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

J. Isaac Miller

Department of Economics, University of Missouri

and

Joon Y. Park

Department of Economics, Texas A&M University and School of Economics, Sungkyunkwan University

Abstract

We consider nonlinear transformations of random walks driven by thick-tailed innovations that may have infinite means or variances. These three nonstandard characteristics: nonlinearity, nonstationarity, and thick tails interact to generate a spectrum of asymptotic autocorrelation patterns consistent with long-memory processes. Such autocorrelations may decay very slowly as the number of lags increases or may not decay at all and remain constant at all lags. Depending upon the type of transformation considered and how the model error is specified, the autocorrelation functions are given by random constants, deterministic functions that decay slowly at hyperbolic rates, or mixtures of the two. Such patterns, along with other sample characteristics of the transformed time series, such as jumps in the sample path, excessive volatility, and leptokurtosis, suggest the possibility that these three ingredients are involved in the data generating processes of many actual economic and financial time series data. In addition to time series characteristics, we explore nonlinear regression asymptotics when the regressor is observable and an alternative regression technique when it is unobservable. To illustrate, we examine two empirical applications: wholesale electricity price spikes driven by capacity shortfalls and exchange rates governed by a target zone.

First Draft: December 2003
This version: January 2008

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

---

[1]We are very grateful to Peter Robinson and three anonymous referees for providing many useful comments. We also thank Yoosoon Chang for many helpful discussions, and participants of seminars at Rice, GSU, and Missouri and of Texas Camp Econometrics IX (SMU), 2004 NASMES (Brown), and the 2004 Symposium of the Center for Computational Finance and Economic Systems (Rice) for helpful comments. Most of the research was completed while we were at Rice University.

1. Introduction

With the improvement of data-collection technology and the availability of high-frequency data, sample sizes of time series have grown tremendously. As more data become available and at higher frequencies, it becomes increasingly difficult to ignore nonstandard time series characteristics displayed in the sample paths of these data. For example, if the spatial distance between two observations does not decrease as the temporal distance decreases, it becomes more evident that the sample path contains a discontinuity. For this reason, time series models in economics and finance have featured Poisson jump processes, discrete switching components, or thick-tailed distributions – all of which may replicated such behavior. Such innovations have followed widespread interest in models featuring nonlinearity and/or nonstationarity, since the fields of econometrics and statistics have strained for decades to move beyond the stationary linear paradigm characterized by simple ARMA modeling approaches.

The aim of this paper is to tie these approaches together in a very simple and intuitive way. Perhaps the most interesting aspect of the classes of models that we introduce is that they may generate diverse patterns of persistency in memory. A random walk of course generates long memory, in the sense that the asymptotic autocorrelation function never decays as the number of lags increase. The autocorrelations of a thick-tailed random walk behave in a qualitatively similar manner. Thus, combining nonstationarity and thick tails creates persistency in memory in the expected way. Similarly, one can expect that the autocorrelations of a (strongly) stationary thick-tailed series will behave much like those of a (weakly) stationary series, and decay quickly. If we instead consider nonlinear transformations of random walks with increments having finite variances (a special case of the model classes we analyze), the autocorrelations become more interesting. Such models may generate time series with long memory, depending on the type of nonlinear function used, but may otherwise possess characteristics consistent with stationary processes. If we combine all three of these ingredients – nonlinearity, nonstationarity, and thick tails – we get models that generate an even wider spectrum of long-memory patterns. Such models may generate time series with additional nonstandard characteristics, such as unusually large, small, or random magnitudes of sample variance, skewness, kurtosis. Moreover, explicitly incorporating thick tails in our model classes allows for excessive outliers and jump behavior.

The theory underlying our model depends crucially on the type of transformation functions involved. We consider separately two types of functions for the underlying transformations: I-regular and H-regular. These are the functions classes introduced by Park and Phillips (1999, 2001) in their studies on nonlinear transformations of integrated time series. We formalize the concept of thick tails below by reviewing some fundamental properties of the α-stable distribution, which plays an integral part in our models. For brevity, we refer to the class of models employing I-regular functions of α-stable random walks as *ITS models* and those that similarly utilize H-regular functions as *AHTS models*. ITS models generate time series that have characteristics similar to those of stationary long-memory processes. More precisely, the transformed processes have asymptotic autocorrelations that decay at hyperbolic rates with the exact rate depending upon the thickness of the tails of the innovations driving the underlying random walks. Such rates are consistent with fractionally integrated I(d) processes with memory parameter d such that $0 < d \leq 1/4$. When model error is present, ITS models may also generate such autocorrelation patterns but with additional or multiplicative stochastic noise, or they may generate autocorrelations that are indistinguishable from 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 beyond the first lag.

We also study the asymptotic properties of the sample variance, skewness, and kurtosis of ITS and AHTS models. Such sample statistics are spurious in the sense that they do not approximate the population counterparts of any distribution in a nonstationary and nonergodic model. However, they still carry meaningful information that allows one to distinguish between conventional stationary or integrated series and those with ITS or AHTS data generating processes (DGP's). This will become clear as our exposition unfolds.

The remainder of the paper is structured as follows. In Section 2, we describe the general model and discuss some of the unusual aspects of α-stable distributions and time series built upon them. We discuss in more detail the function classes that define ITS and AHTS models and present our main theoretical findings in Section 3. In Section 4, we extend the theory developed by Park and Phillips (2001) for nonlinear least squares (NLS) estimation of nonlinear regressions with integrated processes to our models with thick-tailed innovations, and we suggest an alternative estimation technique when (x_t) is unobservable. Finally, in Section 5, we apply our techniques to wholesale electricity prices and target zone exchange rates. Two appendices contain useful lemmas, their proofs, and proofs of the main theoretical results.

Throughout the paper, we suppress the indices of all summations indexed by $t = 1, \ldots, n$ for notational ease.

2. Model and Preliminaries

Let (x_t) be a time series generated by

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

where (v_t) is a sequence of random variables, the densities of which have thick tails, as we specify more concretely below. Let the time series of interest (y_t) be generated by

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

where (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}) is adapted, with $\mathbf{E}\,|\varepsilon_t|^k < \infty$ for some $k \geq 8$, and F is a nonlinear function on $\mathbb{R}$.

We assume that (v_t) and (ε_t) are independent, or equivalently that (x_t) is a strictly exogenous series. This assumption may be relaxed for many of our results but is especially convenient for regression asymptotics. We further assume that the second, third, and fourth conditional moments of (ε_t) are nonstochastic, and accordingly define the notation

$$\sigma_\varepsilon^2 \equiv \mathbf{E}[\varepsilon_t^2|\mathcal{F}_{t-1}], \qquad \tau_\varepsilon^3 \equiv \mathbf{E}[\varepsilon_t^3|\mathcal{F}_{t-1}], \qquad \text{and} \qquad \kappa_\varepsilon^4 \equiv \mathbf{E}[\varepsilon_t^4|\mathcal{F}_{t-1}]$$

for those moments. Let

$$\varepsilon_{2,t} \equiv \varepsilon_t^2 - \sigma_\varepsilon^2, \qquad \varepsilon_{3,t} \equiv \varepsilon_t^3 - \tau_\varepsilon^3, \qquad \text{and} \qquad \varepsilon_{4,t} \equiv \varepsilon_t^4 - \kappa_\varepsilon^4$$

and note that $(\varepsilon_{2,t})$, $(\varepsilon_{3,t})$, and $(\varepsilon_{4,t})$ are also MDS's with respect to the same filtration. All MDS's introduced subsequently are defined with respect to this filtration, unless otherwise noted. As such, we may employ central limit theorems (CLT's) and laws on large numbers (LLN's) of Hall and Heyde (1980), for example, to ascertain the limiting behavior of normalized partial sums of these sequences. We refer to such limit theory without further reference.

We consider two plausible alternative modeling assumptions:

$$\sigma_\varepsilon^2 > 0 \qquad \text{or} \qquad \sigma_\varepsilon^2 = 0,$$

where the former amounts to including modeling error and the latter amounts to omitting it. As such, the time series of interest may be observable only with noise or directly observable. In either case,

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

if this conditional mean is well-defined – which may not be the case if the tails of the innovations are too thick. In words, the time series of interest has a conditional mean given by some function of an integrated series driven by thick-tailed innovations.

Thick-Tailed Innovations: α-Stable Distribution

To solidify the idea of thick tails, we require some technical assumptions about (v_t). We assume that the elements of (v_t) are independent and identically distributed (iid) and have regularly varying tail probabilities – i.e.,

$$\mathbf{P}\{|v_t| > x\} = x^{-\alpha}\ell(x) \tag{3}$$

with $\alpha > 0$ and where ℓ is a slowly varying function at infinity. Moreover, we assume that the tail balancing condition

$$\frac{\mathbf{P}\{v_t > x\}}{\mathbf{P}\{|v_t| > x\}} \to p, \qquad \text{or} \qquad \frac{\mathbf{P}\{v_t < -x\}}{\mathbf{P}\{|v_t| > x\}} \to q \tag{4}$$

holds as $x \to \infty$, for $0 \le p, q \le 1$ and $p + q = 1$. The conditions in (3) and (4) are essential for our subsequent theoretical developments, but the iid assumption may be relaxed at the cost of more involved exposition.

As we show below, the standardized sum of (v_t) satisfying the conditions introduced in (3) and (4) converges to what is known as an *α-stable distribution* or simply *stable distribution*. Formally, a random variable v is said to have a stable distribution $\mathbb{S}_\alpha(\sigma, \beta, \mu)$, for $0 < \alpha \le 2$, $\sigma \ge 0$, $-1 \le \beta \le 1$, and μ real, if it has a 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

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

and sgn(s) is the 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 ($\alpha = 2$), Cauchy ($\alpha = 1$ and $\beta = 0$), and Lévy ($\alpha = 1/2$ and $\beta = 1$). For $0 < \alpha < 2$, v has an infinite (or undefined) variance, and for $0 < \alpha \leq 1$, it has an infinite (or undefined) mean, as well.

Central Limit Theory

Partial sums of iid sequences of α-stable random variables require normalization and centering that are different from that employed by standard central limit theory. Let $0 < \alpha < 2$, and define numerical sequences (a_n) and (b_n) by

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

with

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

as $n \to \infty$. It follows that

$$a_n^{-1} \sum (v_t - b_n) \to_d \mathbb{S}_\alpha(\sigma, \beta, 0), \tag{5}$$

where

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

and $\beta = 2p - 1$. [See Feller (1971, Theorem 3, pg. 580), e.g.[3]]

It is well-known that we may set

$$a_n = n^{1/\alpha}\ell(n), \tag{6}$$

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 (3) holds for large $x > 0$ with $\ell(x) = c$ for some constant $c > 0$, then

$$a_n = c^{1/\alpha} n^{1/\alpha} \tag{7}$$

[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] According to our definition of (a_n), we have $C(2-\alpha)/\alpha = 1$ in his formula. The sign $\mp$ in the formula is in error and should be corrected to $\pm$.

as may easily be verified.

If (5) holds with (6), the law of (v_t) belongs to the *domain of attraction of a stable law.* If (5) holds with (7), it belongs 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) is iid $\mathbb{S}_\alpha(\sigma, \beta, \mu)$, (3) indeed holds with $\ell(x) = c$, where $c > 0$ is defined by Brockwell and Davis (1987, pg. 480). The conditions introduced in (3) and (4) are therefore necessary and sufficient in order that the underlying distribution of (v_t) belongs to the domain of attraction of a stable law.

When $\alpha = 2$, the limit theorem in (5) holds under weaker conditions, with $b_n = \mathbf{E}(v_t)$ for all n. The condition in (3) alone is sufficient to have (5) with (a_n) specified in (6), as shown in Ibragimov and Linnik (1971, Theorem 2.6.2, pg. 79), e.g. Moreover, Ibragimov and Linnik (1971, Theorem 2.6.6, pg. 92) also showed that (5) holds with (a_n) in (7) and with $\alpha = 2$, if and only if the elements of (v_t) have finite variance. Accordingly, the law of (v_t) belongs to the domain of attraction of a *normal* law if (5) holds with (6). If (5) holds with (7), then the law of (v_t) belongs to the domain of *normal* attraction of a *normal* law.

