
MEAN REVERSION IN STOCK INDEX FUTURES MARKETS: A NONLINEAR ANALYSIS

MICHAEL MONOYIOS
LUCIO SARNO*

Several stylized theoretical models of futures basis behavior under nonzero transactions costs predict nonlinear mean reversion of the futures basis towards its equilibrium value. Nonlinearly mean-reverting models are employed to characterize the basis of the S&P 500 and the FTSE 100 indices over the post-1987 crash period, capturing empirically these theoretical predictions and examining the view that the degree of mean reversion in the basis is a function of the size of the deviation from equilibrium. The estimated half lives of basis shocks, obtained using Monte Carlo integration methods, suggest that for smaller shocks to the basis level the basis displays substantial persistence, while for larger shocks the basis exhibits highly nonlinear mean reversion towards its equilibrium value. © 2002 Wiley Periodicals, Inc. *Jrl Fut Mark* 22:285–314, 2002

This paper was begun when Lucio Sarno was on the staff of the University of Oxford and was partly written while he was a Visiting Scholar at the Federal Reserve Bank of St. Louis. The authors are grateful to Abhay Abhyankar, Bernard Dumas, Mark Taylor, and Dick van Dijk for useful conversations or comments on previous drafts. The usual disclaimer applies, meaning that the authors alone are responsible for any errors that may remain and for the views expressed in the paper.

*Correspondence author, Finance Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom; e-mail: lucio.sarno@warwick.ac.uk

Received May 2001; Accepted July 2001

-
- Michael Monoyios is at Brunel University in London, United Kingdom.
 - Lucio Sarno is in the Finance Group at Warwick Business School at the University of Warwick in Coventry, United Kingdom.

1. INTRODUCTION

A large body of conventional finance theory assumes that financial markets are arbitrage-free at all times and also free of transactions costs on trading assets. Although these assumptions can be criticized as unrealistic, they have proved useful to construct benchmark models from which to develop dynamic asset pricing theories. A growing body of the finance literature is developing, however, which focuses on the implications of relaxing some of the conventional assumptions underlying asset pricing theories. In particular, it is obvious that trading a stock does incur costs and that, although transactions costs are expected to be relatively small in mature and liquid markets, they are likely to affect the stochastic process governing asset prices. Even in the presence of transactions costs, however, in real-world financial markets arbitrage opportunities do arise, generating much trading activity aimed at exploiting mispricing. In turn, this trading activity contributes to drive asset prices toward their theoretically fair or equilibrium levels.

In the context of stock index futures markets, a number of empirical studies have focused on the persistence of deviations of the futures basis from the equilibrium level implied by the cost of carry model or variants of it. The cost-of-carry model predicts that spot and futures prices co-move so that their long run equilibrium is essentially defined by the futures basis, which also implies mean reversion in the basis. Several studies record, however, the existence of significant nonlinearities in the dynamics of the futures basis, which may be rationalized on several grounds.¹ Indeed, as discussed below, it is intuitively clear that there are several factors (including, for example, the existence of transactions costs or agents heterogeneity) that generate no-arbitrage bounds and imply a law of motion for the futures basis that is consistent with nonlinear adjustment toward equilibrium. In particular, nonzero transactions costs on trading the underlying asset of the futures contract may lead to the basis displaying a particular form of nonlinear mean reversion such that the basis becomes increasingly mean reverting with the size of the deviation from its equilibrium value. Consequently, one might expect that allowing for nonlinear adjustment toward equilibrium in the empirical modeling of the futures basis may yield a more satisfactory approximation to the true unknown data generating process driving the basis, improving upon linear specifications.

¹See, e.g., Dwyer, Locke, and Yu (1996), Martens, Kofman, and Vorst (1988), and Yadav, Pope, and Paudyal (1994). See also Lo and MacKinlay (1999, Chap. 11).

This paper contributes to the literature on modeling the behavior of the futures basis on several fronts.² Specifically, the paper investigates nonlinearities in basis adjustment toward its equilibrium value and proposes a novel approach to modeling the behavior of the basis inspired by the prediction of the theoretical arguments mentioned above. Using data for the S&P 500 and the FTSE 100 indices during the post-crash period since 1988, the authors provide strong evidence of nonlinear mean reversion in the futures basis for both indices considered. The models indicate that the basis is closer to a unit root process the closer it is to its equilibrium value and becomes increasingly mean-reverting the further it is from its equilibrium value. Moreover, although small shocks to the basis are highly persistent, larger shocks mean-revert much faster, as predicted by the theoretical considerations discussed above and the widely held view that "arbitrage is like gravity": relatively larger deviations from fair values of asset prices (i.e., larger mispricing) induce relatively faster adjustment of asset prices toward their equilibrium values. These results have a natural interpretation, being consistent with standard economic and financial intuition as well as with the argument that there is a tendency of traders to wait for sufficiently large arbitrage opportunities to open up before entering the market and trading.³

The rest of the paper is set out as follows. Section 2 provides an overview of the theoretical arguments that motivate nonlinear mean-reverting behavior in the futures basis. Section 3 discusses the class of nonlinear models employed for modeling the futures basis. Section 4 describes the data set. Section 5 reports the results of summary statistics and univariate unit root tests applied to basis data, cointegration tests applied to a regression involving the spot price and the futures price, linearity tests applied to the basis data, and the estimation results from employing nonlinear models to characterize the basis of the S&P 500 and the FTSE 100 indices. In Section 6 Monte Carlo integration methods are used to calculate the half-lives implied by estimated nonlinear models for the basis, further examining how the nonlinear estimation results can improve the profession's understanding of the dynamics

²The literature related to the present study is very large and includes, among others, Blank (1991), Brennan and Schwartz (1990), Chan (1992), Dwyer et al. (1996), Figlewski (1984), Fujihara and Mougoue (1997a, 1997b), Gao and Wang (1999), Kawaller (1991), Kawaller, Koch, and Koch (1987), Klemkosky and Lee (1991), Lekkos and Milas (2001), MacKinlay and Ramaswamy (1988), Miller, Muthuswamy, and Whaley (1994), Modest and Sundaresan (1983), Parhizgari and de Boyrie (1997), Sarno and Valente (2000), Stoll and Whaley (1990), Yang and Brorsen (1993, 1994), Yadav et al. (1994).

³See, e.g., Dumas (1994) and Sofianos (1993) on this point.

characterizing the major futures markets under investigation. A final section briefly summarizes and concludes.

2. MOTIVATING NONLINEAR MEAN REVERSION IN THE FUTURES BASIS

This section briefly discusses how the effects of transactions costs and other factors characterizing stock index futures markets can induce nonlinear mean reversion of the futures basis toward its equilibrium value such that the speed of adjustment of the basis toward equilibrium is a function of the size of the disequilibrium.

Consider a market containing an asset, a stock index, whose price $S(t)$ under the equivalent martingale measure evolves according to:

$$dS(t) = S(t)(\bar{r} - q)dt + \sigma_S dW_S(t) \quad (1)$$

where $\bar{r}$ is the (constant) risk-free interest rate; q is the (constant) dividend yield on the index; σ_S is the volatility of the index; $W_S(t)$ is a one-dimensional standard Brownian motion in a complete probability space.

Standard derivatives pricing theory gives the futures price $F(t, T)$ at time t for delivery of the stock at time $T \geq t$ as:

$$F(t, T) = E[S(T) | \mathcal{F}(t)] \quad (2)$$

where E denotes the mathematical expectation with respect to the martingale measure $\mathbf{P}$, and $\mathcal{F}(t)$ denotes the information set at time t (e.g., see Karatzas & Shreve, 1998). Given (1)–(2), the futures price has the well-known formula:

$$F(t, T) = S(t) \exp(r_c(T - t)) \quad (3)$$

where $r_c = \bar{r} - q$. This is the familiar expression for a futures price in a nonrandom interest rate environment.

