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

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

Abstract

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

First Draft: December, 2003
This version: April, 2004

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

[1]We are grateful to Yoosoon Chang for many helpful discussions, and we would like to thank participants at the Rice University econometric workshop and at Texas Camp Econometrics IX (Fort Worth) for helpful comments.

1. Introduction

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

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

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

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

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

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

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

2. The Model and Preliminaries

Let (x_t) be the time series generated as

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

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

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

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, and $\mathbf{E}\,|\varepsilon_t|^p < \infty$ for some $p \geq 6$. We further assume that (v_t) and (ε_t) are uncorrelated, or equivalently that (x_t) are strictly exogenous. This assumption may be relaxed for many of our results, but it is especially convenient when dealing with regression asymptotics.

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

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

and

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

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

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

Consequently, the time series (y_t) specified by this model has the conditional mean given as a function of a random walk driven by innovations having thick tails. Our model thus has three ingredients that are commonly observed in many economic and financial time series: nonlinearity, nonstationarity, and thick tails.

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

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

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

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

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

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

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

where

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

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

We first assume $0 < \alpha < 2$. The case $\alpha = 2$ will be considered later. Define numerical sequences (a_n) and (b_n) by

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

as $n \to \infty$, and

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

Then it follows that

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

where

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

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

It is well known that we may set

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

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

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

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

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

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

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

as one may easily check.

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

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

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

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

From now on, we assume that (v_t) are properly centered. For $1 < \alpha \leq 2$, centering simply requires demeaning or assuming zero mean. For $\alpha = 1$, the proper centering can be difficult and more involved unless we assume that the underlying distribution is symmetric. No centering is necessary for the case of $0 < \alpha < 1$. The limiting distribution has the zero shift parameter, i.e., $\mu = 0$ if (v_t) are centered. Furthermore, we let the adjustment for scales be done apriorily so that the normalized sum of (v_t) converges in distribution to a stable distribution with unit scale parameter, i.e., $\sigma = 1$. The scale of the limit distribution only has a trivial effect on our subsequent results, since the rescaling of (v_t) amounts to merely redefining the transformation function F by a constant multiplication of its argument. The skewness parameter β is unrestricted, so we allow for asymmetric limit distributions for (v_t). Finally, the normalizing sequence (a_n) will be assumed to be given by (8) or (9), depending upon whether the distribution of (v_t) belongs the domain of attraction or of normal attraction of a stable law.

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

3. Time Series Properties of ITS and AHTS Models

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

3.1. Classes of Transformation Functions

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

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

for some constants $c > 0$ and $p > 1$. For example, any function that is bounded and has compact support satisfies this condition. Also, all probability density functions (PDF's) and their rescaled and shifted versions belong to the class, as long as they are bounded and decay at faster rates than $|x|^{-1}$ as $|x| \to \infty$. A possible interpretation of such a transformation is that it returns a strong signal when the value of the underlying random walk is near the mode (or modes) of some PDF-like function. We use an integrable transformation to model the relationship between the wholesale electricity price and the capacity utilization rate. Under our specification, we expect to observe a strong price spike whenever system generation nears capacity.

Much of the econometrics literature that deals with persistency in memory hinges on the assumption that a time series with long memory is generated by a fractionally integrated model with well-behaved innovations, while maintaining stationarity. We show, however, that ITS models may generate time series with autocorrelation functions exhibiting rates of decay proportional to stationary fractionally integrated models, but with a very different nonstationary nonlinear data generating mechanism driven by thick-tailed innovations. More precisely, ITS models yield an autocorrelation pattern that is identical to the I(d) model with the memory parameter $d \in (0, 1/4]$. This is just one example. Our results indeed show that other sample characteristics commonly observed from economic and financial time series data can be generated by ITS models.

Following the convention of Park and Phillips (1999), we define an asymptotically homogeneous transformation F such that $F(\lambda x) = \nu(\lambda) H(x) + R(x, \lambda)$ for large λ, where H is locally integrable and R is such that

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

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

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

Park and Phillips (1999) present some useful examples of asymptotically homogeneous transformations. The most common types of asymptotically homogeneous transformations in the literature are homogeneous (especially linear), polynomial, and logarithmic. Obviously, any homogeneous function is also asymptotically homogeneous. Moreover, polynomial functions are asymptotically homogeneous, with asymptotic properties stemming from the term with the highest order. Functions such as $\log x$ and $x^k \log x$ are, as well. The latter have asymptotic orders and LHF's of $\nu(\lambda) = \log \lambda$, $H(x) = 1$ and $\nu(\lambda) = \lambda^k \log \lambda$, $H(x) = x^k$, respectively. The more interesting sub-class of asymptotically homogeneous functions are, however, those that resemble rescaled and shifted cumulative distribution functions (CDF's). Any kind of CDF's have $\nu(\lambda) = 1$ and $H(x) = 1\{x \geq 0\}$, and all their rescaled and shifted versions have the same AO and LHF's given by some affine transformations of the function $1\{x \geq 0\}$.

Any kind of threshold model is essentially a CDF. Falling in this category are artificial neural networks, which frequently use logistic CDF's at the nodes of their hidden layers. If the exogenous signal in such a model follows a random walk, then a feedforward artificial neural network with one hidden layer is an AHTS model. Another example might be a model that aims to capture the price behavior on a regulated market with a price ceiling (such as some electricity and real estate markets). In this context, the limit of the observed price as the "natural" price increases is the price cap itself. Since prices are bounded below by zero, the LHF of such a transformation is essentially the same as that of a rescaled CDF. Still another example of an asymptotically homogeneous model is a target zone exchange rate model, in which policy actions force the observed exchange rate to stay within a fixed band around the target rate. If the underlying fundamental follows a random walk, then the exchange rate is generated by an AHTS model. We use a family of logistic functions that are parametrized appropriately to model this relationship in the empirical section of our analysis.

Asymptotically homogeneous transformations are closely related to the functions that are *regular-at-infinity*. A function F is said to be regular-at-infinity if it satisfies the following

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

conditions:

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

for some number $\kappa > -1$, where c_1 and c_2 are constants such that $|c_1|+|c_2|>0$, and ℓ is slowly varying at infinity, in the sense that $\lim_{\lambda\to\infty}\ell(\lambda x)/\ell(\lambda)=1$ for any $x>0$. The concept of regularity at infinity defines a very broad class of transformations, which includes asymptotically homogeneous transformations. This is shown in the following lemma.

LEMMA 3.1 Asymptotically homogeneous functions are regular at infinity.

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

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

3.2. Asymptotics for ITS Models

Here we investigate the properties of a time series generated by an ITS model. In particular, we look at the asymptotic behaviors of the sample autocorrelation, variance, skewness, and kurtosis of such a series. Computing sample statistics for a nonstationary process may be misleading, because they do not represent those of any well-defined underlying distribution. When the process is nonstationary, these are spurious sample statistics. Nevertheless, our results for these spurious statistics allow the comparison of our model with alternative modeling assumptions about the data generating process for the given time series of interest. Indeed, we show for instance that the autocorrelation pattern of the ITS process is directly comparable with that of a stationary I(d) process.

Our subsequent asymptotic results rely on the following assumptions.

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

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

Since the autocorrelation is the most important for our analysis, we begin with a theorem that gives asymptotics results for that statistic. First, we define the sample autocorrelation as

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

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

THEOREM 3.2 (*Asymptotics for R_{nk} – ITS*). Let Assumption 3.1 hold, and define

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

If $\sigma_\varepsilon^2 = 0$, then we have

$$R_{nk} \to_p R_k$$

Let $\sigma_\varepsilon^2 > 0$. If $1 < \alpha < 2$, then we have

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

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

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

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

We thus expect that the autocorrelation pattern of an ITS process is essentially determined by N_k defined in (13). Note that R_k is also given by a constant multiple of N_k. As this theorem shows, the asymptotic autocorrelation function of an ITS process is given by R_k, possibly with some random scale and shift factors when the transformed series is observed with an error. There is only one exceptional case where $\alpha = 2$ and $\ell(n) \to \infty$. However, even in this case, the second order term, which is of order smaller than the leading term only by $\ell(n)$, is given as a function of N_k. This is shown in the proof of the theorem.

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

$$R_k \to 0$$

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

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

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

as $k \to \infty$.

It is well-known that the sample autocorrelations of stationary fractionally integrated processes also decay at polynomial rates. In particular, such autocorrelations decay at the rate of k^{2d-1} where $d \in (0, 1/2)$ is defined as the *degree of fractional integration* or the *memory parameter*. The autocorrelations of processes generated by fractional Gaussian noise as specified by Mandelbrot decay at the rate of k^{2H-2} where $H \in (1/2, 1)$ is the Hurst coefficient. Geweke and Porter-Hudak (1983) show that any process generated by one of the models can be expressed in terms of the other model, with the expected relationship between these parameters $H = d + 1/2$.

One can see that as k increases, the rate of decay of our autocorrelation function exhibits behavior consistent with that of these other long-memory models (except when the model error dominates the asymptotic distribution). Observationally speaking, there is no difference between the rates of decay of the autocorrelation of these ITS models and that of either an I(d) process with $d \in (0, 1/4]$ or a fractional Gaussian noise process with $H \in (1/2, 3/4]$. Therefore, it would be easy to mistake a time series generated by an ITS model for a process generated by one of these well-known models. If the underlying DGP of an observed time series is in fact an ITS model, then such a misspecification would ignore valuable structural information about the process.

Moving on to the observed sample variance, skewness, and kurtosis of a time series (y_t) generated by an ITS model, we define these statistics as

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

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

and

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

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

In order to compare an ITS process with a stationary process, we introduce the following three theorems, which provide limiting distributions of these statistics.

THEOREM 3.4 (*Asymptotics for* S_n^2 *– ITS*). Let Assumption 3.1 hold. Then we have

$$S_n^2 \to_p \sigma_\varepsilon^2$$

when $\sigma_\varepsilon^2 > 0$, and

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

when $\sigma_\varepsilon^2 = 0$.

THEOREM 3.5 (*Asymptotics for* Q_n^3 *– ITS*). Let Assumption 3.1 hold and define $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$. Then we have

$$Q_n^3 \to_p \tau_\varepsilon^3 / \sigma_\varepsilon^3$$

when $\sigma_\varepsilon^2 > 0$, and

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

when $\sigma_\varepsilon^2 = 0$.

THEOREM 3.6 (*Asymptotics for* K_n^4 *– ITS*). Let Assumption 3.1 hold and define $\kappa_\varepsilon^4 = \mathbf{E}\varepsilon_t^4$. Then we have

$$K_n^4 \to_p \kappa_\varepsilon^4 / \sigma_\varepsilon^4$$

when $\sigma_\varepsilon^2 > 0$, and

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

when $\sigma_\varepsilon^2 = 0$.

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

3.3. Asymptotics for AHTS Models

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

The asymptotics here are based on the following assumptions.

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

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

Again, we start with asymptotics for the sample autocorrelation, which are given by the following theorem.

THEOREM 3.7 (*Asymptotics for R_{nk} – AHTS*). Let Assumption 3.2 hold. Then we have

$$R_{nk} \to_p 1$$

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

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

The following three theorems give us limiting distributions for the remaining statistics.

THEOREM 3.8 (*Asymptotics for S_n^2 – AHTS*). Let Assumption 3.2 hold. Then we have

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

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

THEOREM 3.9 (*Asymptotics for Q_n^3 – AHTS*). Let Assumption 3.2 hold. Then we have

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

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

THEOREM 3.10 (*Asymptotics for K_n^4 – AHTS*). Let Assumption 3.2 hold. Then we have

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

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

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

4. Regressions for ITS and AHTS Models

Having established some tools one may use to distinguish series driven by nonlinear transformations of stable random walks from alternative specifications, we now turn to the issue of regression. Adding an error term to the transformation, which is precisely what we did to create the models with $\sigma_\varepsilon^2 > 0$, naturally leads one to wonder about statistical inference. (Throughout this section, we assume that $\sigma_\varepsilon^2 > 0$.) We first consider the simplest case, in which (x_t) are observable. Nonlinear least squares will generate consistent standard errors as long as the assumption about the uncorrelatedness of (v_t) and (ε_t) is maintained. The asymptotic distributions of the estimators are similar to those derived in Park and Phillips (2001), with rates of convergence consistent with our more general innovations. We also consider regression when (x_t) are unobservable. Naturally, this requires additional assumptions, but we suggest obtaining parameter estimates by way of the extended Kalman filter. We subsequently consider an example of each of these situations in the empirical section of the paper.

4.1. Regression When (x_t) Are Observable

Regressions in which (x_t) are observable yield asymptotics results similar to those explored in detail in Park and Phillips (2001). The difference between that analysis and this one is simply that we allow for non-Gaussian stable innovations, but that analysis focused on Gaussianity. We replace (2) with the following refinement. Let

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

so that the only difference between (2) and (14) is that we are now explicitly incorporating the model parameters θ, which will be estimated by $\hat{\theta}_n$. In this light, we present the following two theorems.

THEOREM 4.1 (*Asymptotics for* $\hat{\theta}_n$ *– ITS*). Consider the time series (y_t) generated by (1) and (14), with integrable F and (v_t) belonging to the domain of attraction of a stable law. Let the conditions of Theorem 5.1 in Park and Phillips (2001) hold. For $1 < \alpha \leq 2$, as $n \to \infty$ the limiting distribution of $\hat{\theta}_n$ is given by

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

where $\dot{F}\left(s, \cdot\right)$ denotes $\partial F / \partial \theta$ and W is standard Brownian motion independent of L.

THEOREM 4.2 (*Asymptotics for* $\hat{\theta}_n$ *– AHTS*). Consider the time series (y_t) generated by (1) and (14), with asymptotically homogeneous F and (v_t) belonging to the domain of attraction of a stable law. Let the conditions of Theorem 5.2 in Park and Phillips (2001) hold. For $0 < \alpha \leq 2$, as $n \to \infty$ the limiting distribution of $\hat{\theta}_n$ is given by

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

where $\dot{H}\left(V\left(r\right), \cdot\right)$ denotes $\partial H / \partial \theta$, $\dot{\nu}\left(\cdot\right)$ denotes the asymptotic order of $\dot{H}$ and $U\left(r\right)$ is limiting stochastic process generated by summing (ε_t) and scaling by $\sqrt{n}$.