From now on, we assume that the elements of (v_t) are properly centered. For $1 < \alpha \leq 2$, centering simply requires demeaning or assuming zero mean. For $\alpha = 1$, the proper centering is more involved unless we assume that the underlying distribution is symmetric. No centering is necessary for $0 < \alpha < 1$. The limiting distribution has a shift parameter of zero – i.e., $\mu = 0$ – if the elements of (v_t) are centered. Furthermore, we let the adjustment for scales be done beforehand 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) merely amounts to redefining the transformation function F by a constant multiple of its argument. The skewness parameter β is not restricted to zero, so that a asymmetric limit distribution of (v_t) is allowed. Finally, the normalizing sequence (a_n) is assumed throughout this paper to be given by (6) or (7), depending upon whether the common distribution of (v_t) belongs the domain of attraction or normal attraction of a stable law.

Invariance Principle and Local Time

The central limit theorem in (5) is not sufficient to establish limit theory for our model. In order to effectively deal with nonstationarity, we need an invariance principle. We construct a stochastic process V_n on $[0, 1]$, defined by

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

where $[x]$ denotes the largest integer not exceeding x, and invoke the invariance principle in Borodin and Ibragimov (1995, pg. 12, hereafter referred to as BI), e.g., so that

$$V_n \to_d V, \tag{8}$$

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 $0 \leq s < t$ and for some $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$. [See also Samorodnitsky and Taqqu (1994, pg. 113) for more details.] The processes V_n and V take values in $D[0,1]$, the space of cadlag functions defined on $[0,1]$, and in (8) we have weak convergence of probability measures in $D[0,1]$. Unlike a Brownian motion, a Lévy motion may have discontinuities in its sample path.

The nonlinearity in our models requires some additional tools. 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* 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 speaking, the local time L characterizes the portion of time the process V spends at x up to time t. As shown in BI (Theorem I.4.1, pg. 18),[4] standard Lévy motions with $\alpha > 1$ have local times that are continuous with respect to both parameters. The local time does not exist if $0 < \alpha \leq 1$.

Serial Correlation of the Innovations

We may consider a more general process (x_t) driven by serially correlated rather than iid innovations. In particular, we may set $x_t = x_{t-1} + u_t$, where

$$u_t \equiv \sum\nolimits_{k=0}^{\infty} c_k v_{t-k} \qquad \text{with} \qquad \sum\nolimits_{k=0}^{\infty} |c_k|^{\delta} < \infty$$

for some $\delta \in (0,\alpha) \cap [0,1]$. Under the summability condition, (u_t) is well-defined a.s., and if the underlying distribution of (v_t) belongs to the domain of normal attraction and (3) holds with $\ell(x) = c$, then

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

as $x \to \infty$. Therefore, condition (3) holds also for (u_t). Clearly, condition (4) can easily be satisfied with $p = q = 1/2$ if we assume that the underlying distribution of (v_t) – and consequently that of (u_t) – is symmetric. [See Brockwell and Davis (1987, Remarks 1 and 2, pg. 481), e.g.]

All of our subsequent results hold – at least qualitatively – for (x_t) generated by the more general linear process (u_t). Some apply without modification. Others just need somewhat obvious modifications and some additional theoretical developments using the Beveridge-Nelson decomposition studied in Phillips and Solo (1992). As this generalization does not add new insight to our results, we assume that (x_t) is generated by (1) without the additional complication of serial correlation.

Function Classes

The time series properties of (y_t) critically depend on the function F in (2). We consider two classes of functions: I-regular and H-regular, defined as follows.

[4] We add the chapter number in Roman numerals to references to theorems of BI, since their enumeration is not unique.

Definition 2.1 (*I-regular Functions*). A transformation F in the class of I-regular functions, is locally Riemann-integrable and satisfies

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

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

Definition 2.2 (*H-Regular Functions*). A transformation F in the class of H-regular functions satisfies $F(\lambda x) = \nu(\lambda) H(x) + R(x, \lambda)$ for large λ, where H is locally Riemann-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 Riemann-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 Riemann-integrable and $Q(x) \to 0$ as $x \to \infty$.

The *asymptotic order* (AO) of an H-regular function is $\nu(\lambda)$, and $H(x)$ is the *limit homogeneous function* (LHF). Intuitively, an H-regular function exhibits an asymptotically dominant component that is homogeneous. We assume throughout this analysis that the LHF of any H-regular function is in fact homogeneous.[5]

For I-regular F, we refer to (1) and (2) with (3) and (4) as an *ITS model*, where ITS denotes "integrable transformation of a stable process." A time series (y_t) generated in such a way is an *ITS process*. A similarly defined model and time series with H-regular F are similarly referred to as an *AHTS model* and an *AHTS process*, where AHTS signifies "asymptotically homogeneous transformation of a stable process."

H-regular functions are closely related to functions that are *regular at infinity*.

Definition 2.3 (*Regular-at-infinity Functions*). A transformation F in the class of regular-at-infinity functions satisfies

$$\lim_{x \to \infty} \frac{F(x)}{x^{\kappa} \ell(x)} = c_1 \qquad \text{and} \qquad \lim_{x \to -\infty} \frac{F(x)}{|x|^{\kappa} \ell(x)} = c_2 \tag{9}$$

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

$$\lim_{\lambda \to \infty} \ell_{\kappa}(\lambda x)/\ell_{\kappa}(\lambda) = 1 \tag{10}$$

for any $x > 0$.

A useful result allows us to tie in general results derived in the mathematics literature for regular-at-infinity functions with the more specific H-regular functions introduced to the econometrics literature by Park and Phillips (1999, 2001).

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

Lemma 2.1 Let F be an H-regular function with LHF H satisfying

$$H(x) = |x|^{\kappa} H(\text{sgn}(x)) \tag{11}$$

for some $\kappa > -1$. Then F is regular at infinity.

Note that the converse is not true, since regular-at-infinity functions are a broader class of functions than H-regular functions.

Any bounded function with compact support provides an example of an I-regular function. Most probability density functions (PDF's) belong to this class, as long as they are bounded and decay faster than $|x|^{-1}$ as $|x| \to \infty$. Scaled and horizontally shifted variations of such PDF's are also I-regular. We may intuitively interpret such a transformation as returning a strong signal when the value of the underlying random walk is near the mode (or modes) of some PDF-like function. For example, in the empirical section of our analysis, we use an I-regular transformation to model the relationship between wholesale electricity prices and the excess capacity. Under our specification, we expect to observe a sudden increase in the price whenever system generation nears capacity, followed by a sharp decrease as either more capacity is brought online or as peak demand diminishes.

The most commonly employed functions that fall within the class of H-regular functions are homogeneous, polynomial (especially linear with an intercept), and logarithmic. A perhaps more interesting sub-class of asymptotically homogeneous functions are smooth transition functions, which resemble rescaled and shifted cumulative distribution functions (CDF's). Any CDF is H-regular with $\nu(\lambda) = 1$ and $H(x) = 1\{x \geq 0\}$, and smooth transition functions have the same AO with an LHF given by some affine transformation of an indicator function. 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 may be 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.

3. Time Series Properties of ITS and AHTS Models

Armed with the tools outlined above, we now turn to the main theoretical analysis of the paper. Specifically, we explore limiting distributions of the sample autocorrelation function, variance, skewness, and kurtosis. Each of these statistics are defined in terms of deviations from the sample mean. As a result, these statistics are asymptotically invariant with respect to a shift by a constant, and ITS and AHTS processes may be characterized by their sample moments only up to a constant term. A transformation comprised of a constant plus an I-regular function is in fact H-regular, but has the same asymptotics as an I-regular function. Our subsequent results for ITS processes therefore also apply to I-regular functions shifted by arbitrary constants, and those for AHTS processes are valid for H-regular functions with nonconstant LHF's.

3.1. Asymptotics for ITS Models

Our subsequent asymptotic results rely on the following assumption.

Assumption 3.1 Let the time series (y_t) be generated by (1) and (2) with I-regular 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 $\varphi(s) \neq 1$ for all $s \neq 0$.

We restrict the order of the limiting stable law of (v_t) to $1 < \alpha \leq 2$, because the asymptotics for ITS models may be greatly simplified when the local time of the limit stable process V exists, which it does *not* when α is unity or smaller. The technical 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, which is not overly restrictive.

The autocorrelation function is the key to unlock the persistency of a time series. Let the sample autocorrelation be

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

where k is any nonnegative integer and $\bar{y}_n \equiv n^{-1}\sum y_t$. Further, let D denote the common underlying PDF of (v_t) with respect to the measure μ on $\mathbb{R}$, and let D_k denote the PDF of $a_k^{-1}(v_1 + \cdots + v_k)$ with respect to the same measure. Clearly, we have $D_k = D$, if (v_t) is an α-stable process.

Theorem 3.1 (*Asymptotics for* R_{nk} *– ITS*). Let Assumption 3.1 hold and define

$$N_k \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x + a_k y)D_k(y)\ dx\ \mu(dy)$$

for $k \geq 1$.

(a) Let $\sigma_\varepsilon^2 = 0$. We have

$$R_{nk} \to_p N_k \bigg/ \int_{-\infty}^{\infty} F^2(x)\,dx$$

as $n \to \infty$.

(b) Let $\sigma_\varepsilon^2 > 0$. We have

$$R_{nk} = n^{-1/2}Z_k + a_n^{-1}\left[L(1,0)/\sigma_\varepsilon^2\right]N_k + o_p(a_n^{-1}) \tag{12}$$

for large n, where (Z_k) is a sequence of independent standard normal random variates that are independent of $L(1,0)$.

The asymptotic autocorrelation structure of (y_t) is most clearly ascertained when the time series of interest is observed without noise – i.e., when $\sigma_\varepsilon^2 = 0$. In this case, the asymptotic autocorrelation function may be expressed (up to constant) by N_k, which we investigate further in the corollary below.

When observation error is present, the asymptotic autocorrelation structure of (y_t) may not seem so straightforward. However, it follows readily from our result (12) in Theorem 3.1 that the sample autocorrelation function is given by N_k also in this case, possibly with some random noise. Note that the leading term $n^{-1/2}Z_k$ in the asymptotic expansion of R_{nk} is pure random noise as a function of k. The autocorrelation structure of (y_t) in large samples is thus given by the second-order term $a_n^{-1}[L(1,0)/\sigma_\varepsilon^2]N_k$, which is nothing but a (random) scalar multiple of N_k. Furthermore, we may well expect that

$$\mathbf{E}R_{nk} = a_n^{-1}\left[\mathbf{E}L(1,0)/\sigma_\varepsilon^2\right]N_k + o(a_n^{-1})$$

under appropriate regularity conditions, which implies that the averaged asymptotic correlation function may be expressed (up to constant) by N_k, exactly as in the case of no observation error.

As can be easily seen, we have

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

in general. Consequently, the random noise term asymptotically dominates the term proportional to N_k in (12). The actual sample autocorrelation structure may therefore appear to be quite different from N_k even in large samples. The dominance, however, becomes weaker as α gets close to 2. Indeed, if we have $\alpha = 2$ with $\ell(n) \to c \neq 0$, then the two terms become equally important asymptotically.

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

$$N_k \to 0$$

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 or to noise. The following corollary extends this result to (v_t) in the domain of attraction of a stable law and only asymptotically stable, and it gives 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 3.2 (*Rate of Decay of R_k – ITS*). Let Assumption 3.1 hold and assume that the elements of (φ_k) are absolutely integrable, $\varphi_k \to \varphi$ in L^1, and D is continuous at the origin. Then we have

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

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 hyperbolic for large k. If the members of (v_t) belong to the domain of *normal* attraction of a stable law, then the rate will be exactly hyperbolic. On the other hand, if they belong only to the domain of attraction of a stable law, then the rate will also depend on the slowly varying function $\ell(k)$.

