambda\to\infty} a(\lambda)/\nu(\lambda) = 0$ and P is locally integrable, or

(b) $|R(x, \lambda)| \leq b(\lambda) Q(\lambda x)$, where $\limsup_{\lambda\to\infty} b(\lambda)/\nu(\lambda) < \infty$ and Q is locally integrable and $Q(x) \to 0$ as $x \to \infty$.

The *asymptotic order* (AO) of an asymptotically homogeneous transformation is $\nu(\lambda)$, and $H(x)$ is the *limit homogeneous function* (LHF). Intuitively, an asymptotically homogeneous transformation exhibits an asymptotically dominant component that is homogeneous. For any asymptotically homogeneous function, we assume throughout this analysis that the LHF is in fact homogeneous.[4]

Asymptotically homogeneous transformations are closely related to functions that are *regular at infinity.*

Definition 3.3 (*Regular-at-Infinity Functions*). A transformation F in the class of regular-at-infinity functions satisfies

$$\lim_{x\to\infty} \frac{F(x)}{x^{\kappa}\ell(x)} = c_1 \quad \text{and} \quad \lim_{x\to-\infty} \frac{F(x)}{|x|^{\kappa}\ell(x)} = c_2$$

for some number $\kappa > -1$, where c_1 and c_2 are constants such that $|c_1| + |c_2| > 0$, and ℓ is slowly varying at infinity, in the sense that $\lim_{\lambda\to\infty} \ell(\lambda x)/\ell(\lambda) = 1$ for any $x > 0$.

The concept of regularity at infinity defines a very broad class of transformations, which includes asymptotically homogeneous transformations, as we show in the proof of the following lemma.

Lemma 3.1 Asymptotically homogeneous functions are regular at infinity.

This useful lemma allows us to tie in general results derived in the mathematics literature for regular-at-infinity functions with the more specific functions discussed in Park and Phillips (1999, 2001) and elsewhere in the econometrics literature. Note that the reverse of this lemma is not true, since regular-at-infinity functions are a broader class of functions than asymptotically homogeneous functions.

An example of an integrable function is a function that is bounded and has compact support. Also, most probability density functions (PDF's) belong to this class, as long as they are bounded and decay at faster rates than $|x|^{-1}$ as $|x| \to \infty$. Variants of such PDF's that are vertically scaled and horizontally shifted are integrable, as well. A possible interpretation of such a transformation is that it returns a strong signal when the value of the underlying random walk is near the mode (or modes) of some PDF-like function. In the empirical section of our analysis, we use an integrable transformation to model the relationship between wholesale electricity prices and the excess utilization. Under our specification, we expect to observe a strong price spike whenever system generation nears capacity.

Park and Phillips (1999) present some useful examples of asymptotically homogeneous transformations. The most common types of asymptotically homogeneous transformations

[4]This is not absolutely necessary, but substantially simplifies our subsequent theory.

in the literature are homogeneous (especially linear), polynomial, and logarithmic. A perhaps more interesting sub-class of asymptotically homogeneous functions are, however, those that resemble rescaled and shifted cumulative distribution functions (CDF's). Any CDF has $\nu(\lambda) = 1$ and $H(x) = 1\{x \geq 0\}$, and rescaled and shifted CDF's have the same AO and an LHF given by some affine transformation of the function $1\{x \geq 0\}$.

Any kind of threshold model is essentially a CDF. For example, if the exogenous signal in a feedforward artificial neural network with one hidden layer follows a random walk, then the model is an AHTS model. Consider a target zone exchange rate model, in which policy actions force the observed exchange rate to stay within a fixed band around the target rate. If the underlying fundamental follows a random walk, then the exchange rate is generated by an AHTS model. We use a family of logistic functions that are parametrized appropriately to model this relationship in the empirical section of our analysis.

In the next subsections, we investigate the time series properties of ITS and AHTS models. More specifically, we develop the asymptotics for the sample autocorrelation function, the sample variance, the sample skewness, and the sample kurtosis. All of these sample statistics are defined in terms of the deviations from the sample mean, and as a result, they are invariant with respect to a shift by a constant. It is therefore obvious that the time series properties of ITS and AHTS models can be characterized by their sample moments only up to a constant term. Consequently, a transformation which is a constant plus an integrable transformation is asymptotically homogeneous but has the same asymptotics as an integrable transformation. For this reason, our subsequent results for ITS models apply also to integrable transformations shifted by arbitrary constants, and those for AHTS models are valid only for asymptotically homogeneous transformations with nonconstant LHF's.

3.2. Asymptotics for ITS Models

We examine the asymptotic behaviors of the sample autocorrelation, variance, skewness, and kurtosis of time series generated by ITS models. As mentioned previously, computing sample statistics for a nonstationary process may be misleading, because they do not represent those of any well-defined underlying distribution. When the process is nonstationary, these are spurious sample statistics. Nevertheless, our results for these spurious statistics allow the comparison of our model with alternative modeling assumptions about the data generating process for the given time series of interest. In fact, the autocorrelation pattern of an ITS process is directly comparable with that of a stationary I(d) process, as we show below.

Our subsequent asymptotic results rely on the following assumptions.

Assumption 3.1 Let the time series (y_t) be generated by (1) and (2) with integrable F, and let (v_t) belong to the domain of attraction of a stable law of order $1 < \alpha \leq 2$ with characteristic function φ satisfying the condition $\varphi(s) \neq 1$ for all $s \neq 0$.

We restrict the order of the limit stable law to $1 < \alpha \leq 2$, because the asymptotics for ITS models crucially rely on the local time of the limit stable process V, which exists only when the stable index of V exceeds unity. The additional condition imposed on the characteristic function of (v_t) merely excludes the possibility of a lattice distribution with support included in the set of integral multiples of some real number. This is not overly restrictive.

We begin with the autocorrelation function, which is the key to determining whether or not a time series exhibits long memory. We define the sample autocorrelation as

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

where k is any nonnegative integer and $\bar{y}_n = \frac{1}{n}\sum_{t=1}^{n} y_t$. We denote by D the PDF of the underlying distribution of (v_t) with respect to the measure μ on $\mathbb{R}$. Moreover, we let D_k be the PDF of $a_k^{-1}(v_1+\cdots+v_k)$ with respect to the same measure. Clearly, we have $D_k = D$, if the process (v_t) itself is α-stable.

Theorem 3.2 (*Asymptotics for R_{nk} – ITS*). Let Assumption 3.1 hold. If $\sigma_\varepsilon^2 = 0$, then

$$R_{nk} \to_p R_k,$$

where

$$R_k = \frac{N_k}{M} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+a_k y)D_k(y)\,dx\,\mu(dy)}{\int_{-\infty}^{\infty} F^2(x)\,dx}. \tag{13}$$

Alternatively, let $\sigma_\varepsilon^2 > 0$. If $1 < \alpha < 2$, then

$$a_n R_{nk} \to_d (1/\sigma_\varepsilon^2)L(1,0)N_k.$$

If on the other hand $\alpha = 2$, then

$$\begin{aligned} n^{1/2}\ell(n)R_{nk} &\to_d \begin{cases} (1/\sigma_\varepsilon^2)L(1,0)N_k & \text{if } \ell(n)\to 0 \\ (1/\sigma_\varepsilon^2)L(1,0)N_k + \mathbb{N}(0,c^2) & \text{if } \ell(n)\to c \text{ for some constant } c \end{cases} \\ n^{1/2}R_{nk} &\to_d \mathbb{N}(0,1) \qquad\qquad\qquad\qquad\qquad\quad \text{if } \ell(n)\to\infty, \end{aligned}$$

where $\mathbb{N}(0,1)$ is a standard normal random variate independent of $L(0,1)$.

This theorem suggests that the autocorrelation pattern of an ITS process is essentially determined by N_k defined in (13). Note that R_k is a constant multiple of N_k. Specifically, the asymptotic autocorrelation function of an ITS process is either R_k when the transformed series is observed directly, or R_k with some random scale and shift when the transformed series is observed with an error. The only exception seems to be when $\alpha = 2$ and $\ell(n) \to \infty$. However, even in this case, the second order term, which is of order smaller than the leading term only by $\ell(n)$, is given as a function of N_k. This is clear from the proof of the theorem.

If (v_t) have an identical stable distribution and $D_k = D$ for all k, then it follows directly from dominated convergence that

$$R_k \to 0$$

as $k \to \infty$, since $a_k \to \infty$ and F is bounded and integrable. The asymptotic autocorrelation of an ITS process thus decreases to zero. The following corollary extends this result to (v_t) in the domain of attraction of a stable law and only asymptotically stable. It also obtains the explicit rate of decay for R_k. We let (φ_k) be the characteristic function of $a_k^{-1}(v_1 + \cdots + v_k)$. It is well-known that if (v_t) belongs to the domain of attraction of a stable law, then $\varphi_k(s) \to \varphi(s)$ pointwise for all $s \in \mathbb{R}$, where φ is the characteristic function of the limiting stable distribution.

Corollary 3.3 (*Rate of Decay of R_{nk} – ITS*). Let Assumption 3.1 hold and assume that (φ_k) are absolutely integrable, $\varphi_k \to \varphi$ in L^1, and D is continuous at the origin. Then we have

$$a_k R_k \to_p D(0) \left(\int_{-\infty}^{\infty} F(x)\, dx \right)^2$$

as $k \to \infty$.

The rate of decay of the asymptotic sample autocorrelation function of an ITS process is therefore a_k^{-1}, which is approximately polynomial for large k. If (v_t) belong to the domain of normal attraction of a stable law, then the rate will be exactly polynomial. On the other hand, if (v_t) only belong to the domain of attraction of a stable law, then the rate will also depend on the slowly varying function $\ell(n)$.

It is well-known that the sample autocorrelations of stationary fractionally integrated processes also decay at polynomial rates. In particular, such autocorrelations decay at the rate of k^{2d-1} where $d \in (0, 1/2)$ is defined as the *degree of fractional integration* or the *memory parameter*. The rates of decay of the autocorrelation of our ITS models and that of an I(d) process with $d \in (0, 1/4]$ will clearly be very similar. On these grounds, it would be easy to mistake a time series generated by an ITS model for a process generated the well-known stationary I(d) model. If the underlying DGP of an observed time series is in fact an ITS model, then such a misspecification would ignore valuable structural information about the process.

We define the observed sample variance, skewness, and kurtosis of a time series (y_t) as

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

$$Q_n^3 = \frac{\frac{1}{n}\sum_{t=1}^{n} (y_t - \bar{y}_n)^3}{\left(\frac{1}{n}\sum_{t=1}^{n} (y_t - \bar{y}_n)^2\right)^{3/2}},$$

and

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

respectively. We would expect that if (y_t) were in fact stationary, with an underlying symmetric distribution with existing fourth moment, then the skewness of that distribution would naturally converge to zero. The variance and kurtosis would converge to some finite number, depending on the rate at which the tails decay, roughly speaking.

In order to compare an ITS process with a stationary process, we introduce the following theorem, which provides limiting distributions of these statistics for (y_t) generated by an ITS model.

Theorem 3.4 (*Asymptotics for* S_n^2, Q_n^3, K_n^4 *– ITS*). Let Assumption 3.1 hold, and define $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$ and $\kappa_\varepsilon^4 = \mathbf{E}\varepsilon_t^4$. Then the asymptotic distribution of the sample variance is

$$S_n^2 \to_p \sigma_\varepsilon^2 \quad \text{for } \sigma_\varepsilon^2 > 0$$

and

$$a_n S_n^2 \to_d L(1,0) \int_{-\infty}^{\infty} F^2(x)\, dx \quad \text{for } \sigma_\varepsilon^2 = 0,$$

that of the sample skewness is

$$Q_n^3 \to_p \tau_\varepsilon^3 / \sigma_\varepsilon^3 \quad \text{for } \sigma_\varepsilon^2 > 0$$

and

$$a_n^{-1/2} Q_n^3 \to_d \frac{\int_{-\infty}^{\infty} F^3(x)\, dx}{\sqrt{L(1,0)} \left(\int_{-\infty}^{\infty} F^2(x)\, dx \right)^{3/2}} \quad \text{for } \sigma_\varepsilon^2 = 0,$$

and that of the sample kurtosis is

$$K_n^4 \to_p \kappa_\varepsilon^4 / \sigma_\varepsilon^4 \quad \text{for } \sigma_\varepsilon^2 > 0$$

and

$$a_n^{-1} K_n^4 \to_d \frac{\int_{-\infty}^{\infty} F^4(x)\, dx}{L(1,0) \left(\int_{-\infty}^{\infty} F^2(x)\, dx \right)^{2}} \quad \text{for } \sigma_\varepsilon^2 = 0$$

as $n \to \infty$.