The ITS asymptotics require that (v_t) and (ε_t) are uncorrelated. While the AHTS asymptotics do not require this condition, the interpretation of the result is more intuitive when it holds. In particular, both results give us Gaussianity from the Brownian motion $W\left(1\right)$ and from the continuous martingale $U\left(r\right)$. This means that standard errors, t-tests, etc. that are generated by a standard regression package will be asymptotically unbiased. Thus, when the (x_t) are observable, inference from regression is straightforward.

4.2. Regression When (x_t) Are Unobservable

When (x_t) are unobservable, we need additional assumptions and tools to get parameter estimates. First, we consider the case in which the innovations (v_t) are Gaussian and thus have finite variance. The traditional method for dealing with linear models in which an exogenous variable is unobservable but assumed to follow an autoregressive process with such innovations is to use the Kalman filter (KF) fed into an MLE routine. This technique assumes values for the model parameters, then creates $\mathbf{E}\left[x_t|\mathcal{F}_t\right]$ and $\mathbf{E}\left[\sigma^2_{x_t}|\mathcal{F}_t\right]$ (where the latter denotes the conditional variance of (x_t) given information available at time t) for each t. These are based on some initial values at time $t = 0$ and the law of iterated projections. Once these series are created, MLE is used to optimize the model parameters. The series of conditional expectations of (x_t) generated by the optimal parameters are subsequently smoothed, in order to take into account information through the end of the sample. Even in the absence of Gaussianity, this method (quasi-MLE) yields consistent and asymptotically normal estimates of the model parameters, according to Hamilton (1994).

Since we are dealing with a nonlinear function F, the Kalman filter will not work. To find an alternative to the traditional Kalman filter, we turn to the engineering literature. The Kalman filter and its variants are widely used in this literature for such applications as tracking satellites and spacecrafts entering Earth's orbit. A common work-around is the extended Kalman filter (EKF), as described in Zarchan and Musoff (2000). The EKF is intuitively appealing, since it approximates $F(x_t)$ by expanding around $\mathbf{E}\left[x_t|\mathcal{F}_{t-1}\right]$, which is "known" at time $t-1$ (albeit unobservable), using a first-order Taylor series expansion. According to Zarchan and Musoff (2000), higher order expansions do not significantly improve the performance of the EKF. Since the EKF is clearly suboptimal, a number of papers have dealt with improving upon this methodology. See, for example, Crassidis and Markley (1997) or Julier and Uhlmann (1997). The econometrics literature also contains alternatives to the EKF. For example, Tanizaki (2000) surveys nonlinear, non-Gaussian state-space modeling using Monte-Carlo techniques.

We use the EKF to estimate $\mathbf{E}\left[x_t|\mathcal{F}_t\right]$ and then smooth these estimates to obtain $\mathbf{E}\left[x_t|\mathcal{F}_n\right]$. We summarize the discrete-time EKF below. Our EKF has a measurement equation given by

$$y_t = F(x_t) + \varepsilon_t$$

and a transition equation of

$$x_t = x_{t-1} + v_t.$$

For convenience of exposition, we use the conventional notation $\cdot_{t|t-1}$ to denote $\mathbf{E}\left[\cdot_t|\mathcal{F}_{t-1}\right]$. Using this notation, we expand F around $x_{t|t-1}$ to get

$$F(x_t) \approx F\left(x_{t|t-1}\right) + \frac{\partial F\left(x_{t|t-1}\right)}{\partial x_{t|t-1}}\left(x_t - x_{t|t-1}\right).$$

This allows us to write

$$y_t \approx \mu_F + \tilde{F}x_t + \varepsilon_t$$

where μ_F is defined as

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

which is constant at time t, and $\tilde{F}$ is simply the derivative of $F\left(x_{t|t-1}\right)$ with respect to its argument. Once the linear approximation is implemented, the EKF works exactly like the linear KF. Defining $\Omega_{t|\cdot} \equiv \mathbf{E}\left[\sigma^2_{x_t}|\mathcal{F}_\cdot\right]$ and $\Sigma_{t|\cdot} \equiv \mathbf{E}\left[\sigma^2_{y_t}|\mathcal{F}_\cdot\right]$ as conditional variances, we replace the usual linear prediction equations of the Kalman filter with

$$x_{t|t-1} = x_{t-1|t-1},$$

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

$$\Omega_{t|t-1} = \Omega_{t-1|t-1} + \sigma^2_v,$$

and

$$\Sigma_{t|t-1} = \tilde{F}^2 \Omega_{t|t-1} + \sigma^2_\varepsilon,$$

where σ^2_v is the variance of (v_t). This is well-defined in the Gaussian case, but in the more general α-stable case, we will use this notation to denote the pseudo-variance of (v_t), since the true variance is infinite. The updating equations become

$$x_{t|t} = x_{t|t-1} + \Omega_{t|t-1} \tilde{F} \Sigma^{-1}_{t|t-1} \left(y_t - y_{t|t-1}\right)$$

and

$$\Omega_{t|t} = \Omega_{t|t-1} - \Omega^2_{t|t-1} \tilde{F}^2 \Sigma^{-1}_{t|t-1}.$$

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

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

See Hamilton (1994) for a more detailed description of the filter in discrete time or Zarchan and Musoff (2000) for continuous time.

Similarly to the KF, in the absence of Gaussianity (but with (v_t) having finite variance), quasi-MLE using the EKF still retains the well-defined projection properties that make it optimal. Allowing for (v_t) having infinite variance, we can no longer make the projection interpretation. Nevertheless, since we are minimizing the sum of squared errors, the resulting parameter estimates and estimates of (x_t) should still be optimal. Once the conditional expectations of the unobserved series are extracted using the EKF, we estimate the stable and scale parameters of the empirical distribution of the innovation. This two-step methodology might be improved by incorporating the stable distribution directly into the log-likelihood function of the EKF procedure and estimating the parameters of the distribution directly. However, such a one-step procedure would be very difficult to implement, since the stable distribution does not have a closed form solution, except in special cases (Gaussian and Cauchy).

Since our theoretical assumptions dictate that the limiting distribution of the innovations is α-stable, possibly with $\alpha \neq 2$, estimates of the pseudo-variance of (v_t) generated by the EKF are not meaningful. Furthermore, omitting higher order terms of the Taylor series expansion biases estimates of the variance of (ε_t). For these reasons, and since these may be considered nuisance parameters, we suggest that estimates of these parameters be dropped. These problems also affect the standard errors of the parameters of interest. In this light, we suggest obtaining confidence intervals for estimates of these parameters by bootstrapping the fitted residuals $(\hat{v}_t)$ and $(\hat{\varepsilon}_t)$ and iteratively re-estimating the parameters.

5. Applications, Simulations, and Empirical Results

We examine two empirical applications of our theoretical models. The first application is a target zone exchange rate model. Theory does not always keep up with practice, and this seems to be true in this case. Target zone models have been used and tested since the 1980's, but the time series properties of such a nonlinear transformation of a nonstationary process were not well-known. With this in mind, we introduce an AHTS model in which (x_t) are unobservable. This illustrates a case in which something very similar to our model has already been used, but the asymptotics were not well-known. Furthermore, we demonstrate the importance of extracting the unobserved fundamental from a target zone model, by testing the long-run purchasing power parity (PPP) hypothesis using this fundamental. Exchange rates under an exchange rate targeting regime are assumed to be generated by a much more complicated DGP than that of a simple integrated process. Since cointegration relies on the assumption that the dependent variable is I(1), cointegration tests using such exchange rates are misspecified. These tests must be conducted on the fundamental and not on the exchange rate.

The second application is a model designed to capture observed price "spikes" on wholesale electricity markets by using an integrable transformation of excess capacity, which is observable. As opposed to the target zone model, which is formulated in the literature and already fits within the framework of one of our econometric models, we propose an electricity price ITS model as an alternative to more conventional approaches.

5.1. Target Zone Exchange Rate Model

Under the European Monetary System (EMS) of the 1980's and 1990's, exchange rates between participating EU countries were allowed to fluctuate within a fixed band around a central parity, which for most participating currencies was $\pm 2.25\%$ (and which was subsequently tightened to $\pm 1.125\%$ in 1990). Under the first phase of the EMS, the target rate could be adjusted by policymakers, if they chose not to defend the bands, but this was only allowed until 1990.

Since the EMS was replaced by the Euro by the majority of EU members in 2002, interest in target zone exchange rate models has waned. Many economists have taken the bipolar view that since enforcement of target zones lacks credibility, monetary authorities around the world are increasingly following either floating or fixed exchange rate regimes. While the bipolar view may be accurate, there are still a large number of countries that have regimes that fall somewhere between the ends of the spectrum. The IMF classifies exchange rate regimes into eight groups in its annual report, two of which are groups that have explicitly announced bands. As of April 2003, there are 10 countries that follow an explicit target zone regime. Among them are some of the European countries that aspire to join the EU, such as Hungary and Cyprus, as well as Denmark, which is already part of the EU but chose not to adopt the Euro. A few developing countries also have explicit targeting regimes. An additional 42 (mostly developing) countries fall into the category of having "other conventional fixed peg arrangements" (other than a currency board). As defined by the IMF, this category includes regimes that allow fluctuations of up to $\pm 1\%$ around a central rate. A target zone model with narrow bands may still be appropriate for some of these countries.[5] Evidently, there are still a large number of countries that neither completely fix there currency to that of another country nor completely let the value of their currency float. For a number of these countries, a target zone exchange rate model may be an appropriate way to capture exchange rate behavior.

Nonlinear Nonstationary Model. Much was written in the economics literature of the 1980's and 1990's about target zone exchange rate models (TZM's), which fall into a class of models that attempt to capture the behavior of exchange rates under this type of regime. Perhaps the most widely known of the target zone models was developed by Krugman (1991). The Krugman model postulates an exchange rate y_t that is a nonlinear function of a fundamental $x_t = m_t + w_t$, where (m_t) represent the endogenous money supply, (w_t) represents exogenous velocity shocks that follows a Brownian motion, and all variables are expressed in logs. Krugman (1991) derives an "S"-shaped function that maps the fundamental x_t onto the realized exchange rate y_t. Specifically,

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

where B and λ are model parameters. The transformation is a result of not only policy intervention, but perhaps even more importantly of rational expectations about policy intervention. These expectations bend the function at the bands 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.

While the literature generally agrees on the random walk assumption about (w_t), there seems to be disagreement on the interpretation of (w_t) and also on how to treat (m_t).

[5] In addition, there are 46 countries that fall into the category of "managed floating with no pre-announced path for the exchange rate." While some of these countries explicitly target inflation or monetary aggregates, many of them do not announce explicit targets. Central banks of these countries may be *de facto* anchoring to an exchange rate (either implicitly or explicitly but unannounced) in such a way that a target zone model would capture the behavior of exchange rates.

Krugman (1991) interprets (w_t) as velocity shocks. Svensson (1990) also interprets them as such, but defines them more precisely in terms of income, the real exchange rate, and the foreign price level and interest rate, among other terms. Mark (2001) defines the fundamental in a different way, in which (w_t) are implicitly defined in terms of the foreign money supply and domestic income relative to foreign income. The literature also differs in treatments of (m_t), which in every case essentially act as horizontal shifts, allowing the implicit targeting of a higher or lower fundamental, while retaining the same explicit target for the exchange rate. Svensson (1990) treats (m_t) in such a way that (x_t) follow a regulated Brownian motion. de Jong (1994) and Mark (2001) assume that (x_t) follow a random walk with a constant drift term, possibly included to reflect the belief that money growth is (on average) constant.

It is not immediately obvious that such a model may be specified as an AHTS model. First, we must assume that (x_t) follow a random walk. This is not an unrealistic assumption. If interventions – shifts in (m_t) – are not large, then they will be captured in the innovation of the random walk. Larger interventions will be captured by outliers of that innovation. In this sense, a thick-tailed α-stable innovation is appropriate to model jumps created by (m_t). Second, note that K is not asymptotically homogeneous. In fact, this is not necessary. The function derived in the Krugman model is not compatible with (x_t) that follow a random walk, for the very simple reason that such a series may take values on $(-\infty, \infty)$ and K loses the "S" shape abruptly beyond the bands.[6]

We postulate an alternative function for the model, which also follows the intuition of the "S" shape. We want a function that bends at the bands as Krugman's does, but does not allow the exchange rate to deviate from the band when the fundamental becomes too large or too small. Such a transformation might resemble a CDF with two horizontal asymptotes. To that end, we propose a generalized version of a logistic CDF as the heart of our TZM. In particular, we choose

$$\begin{aligned} F(x) &= \mu(1-\delta/2) + \delta\mu\left(1+\exp\left\{-\frac{1}{\gamma}(x-\mu)\right\}\right)^{-1} \\ &= \mu - h/2 + h\left(1+\exp\left\{-\frac{1}{\gamma}(x-\mu)\right\}\right)^{-1} \end{aligned}$$

where μ is the shift parameter, γ is the scale parameter, and δ is the width of the band within which the exchange rate is allowed to fluctuate (as a percentage of the shift parameter). Multiplying the CDF by $\delta\mu$ merely squeezes the function vertically to fit within the band. Adding $\mu(1-\delta/2)$ creates a fixed point at μ. We must necessarily use $h = \delta\mu$ to represent the bandwidth, because we modify (x_t) by taking logs and then demeaning. If the bandwidth is expressed as a percentage of the target rate, it will lose its interpretation as a result of the modifications to (x_t). Figure 5.1.1 illustrates this transformation. Even though (x_t) are

[6] In fact, as the fundamental takes arbitrarily large (small) values, $K(x_t)$ takes arbitrarily small (large) values! So, K cannot be used for any fundamental thus specified. If K is to be employed, (x_t) must be limited as in Svensson (1990). This misspecification may account for some of the rejections of the Krugman model in the literature.

demeaned, we retain the shift parameter, since there is no reason why $\mathbf{E}F(x_t)$ should be equal to the target rate. In fact, we can interpret μ as the target rate of the logged-then-demeaned exchange rate,[7] and at that point we would like the function to return the same value as its argument. Thus defined, our TZM is an AHTS model. The AO of the function $F(x)$ is unity and the LHF is

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

which is homogeneous of degree zero for any positive transformation.

Before examining empirical results from our model, we look at some simulated results. Consider a single simulated series of exchange rates (y_t), generated by a series of simulated fundamentals (x_t) following a thick-tailed random walk that are fed through our TZM. Figure 5.1.2 illustrates one such simulation, with parameters based on our estimates below. As expected, we can see that when the fundamental is within the bands, the value of the exchange rate will be close to that of the fundamental. As the fundamental approaches a band, the observed exchange rate is dampened and stays within the band. The exchange rate should stay near that band until the fundamental returns to within the bands. In order to generalize our results, we repeat this simulation $5,000$ times. Figure 5.1.3 shows the average of the sample autocorrelation from such simulations. Obviously, there is a small-sample bias, since our asymptotic result for series generated by an AHTS model suggested that the autocorrelations would not die out at all. We can see from the figure that simulation autocorrelations do in fact decay, albeit at a slow rate.