Defining the logarithmic basis $b(t, T)$ at time t as

$$b(t, T) = \log(F(t, T)/S(t)) \quad (4)$$

then, using (3) and (4) yields

$$b(t, T) = r_c(T - t) \quad (5)$$

A number of studies examining stock index futures prices have catalogued mean reverting deviations of the basis from its equilibrium level defined as in (4) or (5). Also, it can be easily illustrated how incorporating

stochastic factors in (1) (for example by making the cost of carry r_c stochastic) can lead to closed-form expressions for the futures price, which implies mean-reversion in the futures basis (Schwartz, 1997).⁴

This simple model can be extended to determine bounds on stock futures prices under the most commonly investigated market friction: transactions costs are charged on trading the stock. If a share of stock is bought for a price S then the buyer's cash account is debited an amount $S(1 + \nu)$, where $\nu > 0$ is the proportional transactions cost rate for buying the index. Similarly, a stock sale credits the seller's bank account with an amount $S(1 - \mu)$, where $\mu > 0$ is the proportional transactions cost rate for selling stock.

Under these assumptions, the futures price at time t for delivery at T , $F(t, T)$ must lie within the following bounds to prevent arbitrage:⁵

$$S(t)(1 - \mu) \exp(r_c(T - t)) \leq F(t, T) \leq S(t)(1 + \nu) \exp(r_c(T - t)) \quad (6)$$

Given the definition of the logarithmic basis, equation (6) can be rewritten defining the no-arbitrage bounds on the futures price in terms of the basis as follows:

$$\log(1 - \mu) \leq b(t, T) \leq \log(1 + \nu) \quad (7)$$

The above analysis shows that when market frictions such as proportional transactions costs are taken into account, the futures price can fluctuate within a bounded interval without giving rise to any arbitrage profits. In other words, proportional transactions costs create a band for the basis within which the marginal cost of arbitrage exceeds the marginal benefit. Assuming instantaneous arbitrage at the edges of the band then implies that the bounds become reflecting barriers. If the upper bound is violated, for example, arbitrageurs would sell short the futures contract and buy the index, which would drive the basis back within the

⁴It can be shown, for example, that starting from a two-factor model for the stock index price and the interest rate (modelled as an Ornstein–Uhlenbeck process) it is possible to derive a closed-form solution for the futures basis that displays mean reversion (full calculations available from the authors upon request). An alternative, more sophisticated way to rationalize mean reversion in the cost of carry model may be by showing that the volume of arbitrage activity (e.g., sell overpriced futures and buy cheap stocks) is determined within the theoretical structure to be greater when the deviation from equilibrium is greater. This would be a model where the quantity of activity drives the equilibration of price.

⁵To establish the relationship in (6), consider the zero-cost strategy of going long the futures contract (which requires no initial outlay) at time t , then hedging this transaction by selling the index and investing the proceeds in a cash account. At time T the funds in this account amount to $S(t)(1 - \mu) \exp(r_c(T - t))$, and this cannot be more than is needed to buy the index for $F(t, T)$ under the terms of the futures contract, otherwise arbitrage profits would result. This establishes the lower bound in (6), and the upper bound is established by a symmetrical argument.

no-arbitrage bounds (and a similar upward adjustment would occur if the lower bound is violated). Following arguments of this sort, several studies motivate the adoption of threshold-type models of the type originally proposed by Tong (1990) to empirically characterize the behavior of the basis. These threshold models allow for a transactions costs band within which no adjustment takes place—so that deviations from the basis may exhibit unit root behavior—while outside the band the process switches abruptly to become stationary autoregressive. Studies using threshold models provide evidence against linearity in the deviation of the basis from its equilibrium level and in favor of threshold-type behavior (e.g., Yadav et al., 1994; Dwyer et al., 1996; Martens et al., 1998).

Nevertheless, while threshold-type models are appealing in this context, various arguments can be made that rationalize multiple-threshold or smooth, rather than single-threshold or discrete, nonlinear adjustment of the basis toward its equilibrium value. First, some influential studies of arbitrage in financial as well as real markets show that the thresholds should be interpreted more broadly than as simply reflecting proportional transactions costs *per se*, but also as resulting from the tendency of traders to wait for sufficiently large arbitrage opportunities to open up before entering the market and trading (see, e.g., Dumas, 1992, 1994; Neal, 1996; Sofianos, 1993).⁶

Second, if one takes into account fixed as well as proportional costs of arbitrage, this results in a two-threshold model where the basis is reset by arbitrage to an upper or lower inner threshold whenever it hits the corresponding outer threshold. Intuitively, arbitrage will be heavy once it is profitable enough to outweigh the initial fixed cost, but will stop short of returning the basis to the equilibrium value because of the proportional arbitrage costs (see the discussion in Dumas, 1994, in the context of international capital markets).

Third, one may argue that the assumption of instantaneous trade should be replaced with the presumption that it takes some time to observe an arbitrage opportunity and execute transactions and that trade is infrequent. For example, a number of studies have analyzed the effects of arbitrage in futures markets along the lines of models of the type developed by Garbade and Silber (1983), where a continuum of traders induces movements in spot and futures prices such that the basis returns very rapidly to a constant equilibrium level (see Chan, 1992; Kawaller et al., 1987;

⁶Sofianos (1993) and Neal (1996) also find that most arbitrage trades are liquidated before expiration. This finding is consistent with the model of Brennan and Schwartz (1990), which predicts that arbitrage trades are liquidated when the reversal in the mispricing is sufficiently large and trades can be profitable.

Stoll & Whaley, 1990). However, the assumption of a continuum of traders acting in futures markets has often been debated; notably, Miller et al. (1994) argue that much of the mean reversion in basis movements may be explained by infrequent trading in the cash market.

Fourth, in a market with heterogeneous agents who face different levels of transactions costs (or different margin requirements or position limits), agents essentially face no-arbitrage bands of different size. For relatively small deviations of the basis from its equilibrium level, only some traders will be able to effect arbitrage trades. If the bounds are violated by a relatively greater amount, then progressively more agents will enter the market to effect arbitrage trades.^{7,8} Thus, the forces pushing the basis back within the bounds will increase as the deviation from the bounds increases since an increasing number of agents face profitable arbitrage opportunities, implying the possibility of a smooth transition of the basis back towards the bounds such that the speed of mean reversion increases with the degree of violation of the arbitrage bounds (see, e.g., Dumas, 1994).

Overall, the arguments discussed above suggest that the basis should become increasingly mean reverting with the size of the deviation from the equilibrium level. Transactions costs create a band of no arbitrage for the basis, but the basis can stray beyond the thresholds. Once beyond the upper or lower threshold, the basis becomes increasingly mean reverting with the distance from the threshold. Within the transactions costs band, when no trade takes place, the process driving the basis is divergent since arbitrage is not profitable. Under certain restrictive conditions (including, among others identical transactions costs, identical margin requirements and position limits, and homogeneity of

⁷In practice, μ is determined by the transactions costs from the reverse cash-and-carry arbitrage trades of selling the spot good at the bid price, lending the proceeds, and buying a futures contract, whereas ν is determined by the transaction costs from the standard cash-and-carry arbitrage trades of borrowing to buy the spot good at the asked price and selling a futures contract. Moreover, both the lower and upper bounds are made slightly higher by the existence of other carry costs from the reverse (standard) cash-and-carry arbitrage, which should incorporate the arbitrageur's lending (borrowing) rate and, in the case of the lower bound, should also be adjusted to reflect the possibility that the short seller will not earn the full amount of interest on the proceeds from the short sale. In general, depending on their borrowing and lending rates and on the levels of transactions costs they face, different individuals will have the ability to arbitrage at different futures prices.