We can see that ITS models with model error have observed sample statistics that are observationally equivalent to those of stationary processes. This is natural, since the deterministic term or terms of both ITS processes and stationary processes collapse to zero at a faster rate than the error terms. Consequently, if the true DGP of a given process is an ITS model with error, it would be quite easy to confuse it with a stationary process based on these statistics. Again, such a mistake would omit valuable structural information about the DGP that would otherwise enable more accurate inferences.

3.3. Asymptotics for AHTS Models

We derive the same sample statistics for AHTS models as we did for ITS models in the preceding section. The asymptotically homogeneous model is perhaps more important than its integrable counterpart, because the literature is replete with examples of asymptotically homogeneous transformations. If the underlying exogenous variable in such a model is nonstationary and the limiting distribution of the innovations are α-stable (including Gaussian), then our results apply.

In this section, we make the following assumptions.

Assumption 3.2 Let the time series (y_t) be generated by (1) and (2) with asymptotically homogeneous F and (v_t) belonging to the domain of attraction of a stable law.

Note that we do not impose the extra condition on the distribution of the innovation sequence (v_t) that was required for the asymptotics of ITS models. Any lattice distribution is allowed for (v_t) here. Furthermore, the stable parameter for the limit process is allowed to be $0 < \alpha \leq 2$.

Again, we start with an asymptotic result for the sample autocorrelation, which is given by the following theorem.

Theorem 3.5 (*Asymptotics for* R_{nk} *– AHTS*). Let Assumption 3.2 hold. Then we have

$$R_{nk} \to_p 1$$

regardless of whether $\sigma_\varepsilon^2 = 0$ or $\sigma_\varepsilon^2 > 0$.

This result implies that shocks in (y_t) never die out at all, just as shocks in the underlying random walk do not. Given that linear functions are a subset of asymptotically homogeneous transformations, and a linear function of a random walk is itself either a random walk or a random walk with drift, this is not surprising. What is surprising is that this result holds for *any* asymptotically homogeneous transformation, even for those with logarithmic or constant asymptotic orders. It would be impossible to conclude, based on this statistic, that the series (y_t) is stationary. Nevertheless, as we will see in the empirical section of this paper, this asymptotic result does not hold strictly in small samples for the specific functional form discussed there (a rescaled and shifted CDF). As a result of obvious small sample bias, we interpret this result to imply that the rate of decay is very slow, and that relatively large values of R_{nk} may be observed at large values of k.

The following theorem gives us limiting distributions for the remaining statistics.

Theorem 3.6 (*Asymptotics for* S_n^2*,* Q_n^3*,* K_n^4 *– AHTS*). Let Assumption 3.2 hold. Then the asymptotic distribution of the sample variance is

$$\left[\nu^2\left(a_n\right)\right]^{-1} S_n^2 \to_d \int_0^1 H^2\left(V\left(r\right)\right) dr - \left(\int_0^1 H\left(V\left(r\right)\right) dr\right)^2,$$

that of the sample skewness is

$$Q_n^3 \to_d \frac{\int_0^1 \left(H(V(r)) - \int_0^1 H(V(r))dr\right)^3 dr}{\left(\int_0^1 \left(H(V(r)) - \int_0^1 H(V(r))dr\right)^2 dr\right)^{3/2}},$$

and that of the sample kurtosis is

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

as $n \to \infty$, for any $\sigma_\varepsilon^2 \geq 0$.

The implications of the theorem are clear. The observed sample variance of a series generated by an AHTS model diverges at the rate of $\nu^2(a_n)$, which depends not only on the stable parameter α but also on the asymptotic order ν of the transformation. Both the skewness and kurtosis are random, neither converging to zero nor exploding in the limit. In the empirical section of the paper, we simulate a rescaled and shifted CDF to give us a better sense of what the variance, skewness, and kurtosis might look like in that case.

4. Regressions with ITS and AHTS Models

Having established some useful tools for distinguishing series driven by nonlinear transformations of stable random walks from alternative specifications, we now turn to the issue of estimating the transformation function F in (2). Throughout this section, we assume that $\sigma_\varepsilon^2 > 0$ and consider the model given by

$$y_t = F(x_t, \theta) + \varepsilon_t \tag{14}$$

in place of (2), where F is a known function and θ is an unknown parameter vector with true value θ_0. The transformation introduced in (2) is thus parameterized by θ. Here we assume that (ε_t) and (v_t) are independent of each other. We first look at the simpler case, in which (x_t) are observable. In this case, the parameter vector θ can be estimated by nonlinear least squares estimation, as we show below. The asymptotic distributions of the estimators are similar to those derived in Park and Phillips (2001), with rates of convergence consistent with our more general innovations. Subsequently, we consider regression when (x_t) are unobservable. Naturally, this requires additional assumptions, but we suggest obtaining parameter estimates by way of the extended Kalman filter. In the empirical section of the paper, we consider an example of each of these cases.

4.1. Observable Explanatory Variable

Regressions in which (x_t) are observable yield asymptotics results similar to those explored in detail in Park and Phillips (2001). The difference between that analysis and this one is simply that we allow for the regressors to be driven by innovations with thick tails having undefined means or variances, whereas that analysis focused on regressors generated by innovations with finite variances. We let the regression model be given by (14), and denote by $\hat{\theta}_n$ the usual nonlinear least squares estimator of the parameter vector θ given by

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{t=1}^n [y_t - F(x_t, \theta)]^2,$$

where Θ is the parameter set. Moreover, we let $\bar{F}$ denote the vector of partial derivatives of F with respect to θ, i.e.,

$$\bar{F}(x,\theta) = \frac{\partial}{\partial\theta} F(x,\theta).$$

In this light, we present the following two theorems.

Theorem 4.1 (*Asymptotics for* $\hat{\theta}_n$ *– ITS*). Let Assumption 3.1 hold, and let the conditions of Theorem 5.1 in Park and Phillips (2001) be satisfied. Then the limiting distribution of $\hat{\theta}_n$ is given by

$$a_n^{-1/2} n^{1/2} \left(\hat{\theta}_n - \theta_0\right) \to_d \sigma_\varepsilon \left(L\left(1,0\right) \int_{-\infty}^{\infty} \bar{F}\left(x,\theta_0\right) \bar{F}\left(x,\theta_0\right)' dx \right)^{-1/2} W\left(1\right),$$

where W is standard Brownian motion independent of L.

Theorem 4.2 (*Asymptotics for* $\hat{\theta}_n$ *– AHTS*). Let Assumption 3.2 hold, and let the conditions of Theorem 5.2 in Park and Phillips (2001) be satisfied. Then the limiting distribution of $\hat{\theta}_n$ is given by

$$n^{1/2} \bar{\nu}\left(a_n\right)' \left(\hat{\theta}_n - \theta_0\right) \to_d \sigma_\varepsilon \left(\int_0^1 \bar{H}\left(V\left(r\right),\theta_0\right) \bar{H}\left(V\left(r\right),\theta_0\right)' dr \right)^{-1} \int_0^1 \bar{H}\left(V\left(r\right),\theta_0\right) dW\left(r\right),$$

where $\bar{\nu}$ and $\bar{H}$ denote respectively the asymptotic order and limit homogeneous function of $\bar{F}$, and W is standard Brownian motion independent of V.

The asymptotics of $\hat{\theta}_n$ are mixed normal in both cases. In particular, this means that standard errors, t-tests, etc. that are generated by a standard regression package will be asymptotically valid. Thus, when (x_t) are observable, inference from regression is straightforward. Recall that (x_t) are assumed to be strictly exogenous. This is not crucial for the mixed normality of the ITS asymptotics. The same asymptotics hold as long as (x_t) are $(\mathcal{F}_{t-1})$-measurable and a joint invariance principle for (ε_t) and (v_t) holds. However, the mixed normality for the AHTS asymptotics holds only when the two limit stochastic processes V and W are independent, which in turn requires the strict exogeneity of (x_t). Note that the condition $1 < \alpha \leq 2$ is crucial for the ITS asymptotics in Theorem 4.1. As we mentioned earlier, the local time does not exist for $0 < \alpha \leq 1$. The AHTS asymptotics in Theorem 4.2 do not require the existence of local time, and therefore, hold for all $0 < \alpha \leq 2$.

4.2. Unobservable Explanatory Variable

When (x_t) are unobservable, we need additional assumptions and tools to get parameter estimates. First, we consider the case in which (v_t) and (ε_t) are Gaussian. The traditional method for dealing with linear models in which an exogenous variable is unobservable but assumed to follow an autoregressive process with such innovations is to use the Kalman filter (KF) fed into an MLE routine. This technique assumes values for the model parameters, then creates series of $\mathbf{E}\left[x_t|\mathcal{F}_t\right]$ and $\mathbf{var}\left[x_t|\mathcal{F}_t\right]$ for each t, based on some initial values at time $t = 0$ and iterating linear projections. Once these series are created, MLE is used to optimize the model parameters. The series of conditional expectations of (x_t) generated by the optimal parameters are then smoothed in order to take into account information through the end of the sample. It is well-known that the KF yields consistent and asymptotically normal estimates even in the absence of Gaussianity, as long as the underlying models are stationary and the innovations have finite second moments.

Since we are dealing with a nonlinear function F, the KF will not work. To find an alternative to the traditional KF, we turn to the engineering literature. The KF and its variants are widely used in this literature for such applications as tracking satellites and spacecrafts entering Earth's orbit. A common work-around is the extended Kalman filter (EKF), as described in Jazwinski (1970). The EKF is intuitively appealing, since it approximates $F(x_t)$ by expanding around $\mathbf{E}[x_t|\mathcal{F}_{t-1}]$, which is "known" at time $t-1$ (albeit unobservable), using a first-order Taylor series expansion. The econometrics literature also contains alternatives to the EKF. For example, Tanizaki (2000) surveys nonlinear, non-Gaussian state-space modeling using Monte-Carlo techniques.

Implementation of the EKF

We use the EKF to estimate $\mathbf{E}[x_t|\mathcal{F}_t]$ and then smooth these estimates to obtain $\mathbf{E}[x_t|\mathcal{F}_n]$. We summarize the discrete-time EKF below. Our EKF has a measurement equation given by (14) and a transition equation of (1). For convenience of exposition, we use the conventional notation $\cdot_{t|t-1}$ to denote $\mathbf{E}[\cdot_t|\mathcal{F}_{t-1}]$. We also let $\tilde{F}$ be the partial derivative of F with respect to x, i.e., $\tilde{F}(x,\theta) = (\partial/\partial x)F(x,\theta)$. Using this notation, we expand $F(x_t,\theta)$ around $x_{t|t-1}$ to get

$$F(x_t,\theta) \approx F\left(x_{t|t-1},\theta\right) + \tilde{F}\left(x_{t|t-1},\theta\right)\left(x_t - x_{t|t-1}\right). \tag{15}$$

This allows us to write

$$y_t \approx \mu_F(\theta) + \tilde{F}\left(x_{t|t-1},\theta\right)\, x_t + \varepsilon_t,$$

where $\mu_F(\theta)$ is defined as

$$\mu_F(\theta) = F\left(x_{t|t-1},\theta\right) - \tilde{F}\left(x_{t|t-1},\theta\right)\, x_{t|t-1},$$

which is constant at time t. Once the linear approximation is implemented, the EKF works exactly like the linear KF. Defining

$$\omega_{t|\cdot} \equiv \mathbf{var}\left[\left(x_t - x_{t|t-1}\right) \mid \mathcal{F}_\cdot\right]$$

and

$$\Sigma_{t|\cdot} \equiv \mathbf{var}\left[\left(y_t - y_{t|t-1}\right) |\mathcal{F}_\cdot\right]$$

as conditional variances, we replace the usual linear prediction equations of the Kalman filter with

$$x_{t|t-1} = x_{t-1|t-1},$$
$$y_{t|t-1} = F\left(x_{t|t-1},\theta\right),$$

$$\omega_{t|t-1} = \omega_{t-1|t-1} + \sigma_v^2,$$

and

$$\Sigma_{t|t-1} = \omega_{t|t-1} \tilde{F} \left(x_{t|t-1}, \theta \right)^2 + \sigma_\varepsilon^2,$$

where σ_v^2 is the variance of (v_t). The updating equations become

$$x_{t|t} = x_{t|t-1} + \omega_{t|t-1} \tilde{F} \left(x_{t|t-1}, \theta \right) \Sigma_{t|t-1}^{-1} \left(y_t - y_{t|t-1} \right)$$

and

$$\omega_{t|t} = \omega_{t|t-1} - \omega_{t|t-1}^2 \tilde{F} \left(x_{t|t-1}, \theta \right)^2 \Sigma_{t|t-1}^{-1}.$$

MLE is then performed in order to maximize the model parameters, and thus obtain optimal series of $(x_{t|t})$ and $(\omega_{t|t})$. The final step consists of smoothing $(x_{t|t})$ by taking into account information through the end of the sample. This starts at the end of the sample and proceeds back to the beginning of the sample with

$$x_{t|n} = x_{t|t} + \omega_{t|t} \omega_{t+1|t}^{-1} \left(x_{t+1|n} - x_{t+1|t} \right).$$

See Hamilton (1994) or Kim and Nelson (1999) for a more detailed description of the KF and Jazwinski (1970) or Zarchan and Musoff (2000) for the EKF.