Having a functional form for F allows us to simulate the asymptotic distributions of the other sample statistics, since those distributions rely on F. The following table summarizes the mean and median of these simulated asymptotic distributions, as well as the mean and median from the simulations mentioned in the preceding paragraph.

Table 5.1.1

Statistic	*Simulated Asy. Dists.*		*Simulated Sample Stats.*	
	Mean	*Median*	*Mean*	*Median*
Variance	0.0006	0.0006	0.0003	0.0002
Skewness	−0.0811	−0.0784	−0.0178	−0.0893
Kurtosis	99.8315	4.6518	3.2710	2.7952

Clearly, the mean and median of the observed sample variance and skewness in finite samples are adequately represented by their asymptotic distributions. Asymptotically, there appears to be a strong tendency towards leptokurtosis, which is not apparent in small samples. This may be due to the fact that the Gaussian (ε_t) used in simulating the sample statistics have second-order effects that are not completely dominated in small samples. The asymptotic leptokurtosis appears to be neutralized by this small sample bias.

[7] As in Krugman (1991), both exchange rate and underlying fundamental are expressed in logs. This creates a minor empirical problem, namely, asymmetric bands. With the data we use in this analysis, we find that the asymmetry is negligible. Moreover, the derivation of our model does not rely on perfect symmetry, as does that of Krugman (1991).

Figure 5.1.4 shows the asymptotic distributions of the sample variance, skewness, and kurtosis from simulation. The unusual shape of the distribution of the variance comes from the well-known fact that the spatial distribution of an indicator function of a random walk follows the arcsine distribution. Since the variance is essentially the spatial distribution of the square of an indicator function, less a constant, it also follows the arcsine distribution. The distribution of the skewness comes from the facts that a random walk has a symmetric spatial distribution and our transformation is symmetric around the origin. The distribution of the kurtosis has an average that is quite high, but the median reveals another central tendency much closer to 3. Observed platykurtosis is also possible.

Data and Empirical Results. The data used in this empirical exercise are (demeaned log of) daily DEM/FRF exchange rates from March 1, 1979 to December 31, 1989 from OANDA (http://www.oanda.com). The original series (before demeaning and logging) is illustrated in Figure 5.1.5. It spans the early period of the EMS, in which the window was $\pm 2.25\%$. In order to compare periods in which the target rates are different, we "level" the series using publicly available information on the dates and magnitudes of realignments, as shown in Figure 5.1.6. This procedure may introduce some irregularities in exchange rate dynamics near the "fault lines", but we do not expect these irregularities to significantly affect the results.

Revisiting Figure 5.1.3, it is clear that this series has an autocorrelation function that dies out at a slow rate consistent with our simulations. Estimates of the memory parameter range from 0.33 using the technique developed by Mandelbrot and Wallis (1969) based on the Hurst coefficient to 0.50, 0.56, and 0.66 using the techniques of Geweke and Porter-Hudak (1983) and two refinements of those techniques from Andrews and Guggenberger (2003), respectively. While these reveal a significant small-sample bias compared to our asymptotic prediction that autocorrelations generated by an AHTS model do not die out at all, they suggest that the autocorrelations die out more slowly than those of a stationary fractionally integrated process, which has $d \in (0, 1/2)$.

We find an observed sample variance, skewness, and kurtosis of 0.0003, -0.4719, and 4.4049, respectively. These are consistent with our simulations of the AHTS model discussed above, suggesting that it may be an appropriate model for these data. Note that these statistics are not very consistent with a stationary autoregressive series generated by an underlying Gaussian distribution, suggesting that it would be difficult to conclude stationarity and dismiss the more complicated model.

In section 4, we discussed limitations of using the EKF in the context of a nonlinear, nonstationary model with thick-tailed innovations. For these reasons, we do not expect to get reasonable standard errors using this technique. Consequently, we use a bootstrap to estimate confidence intervals around the parameter estimates. The bootstrap confidence intervals were created by bootstrapping the fitted residuals $(\hat{\varepsilon}_t)$ and $(\hat{v}_t)$ generated by the EKF to create a new series $(\hat{y}_t)$ using the parameter estimates. The EKF was then re-run to get new parameter estimates. This was performed iteratively (1,000 times) to get distributions for the parameter estimates, from which the confidence intervals were created.

98% of the parameter estimates lie within the respective intervals, so if zero does not lie within an interval, this may be interpreted as a significance test of size 0.01. Our results are summarized in the following table.

Table 5.1.3

Parameter	*Estimate*	*Boot. Conf. Int.*
μ	-0.0072	$(-0.0139, -0.0011)$
h	0.0700	$(0.0646, 0.0716)$
γ	0.0349	$(0.0312, 0.0403)$

The densities of the parameter estimates, from which the bootstrap confidence intervals are constructed, are shown in Figure 5.1.7. We conclude that all parameters are significant with 99% confidence, since zero does not lie within any of the intervals. Parameter estimates for the variance of (ε_t) and pseudo-variance of (v_t) are not reported for reasons discussed above.

We may interpret μ as the target for the (demeaned log of the) exchange rate. This parameter gives us a *de facto* target of

$$\exp(-0.0072 - 1.2151) = 0.2946 \text{ DEM/FRF}$$

where -1.2151 is the mean of the logged exchange rate and $\hat{\mu} = -0.0072$. The distance between the bands is $\hat{h}$, which yields *de facto* bands of

$$\pm \exp\left(\frac{1}{2} \times 0.0700\right) - 1 = \pm 3.56\%$$

where $\hat{h} = 0.0700$. An alternative to estimating γ is to fix a value of γ such that the fundamental is preserved in a nearly linear transformation between the bands. This would be accomplished by setting γ so that the derivative of $F(x)$ is unity at $x = \mu$. By not fixing γ, we are implicitly allowing the authorities to intervene when the exchange rate is still within the bands and for rational expectations about future policy interventions to bend the "S"-shape further.

Figure 5.1.8 illustrates the leveled exchange rate (y_t), the smoothed conditional expectations of the fundamental $\left(x_{t|n}\right)$, the estimated target $\hat{\mu}$, and the estimated bands $\hat{\mu}\left(1 \pm \hat{\delta}/2\right)$. The fundamental exhibits the expected properties. When the exchange rate approaches one of the bands, the fundamental can be seen to exceed that band. This lends credence to the nonlinear TZM specification. Also, using McCulloch's procedure for estimating α, based on Chambers, et al. (1976), we estimate a stable parameter of approximately 1.54 for the empirical distribution of the innovations. This suggests that thick tails is an appropriate assumption. Since nonstationarity is an assumption in this case and not a testable result, we can neither confirm nor reject it. However, it is commonly believed that free-floating interest rates follow a random walk, which means that the fundamental must also follow a random walk in a free-floating exchange rate regime. There is no reason why this exogenous fundamental should behave differently in a TZM.

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. The random walk assumption is critical to our model, but we do relax the latter two assumptions. Theory and empirical evidence discussed above support using the AHTS model in this application.

Cointegration Test of the PPP Hypothesis. Finally, now that we have extracted (the conditional expectation of) the fundamental driving the exchange rate, we will demonstrate its importance. Suppose we want to test the long-run purchasing power parity (PPP) hypothesis. This long-run relationship is given by

$$y_t = \alpha_0 + \alpha_1 (p_t - p_t^*) + e_t^y$$

where y_t is the log of the exchange rate (DEM/FRF), p_t is the log of the German price index, p_t^* is the log of the French price index, and e_t^y is added to allow for short-run deviations from the PPP equilibrium. Common PPP tests use cointegration or fractional cointegration to test whether or not the series (e_t^y) exhibits long-run mean-reversion. Under a target zone exchange rate regime, such tests are inherently misspecified, since the observed exchange rate is not an I(1) process, as it would be under a free-floating regime. Instead of using the exchange rate, cointegration tests may be performed on the fundamental driving the exchange rate, since the fundamental is I(1) by assumption.

Testing the fundamental instead of the exchange rate can be justified by linearizing the inverse of the target zone function, as follows. We can write

$$\begin{aligned} F\left(x_{t|n}\right) &= \alpha_0 + \alpha_1 (p_t - p_t^*) + e_t^y \\ x_{t|n} &= F^{-1}\left(\alpha_0 + \alpha_1 (p_t - p_t^*) + e_t^y\right) \\ x_{t|n} &= G(z_t) \end{aligned}$$

where x_t is (the conditional expectation of) the fundamental, F is defined as above, $G(\cdot) = F^{-1}(\cdot)$, and $z_t = p_t - p_t^*$. Now, using a Taylor series expansion around z_0 yields

$$\begin{aligned} x_{t|n} &= G(z_0) + \frac{\partial G(z_0)}{\partial z_0}(z_t - z_0) + O\left(z_t^2\right) \\ &= \beta_0 + \beta_1 (p_t - p_t^*) + e_t^x \end{aligned}$$

where

$$\beta_0 = G(z_0) - \frac{\partial G(z_0)}{\partial z_0} z_0,$$

$$\beta_1 = \frac{\partial G(z_0)}{\partial z_0} z_t,$$

and e_t^x is an error term meant to capture the higher-order terms in the Taylor series expansion. Thus, we may perform the same test on the fundamental driving the exchange rate under a target zone regime as we could on the exchange rate under a free-floating regime such as (USD/DEM or USD/JPY).[8] Monthly consumer price indices for each country were obtained from the International Monetary Fund's *International Financial Statistics*. The fundamental extracted using the EKF was transformed by the same vector used to "level" the exchange rate, so that an "unleveled" fundamental could be compared to the original exchange rate. Figure 5.1.9 provides a graphical comparison. The series was then averaged across months to obtain observations at intervals comparable to the CPI data. The results of cointegration tests for long-run PPP between Germany and France using the augmented Dickey-Fuller ρ-test and t-test on e_t^y and e_t^x are presented in Table 5.1.4.

Table 5.1.4

	Misspecified	Well-Specified
Parameter value:	0.85	0.92
ADF ρ-test:	60.07	-12.20
ADF t-test:	-5.36	-2.63

Using the original exchange rate (the misspecified test), provides ambiguous evidence for the long-run PPP hypothesis, since the positive critical value of the ρ-test strongly suggests a unit root, while the t-test does not. On the other hand, the well-specified test shows critical values that suggest rejecting the unit root hypothesis in either case (even though the leading coefficient is closer to unity). Our second test therefore provides unambiguous evidence in support of the long-run PPP hypothesis, demonstrating the importance of extracting the fundamental using our model.

5.2. Electricity Price Spikes

Wholesale electricity markets in most regions of the U.S. and elsewhere are characterized by price "spikes" that occur during peak periods of demand when suppliers are short of capacity. These markets typically feature a Walrasian-type auction to determine the market clearing price. Specific market designs vary by region, but the basic auction mechanism used in energy markets (as opposed to capacity markets or markets for ancillary services) is essentially the same. Bids are ordered from lowest to highest to create a "supply stack", and the intersection of the supply stack with the demand curve determines the wholesale price. The demand curve is usually assumed to be completely or almost completely inelastic, due to heavy regulation of prices on the retail market, as well as time inconsistency issues caused by billing at monthly intervals (since prices change hourly or more frequently on wholesale markets). This is exacerbated by the fact that electricity is not storable in large amounts, so the traditional price-smoothing role of inventories cannot come into play. While the competitive market solution with a deregulated retail market would be the most efficient from the point of view of social welfare, this would not meet policymakers' long-standing goals of equitable distribution of cheap power. Consequently, it is necessary to allow bidding above marginal cost, in order to induce marginal units to produce during peak periods. The

[8] Incidentally, this technique assumes that $\frac{\partial G(z_0)}{\partial z_0}$ is well-defined, which is the case with our TZM. The derivative of the inverse of Krugman's function would not be well-defined, however, since the function has a derivative of zero at the bands.

price must be bid up significantly in order for these marginal units to cover their fixed costs over the short period of time in which they are necessary to maintain supply at the quantity demanded. This allows marginal units to exercise considerable market power during peak periods, through what is sometimes referred to as the "last-man bidding problem".

To illustrate the last-man bidding problem, consider a power system that is operating near full capacity. Suppose an unplanned outage occurs, meaning that the system operator must find additional power. Ignoring imports for the sake of expositional simplicity, it is clear that one or more marginal units must be brought online. If there are only a small number of these units available, or they are all owned by one firm, then clearly there is an incentive to bid up the price as high as possible. Whence the sharp "spikes" that frequently occur in price series from these markets. In light of the California electricity crisis and its aftermath, market power has become a very important issue in the energy literature. Many recent analyses in the literature have focused on forecasting wholesale prices, as such an exercise is valuable not only to market participants who might want to determine when they can employ market power, but also to system regulators who want to try to prevent abuse.[9]

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

$$x_t = 1 - u_t$$

where x_t represents a measure of excess capacity. We cannot assume that (x_t) follow a random walk, because the random variables in this series are bounded on the interval $[0, 1]$ by construction. However, we can and do assume that (x_t) are generated by some function g of a random walk (z_t). Thus, our function F must be a composite function such that $F = f \circ g$, for some f that we assume to be integrable. Since g is bounded, F should also be integrable. Thus $y_t = F(z_t) + \varepsilon_t$ is an ITS model by assumption, and we can get parameter estimates and perform simulations using $y_t = f(x_t) + \varepsilon_t$ since (x_t) are observable. Unfortunately, we cannot directly test the thick-tailed random walk assumption, since (z_t) is unobservable.