It is well-known that the sample autocorrelations of stationary fractionally integrated processes have long memory and decay at hyperbolic rates. These rates are k^{2d-1} where $d \in (0, 1/2)$ is defined as the *degree of fractional integration* or the *memory parameter*. Recalling that $a_k^{-1} = k^{-1/\alpha}/\ell(k)$ with $\alpha \in [1, 2)$, the rates of decay of the autocorrelation of an ITS process and that of an stationary, long-memory I(d) process with $d \in (0, 1/4]$ are clearly quite similar. It would be easy to mistake an ITS process for a more well-known stationary I(d) process, on these grounds. Such a misspecification would ignore valuable structural information about the time series.

Next, we examine the sample variance, skewness, and kurtosis of an ITS process. We define the observed sample variance, skewness, and kurtosis of a time series (y_t) as

$$S_n^2 \equiv \frac{1}{n} \sum (y_t - \bar{y}_n)^2, \qquad Q_n^3 \equiv \frac{\frac{1}{n} \sum (y_t - \bar{y}_n)^3}{(S_n^2)^{3/2}}, \qquad \text{and} \qquad K_n^4 \equiv \frac{\frac{1}{n} \sum (y_t - \bar{y}_n)^4}{(S_n^2)^2},$$

respectively. Sample moments calculated from nonstationary and nonergodic time series are spurious in the sense that they do not approximate the true moments of some underlying distribution, as they would for an iid or (to a lesser extent) stationary series. However, they convey information about the time series and may also provide additional clues to help distinguish an ITS DGP from an alternative one.

If (y_t) were in fact a stationary series with an underlying symmetric distribution with existing fourth moment, 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 of the innovations decay, roughly speaking. In contrast, the limits of these statistics when (y_t) is an ITS process are as follows.

Theorem 3.3 (*Asymptotics for S_n^2, Q_n^3, K_n^4 – ITS*). Let Assumption 3.1 hold.

(a) Let $\sigma_\varepsilon^2 = 0$. We have

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

$$a_n^{-1/2} Q_n^3 \to_d \frac{\int_{-\infty}^{\infty} F^3(x)\, dx}{\sqrt{L(1,0)} \left(\int_{-\infty}^{\infty} F^2(x)\, dx \right)^{3/2}}, \qquad \text{and} \qquad a_n^{-1} K_n^4 \to_d \frac{\int_{-\infty}^{\infty} F^4(x)\, dx}{L(1,0) \left(\int_{-\infty}^{\infty} F^2(x)\, dx \right)^2}$$

as $n \to \infty$.

(b) Let $\sigma_\varepsilon^2 > 0$ and $\tau_\varepsilon \neq 0$. We have

$$S_n^2 \to_p \sigma_\varepsilon^2, \qquad Q_n^3 \to_p \tau_\varepsilon^3/\sigma_\varepsilon^3, \qquad \text{and} \qquad K_n^4 \to_p \kappa_\varepsilon^4/\sigma_\varepsilon^4$$

as $n \to \infty$.

(c) Let $\sigma_\varepsilon^2 > 0$ and $\tau_\varepsilon = 0$.

i. If $1 < \alpha < 2$ or $\alpha = 2$ and $\ell(n) \to \infty$, we have

$$n^{1/2} Q_n^3 \to_d \mathbb{N}\left(0, \frac{\mathbf{E}\varepsilon_t^6}{\sigma_\varepsilon^6} - \frac{6\kappa_\varepsilon^4}{\sigma_\varepsilon^4} + 9\right)$$

as $n \to \infty$.

ii. If $\alpha = 2$ and $\ell(n) \to c$, we have

$$a_n Q_n^3 \to_d \frac{L(1,0)}{\sigma_\varepsilon^3} \int_{-\infty}^{\infty} F^3(x)\, dx + c\mathbb{N}\left(0, \frac{\mathbf{E}\varepsilon_t^6}{\sigma_\varepsilon^6} - \frac{6\kappa_\varepsilon^4}{\sigma_\varepsilon^4} + 9\right)$$

as $n \to \infty$.

ITS processes observed with asymmetric noise have observed sample statistics that are observationally equivalent to those of stationary processes. The nonlinear 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 inference.

The skewness provides a new dimension of information if the modelling error is in fact symmetric. Such symmetry mutes this noise, which in some cases allows a clearer reception of the underlying signal. However, since this underlying signal – involving the local time – is also random, and since the sequence (a_n) is very close to $(n^{1/2})$ in this case, this difference may not offer much practical value. In small samples, these ITS processes may be indistinguishable from stationary processes, based on the latter three sample statistics.

3.2. Asymptotics for AHTS Models

AHTS models are perhaps more important than their I-regular counterpart, because the literature is replete with examples of H-regular transformations. If the underlying exogenous variable in such a model is I(1) and the limiting distribution of the innovations are α-stable (including Gaussian), then our results apply.

We make the following assumption.

Assumption 3.2 Let the time series (y_t) be generated by (1) and (2) with H-regular $F(x)$ with $\inf_{\lambda>0} |\nu(\lambda)| > 0$ 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. A lattice distribution is allowed for (v_t) here. Furthermore, the stable parameter for the limit process is allowed to be $0 < \alpha \leq 2$. The additional assumption merely ensures that the AO is not degenerate.

The following theorem gives the asymptotic result for the sample autocorrelation function.

Theorem 3.4 (*Asymptotics for R_{nk} – AHTS*). Let Assumption 3.2 hold, and define

$$\bar{H}(V(r)) \equiv H(V(r)) - \int_0^1 H(V(r))\,dr.$$

(a) Let $\sigma_\varepsilon^2 = 0$. We have

$$R_{nk} \to_p 1 \tag{13}$$

as $n \to \infty$.

(b) Let $\sigma_\varepsilon^2 > 0$.

i. If $|\nu(\lambda)| \to \infty$ as $\lambda \to \infty$, we have (13) as $n \to \infty$.

ii. If $|\nu(\lambda)| \to 1$ as $\lambda \to \infty$, we have

$$R_{nk} \to_d \frac{\int_0^1 \bar{H}^2(V(r))\,dr}{\int_0^1 \bar{H}^2(V(r))\,dr + \sigma_\varepsilon^2}$$

for $k \geq 1$ as $n \to \infty$.

Shocks in (y_t) do not die out. In most cases, the asymptotic autocorrelation is simply unity at all lags. If the asymptotic order is constant (at least in the limit), there is a one-time decrease at $k = 1$, but no decay thereafter. Such behavior reflects the long-memory of the underlying random walk. Such a result is obvious for affine functions of random walks. It is surprising, however, that it holds for any H-regular transformation, as long as the AO is not decreasing to zero. The persistency in this case is quite extreme, and certainly could not be confused with any stationary process. On the contrary, these could more easily be confused with simple random walks.

The remaining statistics provide additional clues.

Theorem 3.5 (*Asymptotics for S_n^2, Q_n^3, K_n^4 – AHTS*). Let Assumption 3.2 hold, and define

$$\bar{H}(V(r)) \equiv H(V(r)) - \int_0^1 H(V(r))\,dr.$$

(a) Let $\sigma_\varepsilon^2 = 0$. We have

$$\begin{aligned} &\left[\nu^2(a_n)\right]^{-1} S_n^2 \to_d \textstyle\int_0^1 \bar{H}^2(V(r))\,dr \\ &Q_n^3 \to_d \left(\textstyle\int_0^1 \bar{H}^2(V(r))\,dr\right)^{-3/2} \textstyle\int_0^1 \bar{H}^3(V(r))\,dr \\ &K_n^4 \to_d \left(\textstyle\int_0^1 \bar{H}^2(V(r))\,dr\right)^{-2} \textstyle\int_0^1 \bar{H}^4(V(r))\,dr \end{aligned} \tag{14}$$

as $n \to \infty$.

(b) Let $\sigma_\varepsilon^2 > 0$.

i. If $|\nu(\lambda)| \to \infty$ as $\lambda \to \infty$, we have (14) as $n \to \infty$.

ii. If $|\nu(\lambda)| \to 1$ as $\lambda \to \infty$, we have

$$\left[\nu^2(a_n)\right]^{-1} S_n^2 \to_d \int_0^1 \bar{H}^2(V(r))\,dr + \sigma_\varepsilon^2$$
$$Q_n^3 \to_d \left(\int_0^1 \bar{H}^2(V(r))\,dr + \sigma_\varepsilon^2\right)^{-3/2} \left(\int_0^1 \bar{H}^3(V(r))\,dr + \tau_\varepsilon^3\right)$$
$$K_n^4 \to_d \left(\int_0^1 \bar{H}^2(V(r))\,dr + \sigma_\varepsilon^2\right)^{-2} \left(\int_0^1 \bar{H}^4(V(r))\,dr + 6\sigma_\varepsilon^2 \int_0^1 \bar{H}^2(V(r))\,dr + \kappa_\varepsilon^4\right)$$

as $n \to \infty$.

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. Note that $\left[\nu^2(a_n)\right]^{-1} \approx 1$ for large n when $|\nu(\lambda)| \to 1$. Dependence on the transformation is of course a vital characteristic not shared with conventional long-memory approaches. If an observed time series has an autocorrelation function like a random walk, but a sample variance that explodes too fast or not fast enough for a random walk, then it could instead be an AHTS process. Both the skewness and kurtosis are random, neither converging to zero nor exploding in the limit. Their limiting distributions depend explicitly on the LHF of the H-regular function driving the process.

4. Regressions with ITS and AHTS Models

Having established some distinguishing characteristics of time series driven by nonlinear transformations of stable random walks, we now turn to estimation of the function F in (2). Throughout this section, we assume that $\sigma_\varepsilon^2 > 0$ and parameterize (2) by

$$y_t = F(x_t, \theta) + \varepsilon_t, \tag{15}$$

where F is known up to an unknown parameter vector θ with true value θ_0. We also assume in this section that (ε_t) and (v_t) are mutually independent. We first look at the simple case in which the series (x_t) is directly observable. In this case, we show that the parameter vector θ can be estimated using NLS, with asymptotic distributions that are similar to those derived in Park and Phillips (2001), but with more general rates of convergence to allow for the possibility of thick-tailed innovations. We also consider estimation when (x_t) is unobservable. In this case, we observe only (y_t). Naturally, this requires additional assumptions, but we suggest obtaining parameter estimates by way of a nonlinear filter such as the extended Kalman filter (EKF).

4.1. Observable Explanatory Variable

Denote by $\hat{\theta}_n$ the usual nonlinear least squares estimator of the parameter vector θ in the model given by (1) and (15). Specifically,

$$\hat{\theta}_n \equiv \underset{\theta \in \Theta}{\operatorname{argmin}} \sum \left(y_t - F(x_t, \theta)\right)^2, \tag{16}$$

where Θ is the parameter set. For notational convenience, let

$$F_\theta(x,\theta) = \frac{\partial}{\partial\theta} F(x,\theta) \qquad \text{and} \qquad F_{\theta\theta}(x,\theta) = \frac{\partial^2}{\partial\theta\partial\theta'} F(x,\theta)$$

denote the vector of first derivatives and matrix of second derivatives with respect to θ.

For ITS models, we must modify Assumption 3.1.

Assumption 4.1 Let the time series (y_t) be generated by (1) and (15) with $\sigma_\varepsilon^2 > 0$. Let Assumption 2.2 of Park and Phillips hold, with Lévy motion V replacing Brownian motion in Assumption 2.2(a), and omitting the bound on the fourth moment of (v_t) in Assumption 2.2(d). Additionally, assume that

(a) Elements of (v_t) belong to the domain of attraction of a stable law of order $1 < \alpha \le 2$ with characteristic function φ satisfying $\varphi(s) \neq 1$ for all $s \neq 0$,

(b) F, F_θ, and $F_{\theta\theta}$ are I-regular on Θ,

(c) $\int_{-\infty}^{\infty} \left(F\left(x,\theta\right) - F\left(x,\theta_0\right)\right)^2 dx > 0$ for all $\theta \neq \theta_0$, and

(d) $\int_{-\infty}^{\infty} F_\theta\left(x,\theta_0\right) F_\theta\left(x,\theta_0\right)' dx > 0$.

These are essentially the assumptions for the main results of Park and Phillips (2001) for I-regular functions, except that elements of (v_t) no longer need finite variance. The Lévy motion generalizes the Brownian motion to allow for discontinuities in the sample path that may result from the limit of a thick-tailed random walk.