Nonstationarity and Thick Tails

The EKF provides a viable alternative in the presence of nonlinearity, but our models have two other nonstandard features: nonstationarity and thick tails. The development of the rigorous statistical theories of the EKF for models having these features is not simple and certainly beyond the scope of this paper. Therefore, we can only provide intuition and conjectures on why we believe the method should yield sensible (though not necessarily optimal) results for such models. The nonstationarity seems to be easier to deal with. Although theory for the KF for nonstationary models with unit roots is not fully developed yet, we expect that the method would work for such models due to the recent results obtained by Park and Phillips (1999, 2001). Moreover, the EKF based on the linear approximation provides a reasonable method to analyze nonlinear and nonstationary models, as well as nonlinear and stationary models. Note that the expansion in (15) yields a relatively better linear approximation for nonstationary models, since in this case $(x_t - x_{t|t-1})$ is stationary and of a stochastic order smaller than that of (x_t). For stationary models, these series are stochastically of the same order of magnitude.

Obviously, the presence of thick tails in the innovations may affect the validity of the EKF used here in a more fundamental manner. If $0 < \alpha \leq 1$ in our model, the mean of (v_t) does not exist and taking conditional expectations of (x_t) becomes meaningless. We must therefore assume $\alpha > 1$, and as long as this condition is fulfilled, the estimates $(x_{t|t})$ and $(x_{t|n})$ can be meaningfully defined. When $1 < \alpha < 2$, however, the variance of (v_t) is still undefined and the EKF cannot be interpreted as iterated projections. In this case, we view the EKF merely as minimizing the sums of squared errors involved in estimating the conditional expectations of (x_t). Of course, both of σ_v^2 and $\omega_{t|\cdot}$ are not properly defined, so we interpret them as pseudo variance and conditional variance of (v_t). We estimate the stable, skewness, and scale parameters of the empirical distribution of the innovation after the conditional expectations of the unobserved series are extracted using the EKF. This two-step methodology might be improved by incorporating the stable distribution directly into the log-likelihood function of the EKF procedure and estimating the parameters of the distribution directly. However, such a one-step procedure would be very difficult to implement, since the stable distribution does not have a closed form solution, except in special cases (Gaussian and Cauchy).

Applying the EKF to our model yields an estimate of the parameter vector θ, as well as extracting conditional expectations of the unobserved (x_t). The standard errors computed by the EKF are, however, incorrect for our nonstandard models. The limiting distributions of the parameter estimates from the EKF are not known in this situation, but they are likely to be non-Gaussian and involve nuisance parameters. Therefore, we perform simulations to obtain the asymptotic distributions and confidence intervals for the parameter estimates in our empirical analysis involving the EKF. For the simulations, we set the innovations (v_t) to be the stable random variates with the stability, skewness, and scale parameters estimated from the data, which are generated independently of the measurement equation errors (ε_t) drawn from the normal distribution with zero mean and the estimated variance. Note that the bootstrap is not a reasonable alternative here, since it generally becomes inconsistent in the presence of thick tails [see, e.g., Hall (1990)].[5]

5. Applications, Simulations, and Empirical Results

We examine two empirical applications of our theoretical models. The first application is a target zone exchange rate model. Target zone models have been used and tested since the 1980's, but the time series properties of such nonlinear transformations of nonstationary processes were not well-known. With this in mind, we introduce an AHTS model in which (x_t) are unobservable. The second application is a model designed to capture observed price "spikes" on wholesale electricity markets by using an integrable transformation of excess capacity. As opposed to the target zone model, which is formulated in the literature and already fits roughly within the framework of one of our econometric models, we propose an electricity price ITS model as an alternative to other approaches.

5.1. Target Zone Exchange Rate Model

Under the European Monetary System (EMS) of the 1980's and 1990's, exchange rates between participating EU countries were allowed to fluctuate within a fixed band around a central parity, which for most participating currencies was $\pm 2.25\%$ until 1990. During this period, the target rate was sometimes realigned by policymakers to reflect underlying changes in the fundamentals of the EU economies.

[5]The difficulty in bootstrapping thick-tailed distributions can be overcome by using the subsample bootstrap, where the size of a bootstrap resample is an order of magnitude smaller than that of the sample. However, it was compared unfavorably by Hall and Jing (1998) to the simulation method used here.

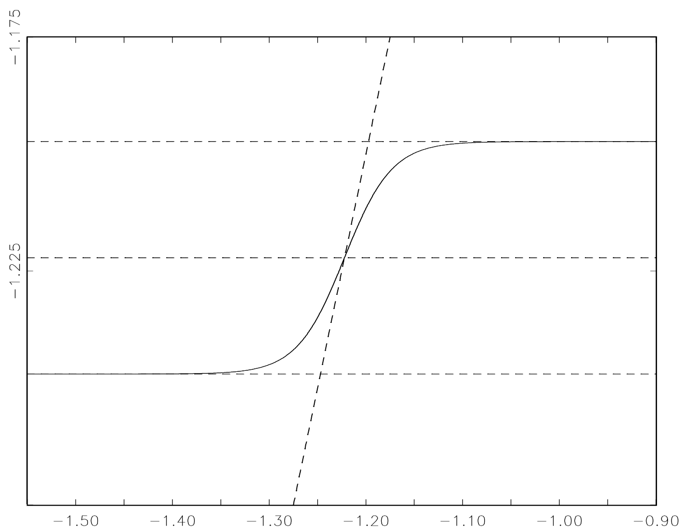

Fig. 5.1.1: Target zone transformation $F(x)$ with our parameter estimates.

Despite the fact that the EMS was replaced by the Euro in the majority of EU countries, There are still a number of countries that follow this type of targeting regime. The IMF classifies exchange rate regimes in its annual report. As of 2004, there are 9 countries that follow an explicit target zone regime. Among them are some of the European countries that aspire to join the EU, such as Hungary and Cyprus, as well as Denmark, which is already part of the EU but chose not to adopt the Euro. A few developing countries also have explicit targeting regimes. An additional 41 (mostly developing) countries fall into the category of having "other conventional fixed peg arrangements" (other than a currency board). As defined by the IMF, this category includes regimes that allow fluctuations of up to $\pm 1\%$ around a central rate. A target zone model with narrow bands may be appropriate for some of these countries.[6] Evidently, even though target zone regimes are not as prevalent as they used to be, they are still important to a number of economies in the world.

Nonlinear Nonstationary Model

Much was written in the economics literature of the 1980's and 1990's about target zone exchange rate models (TZM's), a class of models that attempt to capture the behavior of exchange rates under this type of regime. Perhaps the most widely known of the target zone models was developed by Krugman (1991). The Krugman model postulates that y_t, the log of the exchange rate in such a model, is generated by a nonlinear function of the

[6] In addition, there are 49 countries that fall into the category of "managed floating with no pre-announced path for the exchange rate." While some of these countries explicitly target inflation or monetary aggregates, many of them do not announce explicit targets. Central banks of these countries may be *de facto* anchoring to an exchange rate (either implicitly or explicitly but unannounced) in such a way that a target zone model would capture the behavior of exchange rates.

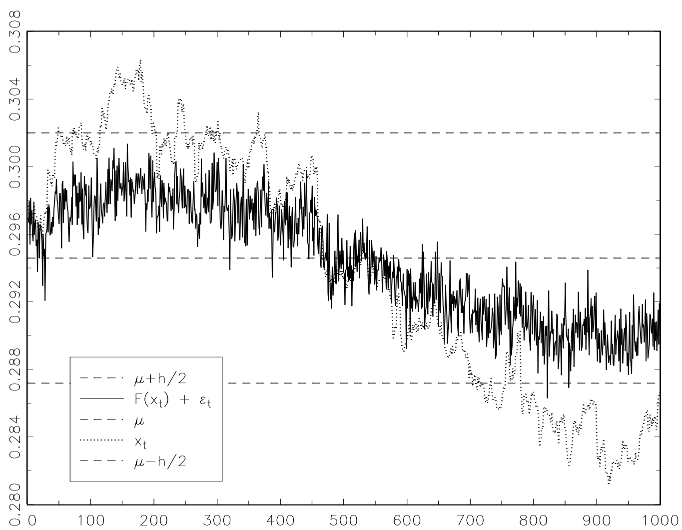

Fig. 5.1.2: Sample simulated exchange rate and fundamental with our parameter estimates.

log of a fundamental x_t. Krugman (1991) derives an "S"-shaped function that maps this fundamental onto the realized exchange rate y_t. Specifically,

$$y_t = K\left(x_t\right) = x_t + B(e^{-\lambda x_t} - e^{\lambda x_t}),$$

where B and λ are model parameters. The transformation is a result of not only policy intervention, but perhaps even more importantly of rational expectations about policy intervention. These expectations bend the function at the edge of the band to create the "S" shape. Stronger expectations of policy intervention correspond to a less steep function – i.e., more deviation from the 45-degree diagonal that maps the fundamental onto the exchange rate under a free floating exchange rate system.

The literature differs on how to treat (x_t). Svensson (1990) assumes that (x_t) follow a regulated Brownian motion. On the other hand, de Jong (1994) and Mark (2001) assume that (x_t) follow a random walk with a constant drift term, possibly included to reflect the belief that money growth is (on average) constant. We assume that (x_t) follow a random walk. This is not unrealistic, as the literature generally agrees on the nonstationarity of the fundamental. Discrete interventions that shift (x_t) should be captured by thick-tailed innovations. Moreover, Dufour and Kurz-Kim (2003) provide evidence that free floating exchange rates follow thick-tailed random walks. In light of the fact such exchange rates follow the fundamental *without* the nonlinear transformation, this supports our assumption that the fundamental itself follows a thick-tailed random walk.