To the extent that we believe (x_t) will behave like a random walk within $[0, 1]$, due to the fact that (u_t) are generated by adding power generation shocks, we can test for integratedness and thick-tailed innovations directly on (x_t). We expect that a unit root test on (x_t) will have a tendency to over-reject the null, since (x_t) are generated by a transformation

[9] For example, McMenamin and Monforte (2000) use a feedforward artificial neural network with one hidden layer to forecast price based on lagged price, quantity, and some other exogenous variables. Knittel and Roberts (2001) explore several continuous-time diffusion specifications for electricity prices. Stevenson (2002) uses a wavelet filtering technique to "denoise" electricity prices, and then forecasts based on linear autoregressive and autoregressive switching models. While the latter two add structure to their reduced-form models, they do not use additional information beyond prices, lagged prices, and variables designed to capture temporal effects. The first analysis mentioned does use additional explanatory variables, but assumes no structure. The artificial neural network specification is unnecessarily flexible, since some structural assumptions can be made to strengthen the forecast.

that flattens the random walk when (z_t) exceed the interval. Using maximum daily load divided by daily scheduled capacity over the period of April 1, 2002 through December 31, 2002 from the Pennsylvania-Jersey Maryland (PJM) power pool (http://www.pjm.com), we test both of these assumptions. Using McCulloch's procedure for estimating α, based on Chambers, et al. (1976), we obtain an estimate of about 1.6. Thus, we find evidence for thick tails. A Dickey-Fuller test fails to find a unit root, however, so nonstationarity is not apparent in (x_t). Of course, this does not rule out a unit root in the unobservable series (z_t).

Finding evidence for nonlinearity is straightforward from observation. Using maximum daily real-time locational marginal price with the capacity data described above (also available at http://www.pjm.com), we perform a Nadaraya-Watson kernel regression of (x_t) onto the electricity price (y_t). The results are illustrated in Figure 5.2.1. Based on observation, we believe that a rescaled PDF is an appropriate function to model price as a function of capacity on the interval $0 \leq x_t < 1$. These endpoints come from the fact that excess capacity must be between 0% and 100%. Our postulated function is

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \theta_1 \exp\left\{-\frac{1}{\theta_2}(x-\theta_3)^2\right\} & \text{if } 0 \leq x < 1 \\ 0 & \text{if } x \geq 1 \end{cases},$$

where θ_1, θ_2, and θ_3 are parameters to be estimated. Parameter estimates using nonlinear least squares are summarized in the following table.

Table 5.2.1

Parameter	*Estimate*	*Std. Error*
θ_1	361.0024	32.9786
θ_2	0.1061	0.0076
θ_3	−0.0001	0.0898

Significant parameter estimates for the first two parameters support our specification. Since the third parameter is just a shift parameter, lack of significance is not a problem. The fitted model using these parameter estimates is also illustrated in Figure 5.2.1. It seems to follow the nonparametric fit quite well, except in the tails, where kernel estimates typically suffer from "empty bin" deficiencies.

Since we have postulated a functional form for f, we can compare observed sample statistics with those calculated from simulation. Unlike in the case of the target zone exchange rate model, we have observable (x_t). The only right-hand side series that must be simulated is (ε_t). Figure 5.2.2 illustrates one such sample simulation using the parameters estimated above compared to the actual price series (y_t). Figure 5.2.3 illustrates $|R_{nk}|$ of the actual price series compared to that of simulation averages. The autocorrelation function of (y_t) clearly dies out at a similar slow rate as that of simulated averages, which also die out at the polynomial rate of approximately $k^{-1/1.6}$, as our theory predicts. Furthermore, estimates of the memory parameter suggest that the process (y_t) is equivalent to a fractionally integrated process of approximately 0.14.[10] Simulation results for the other sample statistics are summarized in the following table.

Table 5.2.2

Statistic	*Mean*	*Median*
Variance	4351.6389	4344.3117
Skewness	0.5948	0.5898
Kurtosis	4.0269	3.9400

Our asymptotic results from Section 3 suggest that the variance converges to the variance of (ε_t), and that the skewness and kurtosis also converge. If we assume that (ε_t) are symmetric disturbances, then we would get a skewness of zero. If we further assume that (ε_t) are Gaussian disturbances, we would get a kurtosis of 3. Our simulations did in fact assume that (ε_t) were Gaussian, which suggests that there are small sample biases that give us positive skewness and leptokurtosis. The observed sample variance, skewness, and kurtosis for the actual series (y_t) are 4392.4926, 7.2313, and 68.6309, respectively, suggesting an even larger skewness and kurtosis, which could be the result of this bias, as well as non-Gaussian (ε_t). While these results do not uniquely point to an ITS model, they do not rule it out, either. Nevertheless, we believe that the fundamental relationship between the wholesale electricity price and excess system capacity supports the ITS specification.

6. Conclusion

We considered nonlinear transformations of random walks driven by thick-tailed innovations in this paper. We showed that such models generate a wide spectrum of patterns of persistency in memory. In particular, they generate time series that have asymptotic autocorrelations that decay very slowly as the number of lags increases or do not even decay at all and remain constant at all lags. The combination of nonlinearity, nonstationarity, and thick tails thus has the potential to generate the persistent memory patterns that are present in many of economic and financial time series data, as well as other prominent properties of many observed time series.

We established various time series properties for ITS and AHTS models. ITS models yield time series that have characteristics similar to those of stationary long-memory processes, with asymptotic autocorrelations decaying at a polynomial rate. In contrast, AHTS models generate time series that have asymptotic autocorrelation functions exhibiting no decay at all. We also derived asymptotics for the sample variance, skewness, and kurtosis. We extended the theories developed by Park and Phillips (2001) for nonlinear regressions with integrated processes to our models driven by stable random walks, in the case in which the explanatory variable is observable. When the explanatory variable is not observable, we discussed the use of the extended Kalman filter.

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

As illustrative examples of empirical applications of our models, we considered a target zone exchange rate model and an electricity price model. We argued that not only did the theoretical underpinnings of these models suggest our econometric framework, but also that both the series generated by these models and the random walk driving the models have empirical characteristics suggested by our theoretical and simulated results. We further tested the long-run PPP hypothesis using the fundamental extracted from the first model, since such tests conducted with the actual exchange rates are misspecified when a target zone regime is in effect.

References

Andrews, D.W.K. and P. Guggenberger (2003). "A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter." *Econometrica*, 71.2, 675-712.

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

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

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

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

Chambers, J.M., C.L. Mallows, and B.W. Stuck (1976). "A Method for Simulating Stable Random Parameters," *Journal of the American Statistical Association*, 71.354, 340-44.

Cheung, Y.-W. and K.S. Lai (1993). "A Fractional Cointegration Analysis of Purchasing Power Parity." *Journal of Business and Economic Statistics*, 11.1, 103-12.

Crassidis, J.L. and F.L. Markley (1997). *AIAA Journal of Guidance, Control, and Dynamics*, 20.6, 1096-103.

de Jong, F. (1994). "A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model," *Journal of Applied Econometrics*, 9.1, 31-45.

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

Geweke, J. and S. Porter-Hudak (1983). "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4.4, 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*, 1.1, 15-29.

Granger, C.W.J. and P. Newbold (1974). "Spurious regression in econometrics," *Journal of Econometrics*, 2.2, 111-20.

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

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

International Monetary Fund (2003). Annual Report 2003: Making the Global Economy Work for All. International Monetary Fund: Washington.

Julier, S.J. and J.K. Uhlmann (1997). "A New Extension of the Kalman Filter to Nonlinear Systems," *Proceedings of the International Society for Optical Engineering*, 3068, 182-93.

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*, 106.3, 669-82.

Mandelbrot, B.B. and J.R. Wallis (1969). "Some Long-Run Properties of Geophysical Records," *Water Resources Research*, 5.2, 321-40.

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

McCulloch, J.H. (1986). "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics: Simulation and Computation*, 15.4, 1109-136.

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

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

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

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

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

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69.1, 117-61.

Phillips, P.C.B. and V. Solo (1992). "Asymptotics for Linear Processes," *The Annals of Statistics*, 20.2, 971-1001.

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

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

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

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

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

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

Appendix A: Useful Lemmas and Their Proofs

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

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

then we have

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

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

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

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

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

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

and consequently, we have

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

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

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

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

then we may write

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

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

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

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

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

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

LEMMA A2 (*Asymptotics for Some Sample Moments – ITS*). Consider (2) with an integrable transformation F, (x_t) generated by (1), an MDS (ε_t) with respect to a filtration $(\mathcal{F}_t)$ to which (x_{t+1}) is adapted, and (v_t) belonging to the domain of attraction of a stable law. Define $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2$ and $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$. The following sample moments have asymptotic distributions and rates of convergence given as follows:

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

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

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

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

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

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

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

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

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

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

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

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

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

(a) Since F is integrable, F^2 must also be integrable. The result thus follows directly from Theorem 2.1 in BI (pg. 143).

(b) Since (ε_t) is assumed to be an MDS with respect to a filtration $(\mathcal{F}_t)$, which is contemporaneously uncorrelated with (v_t), we may apply Lemma 6.2 of Park and Phillips (1999, pg. 279). Hence, there exists a Brownian motion $U(r)$ constructed from compressing (ε_t) to fit the unit interval (by way of r), taking the partial sum over that interval, scaling by $n^{-1/2}$, and letting $n \to \infty$. The result then follows essentially from the proof of Theorem 3.2 in Park and Phillips (2001) with the appropriate substitution for the rate of convergence a_n of a stable process. Note that since we assumed that (v_t) and (ε_t) are uncorrelated, we do not have to worry about σ_{uv} being ill-defined in light of the distributions of the (v_t) having thick tails.

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

$$\sum_{t=2}^{n} F(x_t)F(x_{t-1}) = \sum_{t=2}^{n} (SF)(x_{t-1}) + \sum_{t=2}^{n} F(x_{t-1})u_t \tag{15}$$

where

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

and

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

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

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

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

It is easy to see that S is bounded. Therefore, since F is integrable, SF is also bounded. Furthermore, we have

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

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

$$\begin{aligned} a_n n^{-1}\sum_{t=2}^{n}(SF)(x_{t-1}) \quad &\to_d \quad L(1,0)\int_{-\infty}^{\infty}(SF)(x)dx \\ &= \quad L(1,0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x)F(x+y)D(y)dx\mu(dy) \end{aligned} \tag{16}$$

Now, if we can show

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

then the stated result would be immediate from (15) and (16).

First, note that $(F(x_{t-1})u_t)$ is an MDS, which means that

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

To establish (17), we will further prove that

$$\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F(x_{t-1})u_t\right)^2 \to 0 \tag{18}$$

for any $\alpha > 1$. Using the fact that $(F(x_{t-1})u_t)$ is an MDS, we may deduce that

$$\begin{aligned}\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F(x_{t-1})u_t\right)^2 &= a_n n^{-1}\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F^2(x_{t-1})u_t^2\right)\\ &= a_n n^{-1}\mathbf{E}\left(a_n n^{-1}\sum_{t=2}^{n} F^2(x_{t-1})\mathbf{E}\left(u_t^2|\mathcal{F}_{t-1}\right)\right) \qquad (19)\end{aligned}$$

Moreover, we may write

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

where

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

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

Now we define

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

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

$$N_n \to_d L(1,0)\int_{-\infty}^{\infty}(QF^2 - S^2F^2)(x)dx.$$

Since (N_n) is uniformly integrable as shown in Lemma A1, we have

$$\mathbf{E}\left[N_n\right] \to \mathbf{E}\left[L(1,0)\int_{-\infty}^{\infty}(QF^2 - S^2F^2)(x)dx\right].$$

Consequently, (18) follows from (19) whenever $\alpha > 1$, as was to be shown. The proof for $k = 1$ is now complete. The proof for the general case is obvious and omitted.

(d) From our functional central limit theorem in Section 2, we have

$$V_n(r) = a_n^{-1} \sum_{t=k+1}^{[nr]} v_{t-k}$$

Since we let $n \to \infty$, fixed k is negligible. Therefore, the asymptotics follow exactly the case in which $k = 0$, which was illustrated in part (b) of this lemma.

The proof of part (e) is identical to that of the previous proof, once it is noted that the functional central limit theorem governing the asymptotic distribution of (ε_t) is also invariant with respect to fixed k. The proofs of parts (f), (g), (i), and (j) are obvious by noting that F^2, F^3, and F^4 are integrable.

(h) We can rewrite the sample moment as

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

The distribution of the first term is obvious. To get the stated result, we just need to show that the second term is $o\left(a_n^{-1} n\right)$. We follow the convention of Park (2002) by writing $\varepsilon_{2,t} \equiv \left(\varepsilon_t^2 - \mathbf{E}\varepsilon_t^2\right)$. $(\varepsilon_{2,t}, \mathcal{F}_{t-1})$ is clearly an MDS, since

$$\mathbf{E}\left[\varepsilon_{2,t} | \mathcal{F}_{t-1}\right] = \mathbf{E}\left[\varepsilon_t^2 - \mathbf{E}\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = 0.$$

Since we assume that $\mathbf{E}\left|\varepsilon_t\right|^p < \infty$ for some $p \geq 6$, it is clear that $\mathbf{E}\varepsilon_{2,t}^2 < \infty$. The second term is $O_p\left(a_n^{-1/2} n^{1/2}\right)$ due to part (b) of this lemma, and is therefore $o\left(a_n^{-1} n\right)$ when $\alpha > 1$. The second term is therefore dominated and the asymptotics are determined by the first term.

The proof of part (k) follows directly from that of part (h) and the fact that F^2 is integrable, and that of part (l) also follows directly from that of part (h) by defining $\varepsilon_{3,t} \equiv \left(\varepsilon_t^3 - \mathbf{E}\varepsilon_t^3\right)$ and noting that $\mathbf{E}\varepsilon_{3,t}^2 < \infty$.

For ITS models, it can be shown that parts (h), (k), and (l) hold under more general conditions. In particular, $(\varepsilon_{2,t}, \mathcal{F}_{t-1})$ and $(\varepsilon_{3,t}, \mathcal{F}_{t-1})$ need not be homogeneous MDS's. We can let $p \geq 3$. Again, we just have to show that the second term is $o\left(a_n^{-1} n\right)$, so that the first term dictates the asymptotic distribution of the entire moment. We illustrate with $(\varepsilon_{2,t}, \mathcal{F}_{t-1})$. Define

$$\begin{aligned} N_n(r) &= a_n n^{-1/2} \sum_{t=1}^{j-1} F\left(a_n V_n\left(\frac{t-1}{n}\right)\right)\left(U_2\left(\frac{\tau_{nt}}{n}\right) - U_2\left(\frac{\tau_{n,t-1}}{n}\right)\right) \\ &\quad + a_n n^{-1/2} F\left(a_n V_n\left(\frac{j-1}{n}\right)\right)\left(U_2(r) - U_2\left(\frac{\tau_{n,j-1}}{n}\right)\right) \end{aligned}$$

where $\frac{\tau_{n,j-1}}{n} < r \leq \frac{\tau_{nj}}{n}$ and τ is a stopping time as specified in Park & Phillips 2001 and U_2 is the Brownian motion constructed form $(\varepsilon_{2,t})$. Thus, we may write

$$a_n n^{-1} \sum_{t=1}^{n} F(x_t)\, \varepsilon_{2,t} = N_n\left(\frac{\tau_{nn}}{n}\right).$$

Now, in order to show that this is asymptotically dominated, we need only show that the quadratic variance is degenerate. This is given by

$$\begin{aligned} [N_n]_r &= a_n^2 n^{-1} \sum_{t=1}^{j-1} F\left(a_n V_n\left(\frac{t-1}{n}\right)\right)^2 \left(\frac{\tau_{nt}}{n} - \frac{\tau_{n,t-1}}{n}\right) \\ &\quad + a_n^2 n^{-1} F\left(a_n V_n\left(\frac{j-1}{n}\right)\right)^2 \left(r - \frac{\tau_{n,j-1}}{n}\right) \\ &= a_n^2 n^{-1} \int_0^r F\left(a_n V_n(s)\right)^2 ds\, (1 + o_{a.s.}(1)). \end{aligned}$$

Given the results from part (a) of this lemma – in particular the rate of convergence – it is clear that

$$[N_n]_r \to_p 0,$$

since $a_n n^{-1} \to 0$ for $1 < \alpha \leq 2$. (This holds regardless of how $\ell(n)$ is specified. See the proof of Theorem 3.2.) This gives us the stated result.