With this modest generalization, the result of their theorem holds with slight modification.

Theorem 4.1 (*Asymptotics for* $\hat{\theta}_n$ *– ITS*). Let Assumption 4.1 hold. The limiting distribution of $(\hat{\theta}_n - \theta_0)$ is given by

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

where W is standard Brownian motion independent of L.

There are two differences between our result and that of Park and Phillips (2001). The limiting distribution is defined in terms of a Lévy local time, rather than a Brownian local time. This local time still exists, as long as $\alpha > 1$. More importantly, the rate of convergence to this distribution is *different*. Recalling that $a_n = n^{1/\alpha}\ell(n)$, we obtain their rate of convergence $n^{1/4}$ as a special case when $\alpha = 2$ and $\ell(n) \to c$. If instead $\alpha = 2$ and $\ell(n) \to \infty$, the NLS estimator converges to this distribution at a slightly slower rate. If $1 < \alpha < 2$, the rate of convergence slows down quite dramatically. In the extreme case of $\alpha \approx 1$, the estimator is *almost* inconsistent. It cannot be inconsistent, however, because

$\alpha > 1$ and $\ell(n)$ is a slowly-varying function. Even if $\ell(n) \to \infty$, this is extremely slow and does not offset the remaining $n^{1/2-1/2\alpha}$.[6]

To examine the NLS estimator of AHTS models, we replace Assumption 3.2 with the following.

Assumption 4.2 Let the time series (y_t) be generated by (1) and (15) with $\sigma_\varepsilon^2 > 0$. Let Assumption 2.1 of Park and Phillips hold, with Lévy motion V replacing Brownian motion in Assumption 2.1(a). Additionally, assume that

(a) Elements of (v_t) belong to the domain of attraction of a stable law,

(b) F, F_θ, and $F_{\theta\theta}$ are H-regular on Θ, with $\inf_{\lambda>0} |\nu(\lambda)| > 0$ and ν not a function of θ,

(c) $\int_{|x|\leq\delta} (H(x,\theta) - H(x,\theta_0))^2 \, dx > 0$ for all $\theta \neq \theta_0$ and $\delta > 0$,

(d) $\int_{|x|\leq\delta} H_\theta(x,\theta_0) H_\theta(x,\theta_0)' \, dx > 0$ for all $\delta > 0$, and

(e) $\left\| (\nu_\theta \nu_\theta)^{-1} \nu \nu_{\theta\theta} \right\| < \infty$.

These assumptions again mirror those of Park and Phillips (2001) with obvious modifications. They allow the following result.

Theorem 4.2 (*Asymptotics for* $\hat{\theta}_n$ *– AHTS*). Let Assumption 4.2 hold. The limiting distribution of $(\hat{\theta}_n - \theta_0)$ is given by

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

where ν_θ and H_θ denote respectively the AO and LHF of vector valued F_θ, and W is standard Brownian motion independent of V.

Consistency is obtained at a potentially very fast rate, depending on α and ν.

As a result of the preceding two theorems, the asymptotics of $\hat{\theta}_n$ are mixed normal in *both* cases. Standard errors, t-tests, etc. generated by a standard regression package are therefore asymptotically valid, making inference convenient for both ITS and AHTS models. Recall that (x_t) is 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) is $(\mathcal{F}_{t-1})$-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 requires the strict exogeneity of (x_t).

[6] See the proof of Theorem 3.1 for a more detailed explanation.

4.2. Unobservable Explanatory Variable

When the sequence of explanatory variables (x_t) is unobservable, estimation is not as straightforward. To motivate the discussion, let (v_t) and (ε_t) be Gaussian iid sequences. A 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 a maximum likelihood (ML) routine. This technique assumes values for the model parameters, then creates series of $\mathbf{E}(x_t|\mathcal{F}_t)$ and $\mathbf{var}(x_t|\mathcal{F}_t)$ for each t, based on some initial values at time $t = 0$ and iterating linear projections. Once these series are created, ML is used to estimate θ. The series of conditional expectations of (x_t) generated by these estimates 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 spacecraft 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.

Implementation of the EKF

We summarize the discrete-time EKF below. Our EKF has a measurement equation given by (15) 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 F_x be the partial derivative of F with respect to x, i.e., $F_x(x,\theta) \equiv (\partial/\partial x)F(x,\theta)$. Using this notation, we expand $F(x_t,\theta)$ around $x_{t|t-1}$ to get

$$F(x_t,\theta) \approx F\left(x_{t|t-1},\theta\right) + F_x\left(x_{t|t-1},\theta\right)\left(x_t - x_{t|t-1}\right). \tag{17}$$

This allows us to write

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

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

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

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

$$\omega_{t|\cdot} \equiv \mathbf{var}(x_t - x_{t|t-1} \mid \mathcal{F}_\cdot) \qquad \text{and} \qquad \sigma_{t|\cdot} \equiv \mathbf{var}(y_t - y_{t|t-1} \mid \mathcal{F}_\cdot)$$

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

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

for conditional means and

$$\omega_{t|t-1} = \omega_{t-1|t-1} + \sigma_v^2 \qquad \text{and} \qquad \sigma_{t|t-1} = \omega_{t|t-1} F_x \left(x_{t|t-1}, \theta\right)^2 + \sigma_\varepsilon^2$$

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

$$x_{t|t} = x_{t|t-1} + \frac{\omega_{t|t-1}}{\sigma_{t|t-1}} F_x \left(x_{t|t-1}, \theta\right) \left(y_t - y_{t|t-1}\right)$$

and

$$\omega_{t|t} = \omega_{t|t-1} - \frac{\omega_{t|t-1}^2}{\sigma_{t|t-1}} F_x \left(x_{t|t-1}, \theta\right)^2 ,$$

similarly. By iteratively creating these series, we may estimate the parameters using ML, and thus obtain optimal series $(x_{t|t})$ 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

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

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 rigorous development of the statistical theory of the EKF for models having these features is well beyond the scope of this paper. We can provide only intuition and conjectures on why we believe the method should yield sensible results for such models.

Chang, Miller, and Park (2006) recently developed rigorous theory for the (linear) KF with integrated (x_t). As the (nonlinear) EKF is based on linear approximation, it may provide a reasonable method to analyze nonlinear and nonstationary models, just as it does for nonlinear and stationary models. Note that the expansion in (17) yields a relatively better linear approximation for nonstationary models, since in this case $x_t - \mathbf{E}(x_t|\mathcal{F}_{t-1})$ (which is approximated by $x_t - x_{t|t-1}$ in the nonlinear filter) is stationary and of a stochastic order smaller than that of (x_t). For stationary models, these series have the same order of magnitude.

Obviously, the presence of thick tails in the innovations may affect the validity of the EKF in a more fundamental manner. If $0 < \alpha \leq 1$ in our model, the mean of (v_t) is not well-defined, and taking conditional expectations of (x_t) becomes meaningless. We must therefore assume $\alpha > 1$. As long as this holds, 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 infinite 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 pseudo-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.

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)].[7]

5. Empirical Applications

We examine two empirical applications of our theoretical models. In the first application, we propose a simple ITS model to capture observed price "spikes" on wholesale electricity markets, using excess capacity. Both series are observable, and NLS is used to estimate the unknown parameters of the ITS model. In contrast, the second application features an AHTS model with an unobservable exogenous series. We formulate an AHTS model loosely based on the work of Krugman (1991) to extract the unobservable fundamental driving an exchange rate governed by a target zone exchange rate regime.

5.1. Electricity Price Spikes

Wholesale electricity markets in many parts of the U.S. and around the world are characterized by price "spikes" that occur during peak periods of demand when suppliers fall short of generation capacity, and the price dramatically (but temporarily) increases. Demand is extremely inelastic in such markets, as prices in downstream retail markets tend to be heavily regulated. 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

[7] 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.

marginal units to exercise considerable market power during peak demand periods, particularly when a negative supply shock occurs. The problem is exacerbated by the supply side, since electricity is not storable in large amounts, so the traditional price-smoothing role of inventories is lost. Hence, supply shocks may be more severe than in traditional markets. Price spikes were highlighted in the California electricity crisis of 2001, when some companies passed on the price increases to consumers and others filed for bankruptcy.

Serious financial problems and bankruptcy of not only local electricity companies but also of companies involved in wholesale trading, such as Enron, ensued in the aftermath of the California crisis. As a result, excessive prices and the abuse of market power have become important issues, both in the academic literature and in the mainstream media. In this light, many recent analyses in the energy literature have focused on modeling and forecasting wholesale prices.[8]

Nonlinear Nonstationary Model

Because of the peculiarities of this market, we believe one of the best predictors of price should be excess capacity. Let (y_t) represent electricity prices and (z_t) represent capacity utilization, measured as quantity divided by total system capacity on any given day. We consider $x_t = 1 - z_t$, where (x_t) is a series of measures of excess capacity. Assuming for now that capacity utilization is I(1),[9] we postulate a model using (15) with (1). The function F should have support only over $[0, 1)$, since excess capacity utilization is measured as a rate, and since the lack of storablility means that negative excess capacity is essentially impossible. Moreover, F should be bounded to avoid infinite prices. Setting aside the possibility of thick tails for the moment, an ITS model would be a reasonable candidate for this application.

Since these price spikes tend to be sudden, sharp, and ephemeral, it makes sense to model them with an exponential function. Specifically, we parameterize F by

$$F(x, \theta) = \begin{cases} \varpi \exp(-\gamma x) & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta = (\varpi, \gamma)'$ with scale parameter ϖ and slope parameter γ. Setting the error sequence (ε_t) aside for the moment, we may interpret ϖ as an empirical bound on the price of electricity, which is attained when there is no excess capacity available in the system. (The scale parameter is not a literal bound, unless the error term is strictly non-positive.) The slope parameter γ indicates how quickly the predicted price increases to ϖ. A large γ (relative to unity) indicates a sharp spike, while a small γ indicates a gradual price increase with a capacity shortfall. Because participants in the wholesale market respond quickly to changing market conditions, and because the function should predict very cheap electricity

[8]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 market.

[9]Strictly speaking, (x_t) takes values on $[0, 1)$ and therefore cannot be directly specified as a random walk. It may be more reasonable to model (x_t) as a nonlinear mapping of an I(1) process onto the unit interval – e.g., an H-regular function with unit AO. As our theory suggests, such a process behaves similarly to an I(1) process.

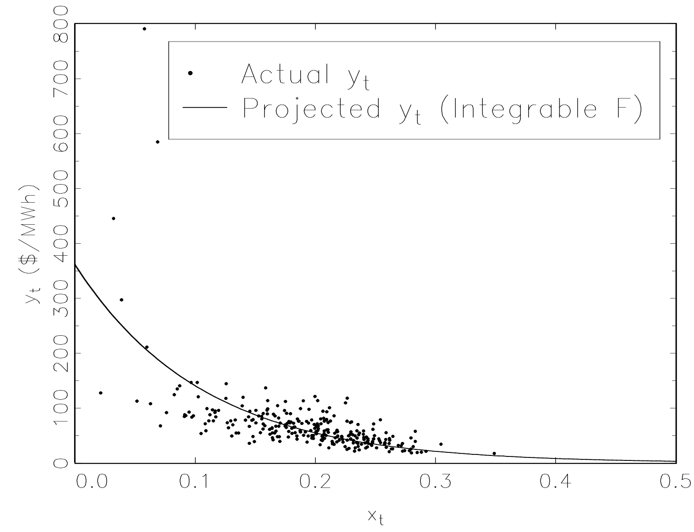

Figure 5.1: Electricity prices vs. excess capacity (April 1, 2002 – December 31, 2002).

when the system has vast amounts of excess capacity (near unity), we expect γ to be no smaller than 2 and more likely greater than 5.

Data and Empirical Results

Our empirical analysis uses maximum daily load divided by daily scheduled 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 (www.pjm.com). Parameter estimates using NLS are summarized in the following table.

Table 5.1

Parameter	*Estimate*	*Std. Error*
ϖ	361.0334	0.4453
γ	9.4278	0.0091

Highly significant parameter estimates support the nonlinear specification. As expected, we estimate a large γ, driving the sharp spikes evident in the data. Fitted estimates of (y_t) using these parameters are shown in Figure 5.1.