The function K derived in the Krugman model is not compatible with (x_t) that follow a random walk, for the very simple reason that such a series may take values on $(-\infty, \infty)$ and K loses the "S" shape abruptly outside of the band.[7] We postulate an alternative function

[7]In fact, as the fundamental takes arbitrarily large (small) values, $K\left(x_t\right)$ takes arbitrarily small (large)

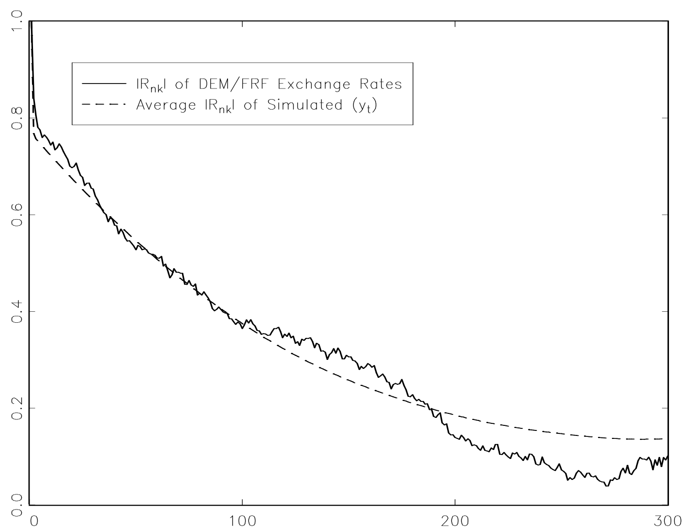

Fig. 5.1.3: $|R_{nk}|$ of actual exchange rate and average $|R_{nk}|$ of simulated exchange rates.

for the model, which follows the intuition of the "S" shape, but allows for a nonstationary fundamental. This function bends at the edge of the band as Krugman's does, but does not allow the exchange rate to deviate from the band when the fundamental becomes too large or too small. In particular, we start with a generalized logistic CDF:

$$F(x) = \nu + h\left(1 + \exp\left\{-\frac{1}{\gamma}(x-\mu)\right\}\right)^{-1},$$

and enforce the restriction $\nu = \mu - h/2$. This restriction serves two purposes. Technically, it is crucial for the identification of μ and γ. Note that they are not identified unless the scale parameter of (v_t) is fixed a priori. Intuitively, it creates a fixed point at μ (the log of the central parity), so that the function returns the same value as its argument at that point. The heart of our TZM is thus given by

$$F(x) = \mu - h/2 + h\left(1 + \exp\left\{-\frac{1}{\gamma}(x-\mu)\right\}\right)^{-1},$$

where μ is the shift parameter, γ is the scale parameter, and h is the bandwidth within which the exchange rate is allowed to fluctuate.[8] Multiplying the CDF by h merely squeezes the function vertically to fit within the band. Figure 5.1.1 illustrates this function. Thus

values! So, K cannot be used for any fundamental thus specified. If K is to be employed, (x_t) must be limited as in Svensson (1990). This misspecification may account for some of the rejections of the Krugman model in the literature.

[8] As in Krugman (1991), both exchange rate and underlying fundamental are expressed in logs. This creates slightly asymmetric bands, but the derivation of our model does not rely on symmetry, as does that of Krugman (1991).

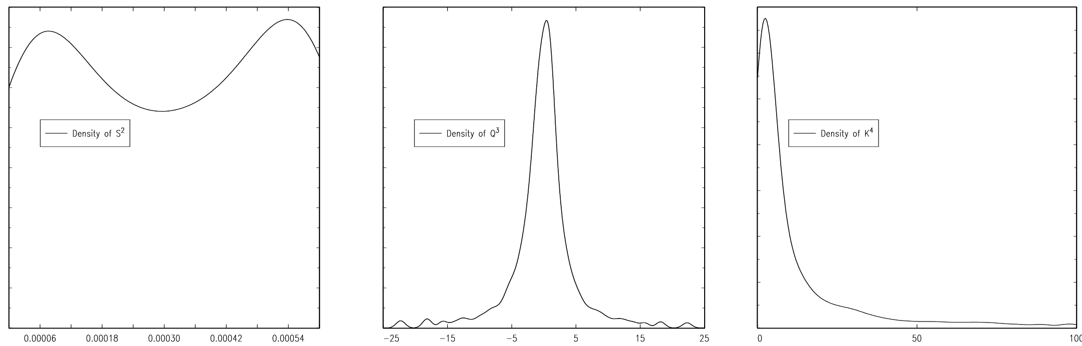

Fig. 5.1.4: Density estimates of the asymptotic distributions of the sample variance, skewness, and kurtosis of (y_t), calculated from an AHTS model generated by the LHF of our TZM with our parameter estimates.

defined, our TZM is an AHTS model. The AO of the function $F(x)$ is unity and the LHF is

$$\left(\mu + \frac{h}{2}\right) \times 1\{x \geq 0\} + \left(\mu - \frac{h}{2}\right) \times 1\{x < 0\},$$

which is homogeneous of degree zero for any positive transformation.

Before examining empirical results from our model, it is useful to examine some results from simulation. Figure 5.1.2 illustrates a single simulated series (y_t) of exchange rates, generated by a series (x_t) of simulated fundamentals following a thick-tailed random walk. This random walk is fed through our TZM with model parameters based on our estimates below. In all of our relevant simulations, pseudo-random numbers drawn from a stable non-Gaussian distribution are obtained using McCulloch's simulation procedure, based on Chambers, et al. (1976). As expected, the observed exchange rate is pushed up or down to stay within the band as the fundamental moves down or up, respectively. In order to generalize our results, we repeat this simulation $5,000$ times with samples of $n = 1,000$. Figure 5.1.3 shows the average of the sample autocorrelation from such simulations. Obviously, there is a small-sample bias, since our asymptotic result for series generated by an AHTS model suggested that the autocorrelations would not die out at all. We can see from the figure that simulated autocorrelations do in fact decay, albeit at a very slow rate.

Figure 5.1.4 shows the asymptotic distributions of the sample variance, skewness, and kurtosis from simulation. The asymptotic distribution of the sample variance somewhat resembles the arcsine distribution. It is well known that the integral of an indicator function on the half real line with a Brownian motion as its argument follows the so-called arcsine distribution [see, e.g., Revuz and Yor (1999, pg. 232)]. Since the asymptotic distribution of the sample variance here is given essentially by the same functional of Levy motion, it is not surprising that we have a similar distribution. A stable random walk has a symmetric spatial distribution and our transformation is symmetric, so the simulated asymptotic distribution of the sample skewness is quite natural. The asymptotic distribution of the sample kurtosis

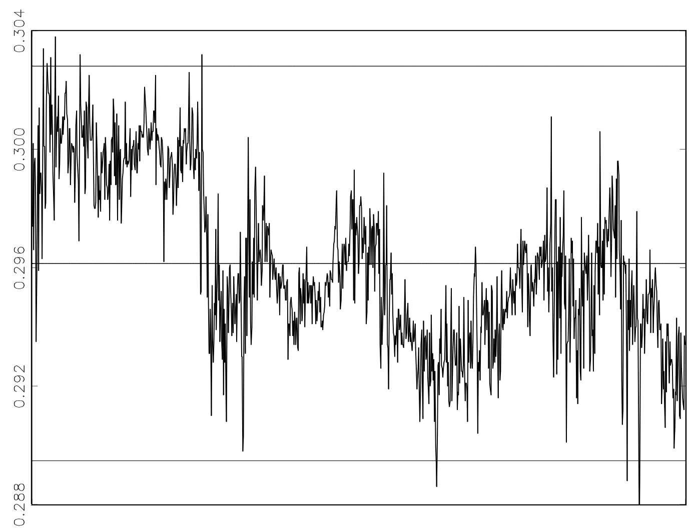

Fig. 5.1.5: DEM/FRF exchange rate (January 12, 1987 - December 31, 1989).

has a mode that appears to be close to 3. We might expect to observe mesokurtosis in such a series, but observed platykurtosis or leptokurtosis of a large degree is also possible.

Data and Empirical Results

We use the log of daily interbank DEM/FRF exchange rates from January 12, 1987 through December 31, 1989 from OANDA (http://www.oanda.com) in this empirical exercise. This is the longest period in which the band was $\pm 2.25\%$ without any realignments of the central parity. The original series (before taking logs) is illustrated in Figure 5.1.5.

Revisiting Figure 5.1.3, it is clear that this series has an autocorrelation function that dies out at a slow rate consistent with our simulations. Estimates of the memory parameter range from 0.35 using the technique developed by Mandelbrot and Wallis (1969) based on the Hurst coefficient to 0.50, 0.54, and 0.60 using the techniques of Geweke and Porter-Hudak (1983) and two refinements of those techniques from Andrews and Guggenberger (2003), respectively. While these reveal a significant small-sample bias compared to our asymptotic prediction that autocorrelations generated by an AHTS model do not die out at all, they suggest that the autocorrelations die out more slowly than those of a stationary fractionally integrated process, which has $d \in (0, 1/2)$.

We find an observed sample variance, skewness, and kurtosis of 0.00009, 0.1996, and 2.3927, respectively. These are consistent with our simulated distributions of the AHTS model discussed above, further suggesting that an AHTS specification may be appropriate for modeling these data. These statistics are not consistent with a stationary series with underlying Gaussian distribution, suggesting that it would be difficult to conclude stationarity and dismiss the richer model.

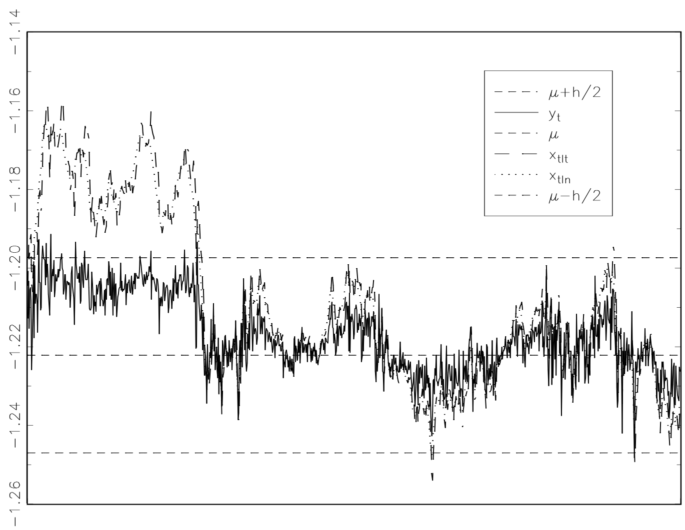

Fig. 5.1.6: Original exchange rate and estimated fundamental.

In Section 4, we discussed limitations of using the EKF in the context of a nonlinear, nonstationary model with thick-tailed innovations. For these reasons, the standard errors generated by the EKF are meaningless, so we do not report them. Our estimates of the model parameters are summarized in the following table. In order to ensure positive values of h, γ, σ_ε^2, and the pseudo-variance of (v_t), we reparameterized the model to estimate the log of these parameters. We allow the autoregressive parameter ρ on (x_t) to vary, in order to test the plausibility of the unit root assumption. The estimated value is in fact very close to unity.

Table 5.1.2

Parameter	*Estimate*
μ	-1.2222
$\ln h$	-3.0021
$\ln \gamma$	-3.7089
ρ	1.0000

Recall that μ may be interpreted as the target for the transformed exchange rate. This parameter gives us a *de facto* target of

$$\exp(\hat{\mu}) = \exp(-1.2222) = 0.2946 \text{ DEM/FRF},$$

where $\hat{\mu}$ is the parameter estimate of μ. Similarly, a *de facto* band of

$$\pm\left\{\exp\left(\frac{1}{2} \times \exp(-3.0021)\right) - 1\right\} = \pm 2.52\%$$

is obtained from the bandwidth estimate $\widehat{\ln h}$, which is clearly very close to the announced bandwidth of ± 2.25.

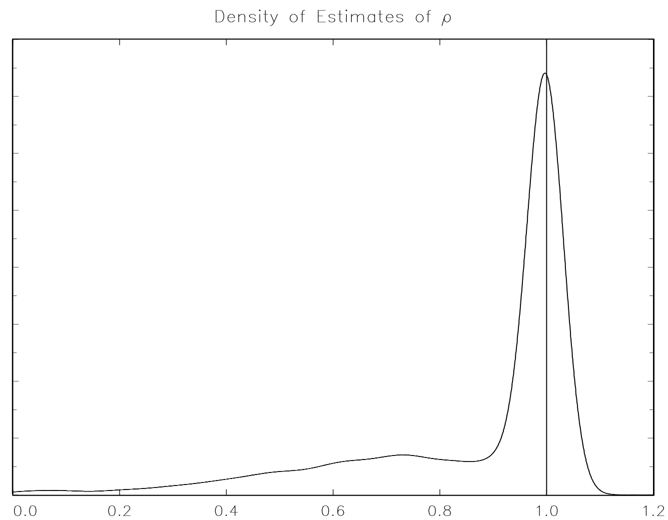

Fig. 5.1.7: Distribution of estimates of ρ obtained through Monte Carlo simulation.