⁸One may be tempted to argue that, once an arbitrage opportunity arises, each arbitrageur will invest as much as possible to exploit the arbitrage opportunity. However, this is obviously not the case in real-world futures markets since arbitrage may be risky for a number of reasons, including the existence of margin requirements and position limits. For example, Liu and Longstaff (2000) demonstrate that, as an effect of the existence of margin requirements, it is not optimal to take unlimited positions in arbitrage and it is often optimal to take smaller positions in arbitrage than margin constraints would allow.

agents) the jump to mean-reverting behavior may be discrete, but in general it is smooth, and Dumas (1994), Teräsvirta (1994), and Granger and Lee (1999) suggest that even in the former case, time aggregation will tend to smooth the transition between regimes. Hence, smooth rather than discrete adjustment may be more appropriate in the presence of proportional transactions costs, and time aggregation and nonsynchronous adjustment by heterogeneous agents is likely to result in smooth aggregate regime switching.

3. NONLINEAR MEAN REVERSION IN THE BASIS: THE EMPIRICAL FRAMEWORK

The time series of interest in the present study is the logarithm of the futures basis, b_t , defined as in (4). Theoretical frameworks inspired by the cost of carry model imply that a long-run relationship must exist between the spot price and the futures price such that the basis is reverting to a stable equilibrium level. In other words, while short-run deviations of the basis from its equilibrium level are allowed for, the basis must be a mean reverting process. Over the last two decades or so, a large body of research focusing on testing the cost-of-carry model or on modeling the basis has developed, initially largely stimulated by the early influential studies of, among others, Modest and Sundaresan (1983) and Figlewski (1984). In particular, a number of empirical studies have focused on the persistence of deviations from the cost of carry using linear econometric methods. Linear methods are valid, however, only under the maintained hypothesis of a linear autoregressive process for the basis, which means that adjustment of the basis toward the equilibrium value is both continuous and of constant speed, regardless of the size of the deviation from the equilibrium value.

As discussed in the previous section, however, the presence of transactions costs, possibly in addition to several other factors, are likely to generate complex nonlinear dynamics in the futures basis, which has important implications for conventional empirical modeling procedures of the basis. Some empirical evidence on the importance of transactions costs is provided by several studies—cited in the introduction—investigating the nonlinear nature of the adjustment process of the basis using threshold models. Threshold models allow for a transactions costs band within which no adjustment takes place (so that deviations from the equilibrium value of the basis may exhibit unit root behavior) while outside the band the process switches abruptly to become stationary autoregressive. Although discrete switching of this kind represents a significant step

ahead relative to the previous literature using stationary models and may be appropriate when considering the effects of arbitrage on individual stocks, discrete adjustment of the basis of a stock index would clearly be most appropriate only when stocks have identical features of, for example, transactions costs, interest rates and liquidity and when agents are homogeneous. Moreover, given the discussion in the previous section, smooth rather than discrete adjustment may be more appropriate in the presence of proportional transactions costs and, as suggested by Dumas (1994), Teräsvirta (1994), and Granger and Lee (1999), time aggregation and, most importantly, nonsynchronous adjustment by heterogeneous agents is likely to result in smooth aggregate regime switching.

A characterization of nonlinear adjustment that allows for smooth rather than discrete adjustment is in terms of a smooth transition autoregressive (STAR) model (Granger & Teräsvirta, 1993; Teräsvirta, 1994). In the STAR model, adjustment takes place in every period but the speed of adjustment varies with the extent of the deviation from equilibrium. A STAR model may be written:

$$b_t = \alpha + \sum_{j=1}^p \beta_j b_{t-j} + \left[\alpha^* + \sum_{j=1}^p \beta_j^* b_{t-j} \right] \Phi[\theta; b_{t-d} - \kappa] + \varepsilon_t \quad (8)$$

where $\{b_t\}$ is a stationary and ergodic process, $\varepsilon_t \sim iid(0, \sigma^2)$ and $(\theta; \kappa) \in \{\mathfrak{N}^+ \times \mathfrak{N}\}$, where $\mathfrak{N}$ denotes the real line $(-\infty, \infty)$ and $\mathfrak{N}^+$ the positive real line $(0, \infty)$. The transition function $\Phi[\theta; b_{t-d} - \kappa]$ determines the degree of mean reversion and is itself governed by the parameter θ , which effectively determines the speed of mean reversion, and the parameter κ which may be seen as the equilibrium level of $\{b_t\}$. A simple transition function suggested by Granger and Teräsvirta (1993) and Teräsvirta (1994), which is particularly attractive in the present context, is the exponential function:

$$\Phi[\theta; b_{t-d} - \kappa] = 1 - \exp[-\theta^2 [b_{t-d} - \kappa]^2] \quad (9)$$

in which case (8) would be termed an exponential STAR or ESTAR model. The exponential transition function is bounded between zero and unity, $\Phi : \mathfrak{N} \rightarrow [0, 1]$, has the properties $\Phi[0] = 0$ and $\lim_{x \rightarrow \pm \infty} \Phi[x] = 1$, and is symmetrically inverse-bell shaped around zero. These properties of the ESTAR model are attractive in the present modeling context because they allow a smooth transition between regimes and symmetric adjustment of the basis for deviations above and below the equilibrium

level.⁹ The transition parameter θ determines the speed of transition between the two extreme regimes, with lower absolute values of θ implying slower transition. The inner regime corresponds to $b_{t-d} = \kappa$, when $\Phi = 0$ and (8) becomes a linear AR(p) model:

$$b_t = \alpha + \sum_{j=1}^p \beta_j b_{t-j} + \varepsilon_t \quad (10)$$

The outer regime corresponds, for a given θ , to $\lim_{(b_{t-d}-\kappa) \rightarrow \pm\infty} \Phi \times [\theta; b_{t-d} - \kappa]$, where (8) becomes a different AR(p) model:

$$b_t = \alpha + \alpha^* + \sum_{j=1}^p (\beta_j + \beta_j^*) b_{t-j} + \varepsilon_t \quad (11)$$

with a correspondingly different speed of mean reversion so long as $\beta_j^* \neq 0$ for at least one value of j .

Granger and Teräsvirta (1993) and Teräsvirta (1994) also suggest the logistic function as a plausible transition function for some applications, resulting in a logistic STAR or LSTAR model. Since, however, the LSTAR model implies asymmetric behavior of b according to whether it is above or below the equilibrium level, that model is considered, a priori, as inappropriate for modeling basis movements. That is to say, it is not straightforward to think of reasons why the speed of adjustment of the basis should vary according to whether the futures price is above or below its fair price.¹⁰ The authors do, however, test for nonlinearities arising from the LSTAR formulation as a test of specification of the estimated models in the section discussing the empirical analysis.

It is also instructive to reparameterize the STAR model (8) as follows:

$$\begin{aligned} \Delta b_t = & \alpha + \rho b_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta b_{t-j} \\ & + \left\{ \alpha^* + \rho^* b_{t-1} + \sum_{j=1}^{p-1} \phi_j^* \Delta b_{t-j} \right\} \Phi[\theta; b_{t-d} - \kappa] + \varepsilon_t \end{aligned} \quad (12)$$

⁹Clearly, the class of nonlinear models is infinite, and this paper focuses on the ESTAR formulation primarily because of these attractive properties, its relative simplicity, and the fact that it seems to be the logical empirical counterpart of the theoretical considerations discussed in Section 2.