We estimate the stable parameter of the innovations of (x_t), and then test this series for integratedness, with critical value based on that estimate. Specifically, our estimate is $\hat{\alpha} = 1.5646$, using the estimation procedure of McCulloch (1986). The following table presents ADF test statistics for different lags, with the initial value subtracted from the

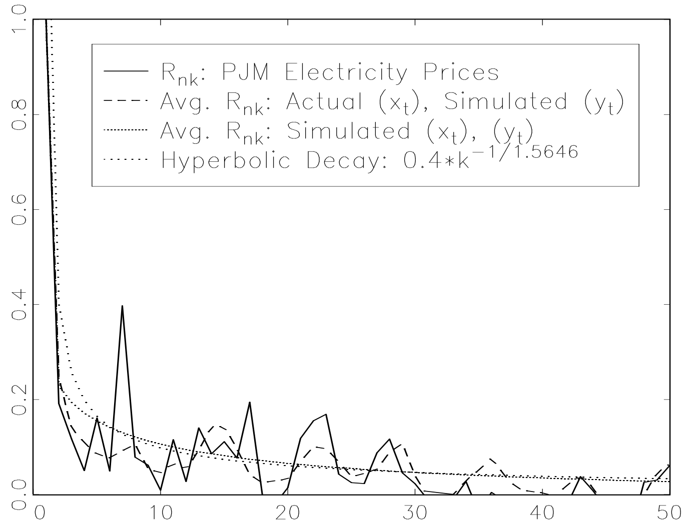

Figure 5.2: R_{nk} of actual prices and average R_{nk} of simulated prices.

series.

Table 5.2

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

Evidently, 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. An explanation lies in the fact that 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. Recall that from a theoretical point of view, the pure random walk assumption may be relaxed in light of our remarks at the end of Section 2. As a result, a series with serially correlated increments, such as this one, is an ITS process, and all of our asymptotic results hold.

Finally, we examine the autocorrelation function of (y_t). For comparative purposes, we average R_{nk} at each k across $50,000$ sets of simulations with a sample size of $1,000$. We use the parameter estimates above as the "true" values,[10] and the sets of simulations include simulated (ε_t) with both actual and simulated (x_t). Figure 5.3 shows these three

[10] In order to simulate (v_t), we draw pseudo-random numbers from a stable non-Gaussian distribution (with α, β, and σ set to the respective estimates above) using McCulloch's simulation procedure, based on Chambers, et al. (1976).

autocorrelations functions, compared to $0.4k^{-1/\hat{\alpha}}$, an example of hyperbolic decay using the estimated value of the stable parameter α. The striking similarity of the rates of decay lends credence to our asymptotic results and to our choice of ITS model for electricity prices.

In order to assess the degree of persistency in the sample autocorrelations, we estimate the memory parameter d using the technique introduced by Geweke and Porter-Hudak (1983) and two refinements by Andrews and Guggenberger (2003). Using GPH and one of the refinements, we obtain 0.14, while the second refinement estimates d to be 0.25. A memory parameter of 0.14 is consistent with a stationary fractionally integrated series with autocorrelations having a rate of decay of $k^{-0.72}$, which is very close to the rate $k^{-0.64}$ suggested by our model and estimate of α. If only d is estimated, it would be tempting to rely on the a conventional long memory model, which would ignore the richer specification of an ITS model.

5.2. Target Zone Exchange Rate Model

According to the 2006 Annual Report of the IMF, there are 67 countries with monetary policy that *de facto* target an exchange rate (or a composite of exchange rates). Seven of these have currency board arrangements, while the remaining 60 follow some sort of target zone policy including three categories. The majority (49) follow a conventional fixed peg policy, which allows deviations of up to ±1% from a central parity (target rate). Another six countries follow a policy (such as the ERM II) that allows fluctuations in excess of ±1%. The remaining five follow a crawling peg. Although exchange rate targeting policies are perhaps not as popular as they once were, such policies are still integral to a large number of countries.[11] 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 ±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.

Nonlinear Nonstationary Model

The most widely known of the target zone models was developed by Krugman (1991). The Krugman model postulates that under such a regime, the series of log exchange rates (y_t) may be modeled using a nonlinear function of the log of an economic fundamental (x_t). Krugman (1991) derives an "S"-shaped function that maps this fundamental onto the realized exchange rate. 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 fundamental (x_t) is generally treated as a regulated Brownian motion in the literature, with some authors adding a drift term. See, e.g., Svensson (1990) and de Jong

[11] To put this into perspective, there are 79 countries that have managed or free floats (more flexible) and another 41 countries – e.g., in the euro area – that have no individual legal tender (less flexible).

(1994). Alternatively, we may let (x_t) follow a random walk (as the discrete-time analog of a Brownian motion), as long as the "S"-shaped function is appropriately modified to allow an unbounded fundamental. A random walk may be a more appropriate assumption with a nonlinear filter, since most filters assume an autoregressive state equation with no regulation. Since there seems to be widespread empirical evidence that free-floating exchange rates may be appropriately modeled using an I(1) process, and since such exchange rates are linear functions of the fundamental, this seems reasonable. The discrete shifts in the domestic money supply that regulate the fundamental in Krugman's model may be approximated by the discontinuous sample paths that are allowed by thick-tailed innovations. Moreover, Dufour and Kurz-Kim (2003) provide evidence that free-floating exchange rates may follow thick-tailed random walks to begin with.

The function derived by Krugman (1991) relies on a bounded fundamental and is therefore inappropriate when (x_t) is allowed to be integrated. We instead use a smooth transition function, which keeps the spirit of the "S" shape, while allowing (x_t) to be unbounded. Specifically, we let

$$F(x, \theta) = \mu - h/2 + h \left(1 + \exp\left(\frac{\mu - x}{\gamma}\right)\right)^{-1}, \tag{18}$$

where μ is the shift parameter, γ is the slope parameter, and h is the bandwidth within which the exchange rate is allowed to fluctuate. It is easy to verify that our target zone model is an AHTS model with an AO of unity and an LHF given by

$$H(x, \theta) = \mu - h/2 + h \times 1(x \geq 0),$$

which is homogeneous of degree zero for any positive transformation.

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 (www.oanda.com). This was the longest period in which the band was $\pm 2.25\%$ with no realignments of the target, and provides a sample size of $1,085$. Figure 5.3 illustrates the sample autocorrelation function of (y_t). As predicted by our theoretical results for an AHTS model with unit AO and nondegenerate modeling error, we see an initial steep drop after $k = 0$. Contrary to our results, the decay continues beyond $k = 1$. Our estimates of d range from 0.5 to 0.6, suggesting that the memory of (y_t) is more persistent than that of a stationary fractionally-integrated process, but not as persistent as that of an I(1) process. Clearly, (y_t) has long memory, but an evident small-sample bias (stemming from second-order terms involving k) causes some decay, as we discuss further below.

Using the EKF, we first estimate $\theta = (\mu, h, \gamma)'$, σ_ε^2, and σ_v^2 – the pseudo-variance of (v_t) – and then estimate the stable parameters using the estimation procedure of McCulloch (1986). This first estimation is somewhat complicated by the nonlinear nature of both the model and the EKF. Since $x_{1|0}$ and $\omega_{1|0}$ are unknown, and since restricting (x_t) to be I(1) may be controversial, we first estimate an unrestricted model with $x_{1|0}$, $\omega_{1|0}$, and the

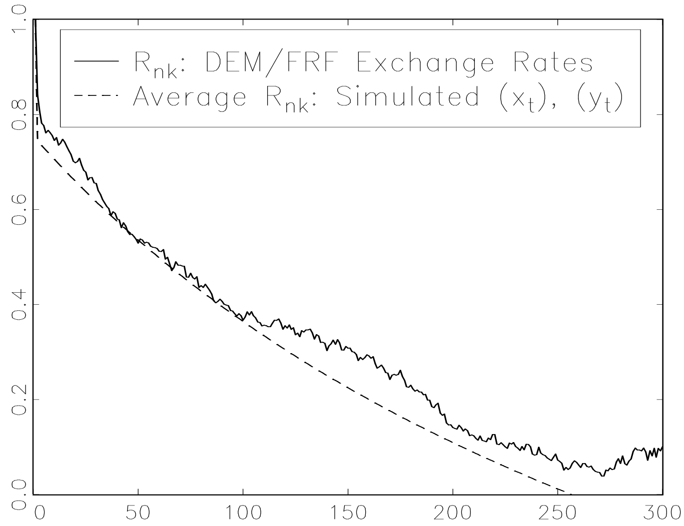

Figure 5.3: R_{nk} of actual exchange rate and average R_{nk} of simulated exchange rates.

autoregressive parameter ρ of (x_t) as additional unrestricted parameters. We restrict these later. In the unrestricted model, we restrict only h, γ, and the variances to be positive, so that the model is feasible. We find that ρ and $x_{1|0}$ are estimated to be close to 1 and y_0, respectively, so we fix these in the restricted model. In a neighborhood of the unrestricted estimates, there appears to be a trade-off between h and $\omega_{1|0}$. Fixing $\omega_{1|0}$ solves this identification problem, and we find that setting $\omega_{1|0} = 0.2$ yields estimates of h that translate into the announced bandwidth of ± 2.25. It is also difficult to uniquely identify σ_v^2 and γ with the EKF. The trade-off in this case is between a more volatile estimated fundamental (both σ_v^2 and γ large) and a less volatile fundamental (both σ_v^2 and γ small). Although the unrestricted model appears to generate a reasonable estimate $\hat{\gamma}$ of γ, subsequent simulations reveal an excessive confidence interval for $\hat{\gamma}$. In order to better identify γ, we restrict the pseudo-variance of (v_t) to be no smaller than the pseudo-variance of $(y_t - y_{t-1})$ and no larger than three times that pseudo-variance. Larger pseudo-variances can generate unrealistically volatile simulated fundamentals, implying nearly continuous monetary interventions to keep the exchange rate within the target zone. Smaller pseudo-variances would imply that the target zone policy actually pushes the exchange rate *away* from its target.

For notational convenience, we denote the parameter vectors estimated in the unrestricted and restricted models by τ^u and τ^r, where

$$\tau^r \equiv \left(\sigma_\varepsilon^2, \sigma_v^2, \theta'\right)' \qquad \text{and} \qquad \tau^u \equiv \left(\tau^{r\prime}, \rho, x_{1|0}, \omega_{1|0}\right)'$$

The table below summarizes the nonlinear transformations employed to enforce the above restrictions.

Table 5.3

Parameter		*Unrestricted*	*Restricted*
σ_ε^2	$=$	$\exp \tau_1^u$	$\exp \tau_1^r$
σ_v^2	$=$	$\exp \tau_2^u$	$\exp\left(-10.4206 + 1.0983/\left(1+\exp(-\tau_2^r)\right)\right)$
μ	$=$	τ_3^u	τ_3^r
h	$=$	$\exp \tau_4^u$	$\exp \tau_4^r$
γ	$=$	$\exp \tau_5^u$	$\exp \tau_5^r$
$x_{1\|0}$	$=$	τ_7^u	y_0
$\omega_{1\|0}$	$=$	τ_8^u	0.02
ρ	$=$	τ_6^u	1

In order to check for robustness, we calculate a likelihood ratio statistic of 4.0654. If we maintain the assumption that (ε_t) is a Gaussian iid sequence independent of (v_t) (which is not Gaussian), it seems reasonable to use a χ_4^2 critical value to evaluate the test statistic. For any reasonable test size, we cannot reject the null that the restricted model is valid, so all of the parameter estimates and confidence intervals from simulations reported below are derived from the restricted model.

As discussed above, the standard errors from ML estimation are not meaningful in the context of thick tails. Instead, we present 95% confidence intervals for our estimates of all parameters based on 50,000 simulations.[12]

Table 5.4

Parameter	*Estimate*	*Confidence Interval*
τ_1^r	-11.0476	$(-11.1412, -10.9435)$
τ_2^r	0.6240	$(-8.8591, 11.5120)$
τ_3^r	-1.2202	$(-1.2296, -1.2094)$
τ_4^r	-3.1123	$(-3.8943, -2.3754)$
τ_5^r	-2.9010	$(-3.7425, -1.5125)$
α	1.6122	$(0.7533, 2.0000)$
β	-0.1703	$(-1.0000, 1.0000)$
σ	0.0018	$(0.0001, 0.0018)$

All EKF intervals except that for the pseudo-variance of (v_t) appear to be reasonably tight. The fact that this pseudo variance cannot be pinned down easily leads to the inaccurate intervals for α, β, and σ estimated in the second estimation step.