LEMMA A.3 (*Asymptotics for Some Sample Moments – AHTS*). Consider (2) with an asymptotically homogeneous transformation F, (x_t) generated by (1), an MDS (ε_t) with respect to a filtration $(\mathcal{F}_t)$ to which (x_{t+1}) is adapted, and (v_t) belonging to the domain of attraction of a stable law. Define $\sigma_\varepsilon^2 = \mathbf{E}\varepsilon_t^2$ and $\tau_\varepsilon^3 = \mathbf{E}\varepsilon_t^3$. The following sample moments have asymptotic distributions and rates of convergence given as follows:

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

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

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

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

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

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

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

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

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

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

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

Proof of LEMMA A.3 (*Asymptotics for Some Sample Moments – AHTS*).

(a) It is easy to see from Theorem 1.6 of BI (pg. 138) that

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

since we show in Lemma 3.1 that all asymptotically homogeneous functions are regular-at-infinity. It remains only to show that $F^2(x)$ is asymptotically homogeneous with AO $\nu^2(\lambda)$ and LHF $H^2(x)$. We may write

$$\begin{aligned}F^2(x) &= \nu^2(\lambda)H^2(x)+\nu(\lambda)H(x)R(x,\lambda)+R^2(x,\lambda)\\ \frac{1}{\nu^2(\lambda)}F^2(x) &= H^2(x)+H(x)\frac{R(x,\lambda)}{\nu(\lambda)}+\frac{R^2(x,\lambda)}{\nu^2(\lambda)}.\end{aligned}$$

From the definition of an asymptotically homogeneous function, the last two terms obviously disappear. This proves the result.

The proof of part (b) is similar to that of Lemma A2(b), with the appropriate substitution for the rate of convergence.

(c) As in the case of the Lemma A2(c), we first consider the case in which $k=1$ and $a_1=1$. Invoking definitions from that lemma, we may again write

$$\sum_{t=1}^{n}F(x_t)F(x_{t-1})=\sum_{t=1}^{n}(SF)(x_{t-1})+\sum_{t=1}^{n}F(x_{t-1})u_t. \tag{20}$$

We claim that $S(x)$ is asymptotically homogeneous with the same AO and LHF as $F(x)$. To verify this claim, write

$$\begin{aligned}S(\lambda x) &= \int_{-\infty}^{\infty}F\left(\lambda\left[x+\frac{y}{\lambda}\right]\right)D(y)\mu(dy)\\ &= \int_{-\infty}^{\infty}\left\{\nu(\lambda)H\left(x+\frac{y}{\lambda}\right)+R\left(x+\frac{y}{\lambda},\lambda\right)\right\}D(y)\mu(dy).\end{aligned}$$

Letting $\lambda\to\infty$ gives us

$$S(\lambda x)\to_p\nu(\lambda)H(x)+R(x,\lambda),$$

since $\int_{-\infty}^{\infty}D(y)\mu(dy)=1$ by definition. We also claim that $F(x+y)\to_p F(x)$. To verify this second claim, write

$$\begin{aligned}F(\lambda(x+y)) &= \nu(\lambda)H\left(x+\frac{y}{\lambda}\right)+R\left(x+\frac{y}{\lambda},\lambda\right)\\ &\to_p \nu(\lambda)H(x)+R(x,\lambda)\end{aligned}$$

as $\lambda\to\infty$.

Now, if we can show that the last term of (20) is $o\left(n\nu^2\left(a_n\right)\right)$, the stated result will obtain. Since (u_t) is an MDS, the expectation of that term obviously collapses to zero. We must further show that

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

Exactly the same techniques may be applied as in the ITS lemma to rewrite the left-hand side of the above expression to obtain

$$\left[n^2\nu^4\left(a_n\right)\right]^{-1}\sum_{t=2}^{n}F^2(x_{t-1})\left(Q\left(x_{t-1}\right)-S^2\left(x_{t-1}\right)\right) \to 0,$$

since

$$\left[n\nu^4\left(a_n\right)\right]^{-1}\sum_{t=2}^{n}F^2(x_{t-1})Q\left(x_{t-1}\right) \to_d \int_0^1 H^4\left(V\left(r\right)\right)dr,$$

and

$$\left[n\nu^4\left(a_n\right)\right]^{-1}\sum_{t=2}^{n}F^2(x_{t-1})S^2\left(x_{t-1}\right) \to_d \int_0^1 H^4\left(V\left(r\right)\right)dr.$$

This yields the desired result for $k = 1$. The proof for general k is now straightforward and omitted.

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

Appendix B: Proofs of the Main Results

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

$$H(x) = |x|^\kappa H(1) \tag{21}$$

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

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

then it follows immediately that

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

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

$$\lim_{\lambda\to\infty} \frac{\ell_\kappa(\lambda x)}{\ell_\kappa(\lambda)} = 1 \tag{22}$$

to finish the proof. However, (22) readily follows from the asymptotic homogeneity of F and (21), since

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

for large $\lambda > 0$, and therefore

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

as $\lambda \to \infty$.

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

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

and therefore, for fixed k,

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

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

Write

$$\sum_{t=k+1}^{n} y_t y_{t-k} = \sum_{t=k+1}^{n} F(x_t)F(x_{t-k}) + \sum_{t=k+1}^{n} F(x_t)\varepsilon_{t-k} + \sum_{t=k+1}^{n} F(x_{t-k})\varepsilon_t + \sum_{t=k+1}^{n} \varepsilon_t \varepsilon_{t-k}. \quad (23)$$

Due to Lemma A2, we have

$$\sum_{t=k+1}^{n} F(x_t)F(x_{t-k}) = O_p(a_n^{-1} n) \quad (24)$$

and

$$\sum_{t=k+1}^{n} F(x_t)\varepsilon_{t-k}, \quad \sum_{t=k+1}^{n} F(x_{t-k})\varepsilon_t = O_p(a_n^{-1/2} n^{1/2}) \quad (25)$$

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

$$\frac{1}{n} \sum_{t=k+1}^{n} \varepsilon_t^2 \to_p \sigma_\varepsilon^2, \quad (26)$$

and for all $k \geq 1$

$$\frac{1}{\sqrt{n}} \sum_{t=k+1}^{n} \varepsilon_t \varepsilon_{t-k} \to_d \mathbb{N}\left(0, \sigma_\varepsilon^4\right), \quad (27)$$

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

We first consider the case $k = 0$ in (23), which also gives us asymptotics for the denominator. It is obvious from (24)–(26) that

$$\frac{1}{n} \sum_{t=k+1}^{n} y_t^2 = \frac{1}{n} \sum_{t=k+1}^{n} \varepsilon_t^2 + o_p(1) \to_p \sigma_\varepsilon^2, \quad (28)$$

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

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

Next, to consider the case $k \geq 1$ in (23), we first note that

$$n^{-\delta} < \ell(n) < n^{\delta} \quad (29)$$

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

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

and therefore,

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

for all large n. Consequently, the terms in (25) are smaller than those in (24) and asymptotically negligible for all $k \geq 0$.

Let $1 < \alpha < 2$. Then it follows from (29) that

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

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

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

due to Lemma A2(c), which together with (27) immediately yields the stated result in this case. Now we let $\alpha = 2$. In this case, the dominant terms would differ depending upon whether $\ell(n) \to 0, c, \infty$. If, for instance, $\ell(n) \to c$ for some constant c, then we have both the first term and the last term in (23) for our asymptotics. As a result, we have

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

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

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

$$D_k(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi_k(s) \, ds \tag{30}$$

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

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

since the characteristic function φ of a stable distribution is absolutely integrable. Therefore, it can be easily deduced from (30) and (31) that

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

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

Now we have

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

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

Proof of THEOREM 3.4 (*Asymptotics for* S_n^2 *– ITS*). The proof follows directly from the asymptotics of the denominator in Theorem 3.3.

Proof of THEOREM 3.5 (*Asymptotics for* Q_n^3 *– ITS*). The proof for the asymptotics of the denominator follows directly from Theorem 3.3. Letting $\sigma_\varepsilon^2 > 0$, we focus on the numerator. It is easy to show that the mean adjustment is asymptotically negligible, as in the proof for the autocorrelation. Expanding the dominant term yields

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

which with the fact that

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

and with Lemma A2 gives us the desired result. When $\sigma_\varepsilon^2 = 0$, only the first term remains.

Proof of THEOREM 3.6 (*Asymptotics for* K_n^4 *– ITS*). The proof is very similar to that of the sample skewness, by expanding the numerator and noting that

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

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

Proof of THEOREM 3.7 (*Asymptotics for* R_{nk} – *AHTS*). Let $\sigma_\varepsilon^2 > 0$. Again, the case in which $\sigma_\varepsilon^2 = 0$ is a special case. Note that

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

which means that the mean adjustment may not be dismissed, as it was in the ITS case.

First, consider the case in which $k = 0$. The above expression reduces to

$$\sum_{t=1}^{n} y_t^2 - \frac{1}{n} \left(\sum_{t=1}^{n} y_t \right)^2 . \tag{32}$$

We may expand the first term of (32) to obtain

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

which by Lemma A3 has the distribution of its first term. We may similarly expand the second term of (32), which also has the distribution of the first term of that expansion. The result for $k = 0$ obviously follows. The result for $k \geq 1$ follows directly from the appropriate parts of Lemma A3, using the same logic.

Proof of THEOREM 3.8 (*Asymptotics for* S_n^2 – *AHTS*). The proof follows directly from the asymptotics in Theorem 3.7 (when $k = 0$).

Proofs of THEOREM 3.9 and 3.10 (*Asymptotics for* Q_n^3 *and* K_n^4 – *AHTS*). The proofs are essentially the same as that for the sample variance, using the appropriate parts of Lemma A3.

Proof of THEOREM 4.1 and 4.2 (*Asymptotics for* $\hat{\theta}_n$ – *ITS*). The proofs follow that of Theorems 5.1 and 5.2 in Park and Phillips (2001), with rates of convergence following from the first and second parts of our Lemmas A2 and A3.

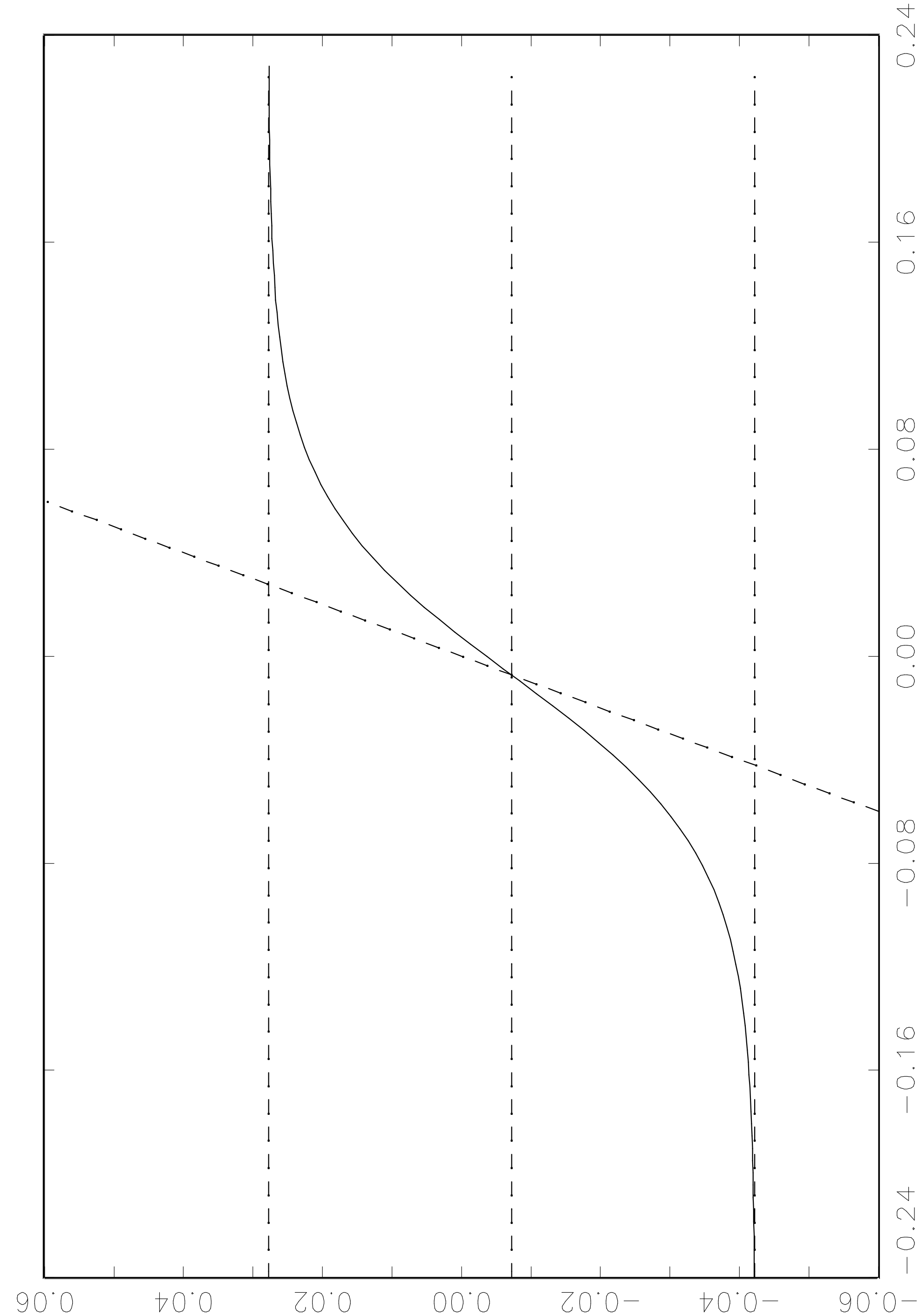

Figure 5.1.1: The target zone transformation $F(x)$ with our parameter estimates..