¹⁰One notable example in the literature of a study proposing asymmetries is due to Brennan and Schwartz (1990), who suggest that if the transactions costs of arbitrage are asymmetric and arbitrage affects the level of the basis then such asymmetry is likely to be reflected in the distribution of the basis. Nevertheless, in general, there is fairly convincing evidence that the distribution of the basis is symmetric—notably, see the evidence provided by Dwyer et al. (1996) using both parametric and nonparametric tests of symmetry applied to data for the S&P 500 index.

where $\Delta b_{t-j} \equiv b_{t-j} - b_{t-j-1}$. In this form, the crucial parameters are ρ and ρ^* . The discussion of the effect of transactions costs in the previous section suggests that the larger the deviation from the equilibrium value of the basis the stronger will be the tendency to move back to equilibrium. This implies that while $\rho \geq 0$ is admissible, one must have $\rho^* < 0$ and $(\rho + \rho^*) < 0$. That is, for small deviations b_t may be characterized by unit root (or even explosive) behavior, but for large deviations the process is mean reverting.¹¹

In empirical applications, Granger and Teräsvirta (1993) and Teräsvirta (1994) suggest choosing the order of the autoregression, p , through inspection of the partial autocorrelation function, PACF; the PACF is to be preferred to the use of an information criterion since it is well known that the latter may bias the chosen order of the autoregression towards low values, and any remaining serial correlation may affect the power of subsequent linearity tests. Granger and Teräsvirta (1993) and Teräsvirta (1994) then suggest applying a sequence of linearity tests to artificial regressions which can be interpreted as second or third-order Taylor series expansions of (8) (see also Luukkonen et al., 1988). This allows detection of general nonlinearity through the significance of the higher-order terms in the Taylor expansions, with the value of d selected as that yielding the largest value of the test statistic. The tests can also be used to discriminate between ESTAR and LSTAR formulations, since third-order terms disappear in the Taylor series expansion of the ESTAR transition function. This method of selecting the order of d and choosing whether an ESTAR or LSTAR formulation is appropriate is termed the Teräsvirta Rule below. In the Monte Carlo study of Teräsvirta (1994), the Teräsvirta Rule worked well in selecting d and also in discriminating between ESTAR and LSTAR unless, understandably, the two models are close substitutes—that is, when most of the observations lie above the equilibrium level κ so that only one half of the inverse-bell shape of the ESTAR transition function is relevant and is well approximated by a logistic curve.

Hence, using results provided by Teräsvirta (1994), prior to modeling the basis using smooth transition models, the tests for linearity are constructed as follows. The following the artificial regression is

¹¹This analysis has implications for conventional unit root tests based on the maintained hypothesis of a linear autoregressive model where the degree of mean reversion is measured by the size of ρ , implicitly assuming $\rho^* = 0$ —e.g., see the discussion in Michael, Nobay, and Peel (1997) in the context of testing for nonstationarity of real exchange rates.

estimated:

$$b_t = \psi_{00} + \sum_{j=1}^p \{ \phi_{0j} z_{t-j} + \phi_{1j} z_{t-j} z_{t-d} + \phi_{2j} z_{t-j} z_{t-d}^2 + \phi_{3j} z_{t-j} z_{t-d}^3 \} \\ + \phi_4 z_{t-d}^2 + \phi_5 z_{t-d}^3 + \text{innovations} \quad (13)$$

where ϕ_4 and ϕ_5 become zero if $d \leq p$. Keeping the delay parameter d fixed, testing the null hypothesis $H_0: \phi_{1j} = \phi_{2j} = \phi_{3j} = \phi_4 = \phi_5 = 0 \forall j \in \{1, 2, \dots, p\}$ against its complement is a general test (LM^G) of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression (i.e., $\phi_{3j} = \phi_5 = 0$ if the true model is an ESTAR, but $\phi_{3j} \neq \phi_5 \neq 0$ if the true model is an LSTAR), however, testing the null hypothesis that $H_0: \phi_{3j} = \phi_5 = 0 \forall j \in \{1, 2, \dots, p\}$ provides a test (LM^3) of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if the restrictions $\phi_{3j} = \phi_5 = 0$ cannot be rejected at the chosen significance level, then a more powerful test (LM^E) for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis $H_0: \phi_{1j} = \phi_{2j} = \phi_4 = 0 \mid \phi_{3j} = \phi_5 = 0 \forall j \in \{1, 2, \dots, p\}$.

In the empirical analysis, therefore, the authors specify a set of values for d in the range $\{1, 2, \dots, D\}$ and employ each of the tests LM^G , LM^3 and LM^E . Subject to LM^G being significant and LM^3 being insignificant, $d = d^*$ is selected such that $LM^E(d^*) = \sup_{d \in \mathfrak{D}} LM^E(d)$ for $\mathfrak{D} = \{1, 2, \dots, D\}$ (for further details, see Teräsvirta, 1994).

4. DATA

The data set comprises daily time series on futures written on the S&P 500 index and the FTSE 100 index, as well as price levels of the corresponding underlying cash indices, over the sample period from January 1, 1988 to December 31, 1998. The data were obtained from Datastream. Given the focus of the present paper on investigating the importance of allowing for nonlinearity and aggregate regime switching in modeling the basis, the authors deliberately choose to use data after the 1987 crash in order to reduce the risk that the nonlinearity detected and modeled in the empirical analysis could be determined by or attributed to a unique and perhaps exceptional event occurred over the sample.

A number of related studies motivated by microstructure considerations or focusing on modeling intraday or short-lived arbitrage have used

intraday data at different intervals.¹² Given that the basic goal of this study is to shed light on the mean-reverting properties of the futures basis and to measure the role of the persistent, low-frequency properties of the basis data, the span of the time series—in terms of years—is much more important than the number of observations per se (e.g., see Shiller & Perron, 1985). Therefore, to reduce the noise element in the futures basis data the authors choose to employ data at daily frequency. The authors did, however, estimate nonlinear models also using intraday data at 5-min intervals. The estimation results were qualitatively identical, suggesting that aggregation and systematic sampling may not have a particularly strong effect on the stochastic mean-reverting behavior of the futures basis.¹³

The data set under examination covers an eleven-year period, which may be sufficiently long to capture some of the main features of the unknown stochastic process driving the basis, while also providing a sufficiently large number of observations $T = 2,870$ to be fairly confident of the estimation results. Also, given that in the UK the futures market ceases trading at 16.10 and the underlying index closes at 16.30, FTSE 100 index levels at 16.10 are used in order to avoid the problems caused by nonsynchronous market closure. Similarly, given that for the S&P 500 the futures market ceases trading at 16.15 EST and the underlying index closes at 16.00 EST, S&P 500 futures index levels at 16.00 EST are used. Obviously, the futures contract is paired up with the spot price comparing the spot price to the contract nearest to maturity. (All times are EST.)

Using these data, the authors construct—for each of the two indices analyzed—the time series of interest in this paper, namely the logarithm of the basis, $b_t \equiv f_t - s_t$, where f_t and s_t denote the logarithm of the futures price and the logarithm of the spot price respectively.

5. EMPIRICAL ANALYSIS

5.1 Preliminary Statistics and Cointegration Analysis

Panel (a) of Table I provides some summary statistics for f_t , s_t , and b_t . For both the S&P 500 and the FTSE 100, the first moment of the futures price is slightly larger than the first moment of the spot price (although it is not the case that $f_t > s_t$ at each point in time), while the second moments of

¹²For example, Chan (1992), Dwyer et al. (1996), and Miller et al. (1994) have used 5-, 15-, and 5-min intervals, respectively.

¹³For an application of nonlinear models to higher frequency data in this context, see Taylor, van Dijk, Franses, and Lucas (2000).