The shift parameter μ may be interpreted as the target for the transformed exchange rate. Our estimate suggests a *de facto* target of

$$\exp(-1.2202) \approx 0.2952 \text{ DEM/FRF},$$

[12]Because of the highly nonlinear nature of both the model itself and of the estimation technique, we omitted simulations where the algorithm failed to converge after 200 iterations.

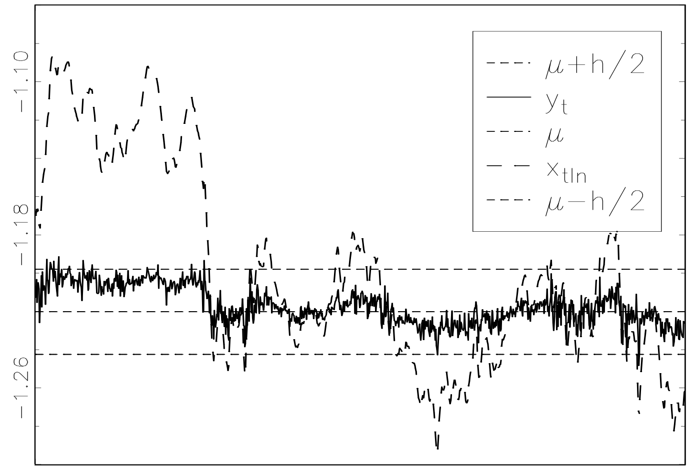

Figure 5.4: Original exchange rate with estimated fundamental, target, and bands.

with a *de facto* band of

$$\pm(\exp(\exp(-3.1123))^{1/2} - 1) \approx \pm 2.2499\%,$$

which is indeed very close to the announced bandwidth of ±2.25. Figure 5.4 illustrates the log of the exchange rate (y_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 estimated fundamental exhibits the expected properties. When the exchange rate approaches one of the bounds, the unconstrained fundamental strays beyond the bound. This lends credence to the nonlinear target zone model specification.

A remaining doubt that the DEM/FRF exchange rate might be generated by an AHTS model is the decay of sample autocorrelation function of (y_t) mentioned above. Although this decay is inconsistent with our large sample theory, simulations of an AHTS model in the family described by (18) provide insight into small sample performance of the sample autocorrelation function. Specifically, we average R_{nk} at each k across 50,000 simulations with a sample size of 1,000, using the parameter estimates above as the "true" values. This average sample autocovariance function is also illustrated in Figure 5.3, along with that of (y_t). The similarity is striking. There is an obvious decay in the autocorrelations of the simulated AHTS model from small sample bias. This result suggests that AHTS processes have shorter memory in small samples than in large sample (but still quite long memory), and that the memory of (y_t) is consistent with such a DGP.

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 reasonable with our modification of Krugman's function. Our empirical evidence supports the use of an AHTS model for target zone exchange rates.

6. Concluding Remarks

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. Our particular focus on persistency in memory leads us to conclude that such models may generate a variety of patterns of decay of asymptotic autocorrelations. 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. Doing so would ignore the richer specification that ITS and AHTS models have to offer.

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.* New York: Chapman & Hall.

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

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

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

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, 340-4.

Chang, Y., J.I. Miller, and J.Y. Park (2006). "Extracting a Common Stochastic Trend: Theory with Some Applications," *Journal of Econometrics*, forthcoming.

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. New York: John Wiley & Sons.

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.

Hall, P. (1990). "Asymptotic Properties of the Bootstrap for Heavy-tailed Distributions," *Annals of Probability*, 18, 1342-60.

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

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: Princeton University Press.

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

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

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

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.

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

McCulloch, J.H. (1986). "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics: Simulation and Computation*, 15, 1109-36.

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.). Boston: Kluwer Academic Publishers, 249-63.

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

Park, J.Y. (2006). "Nonstationary Nonlinearity: An Outlook for New Opportunities," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research* (A. Corbae, S.A. Durlauf, and B.E. Hansen, eds.). New York: Cambridge University Press, 178-211.

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

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

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

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.

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

Appendix A: Ancillary Lemmas with Proofs

Lemma A1 Let Assumption 3.1 hold, and define

$$M_n \equiv a_n n^{-1} \sum F(x_t).$$

We have

$$\sup_{n\geq 1} \|M_n\|_2^2 < \infty,$$

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

Proof of Lemma A1 Our approach is to show that (N_n) is uniformly integrable, where

$$N_n \equiv a_n n^{-1} \sum G(x_t)$$

for some appropriately chosen integrable function G, and then to show that such a function may be constructed close enough to F that (M_n) is also uniformly integrable.

Let $G_n(x) \equiv a_n n^{-1} G(a_n x)$ and $\hat{G}_n(\lambda) \equiv a_n n^{-1} \hat{G}(\lambda)$. Since $a_n n^{-1} = O(1)$ (see the proof of Theorem 3.1), $G_n \in L^1$ and the function $\hat{G}_n$ that coincides with its Fourier transform is also well-defined. Specifically,

$$G_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda x} \hat{G}_n(\lambda) d\lambda, \tag{19}$$

and it follows that

$$\hat{G}(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda x} G(a_n x)\, dx$$

using the inverse Fourier transform and the above definitions. We may choose G such that $\hat{G}$ has compact support. Since $\hat{G}$ has compact support,

$$\hat{G}_n(\lambda) \text{ vanishes outside of the interval } [-ca_n, ca_n] \tag{20}$$

for some constant $c > 0$. Now, let

$$I(G) \equiv \int_{-\infty}^{\infty} G(x) dx,$$

and note that

$$\int_{-\infty}^{\infty} |I(G)|^2 / \left(1 + |\lambda|^2\right) d\lambda < \infty \tag{21}$$

by the Cauchy-Schwarz inequality and integrability of G. Using a change of variables, we have

$$n\hat{G}_n(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda a_n^{-1} x} G(x)\, dx,$$

so that

$$\int_{-\infty}^{\infty} \frac{\left|n\hat{G}_n(\lambda) - I(G)\right|^2}{1 + |\lambda|^2} d\lambda = \int_{-\infty}^{\infty} \frac{\left|\int_{-\infty}^{\infty} (e^{i\lambda a_n^{-1} x} - 1) G(x) d(x)\right|^2}{1 + |\lambda|^2} d\lambda \to 0 \tag{22}$$

as $n \to \infty$, by dominated convergence, since $|G(x)|$ is bounded.

Now, N_n may be rewritten as

$$\begin{aligned} N_n &= n \int_0^1 G_n(V_n(r))\, dr \\ &= n \int_0^1 \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda V_n(r)} \hat{G}_n(\lambda) d\lambda dr \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} a_n \hat{G}(\lambda) \int_0^1 e^{-i\lambda V_n(r)} dr d\lambda \end{aligned}$$

using the above definitions, the definition of $V_n(r)$, and (19). The conditions for Theorem III.2.1 of BI (pg. 85) are satisfied by equations (20), (21), and (22). Following the proof of Theorem 2.1 of BI (pp. 87-88), it is now straightforward to deduce that

$$\|N_n\|_2^2 \leq c \left(\int_{-\infty}^{\infty} \frac{|I(G)|^2}{1+|\lambda|^\alpha} d\lambda \right)^{1/2}$$

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

Finally, for any F satisfying our assumptions, it is possible to create a function G as described above, such that

$$\int_{-\infty}^{\infty} |F(x) - G(x)|\, dx \leq \varepsilon \qquad \text{and} \qquad |F(x) - G(x)| \leq \varepsilon$$

for $\varepsilon > 0$. This is shown in the proof of Theorem IV.2.1 of BI (pg. 143). It follows from the triangle inequality that

$$|N_n - M_n| \leq a_n \int_0^1 |G(a_n V_n(r)) - F(a_n V_n(r))|\, dr$$

and that

$$\begin{aligned} a_n \int_0^1 |G(a_n V_n(r)) - F(a_n V_n(r))|\, dr &= a_n \int_{-\infty}^{\infty} L(1,x)\, |G(a_n x) - F(a_n x)|\, dx \\ &= L(1,0) \int_{-\infty}^{\infty} |G(x) - F(x)|\, dx \end{aligned}$$

by the occupation time formula (see BI, pg. 19, e.g.), a change of variables, and dominated convergence. Since $L(1,0)$ is bounded in L^2 (see Lemma I.4.2 of BI, e.g.),

$$\|N_n - M_n\|_2 \leq c\varepsilon$$

for some constant c. Finally, the inverse triangle inequality implies that

$$\left| \|N_n\|_2 - \|M_n\|_2 \right| \leq \|N_n - M_n\|_2 \leq c\varepsilon$$

which yields that stated result. □

Lemma A2 (*Asymptotics for Some Sample Moments – ITS*). Let Assumption 3.1 hold. We have

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

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

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

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

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

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

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

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

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

as $n \to \infty$.

Proof of Lemma A2 (*Asymptotics for Some Sample Moments – ITS*). We present four proofs from which the proofs of the remaining six parts easily follow. For the proof of mean asymptotics, we need only note that F^2, F^3, F^4 are I-regular. Results in parts (a), (d), and (e) then follow directly from Theorem IV.2.1 of BI (pg. 143).

The proof of part (b) follows along the similar lines as that of Theorem 6.3 in Park and Phillips (1999) and Theorem 3.2 of Park and Phillips (2001), except that $V_n(r)$ converges to a more general Lévy rather than to a Brownian motion, so that we allow different rates of convergence. We present a sketch of the proof. The reader is referred to Park and Phillips (1999, 2001) for more details. There exist a Brownian motion $U(r)$ with long-run variance σ_ε^2 and an increasing sequence of stopping times (τ_{nj}) such that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{j} \varepsilon_t =_d U\left(\frac{\tau_{nj}}{n}\right) \tag{23}$$

as $n \to \infty$. [See Park and Phillips (1999, 2001) or Hall and Heyde (1980).] Define the continuous martingale

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

where $\tau_{n,j-1}/n < r \leq \tau_{nj}/n$. Note that

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

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. We thus obtain the stated result in part (b), and the proofs of parts (f) and (g) clearly follow from along the same lines.

For the proof of part (c), we first let $k = 1$ and $a_1 = 1$. Write

$$\sum_{t=2}^n F(x_t)F(x_{t-1}) = \sum_{t=2}^n (GF)(x_{t-1}) + \sum_{t=2}^n F(x_{t-1})u_t, \tag{24}$$

where

$$G(x) \equiv \int_{-\infty}^{\infty} F(x+y)D_1(y)\mu(dy)$$

and

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

for $t \geq 1$. Obviously, G is well-defined for all $x \in \mathbb{R}$, since F and D_1 are integrable. Note that

$$\mathbf{E}\left(F(x_t)|\mathcal{F}_{t-1}\right) = \int_{-\infty}^{\infty} F(x+y)D_1(y \mid x)\mu(dy)$$

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 $t \geq 1$. Since the sequence (v_t) is iid, this is equal to $G(x_{t-1})$. 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

$$\int_{-\infty}^{\infty} (GF)(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D_1(y)dx\mu(dy)$$

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

$$a_n n^{-1} \sum_{t=2}^n (GF)(x_{t-1}) \to_d L(1,0) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D_1(y)dx\mu(dy). \tag{25}$$

Now, if we can show

$$a_n n^{-1} \sum_{t=2}^n F(x_{t-1})u_t = o_p(1), \tag{26}$$

then the stated result follows immediately from (24) and (25).