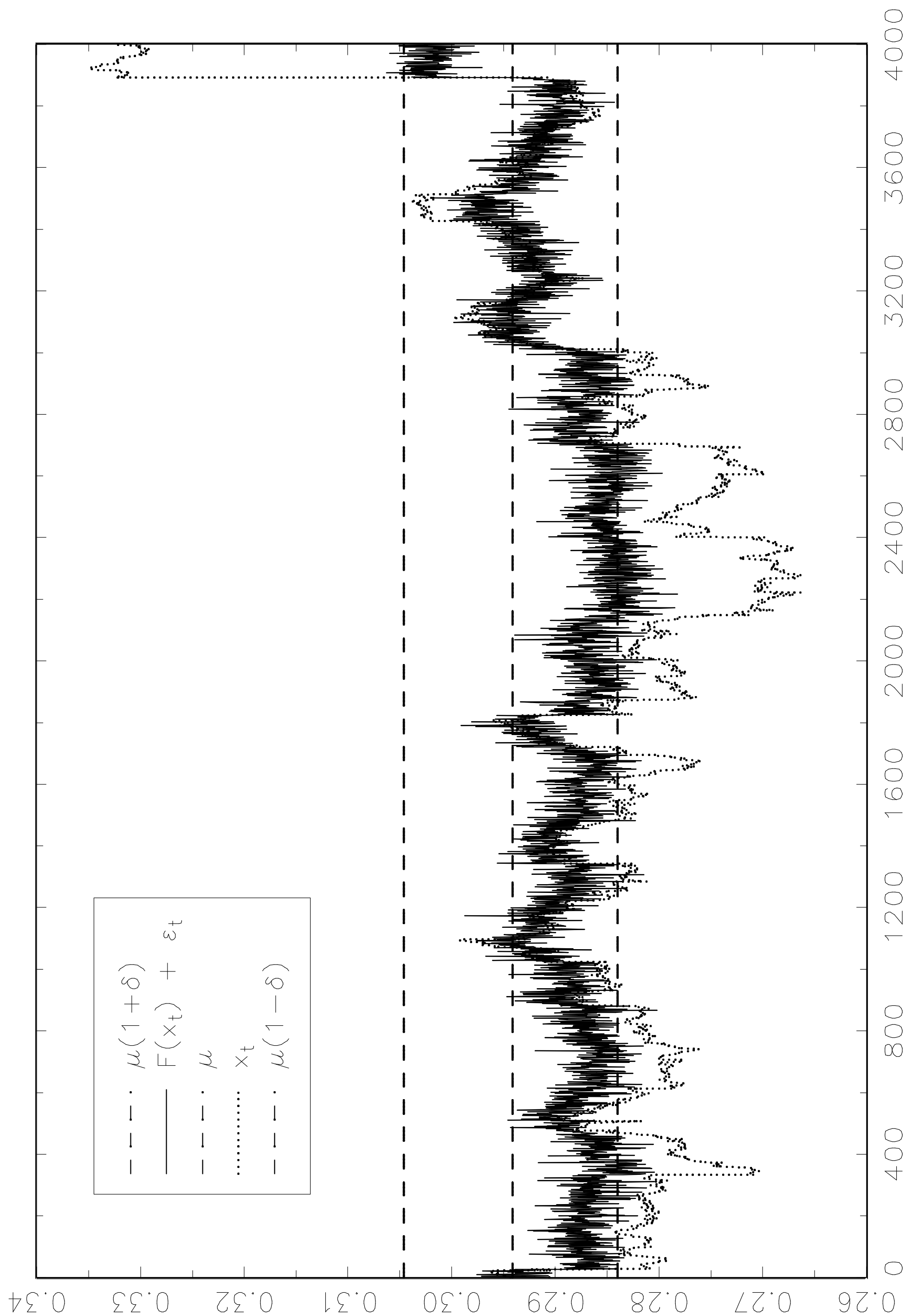

Figure 5.1.2: Sample simulated exchange rate and fundamental with our parameter estimates.

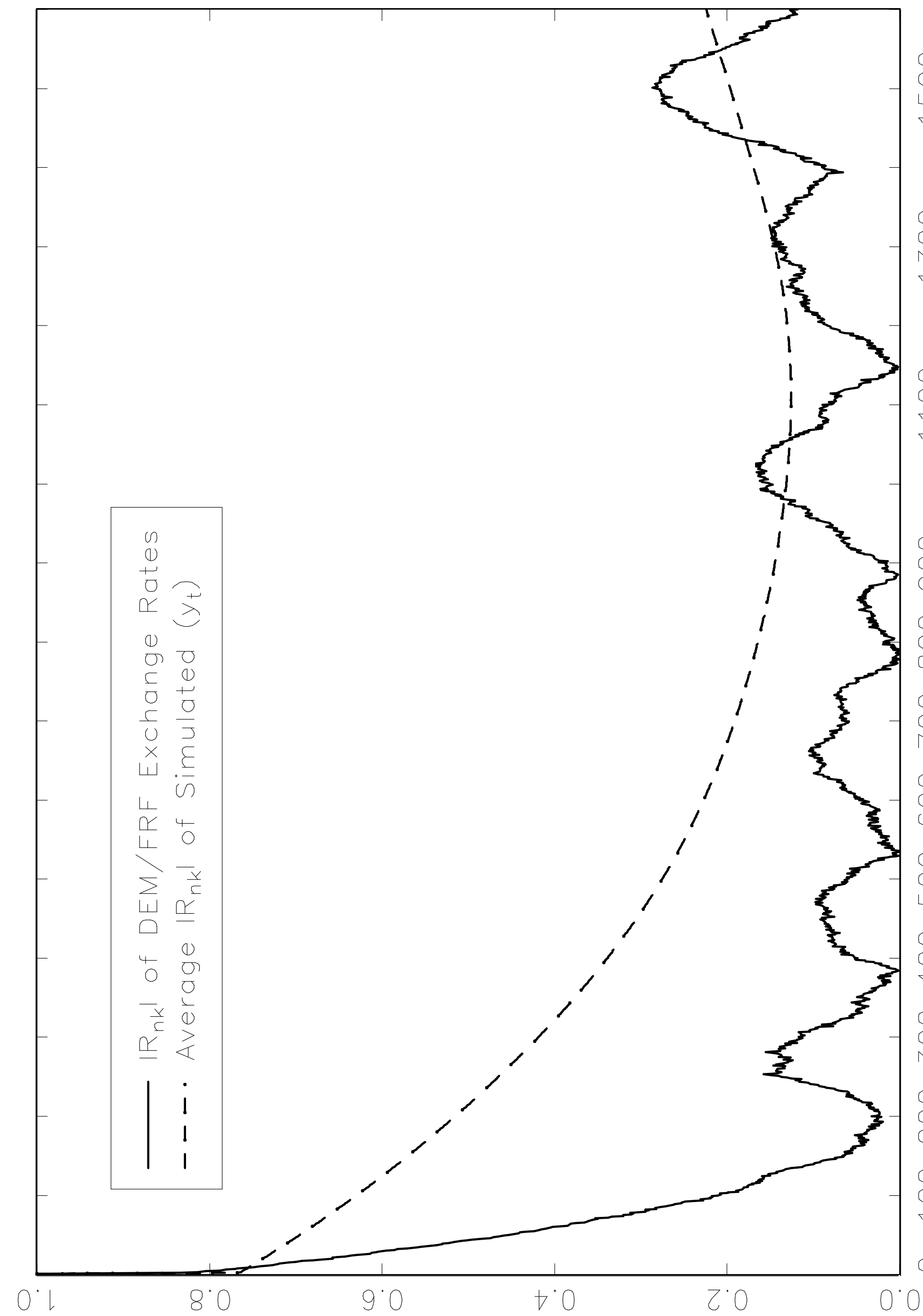

Figure 5.1.3: $|R_{nk}|$ of actual exchange rate and average $|R_{nk}|$ of simulated exchange rates.

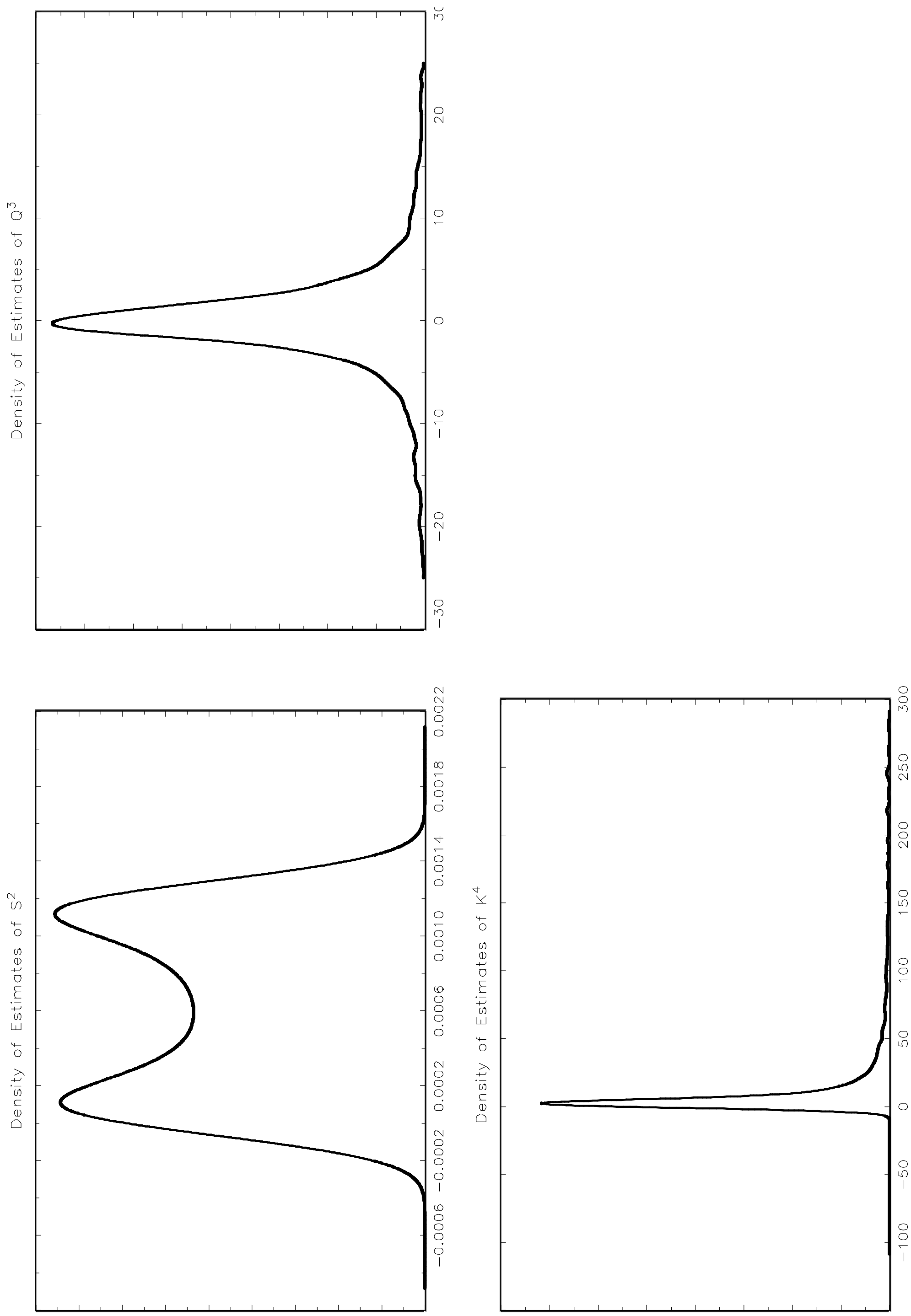

Figure 5.1.4: Density estimates of the asymptotic distributions of the sample variance, skewness, and kurtosis of (y_t), calculated from an AHTS model generated by the LHF of our TZM with our parameter estimates.

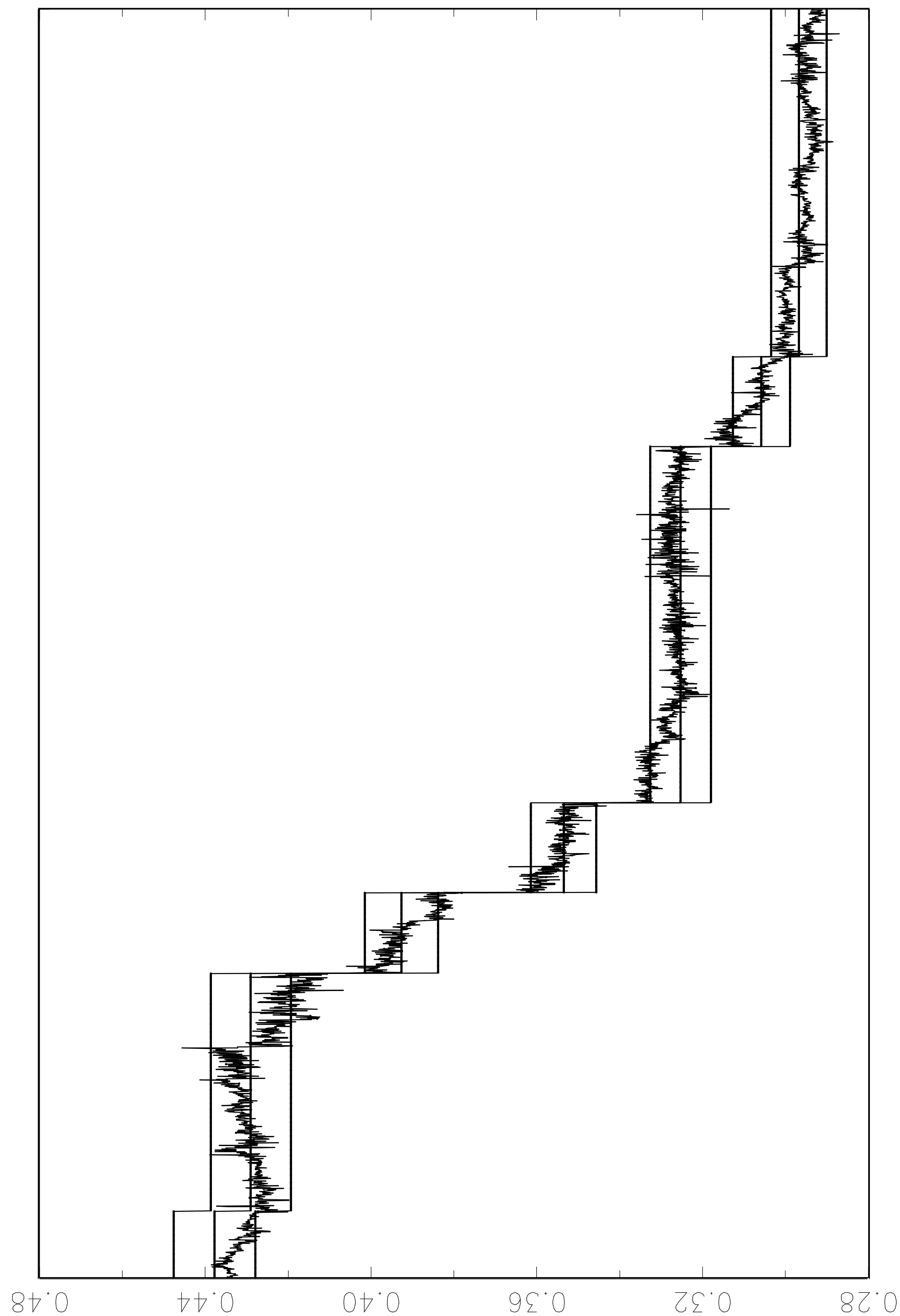

Figure 5.1.5: DEM/FRF exchange rate (3/1/79-12/31/89).