TABLE I
Preliminary Data Analysis

	S&P 500			FTSE 100		
	f_t	s_t	b_t	f_t	s_t	b_t
Panel (a): Summary Statistics						
Minimum	5.4837	5.4915	-0.0195	7.4277	7.4351	-0.0259
Maximum	7.1346	7.1243	0.0221	8.7358	8.7289	0.0316
Mean	6.1802	6.1752	0.0000	8.0113	8.0033	0.0000
Variance	0.1734	0.1730	1.77E-5	0.1096	0.1109	6.24E-5
PACF:						
lag 1	0.9984	0.9984	0.7717	0.9983	0.9984	0.8959
lag 2	0.0285	0.0119	0.3284	0.0111	-0.0074	0.2304
lag 3	0.0012	0.0003	0.1635	0.0149	0.0103	0.1114
lag 4	0.0106	0.0129	0.0626	0.0177	0.0107	0.0740
lag 5	0.0075	0.0045	0.0645	-0.0040	-0.0050	0.0477
lag 6	-0.0390	-0.0323	0.0244	-0.0044	-0.0041	-0.0005
lag 7	0.0139	0.0108	0.0113	-0.0101	-0.0057	0.0097
lag 8	0.0188	0.0165	0.0029	0.0242	0.0130	0.0061
lag 9	0.0008	0.0002	0.0065	-0.0131	-0.0064	0.0021
lag 10	0.0061	0.0037	0.0067	0.0060	0.0116	0.0007
lag 15	-0.0179	-0.0173	-0.0099	-0.0048	-0.0054	0.0020
lag 20	0.0047	0.0091	-0.0097	0.0124	0.0107	-0.0074

Panel (b): Unit Root Tests

Futures Price				
	$f_t^{(c)}$	$f_t^{(c,\tau)}$	$\Delta f_t^{(c)}$	$\Delta^2 f_t^{(c)}$
S&P 500	-0.0697	-1.6101	-40.2570	-66.6875
FTSE 100	-0.2877	-2.9778	-39.1603	-64.2835
Spot Price				
	$s_t^{(c)}$	$s_t^{(c,\tau)}$	$\Delta s_t^{(c)}$	$\Delta^2 s_t^{(c)}$
S&P 500	-0.0832	-1.4489	-23.5721	-63.7417
FTSE 100	-0.2356	-2.7981	-49.3863	-62.8683
Basis				
	b_t	$b_t^{(c)}$	$\Delta b_t^{(c)}$	$\Delta^2 b_t^{(c)}$
S&P 500	-21.2437	-8.9122	-55.4507	-80.8012
FTSE 100	-14.8101	-7.2860	-20.5121	-73.0704

Notes. f_t , s_t , and b_t denote the log-level of the futures price, the log-level of the spot price, and the demeaned log-level of the basis, respectively. In Panel (a), PACF is the partial autocorrelation function, and its standard deviation can be approximated by the square root of the reciprocal of the number of observations, $T = 2,870$, hence being about 0.0187. In Panel (b), statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process; the (c) (c , τ) superscript indicates that a constant (a constant and a linear trend) was (were) included in the augmented Dickey-Fuller regression, while absence of a superscript indicates that neither a constant nor a trend were included; Δ is the first-difference operator. The critical value at the five percent level of significance is -1.95 to two decimal places if neither a constant nor a time trend is included in the regression, -2.86 if a constant only is included, and -3.41 if both a constant and a linear trend are included (Fuller, 1976; MacKinnon, 1991).

the spot price and the futures price are virtually the same. The summary statistics for b_t (demeaned prior to the empirical analysis, hence generating a mean exactly equal to zero) indicate much lower variability relative to futures and spot prices. Nevertheless, it is interesting to note that, while the variance of the S&P 500 futures and spot prices is (about 60%) higher than the variance of the FTSE 100 futures and spot prices, the variance of the S&P 500 basis is (about 70%) lower than the variance of the FTSE 100 basis. The partial autocorrelation functions, reported in Panel (a) of Table I, suggest that both futures and spot prices display very strong first-order serial correlation but they do not appear to be significantly serially correlated at higher lags; the basis appears to be less strongly serially correlated but displays a slower decay of the partial autocorrelation function at higher lags.

As a preliminary exercise, the authors test for unit root behavior of each of the futures price and spot price time series by calculating standard augmented Dickey–Fuller (ADF) test statistics, reported in Panel (b) of Table I. In each case, the number of lags is chosen such that no residual autocorrelation was evident in the auxiliary regressions.¹⁴ In keeping with the very large number of studies of unit root behavior for these time series and conventional finance theory, the authors are in each case unable to reject the unit root null hypothesis applied to each of the futures price and the spot price for both indices at conventional nominal levels of significance. On the other hand, differencing the price series appears to induce stationarity in each case, clearly indicating that both f_t and s_t are realizations from stochastic processes integrated of order one. Nevertheless, the results suggest strongly a rejection the unit root null hypothesis applied to b_t in levels as well as in differences, suggesting stationarity of the basis and possibly the existence of a cointegrating relationship between the futures price and the spot price for each of the S&P 500 and the FTSE 100.¹⁵

To complete the analysis of the long-run properties of the data, the authors test for cointegration between f_t and s_t , employing the well-known

¹⁴Moreover, using non-augmented Dickey–Fuller tests or augmented Dickey–Fuller tests with any number of lags in the range from 1 to 20 yielded results qualitatively identical to those reported in Panel (b) of Table I, also regardless of whether a constant or a deterministic trend was included in the regression. Also, note that a deterministic trend was found to be statistically significantly different from zero at conventional nominal levels of significance in the auxiliary regressions for both the futures price and the spot price (not for the basis), consistent with a large empirical literature in this context.

¹⁵In addition to the ADF tests, the authors also execute unit root tests of the type proposed by Phillips and Perron (1988) as well as Johansen likelihood ratio tests (Johansen, 1988, 1991) in a vector autoregression with one series. The results were perfectly consistent with the ADF tests results, indicating that both f_t and s_t are $I(1)$, whereas b_t is $I(0)$ (results not reported but available from the authors upon request).

Johansen (1988, 1991) maximum likelihood procedure in a vector autoregression comprising f_t and s_t , and allowing for a lag length of 5 and an unrestricted constant term.¹⁶ Both Johansen likelihood ratio (LR) test statistics (based on the maximum eigenvalue and on the trace of the stochastic matrix respectively) clearly suggest that a cointegrating relationship exists for both indices under investigation. Also, the restriction suggested by conventional finance theory in the spirit of the cost of carry model that the cointegrating parameter equals unity could not be rejected at conventional nominal levels of significance for both the S&P 500 and the FTSE 100. In fact, estimation of the vector autoregression with the cointegrating parameter left unrestricted produced estimated values of the cointegrating parameter equal to 1.0001 for the S&P 500 and equal to 1.0003 for the FTSE 100. Then, estimation of the VAR with the imposition of the restriction that the cointegrating parameter equals unity produced the results given in Table II, suggesting that a unique cointegrating relationship exists between f_t and s_t for both the S&P 500 (Panel (a)) and the FTSE 100 (Panel (b)).^{17,18}

5.2 Linearity Tests

As a preliminary to employing nonlinear stochastic models to characterize the basis, the authors carry out both a general test for linearity of the residuals from an autoregressive process for the basis as well as the linearity tests discussed in Section 3 do discriminate between a linear model, an ESTAR model and an LSTAR model.

The first linearity test employed is a RESET (Ramsey, 1969) test of the null hypothesis of linearity of the residuals from an AR(5) for b_t

¹⁶The authors were very careful in selecting the number of lags in the vector autoregression, being aware of the sensitivity of vector autoregression analysis to the lag length in this context. Both the Akaike information criterion (AIC) and the Schwartz information criterion (SIC) suggested a lag length of 5 for both systems. Nevertheless, in the present application, the results were found to be very robust to the choice of the lag length. Also, note that the VAR considered is essentially model $H_1^*(r)$ in Johansen (1995, p. 83) notation, where a linear deterministic trend is implicitly allowed for but this can be eliminated by the cointegrating relations and the process contains no trend-stationary components; hence the model allows for a linear trend in each variable but not in the cointegrating relations.

¹⁷LR tests of the hypothesis that the cointegrating parameter equals unity could not be rejected with p -values equal to .503 and .582 for the S&P 500 and the FTSE 100 respectively. Also, note that testing for cointegration between f_t and s_t under the unity restriction on the cointegrating parameter is tantamount to testing for nonstationarity of the basis.

¹⁸The finding of cointegration between the spot and futures price is expected in the light of the cost of carry model but it is not a trivial result. Several authors have not been able to detect cointegration between the spot and the futures price or mean reversion in the basis. Notably, Miller et al. (1994) find that the S&P 500 index basis appears nonstationary (using intraday data), implying that the spot and the futures price may not cointegrate.