Figure 5.1.6 illustrates the log of the exchange rate (y_t), the conditional expectations of the fundamental ($x_{t|t}$), the smoothed conditional expectations of the fundamental ($x_{t|n}$), the estimated target $\hat{\mu}$, and the estimated band $\hat{\mu} \pm \hat{h}/2$. The fundamental exhibits the expected properties. When the exchange rate approaches one of the edges, the unconstrained fundamental strays from the band. This lends credence to the nonlinear TZM specification. We estimate stable parameters of approximately 1.51 (using the unsmoothed series) and 1.60 (using the smoothed series) for the empirical distribution of the innovations, using the estimation procedure of McCulloch (1986). This suggests that thick tails is an appropriate assumption. Moreover, the 95% confidence interval $(0.1953, 1.0005)$ around $\hat{\rho} = 1.0000$ that we obtain through Monte Carlo simulations suggest that the unit root hypothesis is certainly tenable. The distribution obtained through these simulations and illustrated in Figure 5.1.7 further suggests the plausibility of the unit root hypothesis with a parameter estimate so close to unity.

When de Jong (1994) tested the Krugman model, he concluded that it was misspecified, and the misspecification was specifically blamed on three assumptions: 1) the fundamental follows a random walk, 2) the random walk has Gaussian innovations, and 3) the model does not allow for interventions within the band. We relax the latter two assumptions, but the first seems quite reasonable. Theory and the empirical evidence discussed above support the AHTS model in this situation.

5.2. Electricity Price Spikes

Wholesale electricity markets in most regions of the U.S. and elsewhere are characterized by price "spikes" that occur during peak periods of demand when suppliers are short of

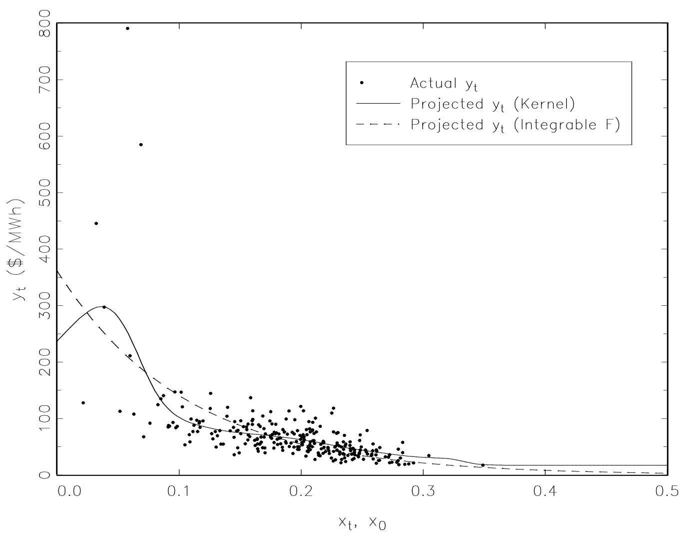

Fig. 5.2.1: Electricity prices vs. excess capacity (April 1, 2002 – December 31, 2002).

capacity. The demand curve is usually assumed to be completely or almost completely inelastic. This is exacerbated by the fact that electricity is not storable in large amounts, so the traditional price-smoothing role of inventories cannot come into play. Policymakers' long-standing goals of equitable and reliable distribution of power necessitate allowing generating units to price above marginal cost, in order to induce marginal units to produce during peak periods. The price may increase significantly in order for these marginal units to cover their fixed costs over the short period of time in which they are necessary to maintain supply at the quantity demanded. This allows marginal units to exercise considerable market power during peak demand periods, particularly when a negative supply shock occurs. Hence we observe the sharp "spikes" that frequently occur in price series from these markets. In light of the California electricity crisis and its aftermath, excessive prices and the abuse of market power have become important issues, with many recent analyses in the energy literature focusing on modeling and forecasting wholesale prices.[9]

Nonlinear Nonstationary Model

Because of the peculiarities of this market, we believe one of the best predictors of price should be excess capacity. Let (u_t) represent capacity utilization, measured as quantity divided by total system capacity on any given day. We consider

$$x_t = 1 - u_t,$$

where x_t represents a measure of excess capacity. Allowing (y_t) to signify electricity prices, we may write the measurement equation of our model in the form of (2). Our empirical analysis of electricity price employs maximum daily load divided by daily scheduled

[9]See, for example, McMenamin and Monforte (2000), Knittel and Roberts (2001), and Stevenson (2002) for a wide variety of statistical and structural techniques applied to this end.

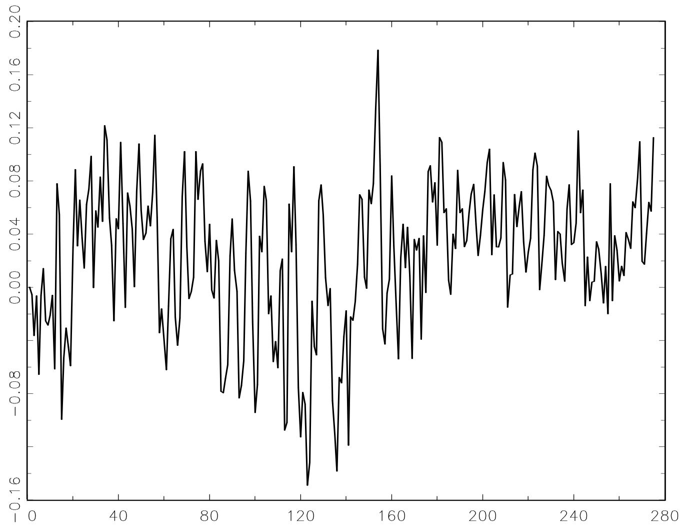

Fig. 5.2.2: Excess capacity (April 1, 2002 – December 31, 2002).

capacity and maximum daily real-time locational marginal price over the period of April 1, 2002 through December 31, 2002 from the Pennsylvania-Jersey-Maryland power pool (http://www.pjm.com). We perform a Nadaraya-Watson kernel regression of (y_t) onto (x_t), which is illustrated in Figure 5.2.1. There is clearly a nonlinear relationship between prices and excess capacity, as even casual observation suggests. We use a rescaled Gaussian density to model price as a function of capacity on the interval $0 \leq x_t < 1$. These endpoints result from the fact that excess capacity must be between 0% and 100%.[10] Our postulated function is

$$F(x) = \begin{cases} \varpi \exp\left\{-\frac{1}{\gamma}(x-\mu)^2\right\} & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where ϖ, γ, and μ are parameters to be estimated.

Data and Empirical Results

Parameter estimates using maximum likelihood estimation are summarized in the following table.

Table 5.2.1

Parameter	*Estimate*	*Std. Error*
ϖ	361.0024	32.9786
γ	0.1061	0.0076
μ	−0.0001	0.0898

[10] Strictly speaking, (x_t) takes values on $[0, 1)$ and therefore cannot be directly specified as a random walk. In particular, it appears to be more reasonable to model (x_t) as a random walk nonlinearly transformed into the unit interval. Such a specification, however, would not affect our subsequent analyses in any critical manner.

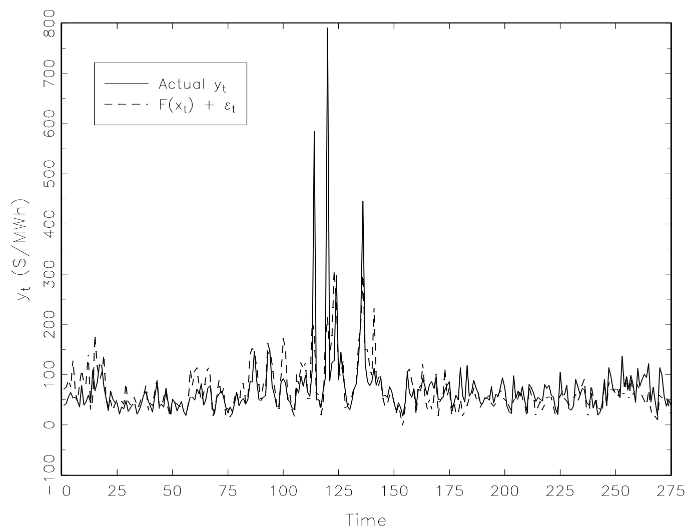

Fig. 5.2.3: Actual and sample simulated electricity prices.

Significant parameter estimates for the first two parameters support our specification. Since the third parameter is just a shift parameter, lack of significance is not a problem. Fitted estimates of (y_t) using these parameters are shown in Figure 5.2.1. This fitted series seems to follow the nonparametric fit quite well, except in the tails, where kernel estimates typically suffer from "empty bin" deficiencies.

We estimate the stable parameter of the empirically observed innovations of (x_t), and then test this series for integratedness, with critical value based on that estimate. Specifically, our estimate is approximately 1.6, using the estimation procedure of McCulloch (1986). It is well-known that the asymptotic distributions of unit root tests in the presence of α-stable innovations depend on that parameter. Chan and Tran (1989), Phillips (1990), Rachev, Mittnik, and Kim (1998), and Rachev and Mittnik (2000) derive the asymptotic distributions of the simple Dickey-Fuller test statistics in the presence of thick tails. Miller (2004) derives the asymptotic distributions in the presence of serial correlation (using ADF test statistics), and tabulates the results for different values of the stable parameter. Table 5.2.2 presents ADF test statistics for different lags, with the initial value subtracted from the series.

Table 5.2.2

Lags	ρ-*test*	t-*test*
4	−44.380	−4.107
8	−31.116	−3.164
12	−9.365	−1.763
16	−14.606	−2.037
20	−5.174	−1.238

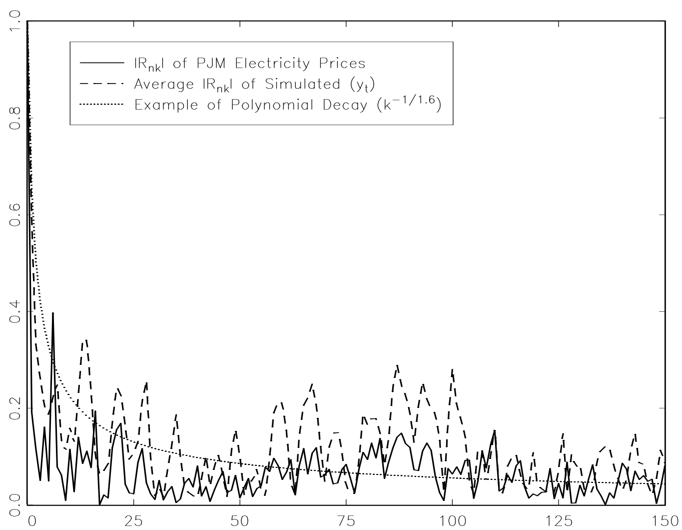

Fig. 5.2.4: $|R_{nk}|$ of actual prices and average $|R_{nk}|$ of simulated prices.

Clearly the decision to reject is highly dependent upon the number of lags considered. As more lags are included, the hypothesis becomes more difficult to reject. The explanation for this seemingly contrary result is apparent from the data. As Figure 5.2.2 shows, capacity utilization is extremely volatile in the short run, but has a clearly nonstationary path over a longer period of time. Hence, stationarity cannot be assumed. The purely random walk AR(1) assumption that we make is a necessary simplification for the purposes of technical ease, but should not substantially alter the results in light of our remarks at the end of Section 2.

Since we have postulated a functional form for F, we can compare observed sample statistics with those calculated from simulation. Unlike the case of the target zone exchange rate model, we have observable (x_t). The only right-hand side series that must be simulated is (ε_t). Figure 5.2.3 illustrates one such sample simulation using the parameters estimated above compared to the actual price series (y_t). Figure 5.2.4 illustrates $|R_{nk}|$ of the actual price series compared to that of simulation averages. The autocorrelation function of (y_t) clearly dies out at a slow rate that is similar to that of simulated averages. Its rate of decay seems to be approximately $k^{-1/1.6}$ as our theory and estimate of the stable parameter predict. Furthermore, estimates of the memory parameter suggest that the rate of decay of the process (y_t) is equivalent to that of a fractionally integrated process with $d \approx 0.14$.[11] Note that this would suggest a rate of decay of $k^{-0.72}$, which is very close to the rate $k^{-0.63}$ suggested by our model. Consequently, if the true DGP were in fact an ITS model, but this spurious parameter were estimated, one might mistakenly ignore the richer specification of the ITS model in favor of a simpler stationary fractionally integrated process.