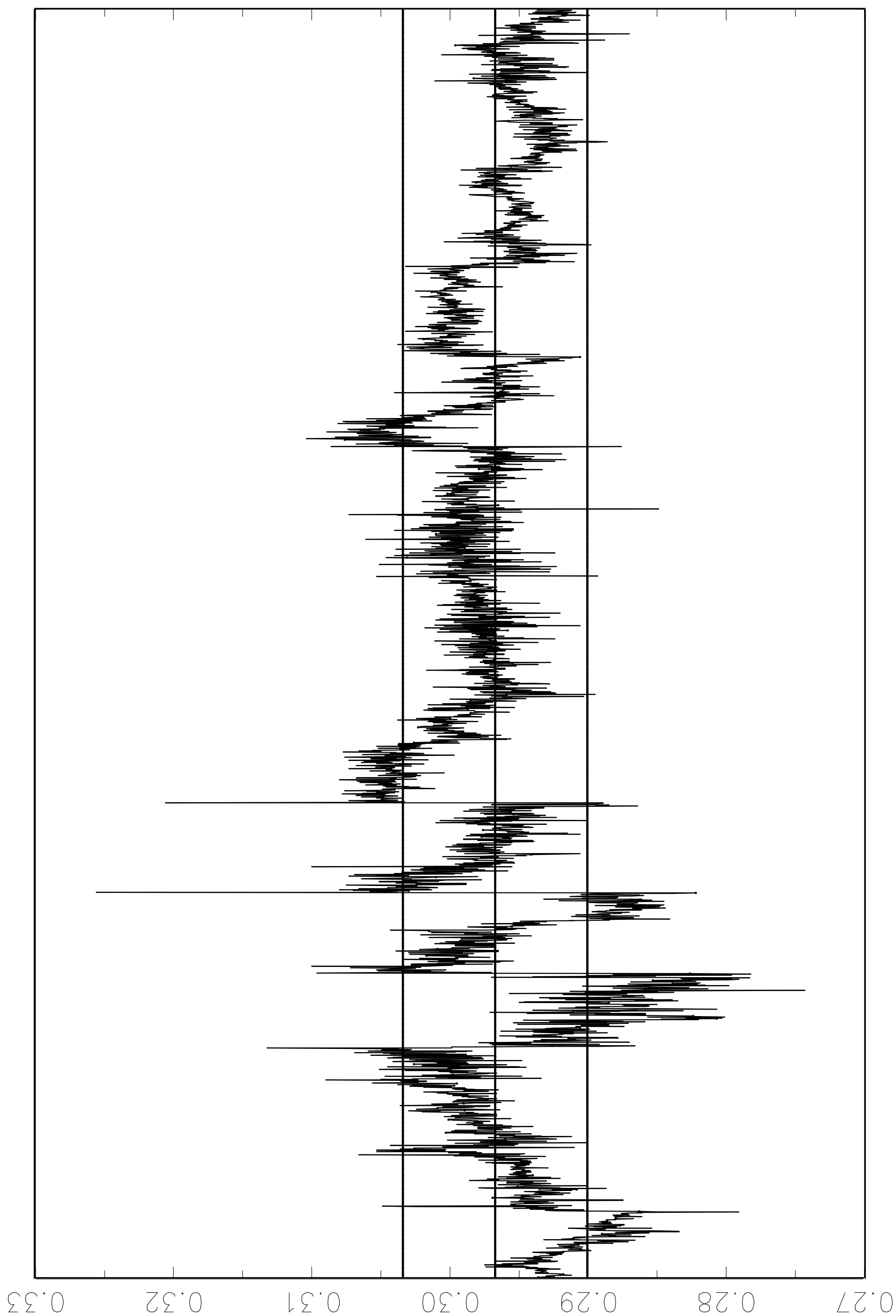

Figure 5.1.6: Leveled DEM/FRF exchange rate (vertically shifted to allow for actual realignments).

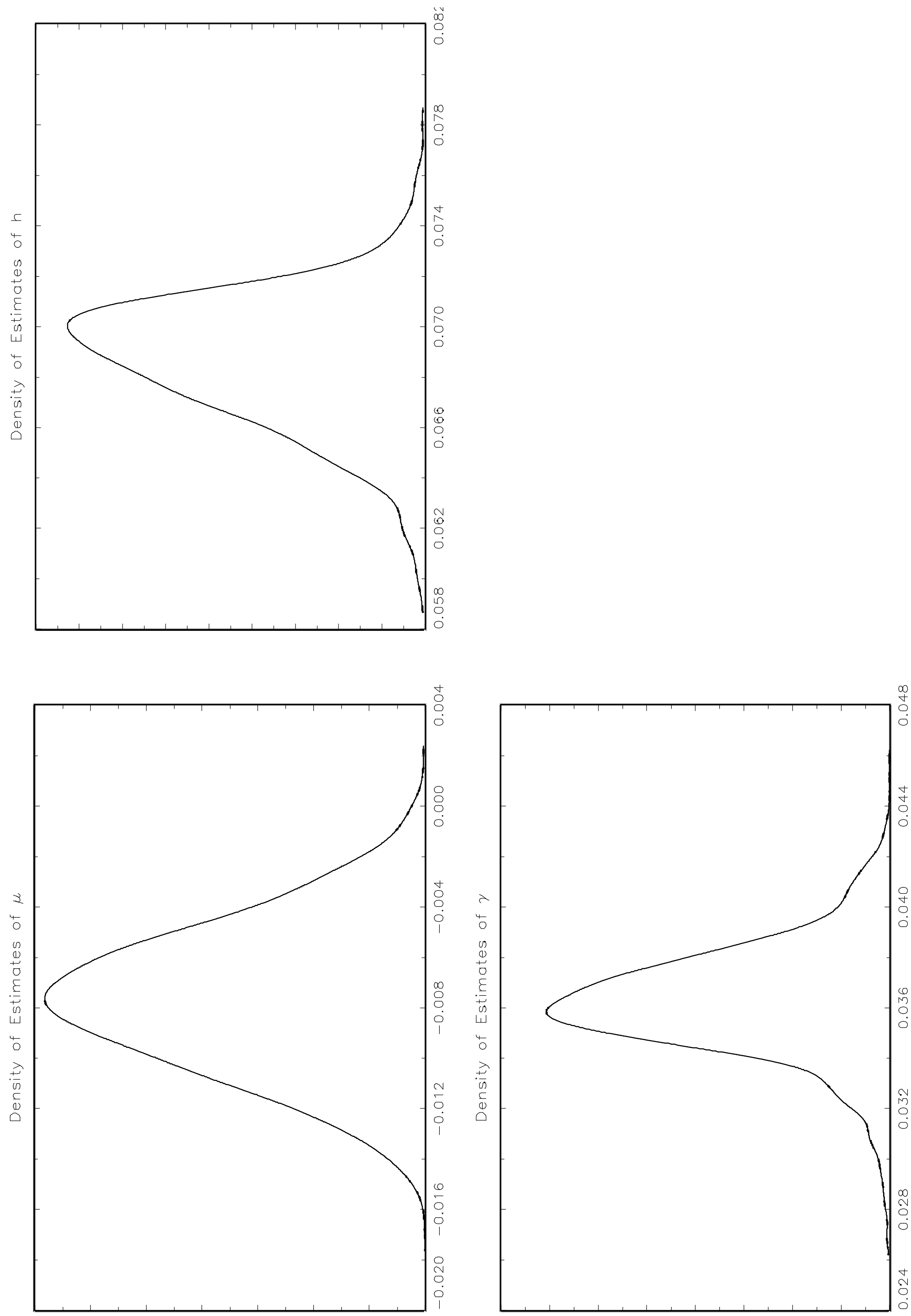

Figure 5.1.7: Densities of the model parameter estimates from the EKF used to construct the bootstrap confidence intervals.

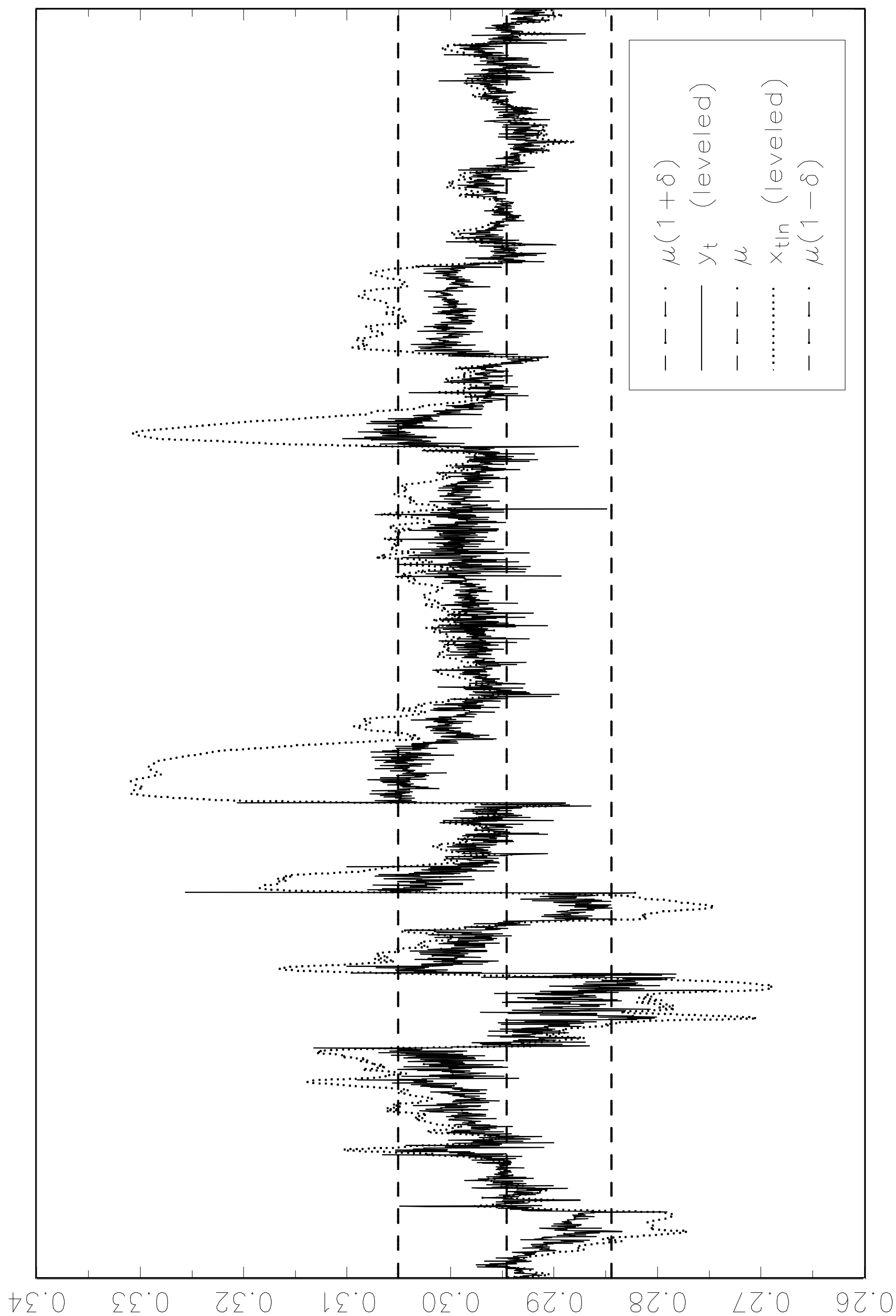

Figure 5.1.8: Leveled exchange rate and estimated fundamental (3/1/79-12/31/89).