TABLE II
Johansen Maximum Likelihood Cointegration Procedure

H_0	H_1	LR	5% Critical Value
<i>Panel (a): S&P 500</i>			
<i>LR Tests Based on the Maximum Eigenvalue of the Stochastic Matrix</i>			
$g = 0$	$g = 1$	380.12	14.06
$g \leq 1$	$g = 2$	0.30	3.84
<i>LR Tests Based on the Trace of the Stochastic Matrix</i>			
$g = 0$	$g \geq 1$	380.04	15.41
$g \leq 1$	$g = 2$	0.30	3.84
<i>Panel (b): FTSE 100</i>			
<i>LR Tests Based on the Maximum Eigenvalue of the Stochastic Matrix</i>			
$g = 0$	$g = 1$	59.85	14.06
$g \leq 1$	$g = 2$	58E-3	3.84
<i>LR Tests Based on the Trace of the Stochastic Matrix</i>			
$g = 0$	$g \geq 1$	59.86	15.41
$g \leq 1$	$g = 2$	58E-3	3.84

Notes. The vector autoregression under consideration comprises the futures price and the spot price, as described in the text. H_0 and H_1 denote the null hypothesis and the alternative hypothesis, respectively; g denotes the number of cointegrating vectors; the 5% critical values reported in the last column are taken from Osterwald-Lenum (1992).

against the alternative hypothesis of general model misspecification. Under the RESET test statistic, the alternative model involves a higher-order polynomial to represent a different functional form; under the null hypothesis, the statistic is distributed as $\chi^2(q)$ with q equal to the number of higher-order terms in the alternative model. Table III reports the results from executing RESET test statistics where in the alternative model a quadratic and a cubic term are included; the null hypothesis is very strongly rejected for the basis of both indices considered with p -values of virtually zero, clearly suggesting that a linear autoregressive process for the basis is misspecified.

Upon inspection of the partial autocorrelation functions of the basis, for both the S&P 500 and the FTSE 100, the authors consider a lag length of 5 for executing the linearity tests discussed in Section 3. Table IV reports values of the test statistics LM^G , LM^3 , and LM^E . The authors consider $d \in \{1, 2, \dots, 10\}$ as plausible values for the delay parameter, although it seems plausible to expect a rather fast reaction of agents to deviations of the basis from the equilibrium value and, hence, a relatively low value of d . From Table IV it can be seen that linearity is

TABLE III
RESET Tests on the Basis

S&P 500	FTSE 100
532.3466 {0.0000}	614.6272 {0.0000}

Notes. RESET test statistics are computed considering a linear ADF regression for the basis with four lags with no constant and no time trend against an alternative model with a quadratic and a cubic term. In constructing the tests the F -statistic form is used since it is well known that in finite samples the actual test size of the F approximation may be closer to the nominal significance level than the actual size of the χ^2 approximation. Figures in braces denote marginal significance levels (p -values).

TABLE IV
Linearity Tests on the Basis

	LM^G	LM^3	LM^E
Panel (a): S&P 500			
$d = 1$	271.616 {0.0}	6.5301 {0.2580}	90.7058 {0.0}
$d = 2$	88.6367 {0.0}	5.4172 {0.3671}	82.2928 {0.0}
$d = 3$	62.2970 {1.0E-7}	4.1220 {0.5320}	44.1118 {3.14E-6}
$d = 4$	103.864 {0.0}	8.9173 {0.1124}	56.0570 {0.2.0E-8}
$d = 5$	112.989 {0.0}	7.1421 {0.2103}	78.9965 {0.0}
$d = 6$	105.218 {0.0}	5.1217 {0.4012}	76.4296 {0.0}
$d = 7$	79.3125 {0.0}	2.9515 {0.7075}	64.2786 {0.0}
$d = 8$	60.8880 {1.8E-7}	3.1005 {0.6845}	49.0440 {4.0E-7}
$d = 9$	66.1436 {2.0E-8}	4.7150 {0.4516}	42.1225 {7.13E-6}
$d = 10$	73.6207 {0.0}	6.1023 {0.2964}	45.8093 {1.55E-6}
Panel (b): FTSE 100			
$d = 1$	135.742 {0.0}	3.9694 {0.5538}	66.3531 {0.0}
$d = 2$	34.4171 {0.0030}	4.5238 {0.4767}	27.2210 {0.0024}
$d = 3$	80.9632 {0.0}	7.6144 {0.1788}	28.5829 {0.0014}
$d = 4$	24.3906 {0.0587}	4.4830 {0.4822}	18.6640 {0.0447}
$d = 5$	32.2180 {0.0060}	3.0715 {0.6890}	20.9364 {0.0215}
$d = 6$	43.2861 {0.0001}	6.3820 {0.2708}	29.6207 {0.0010}
$d = 7$	29.2760 {0.0148}	8.4462 {0.1333}	19.8501 {0.0307}
$d = 8$	18.3425 {0.2451}	4.6442 {0.4608}	15.5808 {0.1123}
$d = 9$	22.2351 {0.1018}	4.8443 {0.4352}	17.3302 {0.0674}
$d = 10$	35.1728 {0.0023}	6.6501 {0.1556}	23.4487 {0.0092}

Notes. The statistics LM^G , LM^3 , and LM^E are Lagrange multiplier test statistics for linearity constructed as discussed in the text for different delays $d \in \{1, 2, \dots, 10\}$. The order of the autoregression, p equals five in each case. In constructing the tests the F -statistic form is used since it is well known that in finite samples the actual test size of the F approximation may be closer to the nominal significance level than the actual size of the χ^2 approximation. Figures in braces denote marginal significance levels (p -values); p -values equal to zero to the eight decimal place are considered as virtually zero and reported as {0.0}.

easily rejected at the 5% significance level for the S&P 500 for each value of the delay parameter considered and for the FTSE 100 for all values of the delay parameter except 5, 8, and 9. Also, the rejections are particularly strong (p -values are very low) for $d = 1$. LM^3 is always insignificantly different from zero at conventional significance levels for any value of

the delay parameter considered and for both indices under investigation, implying that greater power may be obtained using LM^E . LM^E is, in fact, statistically significant for each value of d for the S&P 500 index and for 8 out of 10 values of d for the FTSE 100 index. Following the Teräsvirta Rule—i.e., minimize the significance level of LM^E —an ESTAR model with delay parameter equal to unity is selected for both indices.¹⁹

5.3 Nonlinear Estimation Results

The results reported and discussed in the previous section led to the choice of an ESTAR model for each of the two bases examined, with the lag length set equal to 5 and the delay parameter set equal to unity. Hence, for each of the bases, ESTAR models are estimated in first-difference form as in equation (12) with $p = 5$ and $d = 1$.

Experimentation with various starting values for the parameters yielded identical results, indicating the location of a global optimum. For each of the estimated ESTAR models, the authors could not reject, at the 5% significance level, the hypothesis of no remaining nonlinearity for values of d ranging from 2 to 10, on the basis of Lagrange multiplier tests (Table V reports only the maximum value of the LM statistic testing for no remaining ESTAR nonlinearity, $NLES_{MAX}$). Neither could the authors reject at the 5% level the hypothesis of no remaining nonlinearity of the LSTAR variety with values of the delay parameter in the range from 1 to 10 ($NLLS_{MAX}$ in Table V). This procedure therefore provides support for setting $d = 1$ and for using a symmetric nonlinear transition function.

With $p = 4$, the delay parameter, d , is also estimated directly together with the other model parameters, by nonlinear least squares involving a grid search over values of d from 1 to 10. A value of $d = 1$ was again implied for each of the bases. It is significant that $d = 1$ is the least squares estimate because, as noted by Hansen (1997), since the parameter space for d is discrete, its least squares estimate is super-consistent and can be treated as known for the purposes of further inference.

Hence, the Teräsvirta Rule appears to be very robust in the present application and an ESTAR model with $p = 5$ and $d = 1$ is the preferred model for both series. The resulting ESTAR models, estimated by nonlinear least squares,²⁰ are reported in Table V.