[11] Techniques based on Mandelbrot and Wallis (1969) obtain 0.15, Geweke and Porter-Hudak (1983) obtain 0.14, and Andrews and Guggenberger (2003) obtain 0.14 or 0.25.

6. Conclusion

In the economics and financial literature, there has been an increasing interest in models that explain observed phenomena such as nonstationarity, persistency in memory, jumps in sample paths, leptokurtosis, and many others. Conventional models may deal with some of these characteristics, but not many conventional models are flexible enough to capture more than a few of these characteristics. We introduce two classes of models, ITS and AHTS models, that embrace three of these attributes – nonlinearity, nonstationarity, and thick tails – and demonstrate that this triad may generate many of these and other observed phenomena. Specifically, our models are generated by nonlinear functions of thick-tailed random walks. Our particular focus on persistency in memory leads us to conclude based on our results that such models may generate a variety of patterns of decay of asymptotic autocorrelations. Two empirical examples illustrate how such models may be applied in practice. While the first utilizes an extension of the Kalman filter to extract an unobservable explanatory variable, the second relies on our theoretical results on nonlinear least squares regression in this context. We assert that due to the time series properties of data generated by our models, it would be easy for a researcher to mistakenly use a more conventional approach – a stationary fractionally integrated model, e.g. – to make inferences.

Our work along these lines suggests many possibilities for future research in this area. It should be possible to generalize such a model to allow for serial correlation in the innovations of explanatory variable, as we discussed in Section 2, and also for ARCH-type measurement errors to be incorporated in the model. Furthermore, more work could be done in the area of nonlinear, nonstationary Kalman filtering. In light of recent work by Park and Phillips (2001), it may be possible to work out the asymptotic properties of ML estimators in such a framework. Extending the theory of this technique to ITS and AHTS models, in which the innovations have infinite variances, might prove to be quite difficult. We leave these challenges to future research.

References

Andrews, D.W.K. and P. Guggenberger (2003). "A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter." *Econometrica*, 71, 675-712.

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

Box, G.E.P. and G.M. Jenkins (1970). *Time Series Analysis, Forecasting and Control.* Holden Data: San Francisco.

Brockwell, P.J. and R.A. Davis (1987). *Time Series: Theory and Methods.* Springer-Verlag: New York.

Borodin, A.N. and I.A. Ibragimov (1995). *Limit Theorems for Functionals of Random Walks.* American Mathematical Society: Providence.

Chambers, J.M., C.L. Mallows, and B.W. Stuck (1976). "A Method for Simulating Stable Random Parameters," *Journal of the American Statistical Association*, 71, 340-44.

Chan, N.H. and L.T. Tran (1989). "On the First-Order Autoregressive Process with Infinite Variance," *Econometric Theory*, 5, 354-362.

Chang, Y. and J.Y. Park (2004). "Endogeneity in Nonlinear Regressions with Integrated Time Series," mimeograph, Department of Economics, Rice University.

de Jong, F. (1994). "A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model," *Journal of Applied Econometrics*, 9, 31-45.

Dufour, J.-M. and J.R. Kurz-Kim (2003). "Exact Tests and Confidence Sets for the Tail Coefficient of α-Stable Distributions," Deutsche Bundesbank Discussion Paper, 16/2003.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II. John Wiley & Sons: New York.

Geweke, J. and S. Porter-Hudak (1983). "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-38.

Granger, C.W.J. (1980). "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227-38.

Granger, C.W.J. and R. Joyeux (1980). "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-29.

Granger, C.W.J. and P. Newbold (1974). "Spurious Regression in Econometrics," *Journal of Econometrics*, 2, 111-20.

Hall, P. (1990). "Asymptotic Properties of the Bootstrap for Heavy-tailed Distributions," *Annals of Probability*, 18, 1342-1360.

Hall, P. and B.-Y. Jing (1998). "Comparison of Bootstrap and Asymptotic Approximations to the Distribution of a Heavy-tailed Mean," *Statistica Sinica*, 8, 887-906.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press: Princeton.

Ibragimov, I.A. and Yu.V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhof: Groningen.

International Monetary Fund (2004). "Annual Report of the Executive Board for the Financial Year Ended April 30, 2004," International Monetary Fund: Washington.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press: New York.

Kim, C.-J. and C.R. Nelson (1999). *State-Space Models with Regime Switching.* MIT Press: Cambridge.

Knittel, C.R. and M.R. Roberts (2001). "An Empirical Examination of Deregulated Electricity Prices," POWER working paper, 87.

Krugman, P.R. (1991). "Target Zones and Exchange Rate Dynamics," *The Quarterly Journal of Economics*, 106, 669-82.

Mandelbrot, B.B. and J.R. Wallis (1969). "Some Long-Run Properties of Geophysical Records," *Water Resources Research*, 5, 321-40.

Mark, N.C. (2001). *International Macroeconomics and Finance: Theory and Econometric Methods*. Blackwell Publishers: Malden, MA.

McCulloch, J.H. (1986). "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics: Simulation and Computation*, 15, 1109-136.

McCulloch, J.H. (1994), "Numerical Approximation of the Symmetric Stable Distributions and Densities," mimeograph, Department of Economics, Ohio State University.

McMenamin, J.S. and F.A. Monforte (2000). "Statistical Approaches to Electricity Price Forecasting," in *Pricing in Competitive Electricity Markets* (A. Faruqui and K. Eakin, eds.), Kluwer Academic Publishers: Boston, 249-63.

Miller, J.I. (2004). "Unit Root Tests on Time Series with Serially Correlated, Thick-Tailed Innovations," mimeograph, Department of Economics, Rice University.

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

Park, J.Y. (2003). "Nonstationary Nonlinearity: An Outlook for New Opportunity," mimeograph, Department of Economics, Rice University.

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for Nonlinear Transformation of Integrated Time Series," *Econometric Theory*, 15, 269-98.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117-61.

Phillips, P.C.B. (1990)."Time Series Regression With a Unit Root and Infinite-Variance Errors," *Econometric Theory*, 6, 44-62.

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

Rachev, S.T., S. Mittnik, and J.-R. Kim (1998). "Time Series with Unit Roots and Infinite-Variance Disturbances," *Applied Mathematics Letters*, 11.5, 69-74.

Rachev, S.T., and S. Mittnik (2000). *Stable Paretian Models in Finance.* John Wiley & Sons: New York.

Revuz, D. and M. Yor (1999). *Continuous Martingales and Brownian Motion.* Springer-Verlag: New York.

Samorodnitsky, G. and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes.* Chapman & Hall: New York.

Stevenson, M. (2002). "Filtering and Forecasting Spot Electricity Prices in the Increasingly Deregulated Australian Electricity Market," presented at the Tenth Annual Symposium of The Society for Nonlinear Dynamics and Econometrics, Atlanta.

Svensson, L.E.O. (1991). "The Term Structure of Interest Rate Differentials in a Target Zone: Theory and Swedish Data," *Journal of Monetary Economics*, 28, 87-116.

Tanizaki, H. (2003). "Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Techniques: A Survey and Comparative Study," in *Handbook of Statistics: Stochastic Processes: Modeling and Simulation* (C.R. Rao and D.N. Shanbhag, eds.), North-Holland: Amsterdam, 871-929.

Zarchan, P. and H. Musoff (2000). *Fundamentals of Kalman Filtering: A Practical Approach.* American Institute of Aeronautics and Astronautics: Reston, VA.

Appendix A: Useful Lemmas and Their Proofs

Lemma A1 Consider integrable F and (v_t) belonging to the domain of attraction of a stable law. If we define

$$M_n = a_n n^{-1} \sum_{t=1}^{n} F(x_t),$$

then we have

$$\sup_{n\geq 1} \mathbf{E} \left| M_n \right|^2 < \infty,$$

and therefore, in particular, (M_n) is uniformly integrable.

Proof of Lemma A1 Let $\hat{F}$ be the Fourier transform of F, i.e.,

$$\hat{F}(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} F(x) dx.$$

As in the proof of Theorem 2.1 of BI (pg. 143), we may assume without loss of generality that $\hat{F}$ has compact support. Moreover, since F is bounded, so is $\hat{F}$. Therefore, we may write

$$\begin{aligned} F(x_t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda x_t} \hat{F}(\lambda) d\lambda \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda (a_n^{-1} x_t)} \hat{F}(a_n^{-1}\lambda) d(a_n^{-1}\lambda) \end{aligned}$$

and consequently, we have

$$\begin{aligned} M_n &= a_n n^{-1} \sum_{t=1}^{n} F(x_t) \\ &= a_n \int_0^1 F\left(a_n V_n\left(r\right)\right) dr \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{F}(a_n^{-1}\lambda) \int_0^1 e^{-i\lambda V_n(r)} dr d\lambda \end{aligned}$$

as one may easily see. The last line follows from Fubini's Theory.

Now note that $\hat{F}(a_n^{-1}\cdot)$ vanishes outside the interval $[-ca_n, ca_n]$ for some constant $c > 0$, since we have assumed that $\hat{F}$ has compact support. Moreover, if we let

$$I(F) = \int_{-\infty}^{\infty} F(x) dx,$$

then we may write

$$\int_{-\infty}^{\infty} \frac{\left|\hat{F}(a_n^{-1}\lambda) - I(F)\right|^2}{1+|\lambda|^2} d\lambda = \int_{-\infty}^{\infty} \frac{\left| \int_{-\infty}^{\infty} e^{i(a_n^{-1}\lambda x)} F\left(x\right) dx - \int_{-\infty}^{\infty} F\left(x\right) dx \right|^2}{1+|\lambda|^2} d\lambda \to 0$$

as $n \to \infty$ by dominated convergence, since $|F(x)|$ is bounded. Also, note that

$$\int_{-\infty}^{\infty} \frac{|I(F)|^2}{1+|\lambda|^2} d\lambda < \infty.$$

The conditions for Theorem 2.1 of BI (pg. 85) are thus satisfied. Following the proof of Theorem 2.1 of BI (pp. 87-88), we may now readily deduce that

$$\begin{aligned}\mathbf{E}\left|M_n\right|^2 &= \mathbf{E}\left|\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{F}(a_n^{-1}\lambda)\int_0^1 e^{-i\lambda V_n(r)} dr d\lambda\right|^2 \\ &\leq c\left(\int_{-\infty}^{\infty} \frac{|I(F)|^2}{1+|\lambda|^{\alpha}} d\lambda\right)^{1/2}\end{aligned}$$

for some constant $c > 0$. See Equation (2.14) of BI (pg. 88). This completes the proof.

Lemma A2 (*Asymptotics for Some Sample Moments – ITS*). Let Assumption 3.1 hold, and define $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2$ and $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$. The following sample moments have asymptotic distributions and rates of convergence given as follows:

(a) $a_n n^{-1}\sum_{t=1}^n F^2(x_t) \to_d L(1,0)\int_{-\infty}^{\infty} F^2(x)\,dx$

(b) $a_n^{1/2} n^{-1/2}\sum_{t=1}^n F(x_t)\varepsilon_t \to_d MN\left(0, \sigma_\varepsilon^2 L(1,0)\int_{-\infty}^{\infty} F^2(x)\,dx\right)$

(c) $a_n n^{-1}\sum_{t=k+1}^n F(x_t)F(x_{t-k}) \to_d L(1,0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+a_k y)D_k(y)\,dx\,\mu(dy)$

(d) $a_n n^{-1}\sum_{t=1}^n F^3(x_t) \to_d L(1,0)\int_{-\infty}^{\infty} F^3(x)\,dx$

(e) $a_n n^{-1}\sum_{t=1}^n F^4(x_t) \to_d L(1,0)\int_{-\infty}^{\infty} F^4(x)\,dx$

(f) $a_n^{1/2} n^{-1/2}\sum_{t=1}^n F^2(x_t)\varepsilon_t \to_d MN\left(0, \sigma_\varepsilon^2 L(1,0)\int_{-\infty}^{\infty} F^4(x)\,dx\right)$

(g) $a_n^{1/2} n^{-1/2}\sum_{t=1}^n F^3(x_t)\varepsilon_t \to_d MN\left(0, \sigma_\varepsilon^2 L(1,0)\int_{-\infty}^{\infty} F^6(x)\,dx\right)$

(h) $a_n n^{-1}\sum_{t=1}^n F(x_t)\varepsilon_t^2 \to_d \sigma_\varepsilon^2 L(1,0)\int_{-\infty}^{\infty} F(x)\,dx$

(i) $a_n n^{-1}\sum_{t=1}^n F^2(x_t)\varepsilon_t^2 \to_d \sigma_\varepsilon^2 L(1,0)\int_{-\infty}^{\infty} F^2(x)\,dx$

(j) $a_n n^{-1}\sum_{t=1}^n F(x_t)\varepsilon_t^3 \to_d \tau_\varepsilon^3 L(1,0)\int_{-\infty}^{\infty} F(x)\,dx$

Proof of Lemma A2 (*Asymptotics for Some Sample Moments – ITS*).