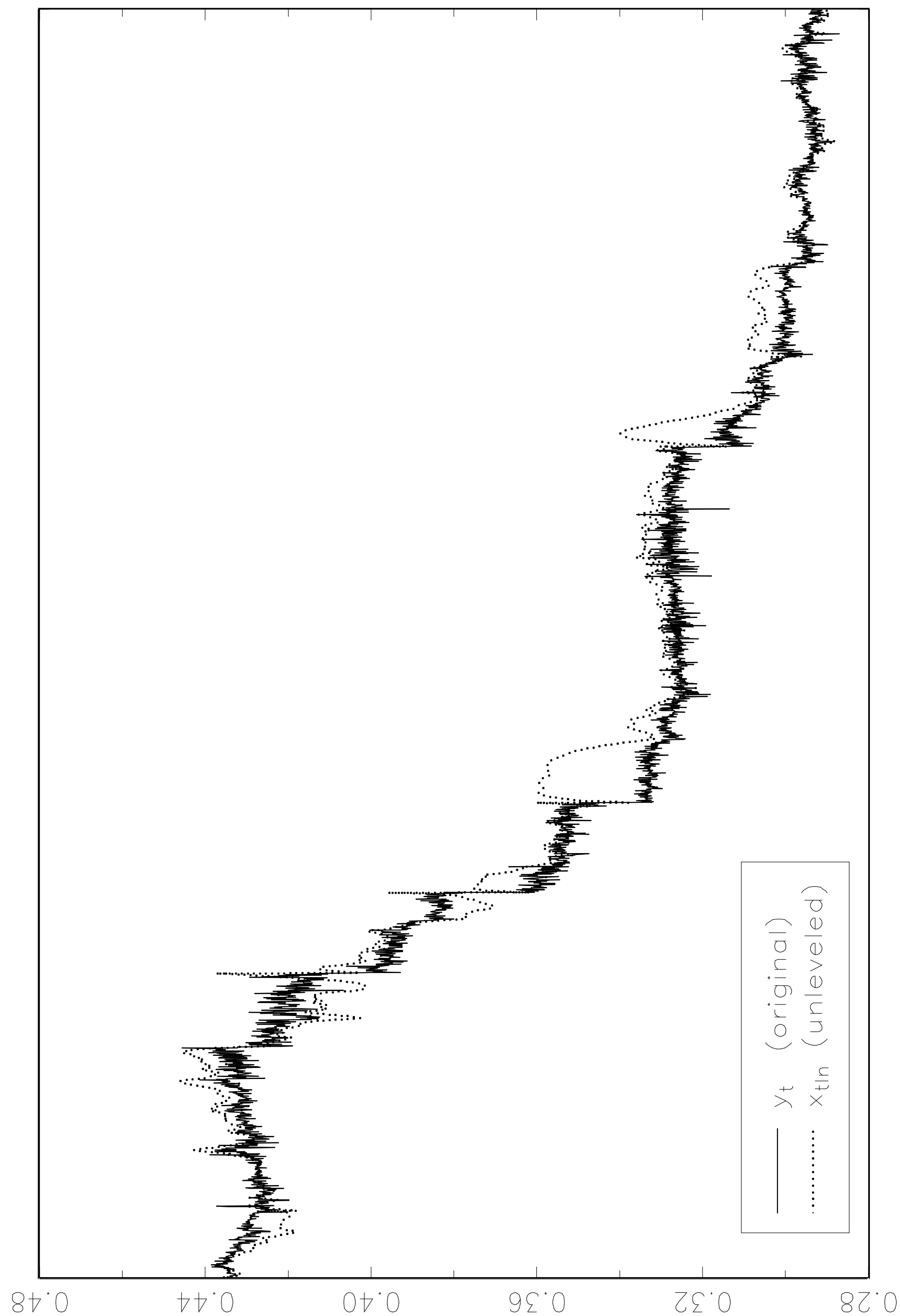

Figure 5.1.9: Original exchange rate and unleveled fundamental (3/1/79-12/31/89).

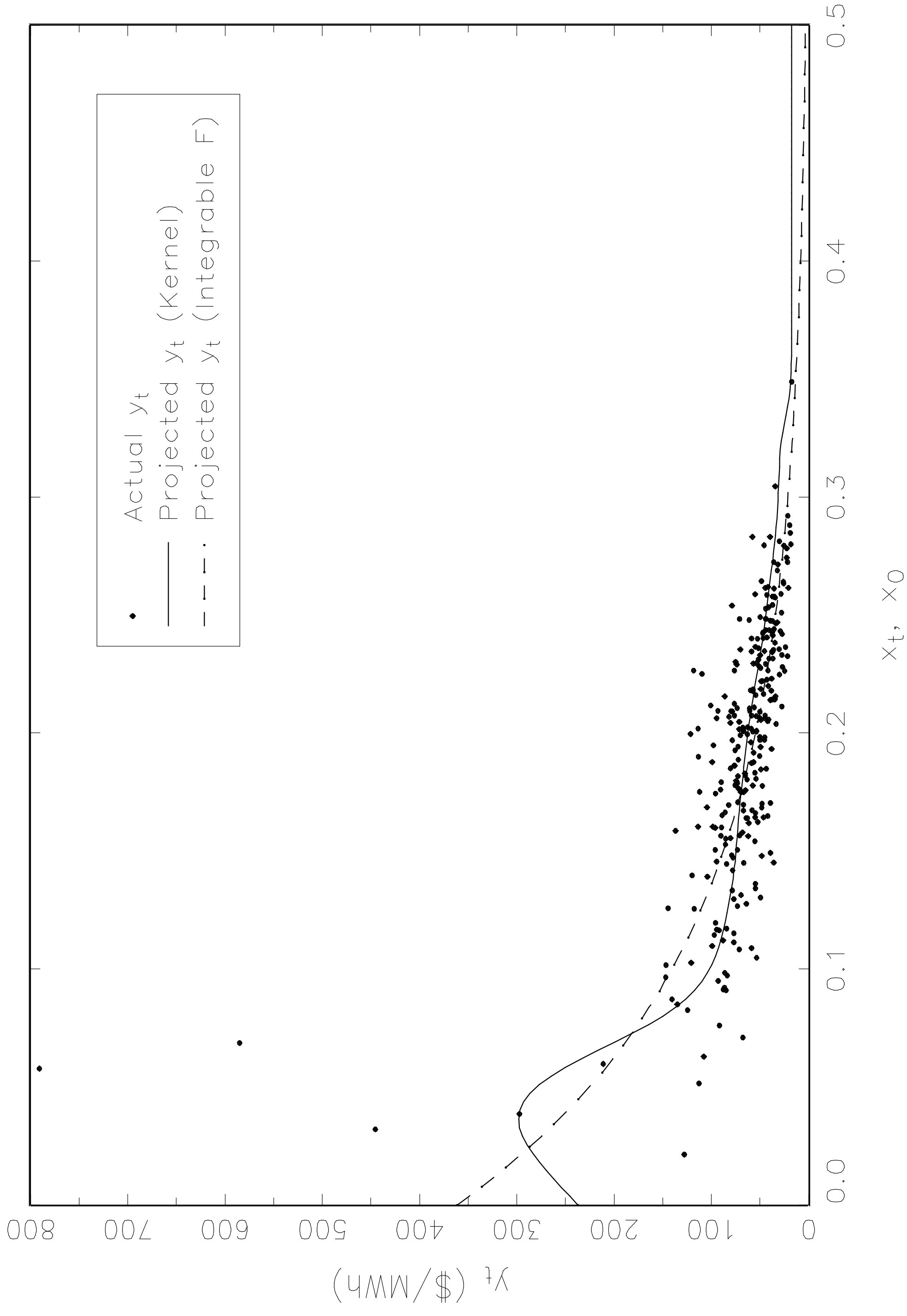

Figure 5.2.1: Electricity prices vs. excess capacity (4/1/02-12/31/02).

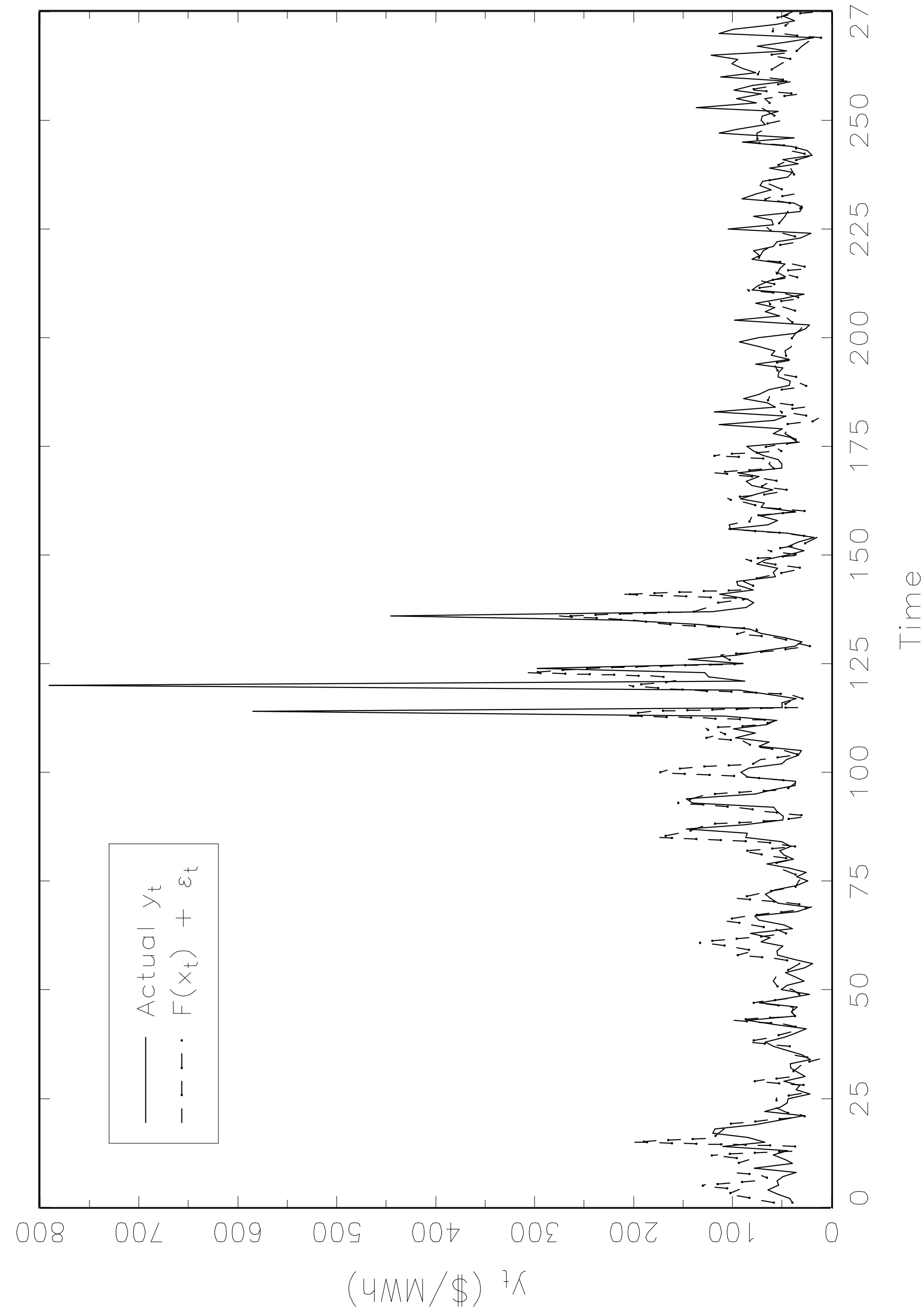

Figure 5.2.2: Actual and sample simulated electricity prices.

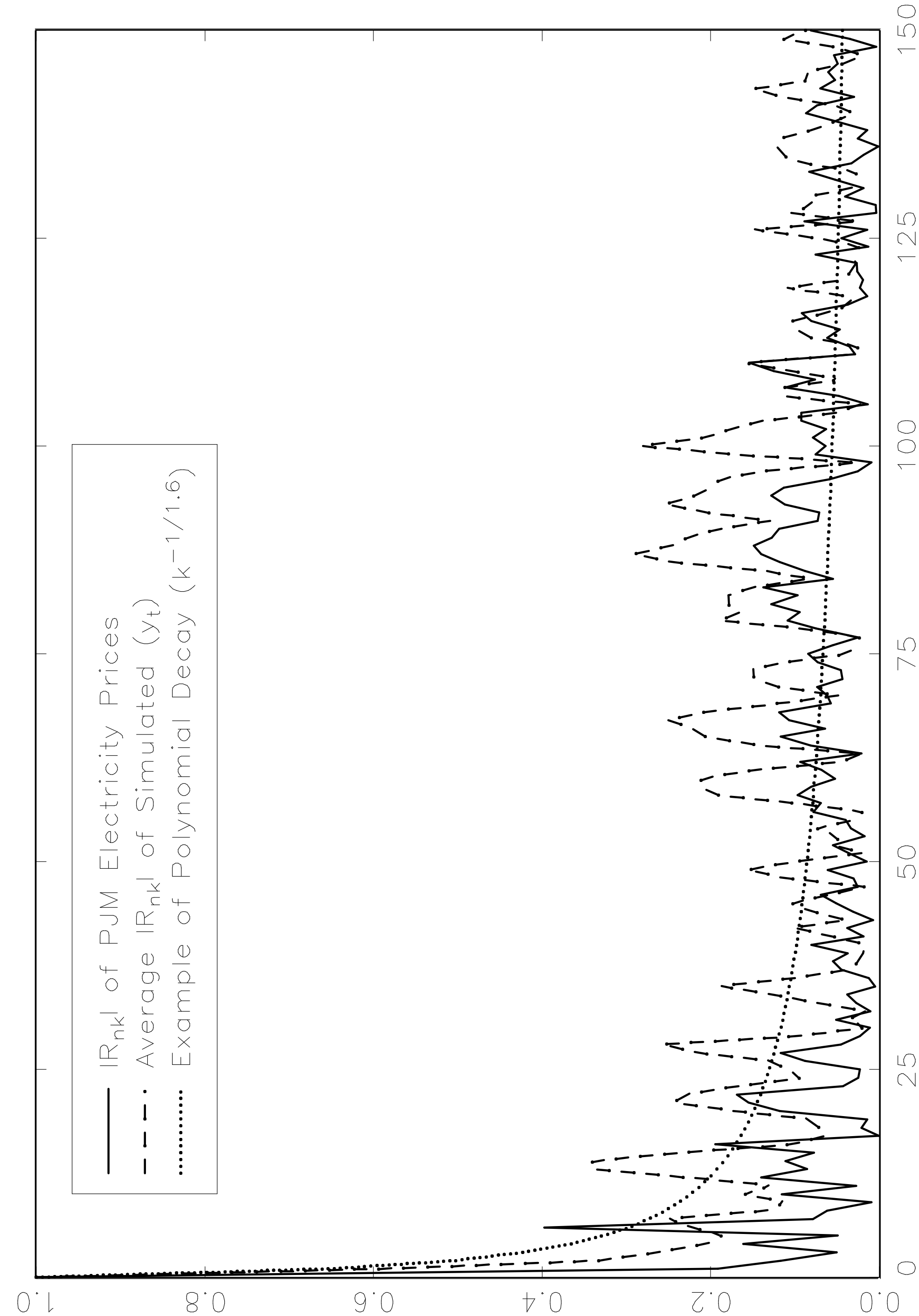

Figure 5.2.3: $|R_{nk}|$ of actual prices and average $|R_{nk}|$ of simulated prices.