To establish (26), it is sufficient to show that

$$\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F(x_{t-1})u_t\right)^2 \to 0. \tag{27}$$

as $n \to \infty$. Using the fact that $(F(x_{t-1})u_t, \mathcal{F}_t)$ is an MDS and applying the law of iterated expectations, we may deduce that the LHS of (27) is equal to

$$a_n n^{-1}\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F^2(x_{t-1})u_t^2\right) = a_n n^{-1}\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F^2(x_{t-1})\mathbf{E}\left(u_t^2|\mathcal{F}_{t-1}\right)\right). \tag{28}$$

Defining

$$H(x) \equiv \int_{-\infty}^{\infty} F^2(x+y)D_1(y)\mu(dy).$$

allows

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

since $G(x_{t-1})$ is $\mathcal{F}_{t-1}$-measurable. Similarly to G, H is well-defined and bounded.

Finally, we define

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

and, again due to Theorem IV.2.1 of BI (pg. 143), we have

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

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. Note that

$$\int_{-\infty}^{\infty}(HF^2 - G^2F^2)(x)dx$$

is nonstochastic and finite. Moreover, since the expectation of the local time $L(1,0)$ is finite by Lemma I.4.2 of BI, (27) follows from (28), which completes the proof for $k = 1$.

More generally, let $k \geq 1$. Recall that D_k is defined as the PDF of $a_k^{-1}(v_1 + \cdots + v_k)$. In this case,

$$G(x) \equiv \int_{-\infty}^{\infty} F(x + a_k y)D_k(y)\, dy$$

and boundedness again follows from the integrability of F and D_k. The result thus follows in the same way as for $k = 1$.

The subsequent proof of part (h) extends to part (i) and (j) in an obvious way. Rewrite the sample moment in part (h) as

$$\sum F(x_t)\,\varepsilon_t^2 = \sigma_\varepsilon^2 \sum F(x_t) + \sum F(x_t)\,\varepsilon_{2,t},$$

and recall that $(\varepsilon_{2,t})$ is an MDS independent of (v_t) with bounded second moment, since we assume that $\mathbf{E}\,|\varepsilon_t|^4 < \infty$. The limiting distribution of the first term follows from Theorem IV.2.1 of BI (pg. 143). The second term is $O_p(a_n^{-1/2}n^{1/2})$ due to part (b) of this lemma, and is therefore, $o_p(a_n^{-1}n)$ since $\alpha > 1$ (see the proof of Theorem 3.1). The result immediately follows.

We may easily obtain the joint convergence of (a) – (j). Indeed, the joint mean asymptotics of (a), (c), (d), (e), (h), (i) and (j) and the joint covariance asymptotics of (b), (f) and (g) follow respectively in a straightforward manner from the usual Cramér-Wold device. Now, we embed the stochastic process U introduced in (23) into an extended probability space, where the distributionally equivalent copies of V_n and V in (8) (which we continue to denote by V_n and V respectively to avoid introducing unnecessary additional notations) are defined and $V_n \to_{a.s} V$. This is clearly possible due to the well-known Skorokhod representation and independence of V_n and V from U. The joint convergence to the mean and covariance asymptotics then follows readily as in Park and Phillips (2001). The reader is referred to Park and Phillips (2001) for more details. □

Lemma A3 (*Asymptotics for Some Sample Moments – AHTS*). Let Assumption 3.2 hold. We have

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

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

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

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

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

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

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

(h) $[n\nu(a_n)]^{-1} \sum F(x_t)\,\varepsilon_t^2 \to_d \sigma_\varepsilon^2 \int_0^1 H(V(r))\,dr$

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

(j) $[n\nu(a_n)]^{-1} \sum F(x_t)\,\varepsilon_t^3 \to_d \tau_\varepsilon^3 \int_0^1 H(V(r))\,dr$

jointly as $n \to \infty$.

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

As with the previous lemma, there are essentially four proofs. For the mean asymptotics, note that F^2, F^3, and F^4 are H-regular with AO's ν^2, ν^3, and ν^4 and LHF's H^2, H^3, and H^4, respectively. These functions are therefore regular at infinity due to Lemma 2.1, and the stated results in parts (a), (d), and (e) follow directly from Theorem IV.1.6 of BI (pg. 138).

For the proof of part (b), we may write

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

along the same lines as part (b) of the previous lemma. The stated result then follows in the same way. The proofs of part (f) and (g) are completely analogous.

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

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

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 H-regular function, note that

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

for any finite k. This approximation holds without differentiability of the H-regular function. A rigorous proof of the second equality is essentially identical to the lengthy and tedious proof of Lemma 3.2 in Chang and Park (2004), and is therefore omitted.

The proofs of parts (h), (i) and (j) are completely analogous to the proofs of the corresponding parts of Lemma A2, with substitutions for the appropriate rates of convergence, and are therefore omitted. Finally, the joint convergence of (a) – (j) follows as in the proof of Lemma A2. □

Appendix B: Proofs of the Main Results

Proof of Lemma 2.1 We focus on the case $x > 0$. The proof for $x < 0$ is entirely analogous. If we define

$$\ell_\kappa(x) \equiv (1/c_1)x^{-\kappa}F(x),$$

it follows immediately that

$$\lim_{x\to\infty}\frac{F(x)}{x^\kappa \ell_\kappa(x)} = c_1,$$

so it suffices to show that ℓ_κ is slowly varying at infinity – i.e., that (10) holds – in order to show that $F(x)$ is regular at infinity – i.e., that (9) holds with some constant c_1. However, (10) readily follows from the asymptotic homogeneity of F and (11), since

$$F(\lambda x) = \nu(\lambda)\,(H(x) + o(1)) \qquad \text{and} \qquad F(\lambda) = \nu(\lambda)\,(H(1) + o(1))$$

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 3.1 (*Asymptotics for R_{nk} – ITS*). First, let $\sigma_\varepsilon^2 > 0$. Expanding the numerator of $((n-k)/n)\,R_{nk}$ yields

$$\frac{1}{n}\sum_{t=k+1}^{n} y_t y_{t-k} + \left(1 - \frac{k}{n}\right)\bar{y}_n^2 - \bar{y}_n\left(\frac{1}{n}\sum_{t=k+1}^{n}(y_{t-k} + y_t)\right), \tag{29}$$

the first term of which is

$$\frac{1}{n}\sum_{t=k+1}^{n} F(x_t)F(x_{t-k}) + \frac{1}{n}\sum_{t=k+1}^{n} F(x_{t-k})\varepsilon_t + \frac{1}{n}\sum_{t=k+1}^{n} F(x_t)\varepsilon_{t-k} + \frac{1}{n}\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-k}. \tag{30}$$

Now, the first term of (30) is $O_p(a_n^{-1})$ with a limiting distribution given by Lemma A2(c), and the second two terms are $O_p(a_n^{-1/2}n^{-1/2})$ for all $k \geq 0$ by Lemma A2(b).[13] Note that

$$\frac{1}{n}\sum \varepsilon_t^2 \to_p \sigma_\varepsilon^2, \qquad \text{and} \qquad \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-k} \to_d \mathbb{N}\left(0, \sigma_\varepsilon^4\right)$$

for $k = 0$ and all $k \geq 1$, respectively, by a law of large numbers and central limit theorem for the MDS (ε_t). [See Hall and Heyde (1980), e.g.] Consequently, the asymptotic order of the final term of (30) is $O_p(1)$ for $k = 0$ and $O_p(n^{-1/2})$ for $k \geq 1$. Note that by Lemma A1 and the assumption that (ε_t) is an MDS,

$$\bar{y}_n = \frac{1}{n}\sum F(x_t) + \frac{1}{n}\sum \varepsilon_t = O_p(a_n^{-1}) + O_p(n^{-1/2}), \tag{31}$$

so that the remaining terms of (29) are

$$O_p(a_n^{-2}) + O_p(a_n^{-1}n^{-1/2}) + O_p(n^{-1}) = o_p(a_n^{-1}) + o_p(a_n^{-1/2}n^{-1/2}) + o_p(n^{-1/2})$$

since $a_n \to \infty$, and thus asymptotically negligible. Letting $k = 0$ gives the asymptotics for the denominator of R_{nk}. Again since $a_n \to \infty$, the dominant term of (30) is clearly the final term, so that the denominator of R_{nk} converges in probability to σ_ε^2.

[13] By construction, the test statistic is unity when $k = 0$, but we allow $k = 0$ in the numerator in order to get the asymptotics for the denominator.

When $k \geq 1$, the final term of (30) is $O_p(n^{-1/2})$, which still dominates the middle terms. Note that

$$n^{-\delta} < \ell(n) < n^{\delta} \tag{32}$$

for any $\delta > 0$ and for all n sufficiently large. This is well-known [see for example Feller (1971, Lemma 2, pg. 277)]. Together with our assumption that $\alpha > 1$, (32) implies that

$$a_n^{-1/2} n^{-1/2} = n^{-(1/2\alpha+1/2)} \ell(n)^{-1/2} = o\left(n^{-1/\alpha} \ell(n)^{-1}\right) = o\left(a_n^{-1}\right),$$

so that the first term of (30) also dominates the middle terms. Consequently, and since $(n-k)/n \to 1$ as $n \to \infty$, (12) is obtained. The limiting distributions of the respective terms come from Lemma A2(c) and an MDS central limit theorem.

Finally, note that

$$\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \varepsilon_t \varepsilon_{t-p} \qquad \text{and} \qquad \frac{1}{\sqrt{n}} \sum_{t=q+1}^{n} \varepsilon_t \varepsilon_{t-q}$$

are asymptotically independent for any $p, q > 0$ such that $p \neq q$.

Letting $\sigma_\varepsilon^2 = 0$ means that all but the first term of (30) drop out. The limiting distributions of numerator and denominator of R_{nk} are thus given by Lemma A1(c) and (a), respectively. □

Proof of Corollary 3.2 (*Rate of Decay of* R_k *– ITS*). Since we assume that the elements of (φ_k) are absolutely integrable, and since the characteristic function φ of a stable distribution is absolutely integrable, we may write

$$D_k(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi_k(s)\, ds \qquad \text{and} \qquad D(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi(s)\, ds,$$

due to the Fourier inversion formula. It can be deduced from these equations that

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

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

Note that absolute integrability of (φ_k) implies that the distribution of (v_t) is absolutely continuous with respect to the Lebesgue measure, so we use the notation dy in place of $\mu(dy)$. We have

$$\int_{-\infty}^{\infty} F(x + a_k y) D_k(y)\, dy = a_k^{-1} \int_{-\infty}^{\infty} F(x+y) D_k(a_k^{-1} y)\, dy$$

due to a change of variables,

$$a_k^{-1} \int_{-\infty}^{\infty} F(x+y) D_k(a_k^{-1} y)\, dy = a_k^{-1} \int_{-\infty}^{\infty} F(x+y) D(a_k^{-1} y)\, dy + o(a_k^{-1})$$

for large k, due to uniform convergence of D_k to D. Finally,

$$a_k^{-1} \int_{-\infty}^{\infty} F(x+y) D(a_k^{-1} y)\, dy + o(a_k^{-1}) = a_k^{-1} D(0) \int_{-\infty}^{\infty} F(x+y)\, dy + o(a_k^{-1})$$

due to dominated convergence and continuity of D at the origin. The stated result immediately follows. □

Proof of Theorem 3.3 (*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 3.1.