For the proof of part (a), we only need to note that F^2 is integrable. The stated result then follows directly from Theorem 2.1 in BI (pg. 143).

The result in part (b) follows essentially from the proof of Theorem 3.2 in Park and Phillips (2001) with the appropriate substitution for the rate of convergence a_n of a stable process. Here we just give a sketch of the proof. Define

$$\begin{aligned}M_n(r) &= a_n^{1/2} n^{-1/2}\sum_{t=1}^{j-1} F\left(a_n V_n\left(\frac{t-1}{n}\right)\right)\left(U\left(\frac{\tau_{nt}}{n}\right) - U\left(\frac{\tau_{n,t-1}}{n}\right)\right) \\ &\quad + a_n^{1/2} n^{-1/2} F\left(a_n V_n\left(\frac{j-1}{n}\right)\right)\left(U(r) - U\left(\frac{\tau_{n,j-1}}{n}\right)\right),\end{aligned}$$

where $\tau_{n,j-1}/n < r \leq \tau_{nj}/n$, (τ_{ni}) is a time change as specified in Park and Phillips (2001), and U is the Brownian motion constructed from (ε_t). Then we may write

$$a_n^{1/2} n^{-1} \sum_{t=1}^{n} F(x_t)\,\varepsilon_t = M_n\left(\frac{\tau_{nn}}{n}\right), \tag{16}$$

and it follows that

$$\begin{aligned}
[M_n](1) &= a_n n^{-1} \sum_{t=1}^{j-1} F\left(a_n V_n\left(\frac{t-1}{n}\right)\right)^2 \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n}\right) \\
&\quad + a_n n^{-1} F\left(a_n V_n\left(\frac{j-1}{n}\right)\right)^2 \left(r - \frac{\tau_{n,j-1}}{n}\right) \\
&= \sigma_\varepsilon^2 a_n \int_0^r F\left(a_n V_n(s)\right)^2 ds\,(1 + o_p(1)) \\
&\to_d \sigma_\varepsilon^2 L(r,0) \int_{-\infty}^{\infty} F^2(x)\,dx,
\end{aligned}$$

uniformly in $r \in [0,1]$. Due to the independence of U and V_n, M_n becomes asymptotically independent of V. The stated result thus follows exactly as in Park and Phillips (2001).

The proof for part (c) is much more involved. For the sake of clarity, we first consider the case in which $k = 1$ and $a_1 = 1$. We also use the notation D for D_1 to simplify the notation. Write

$$\sum_{t=2}^{n} F(x_t)F(x_{t-1}) = \sum_{t=2}^{n} (GF)(x_{t-1}) + \sum_{t=2}^{n} F(x_{t-1})u_t, \tag{17}$$

where

$$G(x) = \int_{-\infty}^{\infty} F(x+y)D(y)\mu(dy)$$

and

$$u_t = F(x_t) - G(x_{t-1})$$

for $t \geq 1$. Obviously, G is well-defined for all $x \in R$, since F is bounded. Note that

$$\mathbf{E}\left(F(x_t)|\mathcal{F}_{t-1}\right) = G(x_{t-1}),$$

where $(\mathcal{F}_t)$ is a filtration such that $\mathcal{F}_t$ is defined by the σ-field generated by $(x_s)_{s=1}^t$ for each $t \geq 1$. Consequently, $(u_t, \mathcal{F}_t)$ is an MDS.

It is easy to see that G is bounded. Therefore, since F is integrable, so is GF. Furthermore, we have

$$\int_{-\infty}^{\infty} (GF)(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D(y)dx\mu(dy)$$

due to the Fubini's theorem. It therefore follows from Theorem 2.1 in BI (pg. 143) that

$$\begin{aligned} a_n n^{-1} \sum_{t=2}^{n} (GF)(x_{t-1}) &\to_d L(1,0) \int_{-\infty}^{\infty} (GF)(x)dx \\ &= L(1,0) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x)F(x+y)D(y)dx\mu(dy). \end{aligned} \tag{18}$$

Now, if we can show

$$a_n n^{-1} \sum_{t=2}^{n} F(x_{t-1})u_t = o_p(1), \tag{19}$$

then the stated result would be immediate from (17) and (18).

To establish (19), we will prove that

$$\mathbf{E}\left(a_n n^{-1} \sum_{t=2}^{n} F(x_{t-1})u_t\right)^2 \to 0. \tag{20}$$

Using the fact that $(F(x_{t-1})u_t, \mathcal{F}_t)$ is an MDS, and subsequently applying the law of iterated expectations, we may deduce that

$$\begin{aligned} \mathbf{E}\left(a_n n^{-1} \sum_{t=2}^{n} F(x_{t-1})u_t\right)^2 &= a_n n^{-1} \mathbf{E}\left(a_n n^{-1} \sum_{t=2}^{n} F^2(x_{t-1})u_t^2\right) \\ &= a_n n^{-1} \mathbf{E}\left(a_n n^{-1} \sum_{t=2}^{n} F^2(x_{t-1}) \mathbf{E}\left(u_t^2 | \mathcal{F}_{t-1}\right)\right). \end{aligned} \tag{21}$$

Moreover, we may write

$$\mathbf{E}\left(u_t^2 | \mathcal{F}_{t-1}\right) = H(x_{t-1}) - G^2(x_{t-1}),$$

where

$$H(x) = \int_{-\infty}^{\infty} F^2(x+y)D(y)\mu(dy).$$

It is easy to see that H is well defined and bounded, just like G introduced above.

Now we define

$$\begin{aligned} M_n &= a_n n^{-1} \sum_{t=2}^{n} F^2(x_{t-1}) \mathbf{E}\left(u_t^2 | \mathcal{F}_{t-1}\right) \\ &= a_n n^{-1} \sum_{t=2}^{n} (HF^2 - G^2F^2)(x_{t-1}). \end{aligned}$$

Then we have, again due to Theorem 2.1 in BI (pg. 143),

$$M_n \to_d L(1,0) \int_{-\infty}^{\infty} (HF^2 - G^2F^2)(x)dx,$$

since $HF^2 - G^2F^2$ is integrable. We may therefore deduce that

$$\mathbf{E}\, M_n \to \mathbf{E}\left[L(1,0)\int_{-\infty}^{\infty}(HF^2 - G^2F^2)(x)dx\right],$$

since (M_n) is uniformly integrable as shown in Lemma A1. Consequently, (20) follows from (21), as was to be shown. The proof for $k = 1$ is now complete. The proof for the general case is obvious and omitted.

The proofs of parts (d) and (e) are identical to that of part (a), and the proofs of parts (f) and (g) are the same as that of part (b).

To prove part (f), we rewrite the sample moment as

$$\sum_{t=1}^{n} F\left(x_t\right)\varepsilon_t^2 = \sum_{t=1}^{n} F\left(x_t\right)\mathbf{E}\varepsilon_t^2 + \sum_{t=1}^{n} F\left(x_t\right)\left(\varepsilon_t^2 - \mathbf{E}\varepsilon_t^2\right).$$

The distribution of the first term is obvious. To get the stated result, we just need to show that the second term is $o_p(a_n^{-1}n)$. If we define $\varepsilon_{2t} = \left(\varepsilon_t^2 - \mathbf{E}\varepsilon_t^2\right)$, then (ε_{2t}) is iid and independent of (v_t). Moreover, it is clear that $\mathbf{E}\varepsilon_{2t}^2 < \infty$, since we assume that $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ for some $p \geq 6$. Now we may easily see that the second term is $O_p(a_n^{-1/2}n^{1/2})$ due to part (b) of this lemma, and is therefore, $o_p(a_n^{-1}n)$ when $\alpha > 1$. The second term is therefore dominated and the asymptotics are determined by the first term.

The proofs of parts (i) and (j) are essentially the same as that of part (h). For part (j), note that we assume $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ for some $p \geq 6$, and therefore, $\mathbf{E}\varepsilon_{3t}^2 < \infty$, where $\varepsilon_{3t} = \varepsilon_t^3 - \mathbf{E}\varepsilon_t^3$.

Lemma A3 (*Asymptotics for Some Sample Moments – AHTS*). Let Assumption 3.2 hold, and define $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2$ and $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$. The following sample moments have asymptotic distributions and rates of convergence given as follows:

(a) $\left[n\nu^2\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^2\left(x_t\right) \to_d \int_0^1 H^2\left(V\left(r\right)\right)dr$

(b) $\left[n^{1/2}\nu\left(a_n\right)\right]^{-1}\sum_{t=1}^n F\left(x_t\right)\varepsilon_t \to_d \int_0^1 H\left(V\left(r\right)\right)dU\left(r\right)$

(c) $\left[n\nu^2\left(a_n\right)\right]^{-1}\sum_{t=k+1}^n F\left(x_t\right)F\left(x_{t-k}\right) \to_d \int_0^1 H^2\left(V\left(r\right)\right)dr$

(d) $\left[n\nu^3\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^3\left(x_t\right) \to_d \int_0^1 H^3\left(V\left(r\right)\right)dr$

(e) $\left[n\nu^4\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^4\left(x_t\right) \to_d \int_0^1 H^4\left(V\left(r\right)\right)dr$

(f) $\left[n^{1/2}\nu^2\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^2\left(x_t\right)\varepsilon_t \to_d \int_0^1 H^2\left(V\left(r\right)\right)dU\left(r\right)$

(g) $\left[n^{1/2}\nu^3\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^3\left(x_t\right)\varepsilon_t \to_d \int_0^1 H^3\left(V\left(r\right)\right)dU\left(r\right)$

(h) $\left[n\nu\left(a_n\right)\right]^{-1}\sum_{t=1}^n F\left(x_t\right)\varepsilon_t^2 \to_d \sigma_\varepsilon^2\int_0^1 H\left(V\left(r\right)\right)dr$

(i) $\left[n\nu^2\left(a_n\right)\right]^{-1}\sum_{t=1}^n F^2\left(x_t\right)\varepsilon_t^2 \to_d \sigma_\varepsilon^2\int_0^1 H^2\left(V\left(r\right)\right)dr$

(j) $\left[n\nu\left(a_n\right)\right]^{-1}\sum_{t=1}^n F\left(x_t\right)\varepsilon_t^3 \to_d \tau_\varepsilon^3\int_0^1 H\left(V\left(r\right)\right)dr$

Proof of Lemma A3 (*Asymptotics for Some Sample Moments – AHTS*).

For the proof of part (a), we note that F^2 is asymptotically homogeneous with AO ν^2 and LHF H^2. This is shown in Park and Phillips (2001). In particular, F^2 is regular at infinity due to Lemma 3.1, and therefore, the stated result follows directly from Theorem 1.6 of BI (pg. 138).

For part (b), we write

$$\left[n^{1/2}\nu(a_n)\right]^{-1}\sum_{t=1}^{n} F(x_t)\varepsilon_t = \int_0^1 H(V_n(r))\,dU_n + o_p(1),$$

where U_n is the partial sum process defined from (ε_t), i.e., $U_n(r) = n^{-1/2}\sum_{t=1}^{[nr]}\varepsilon_t$. The stated result now follows exactly as in Park and Phillips (1999).