¹⁹As a check of model adequacy against nonlinearity with other delays, however, the authors report below a Lagrange multiplier test for no remaining nonlinearity in the estimated ESTAR models, as suggested by Eitrheim and Teräsvirta (1996).

²⁰Regularity conditions for the consistency and asymptotic normality of the nonlinear least squares estimator are discussed in this context by Tjøstheim (1986).

TABLE V
Nonlinear Estimation Results

	S&P 500	FTSE 100
$\hat{\phi}_1 (= -\hat{\phi}_1^*)$	-0.456 (0.032)	-0.408 (0.029)
$\hat{\phi}_2 (= -\hat{\phi}_2^*)$	-0.243 (0.033)	-0.229 (0.027)
$\hat{\phi}_3 (= -\hat{\phi}_3^*)$	-0.152 (0.032)	-0.139 (0.026)
$\hat{\phi}_4 (= -\hat{\phi}_4^*)$	-0.107 (0.027)	-0.079 (0.024)
$\hat{\rho}^*$	0.303 (0.016)	0.120 (0.015)
$\hat{\theta}^2$	0.205 (0.031) {2E-8}	0.272 (0.067) {8E-5}
$\bar{R}^2$	0.645	0.820
s	2.51E-3	3.35E-3
Diagnostic tests:		
AR(10)	{0.674}	{0.349}
ARCH(10)	{0.750}	{0.752}
NLES _{MAX}	{0.456}	{0.551}
NLLS _{MAX}	{0.393}	{0.642}
JB	{0.436}	{0.368}
LR	{0.430}	{0.542}
SKALIN	{6E-8}	{1E-4}

Notes. The model estimated is of the form (12) with the restrictions discussed in Section 5.3. $\bar{R}^2$ denotes the adjusted coefficient of determination, and s is the standard error of the regression. AR(10) and ARCH(10) are Lagrange multiplier test statistics for up to tenth-order serial correlation in the residuals constructed as in Eitrheim and Teräsvirta (1996) and test statistics for up to tenth-order autoregressive conditional heteroskedasticity constructed as in Engle (1982), respectively. NLES_{MAX} is the maximal Lagrange multiplier test statistic for no remaining ESTAR nonlinearity with delays in the range from 2 to 10; NLLS_{MAX} is the maximal Lagrange multiplier test statistic for no remaining LSTAR nonlinearity with delay in the range from 1 to 10 (Eitrheim & Teräsvirta, 1996). JB is a Jarque-Bera test statistic for normality of the residuals (Jarque & Bera, 1987). LR is a likelihood ratio statistic for the parameter restrictions on the unrestricted ESTAR model. SKALIN is a parametric bootstrap likelihood ratio test of linearity constructed as suggested by Skalin (1998). Figures in parentheses below coefficient estimates denote estimated standard errors. Figures in braces denote marginal significance levels (p -values). For test statistics which are distributed as central χ^2 (the LR and JB statistics) or F (the AR, ARCH, NLES_{MAX}, and NLLS_{MAX} statistics) under the null hypothesis, only the marginal significance level is reported. Marginal significance levels for the estimated transition parameters and for SKALIN are calculated by simulation methods as described in the text.

Table V in fact reports only the most parsimonious form of the estimated equations, since in no case the restrictions that $\rho = 0$, $\phi_j = -\phi_j^*$ and $\alpha = \alpha^* = \kappa = 0$ could be rejected at the 5% significance level (see the likelihood ratio statistic, LR, in Table V). These restrictions imply an equilibrium log-level of the basis of zero, in the neighborhood of which b_t is nonstationary, becoming increasingly mean reverting with the absolute size of the deviation from equilibrium.

The residual diagnostic statistics are satisfactory in all cases (Eitrheim & Teräsvirta, 1996). The estimated transition parameter (standardized by dividing it by the estimated variance of the dependent

variable, as suggested by Teräsvirta, 1994, 1998) in each case appears to be strongly significantly different from zero on the basis of the individual "t-ratios." It should be clear on reflection, however, that the "t-ratios" for these parameters should be interpreted with caution since, under the null hypothesis $H_0: \theta^2 = 0$, the basis follows a unit root process. Hence, the presence of a unit root under the null hypothesis complicates the testing procedure analogously to the way in which the distribution of a Dickey-Fuller statistic cannot be assumed to be Student's t . The authors therefore calculate the empirical marginal significance levels of these test statistics by Monte Carlo methods assuming that the true data generating process for the logarithm of each of the basis series was a random walk, with the parameters of the data generating process calibrated using the actual basis data over the sample period.²¹ From these empirical marginal significance levels (reported in braces next to the coefficient estimates in Table V), the estimated transition parameter is found to be significantly different from zero at the one percent significance level in each case. Since these tests may be thought of as nonlinear unit root tests, the results indicate strong evidence of nonlinear mean reversion for each of the basis examined over the sample.²²

The strongly nonlinear behavior implied by the empirical results is made clear by Figure 1, which displays the plot of the estimated transition function against the transition variable b_{t-1} , showing that the limiting case of $\Phi[\cdot] = 1$ is attained for both series, although—*ceteris paribus*—the S&P 500 basis appears to adjust toward equilibrium faster than the FTSE 100 basis.

Overall, the nonlinear estimation results are encouraging, uncovering statistically significant evidence of nonlinear mean reversion for each of the two bases examined over the post-1987 sample period. The estimated models are in every case statistically well determined, provide good fits to the data and pass a battery of diagnostic tests.

²¹The empirical significance levels were based on 5,000 simulations of length 2,970, initialized at $b_1 = 0$, from which the first 100 data points were in each case discarded. At each replication a system of ESTAR equations identical in form to those reported in Table V was estimated. The percentage of replications for which a "t-ratio" for the estimated transition parameters greater in absolute value than that reported in Table V was obtained was then taken as the empirical significance level in each case.

²²In addition, to provide corroborating evidence in favor of significant ESTAR nonlinearity in b and its nonlinear mean reversion to a stable equilibrium level, the authors also test for the significance of θ^2 using Skalin's (1998) parametric bootstrap likelihood ratio test. The resulting likelihood ratio statistic for the null hypothesis that $\theta^2 = 0$ for each ESTAR model is very large, with a marginal significance level, similarly calculated by Monte Carlo methods, of virtually zero.

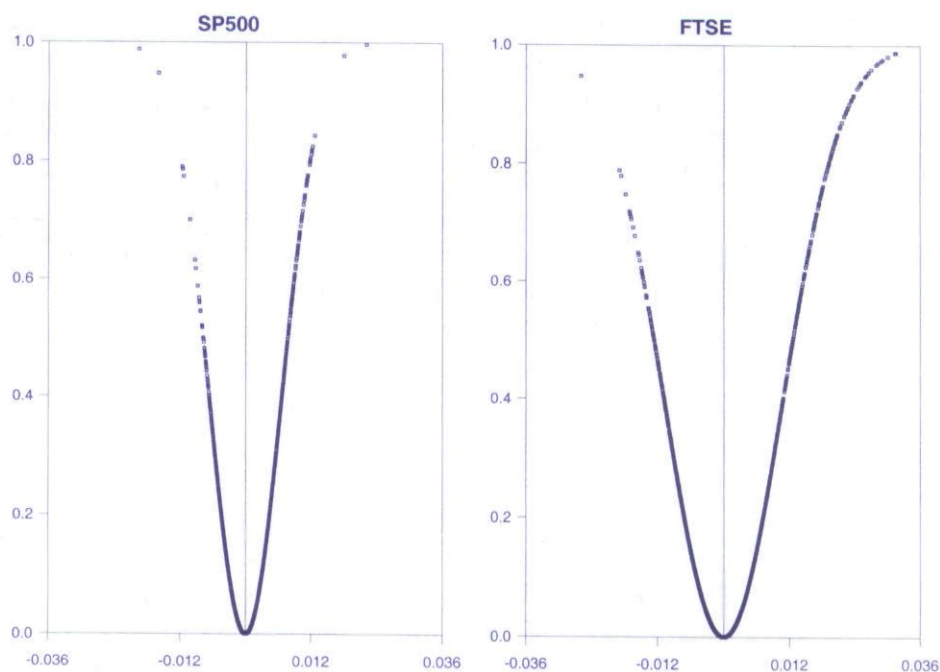


FIGURE 1
Estimated transition functions.