Letting $\sigma_\varepsilon^2 > 0$, we now turn to the numerator of Q_n^3. This may be rewritten as

$$\frac{1}{n}\sum(y_t - \bar{y}_n)^3 = \frac{1}{n}\sum y_t^3 - \bar{y}_n \frac{3}{n}\sum y_t^2 + 2\bar{y}_n^3, \tag{33}$$

and, using (31) and the proof for S_n^2, the final term of (33) is strictly dominated by the second term. The first two terms of (33) may be expanded and grouped as

$$\frac{1}{n}\sum F^3(x_t) + \frac{1}{n}\sum\left(\varepsilon_{3t} - 3\varepsilon_t\left(\frac{1}{n}\sum\varepsilon_{2t} + \sigma_\varepsilon^2\right)\right) + \tau_\varepsilon^3 \tag{34}$$
$$+\frac{3}{n}\sum F^2(x_t)(\varepsilon_t - \bar{y}_n) + \frac{3}{n}\sum F(x_t)\left(\varepsilon_{2t} - \frac{1}{n}\sum\varepsilon_{2t}\right) - 6\bar{y}_n\frac{1}{n}\sum F(x_t)\varepsilon_t$$

where $(\varepsilon_{2,t})$ and $(\varepsilon_{3,t})$ are defined above. The first term of (34) is $O_p\left(a_n^{-1}\right)$ with limiting distribution given by Lemma A2(d). Note that the second term may be rewritten as

$$\frac{1}{n}\sum\left(\varepsilon_{3t} - 3\sigma_\varepsilon^2\varepsilon_t\right) - 3\frac{1}{n}\sum\varepsilon_t\frac{1}{n}\sum\varepsilon_{2t}. \tag{35}$$

Since $\left(\varepsilon_{3t} - 3\sigma_\varepsilon^2\varepsilon_t\right)$, $(\varepsilon_{2,t})$, and $(\varepsilon_{3,t})$ are MDS's, each summation in (35) is $O_p\left(n^{-1/2}\right)$ using a CLT, and the whole term is therefore $O_p\left(n^{-1/2}\right)$ since the second term in (35) is $O_p\left(n^{-1}\right)$ by the continuous mapping theorem. Moreover,

$$\lim_{n\to\infty}\mathbf{var}\left(n^{-1/2}\sum\left(\varepsilon_{3t} - 3\sigma_\varepsilon^2\varepsilon_t\right)\right) = \mathbf{E}\varepsilon_t^6 - \tau_\varepsilon^6 - 6\sigma_\varepsilon^2\kappa_\varepsilon^4 + 9\sigma_\varepsilon^6,$$

yields the limiting variance for the CLT applied to the dominant term of (35). The third term of (34) is $O(1)$. The fourth term of (34) is

$$O_p(a_n^{-1/2}n^{-1/2}) + O_p(a_n^{-1})O_p\left(\max(a_n^{-1}, n^{-1/2})\right)$$

by Lemma A2(f), by Lemma A2(a) and a CLT, and by Lemma A2(a) and (31), respectively. The fifth term is

$$O_p((a_n^{-1/2}n^{-1/2}) + O_p(a_n^{-1})o_p\left(n^{-1/2}\right)$$

by an ancillary result in the proof of Lemma A2(h), and by Lemma A1 and a CLT. The final term of (34) is

$$O_p\left(\max(a_n^{-1}, n^{-1/2})a_n^{-1/2}n^{-1/2}\right)$$

by (31) and Lemma A2(b). It is clear that the last three terms of (34) are $o_p\left(\max\left(a_n^{-1}, n^{-1/2}\right)\right)$, so we may focus on the first three terms. The dominant term among these is τ_ε^3, unless τ_ε^3 happens to be zero. If so, then we must compare a_n with $n^{1/2}$, as we did in the proof of Theorem 3.1. Dominance of either the first or second term of (34) depends on α and the limit of $\ell(n)$.

The asymptotics for Q_n^3 when $\sigma_\varepsilon^2 = 0$ are essentially a special case, except that terms with (ε_t) are omitted. In this case, the asymptotics of the numerator follow directly from Lemma A2(d).

The asymptotics for K_n^4 when $\sigma_\varepsilon^2 > 0$ are more straightforward than for Q_n^3, since $\kappa_\varepsilon^4 > 0$. Expanding the numerator yields

$$\frac{1}{n}\sum (y_t - \bar{y}_n)^4 = \frac{1}{n}\sum y_t^4 + 6\bar{y}_n^2 \frac{1}{n}\sum y_t^2 - 4\bar{y}_n \frac{1}{n}\sum y_t^3 - 3\bar{y}_n^4, \tag{36}$$

and the first term is dominated by κ_ε^4 just as the first term of (33) is dominated by τ_ε^3 when $\tau_\varepsilon^3 > 0$, as may easily be verified using Lemma A2(e), (g), (i), (j), and an LLN. We only need to show that the remaining terms are $o_p(1)$. It is clear from (31) that $\bar{y}_n = o_p(1)$, so that positive powers of $\bar{y}_n$ are also $o_p(1)$. Moreover, we have

$$\frac{1}{n}\sum y_t^2, \frac{1}{n}\sum y_t^3 = O_p(1)$$

from the proofs corresponding to S_n^2 and Q_n^3. As a result, the second, third, and fourth terms of (33) are $o_p(1)$.

The case of $\sigma_\varepsilon^2 = 0$ is again straightforward, with the limiting distribution coming from Lemma A2(e). □

Proof of Theorem 3.4 (*Asymptotics for R_{nk} – AHTS*). Let $\sigma_\varepsilon^2 > 0$. First, note that

$$\bar{y}_n = \frac{1}{n}\sum F(x_t) + \frac{1}{n}\sum \varepsilon_t = O_p(\nu(a_n)) + O_p(n^{-1/2}) = O_p(\nu(a_n)) \tag{37}$$

from Lemma 3.1 in conjunction with Theorem 1.6 of BI (pg. 138), from a CLT since (ε_t) is an MDS, and since $\inf_{\lambda>0} |\nu(\lambda)| > 0$. Second, note that

$$\frac{1}{n-k}\sum_{t=k+1}^{n} y_t = \bar{y}_n + \left(\frac{k}{n}\right)\frac{1}{n-k}\sum_{t=k+1}^{n} y_t - \frac{1}{n}\sum_{t=1}^{k} y_t = \bar{y}_n + o_p(\nu(a_n))$$

since the latter two terms are $O_p\left(n^{-1}\nu(a_n)\right)$ and $o_p(1)$ for $k = o(n)$. And, third, note that

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

using similar arguments. Now, the numerator of R_{nk} may be written simply as

$$\frac{1}{n-k}\sum_{t=k+1}^{n} y_t y_{t-k} - \bar{y}_n^2 + o_p\left(\nu^2(a_n)\right) \tag{38}$$

using the above expressions. Expanding the first term of (38) and premultiplying by $(n-k)/n$ yields (30), as in the proof of Theorem 3.1. Now, the first term of (30) is $O_p\left(\nu^2(a_n)\right)$ with distributions given by Lemma A3(a) for $k = 0$ and (c) for $k \geq 1$, the middle terms are both $O_p\left(n^{-1/2}\nu\left(a_n\right)\right)$ by Lemma A3(b), and the final term is $O_p\left(1\right)$ for $k=0$ or $O_p\left(n^{-1/2}\right)$ for $k\geq 1$, similarly to the proof of Theorem 3.2. In the H-regular case, the first term of (30) dominates the remaining terms when either $k\geq 1$ or when $k=0$ if $\left|\nu\left(\lambda\right)\right|\to\infty$. If $\left|\nu\left(\lambda\right)\right|=O\left(1\right)$, final term of (30) has the same asymptotic order as the first when $k=0$. The asymptotic distribution of the second term of (38) follows from Lemma 3.1 in conjunction with Theorem 1.6 of BI (pg. 138). Finally, note from the proof of Lemma A3(c) that the first term of (30) for any k differs from itself for $k=0$ only by an $o_p\left(1\right)$ term. Together with the fact that $\left(n-k\right)/n\to 1$, this means that (38) may be rewritten as

$$\frac{1}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)^2+\frac{1}{n-k}\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-k}+o_p\left(1\right),$$

where $\bar{F}\left(x_t\right)\equiv n^{-1}\sum F\left(x_t\right)$. The second term is negligible when either $k\geq 1$ or when $\left|\nu\left(\lambda\right)\right|\to\infty$. In these cases, the whole statistic may be written simply as $1+o_p\left(1\right)$. Otherwise, the stated distribution is given by applying the continuous mapping theorem.□

Proof of Theorem 3.5 (*Asymptotics for S_n^2, Q_n^3, K_n^4 – AHTS*). The proof for S_n^2 follows directly from the asymptotics in denominator in Theorem 3.5. As in the case of Theorem 3.5, and contrary to the ITS case, the mean adjustment may not be ignored in the AHTS case.

The numerator of Q_n^3 may be rewritten as

$$\frac{1}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)^3+\frac{3}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)\varepsilon_t^2+\frac{1}{n}\sum\varepsilon_t^3+o_p\left(\nu^2\left(a_n\right)\right)$$

using algebraic manipulations, (37), Lemma A3(b) and (f), and the assumption that (ε_t^q) for $q=1,\ldots,4$ is an MDS. Note that the second term may be written as

$$\frac{3}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)\sigma_\varepsilon^2+\frac{3}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)\varepsilon_{2,t},$$

the first term of which is simply zero, and the second term of which is $o_p\left(\nu\left(a_n\right)\right)$. The limiting distribution of this expression then follows from (37), Lemma A3(d), (h), and again the assumption that (ε_t^q) is an MDS.

Similarly, the numerator of K_n^4 may be rewritten as

$$\begin{aligned}&\frac{1}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)^4+\frac{1}{n}\sum\varepsilon_t^4+\frac{4}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)\varepsilon_t^3\\&+\frac{6}{n}\sum\left(F\left(x_t\right)-\bar{F}\left(x_t\right)\right)^2\varepsilon_t^2+o_p\left(\nu^3\left(a_n\right)\right)\end{aligned}$$

using algebraic manipulations, (37), Lemma A3(a), (b), (d), (f), (g), (h), and the assumption that (ε_t^q) is an MDS. As with Q_n^3, the term involving the sum of $F\left(x_t\right)-\bar{F}\left(x_t\right)$ is zero plus $o_p\left(\nu\left(a_n\right)\right)$. The limiting distribution of this expression then follows from (37), Lemma A3(e), (i), (j), and again the assumption that (ε_t^q) is an MDS. □

Proof of Theorem 4.1 (*Asymptotics for $\hat{\theta}_n$ – ITS*). Under our assumptions, and similarly to Park and Phillips (2001), the NLS estimator in (16) may be rewritten as

$$\varpi_n' \left(\hat{\theta}_n - \theta_0\right) = -\left(\varpi_n^{-1} \sum F_\theta (x_t, \theta_0) F_\theta (x_t, \theta_0)' \varpi_n^{-1\prime}\right)^{-1} \varpi_n^{-1} \sum F_\theta (x_t, \theta_0) \varepsilon_t + o_p (1) \tag{39}$$

where (ϖ_n) is a matrix of appropriately-chosen normalization sequences. For the case considered here, in which all element of $F_\theta (x_t, \theta_0)$ are I-regular, we may choose ϖ_n to be equal to an identity matrix times $a_n^{-1/2} n^{1/2}$. According to Lemma A1 of Park and Phillips, the class of I-regular functions is closed under multiplication, so each element of the matrix $F_\theta F_\theta'$ is also I-regular. The limiting distribution of (39) thus follows along the same lines as Theorems 3.2 and 5.1 in Park and Phillips (2001), with the substitution of our Lemma A2(a), (b) for Theorem 5.1 of Park and Phillips (1999) used in Theorem 3.2 of Park and Phillips (2001). In particular, the rates of convergence $n^{-1/2}$ for mean asymptotics and $n^{-1/4}$ for covariance asymptotics are replaced by our more general $a_n n^{-1}$ and $a_n^{1/2} n^{-1/2}$, respectively. □

Proof of Theorem 4.2 (*Asymptotics for $\hat{\theta}_n$ – AHTS*). Similarly to Park and Phillips (2001) and the proof of Theorem 4.1, we may approximate (16) with (39) under our assumptions. Choosing (ϖ_n) in this case is not as simple, because although we assume that the elements of $F^{(\theta)}$ are H-regular, they need not have the same asymptotic orders. Again, $F_\theta F_\theta'$ is also H-regular by Lemma A1 of Park and Phillips (2001), with the AO of the ij^{th} element given by the AO of the i^{th} element of F_θ times the AO of the j^{th} element of F_θ, which is straightforward from the definition. The limiting distribution follows along the same lines as Theorems 3.3 and 5.2 in Park and Phillips (2001), using our Lemma A3(a) and (b). In this case, the rates of convergence $[n\nu^2(\sqrt{n})]^{-1}$ for mean asymptotics and $[n^{1/2}\nu(\sqrt{n})]^{-1}$ for covariance asymptotics are replaced by our more general $[n\nu^2 (a_n)]^{-1}$ and $[n^{1/2}\nu (a_n)]^{-1}$, respectively. □