To prove part (c), it suffices to show that

$$\begin{aligned}\left[n\nu^2(a_n)\right]^{-1}\sum_{t=k+1}^{n} F(x_t)F(x_{t-k}) &= \frac{1}{n}\sum_{t=k+1}^{n} H\left(\frac{x_t}{a_n}\right)H\left(\frac{x_{t-k}}{a_n}\right) + o_p(1) \\ &= \frac{1}{n}\sum_{t=k+1}^{n} H^2\left(\frac{x_t}{a_n}\right) + o_p(1) \\ &= \int_0^1 H^2(V_n(r))\,dr + o_p(1),\end{aligned}$$

from which the stated result follows immediately. We may easily deduce the first equality from the asymptotic homogeneity of F. The second equality is somewhat harder to prove. For a smooth asymptotically homogeneous function, it is well expected, since

$$H\left(\frac{x_t}{a_n}\right) = H\left(\frac{x_{t-k}}{a_n} + \frac{v_t + \cdots + v_{t-k+1}}{a_n}\right) \approx H\left(\frac{x_{t-k}}{a_n}\right)$$

for any finite k. The approximation indeed holds without differentiability of the asymptotically homogeneous function. The rigorous proof for the second inequality, however, will not be given, since it is too long, yet tedious and essentially identical to the proof of Lemma 3.2 in Chang and Park (2004) for their study of asymptotics for nonlinear regressions with integrated time series under endogeneity.

The proofs of parts (h), (i) and (j) of the lemma are completely analogous to the proofs of the corresponding parts of Lemma A2, and are therefore omitted. Noting that F^3 and F^4 are asymptotically homogeneous with appropriate AO's and LHF's trivializes the proofs of parts (d), (e), (f), and (g), so we omit those proofs, as well.

Appendix B: Proofs of the Main Results

Proof of Lemma 3.1 Let F be asymptotically homogeneous with LHF H satisfying

$$H(x) = |x|^{\kappa}H(1) \tag{22}$$

for some $\kappa > -1$. If we define

$$\ell_\kappa(x) = |x|^{-\kappa} F(x),$$

then it follows immediately that

$$\lim_{|x|\to\infty} \frac{F(x)}{|x|^\kappa \ell_\kappa(x)} = 1$$

Therefore, it suffices to show that ℓ_κ is slowly varying at infinity, i.e.,

$$\lim_{\lambda\to\infty} \frac{\ell_\kappa(\lambda x)}{\ell_\kappa(\lambda)} = 1 \tag{23}$$

to finish the proof. However, (23) readily follows from the asymptotic homogeneity of F and (22), since

$$\begin{aligned} F(\lambda x) &= \nu(\lambda)[H(x) + o(1)], \\ F(\lambda) &= \nu(\lambda)[H(1) + o(1)] \end{aligned}$$

for large $\lambda > 0$, and therefore

$$\frac{\ell_\kappa(\lambda x)}{\ell_\kappa(\lambda)} \to \frac{H(x)}{|x|^\kappa H(1)} = 1$$

as $\lambda \to \infty$.

Proof of Theorem 3.2 (*Asymptotics for* R_{nk} – *ITS*). We let $\sigma_\varepsilon^2 > 0$. The result for the model with $\sigma_\varepsilon^2 = 0$ may simply be derived as a special case for which $(\varepsilon_t) \equiv 0$ in what follows. Note that

$$\sum_{t=1}^n y_t = \sum_{t=1}^n F(x_t) + \sum_{t=1}^n \varepsilon_t = O_p(a_n^{-1} n) + O_p(n^{1/2}),$$

and therefore, for fixed k,

$$\sum_{t=k+1}^n (y_t - \bar{y}_n)(y_{t-k} - \bar{y}_n) = \sum_{t=k+1}^n y_t y_{t-k} + o\left(a_n^{-1} n\right),$$

due in particular to Lemma A2. As a consequence, the mean adjustment in the definition of the sample correlation becomes negligible and does not affect the asymptotics, as long as $a_n n^{-1} \to 0$. This will be seen clearly in the subsequent proof.

Write

$$\sum_{t=k+1}^n y_t y_{t-k} = \sum_{t=k+1}^n F(x_t)F(x_{t-k}) + \sum_{t=k+1}^n F(x_t)\varepsilon_{t-k} + \sum_{t=k+1}^n F(x_{t-k})\varepsilon_t + \sum_{t=k+1}^n \varepsilon_t \varepsilon_{t-k}. \tag{24}$$

Due to Lemma A2, we have

$$\sum_{t=k+1}^{n} F(x_t)F(x_{t-k}) = O_p(a_n^{-1}n) \tag{25}$$

and

$$\sum_{t=k+1}^{n} F(x_t)\varepsilon_{t-k}, \ \sum_{t=k+1}^{n} F(x_{t-k})\varepsilon_t = O_p(a_n^{-1/2}n^{1/2}) \tag{26}$$

for all $k \geq 0$. Moreover, we have

$$\frac{1}{n}\sum_{t=k+1}^{n} \varepsilon_t^2 \to_p \sigma_\varepsilon^2, \tag{27}$$

and for all $k \geq 1$

$$\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n} \varepsilon_t\varepsilon_{t-k} \to_d \mathbb{N}\left(0, \sigma_\varepsilon^4\right), \tag{28}$$

by the standard law of large numbers and central limit theorem.

We first consider the case of $k = 0$ in (24), which also gives us asymptotics for the denominator. It is obvious from (25)–(27) that

$$\frac{1}{n}\sum_{t=k+1}^{n} y_t^2 = \frac{1}{n}\sum_{t=k+1}^{n} \varepsilon_t^2 + o_p(1) \to_p \sigma_\varepsilon^2, \tag{29}$$

since $a_n \to \infty$, and hence,

$$a_n^{-1}n, \ a_n^{-1/2}n^{1/2} = o(n).$$

Next, to consider the case of $k \geq 1$ in (24), we first note that

$$n^{-\delta} < \ell(n) < n^{\delta} \tag{30}$$

for any $\delta > 0$ and for all n sufficiently large. This is well-known [see for example Feller (1971, Lemma 2, pg. 277)]. Since we assume $\alpha > 1$, this implies that

$$a_n^{-1}n \to \infty,$$

and therefore,

$$a_n^{-1/2}n^{1/2} = o(a_n^{-1}n)$$

for all large n. Consequently, the terms in (26) are smaller than those in (25) and asymptotically negligible for all $k \geq 0$.

Let $1 < \alpha < 2$. Then it follows from (30) that

$$n^{1/2} = o(a_n^{-1}n),$$

and therefore we have for all $k \geq 1$

$$\begin{aligned} a_n n^{-1} \sum_{t=k+1}^{n} y_t y_{t-k} &= a_n n^{-1} \sum_{t=k+1}^{n} F(x_t) F(x_{t-k}) + o_p(1) \\ &= L(1,0) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) F(x + a_k y) D_k(y)\, dx\, \mu(dy), \end{aligned}$$

due to Lemma A2(c), which together with (28) immediately yields the stated result in this case. Now we let $\alpha = 2$. In this case, the dominant terms differ depending upon whether $\ell(n) \to 0, c, \infty$. If, for instance, $\ell(n) \to c$ for some constant c, then we have both the first term and the last term in (24) for our asymptotics. As a result, we have

$$n^{-1} a_n \sum_{t=k+1}^{n} y_t y_{t-k} = n^{-1} a_n \sum_{t=k+1}^{n} F(x_t) F(x_{t-k}) + n^{-1/2} \ell(n) \sum_{t=k+1}^{n} \varepsilon_t \varepsilon_{t-k} + o_p(1),$$

and the stated result easily follows. The result for each of the cases $\ell(n) \to 0$ and $\ell(n) \to \infty$ can also be readily deduced upon noticing that the first or the last term dominates the other in each case.

Proof of Corollary 3.3 (*Rate of Decay of R_{nk} – ITS*). Since we assume that (φ_k) are absolutely integrable, we may have

$$D_k(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi_k(s)\, ds \tag{31}$$

due to the Fourier inversion formula. By the same token, we may also have

$$D(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi(s)\, ds, \tag{32}$$

since the characteristic function φ of a stable distribution is absolutely integrable. Therefore, it can be easily deduced from (31) and (32) that

$$\sup_{x \in \mathbb{R}} |D_k(x) - D(x)| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi_k(s) - \varphi(s)|\, ds \to 0$$

as $k \to \infty$, since $\varphi_k \to \varphi$ in L^1. The sequence of PDF's (D_k) thus converge uniformly.

Now we have

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) F(x + a_k y) D_k(y)\, dx\, dy \\ &= a_k^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) F(x + y) D_k(a_k^{-1} y)\, dx\, dy \\ &= a_k^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) F(x + y) D(a_k^{-1} y)\, dx\, dy + o(a_k^{-1}) \\ &= a_k^{-1} D(0) \left(\int_{-\infty}^{\infty} F(x)\, dx \right)^2 + o(a_k^{-1}) \end{aligned}$$

for large k, by the change of variables, the uniform convergence of D_k to D and the continuity of D at the origin. Note that the absolute integrability of (φ_k) implies that the distribution of (v_t) is absolutely continuous with respect to Lebesgue measure, and for this reason, we use the notation dy in place of $\mu(dy)$. The stated result now follows immediately and the proof is complete.

Proof of Theorem 3.4 (*Asymptotics for* S_n^2, Q_n^3, K_n^4 *– ITS*). The distributions for S_n^2 and the denominators of the other two sample statistics follow directly from the asymptotics of the denominator in Theorem 3.2. Letting $\sigma_\varepsilon^2 > 0$, we now turn to the numerator of Q_n^3. It is easy to show that the mean adjustment is asymptotically negligible, as in the proof for the autocorrelation. Expanding the dominant term yields

$$\frac{1}{n}\sum_{t=1}^{n} y_t^3 = \frac{1}{n}\sum_{t=1}^{n} F^3(x_t) + \frac{3}{n}\sum_{t=1}^{n} F^2(x_t)\,\varepsilon_t + \frac{3}{n}\sum_{t=1}^{n} F(x_t)\,\varepsilon_t^2 + \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^3,$$

which with the fact that

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^3 \to_p \tau_\varepsilon^2$$

and with Lemma A2 gives us the desired result. When $\sigma_\varepsilon^2 = 0$, only the first term remains. The proof for the sample kurtosis is very similar to that of the sample skewness, by expanding the numerator and noting that

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^4 \to_p \kappa_\varepsilon^4,$$

which determines the probability limit of the numerator when $\sigma_\varepsilon^2 > 0$. Again, the case of $\sigma_\varepsilon^2 = 0$ is trivial.

Proof of Theorem 3.5 (*Asymptotics for* R_{nk} *– AHTS*). The case in which $\sigma_\varepsilon^2 = 0$ is a special case, so let $\sigma_\varepsilon^2 > 0$. Note that

$$\begin{aligned}&\sum_{t=k+1}^{n}(y_t - \bar{y}_n)(y_{t-k} - \bar{y}_n)\\ &= \sum_{t=k+1}^{n} y_t y_{t-k} - \frac{1}{n}\sum_{t=k+1}^{n} y_{t-k}\sum_{t=1}^{n} y_t - \frac{1}{n}\sum_{t=k+1}^{n} y_t\sum_{t=1}^{n} y_t + \sum_{t=k+1}^{n}\left(\frac{1}{n}\sum_{t=1}^{n} y_t\right)^2,\end{aligned}$$

which means that the mean adjustment may not be dismissed, as it was in the ITS case. First, consider the case in which $k = 0$. The above expression reduces to

$$\sum_{t=1}^{n} y_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n} y_t\right)^2. \tag{33}$$

We may expand the first term of (33) to obtain

$$\sum_{t=1}^{n} F^{2}\left(x_{t}\right)+2 \sum_{t=1}^{n} F\left(x_{t}\right) \varepsilon_{t}+\sum_{t=1}^{n} \varepsilon_{t}^{2}.$$

which by Lemma A3 has the distribution of its first term. We may similarly expand the second term of (33), which also has the distribution of the first term of that expansion. The result for $k = 0$ obviously follows. The result for $k \geq 1$ follows directly from the appropriate parts of Lemma A3, using the same logic.

Proof of Theorem 3.6 (*Asymptotics for* S_n^2, Q_n^3, K_n^4 – *AHTS*). The proof for S_n^2 follows directly from the asymptotics in Theorem 3.5 (by letting $k = 0$), and the other proofs are essentially the same as that for the sample variance, using the appropriate parts of Lemma A3.

Proofs of Theorems 4.1 and 4.2 (*Asymptotics for* $\hat{\theta}_n$). Given the invariance principle (10) and Lemma 3.1, the proofs follow that of Theorems 5.1 and 5.2 in Park and Phillips (2001), with rates of convergence following from our Lemmas A2 and A3.