6. THE EMPIRICAL IMPORTANCE OF NONLINEARITY IN THE BASIS: HOW MUCH DOES THE SIZE OF THE SHOCK MATTER?

While the estimated ESTAR models given in Table V impart some idea of the degree of mean reversion exhibited by the basis, a sensible way to gain a full insight into the mean-reverting properties of the estimated nonlinear models is through dynamic stochastic simulation. In particular, an analysis of the impulse response functions will allow the half life of shocks to the basis models to be gauged and these can give a clearer understanding of the importance of nonlinear dynamics in the basis and the validity of the prior that the degree of mean reversion is stronger the bigger the shock to the basis, that is, the larger the deviation of the basis from its equilibrium value.

Thus, the authors examine the dynamic adjustment in response to shocks through impulse response functions which record the expected effect of a shock at time t on the model at time $t + j$. For a linear model, the impulse response function is equivalent to a plot of the coefficients of the moving average representation (e.g., Hamilton, 1994, p. 318). Estimating the impulse response function for a nonlinear

model raises special problems both of interpretation and of computation (Gallant et al., 1993; Koop et al., 1996). In particular, with nonlinear models, the shape of the impulse response function is not independent with respect to either the history of the model at the time the shock occurs, the size of the shock considered, or the distribution of future exogenous innovations. Exact estimates can only be produced—for a given shock size and initial conditions—by multiple integration of the non-linear function with respect to the distribution function of each of the j future innovations, which is computationally impracticable for the long forecast horizons required in impulse response analysis. In this paper, the impulse response functions are calculated, conditional on average initial history, using the Monte Carlo integration method discussed by Gallant et al. (1993).

Specifically, Monte Carlo methods are employed to forecast a path for b_{t+j} given its average history, with and without a shock of size k at time t . Starting at the first data point, b_{t-1} is set equal to $\{|b(1988 : 01 : 01) - \hat{\kappa}| + \hat{\kappa}\}$. If $b(1988 : 01 : 01) - \hat{\kappa}$ is positive, this is just $b(1988 : 01 : 01)$ itself; however, if $b(1988 : 01 : 01) - \hat{\kappa}$ is negative, then $\{|b(1988 : 01 : 01) - \hat{\kappa}| + \hat{\kappa}\}$ is the value which is an equal absolute distance above the estimated equilibrium value $\hat{\kappa}$. This transformation is necessary because only positive shocks are considered, and it is innocuous because of the symmetric nature of ESTAR adjustment below and above equilibrium. Two hundred simulations of length two hundred, with and without a positive shock of size k at time t are then generated using the estimated ESTAR model, and realizations of the differences between the two simulated paths are calculated and stored as before. The authors then move up one data point (hence setting $t - 1 = 1988 : 01 : 02$), and repeat this procedure. Once this has been done for every data point in the sample up to the last sample observation, an average over all of the simulated sequences of differences in the paths of the basis with and without the shock at time t is taken as the estimated impulse response function conditional on the average history of the given basis and for a given shock size. In all, this procedure requires $2,870 \times 200 = 574,000$ simulations for each basis and each shock size.

For linear time series models the size of shock used to trace out an impulse response function is not of particular interest since it serves only as a scale factor, but it is of crucial importance in the nonlinear case. In the present application the authors are particularly concerned with the effect of shocks to the level of the basis. The estimated impulse response functions, obtained from implementing the method discussed above, are graphed in Figure 2 for each of the two bases examined and various

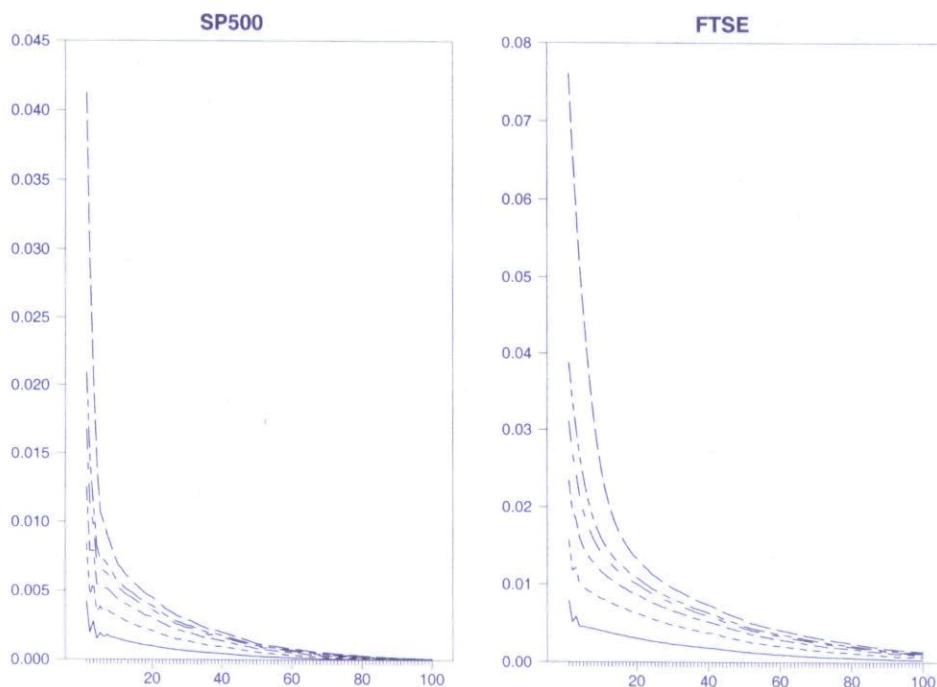


FIGURE 2
Nonlinear impulse response functions.

shock sizes k . Precisely, denoting $\hat{\sigma}_b$ the sample standard deviation of the basis, $k \in \{1 \times \hat{\sigma}_b, 2 \times \hat{\sigma}_b, 3 \times \hat{\sigma}_b, 4 \times \hat{\sigma}_b, 5 \times \hat{\sigma}_b, 10 \times \hat{\sigma}_b\}$. These graphs illustrate very clearly the nonlinear nature of the adjustment, with the impulse response functions for larger shocks decaying much faster than those for smaller shocks.

The estimated quarter lives and half lives of the two basis models, reported in Table VI also illustrate the nonlinear nature of the estimated models, with larger shocks displaying much less persistence than smaller shocks for both indices examined.²³ The S&P 500 basis model shows much faster adjustment in terms of the half life than the FTSE 100 basis model, consistent with the impression given by the plots of the estimated transition functions discussed in Section 5. In fact, for the S&P 500 basis, the model indicates quite fast mean reversion, ranging from a half life of one day for the largest shock size of ten standard deviations to

²³Given a particular value of the log basis at time t , b_t , a shock of k percent to the level of the basis involves augmenting b_t additively by $\log(1 + k/100)$. Hence, a natural measure for the half life is the discrete number of days taken until the shock to the level of the basis has dissipated by a half—i.e., when the impulse response function falls below $\log(1 + k/200)$.

TABLE VI
Estimated Lives of Basis Shocks

Shock Size (%)	S&P 500		FTSE 100	
	1/4 Life	1/2 Life	1/4 Life	1/2 Life
$1 \times \hat{\sigma}_b$	21	7	39	13
$2 \times \hat{\sigma}_b$	20	6	37	11
$3 \times \hat{\sigma}_b$	19	4	35	9
$4 \times \hat{\sigma}_b$	15	3	28	7
$5 \times \hat{\sigma}_b$	11	2	23	5
$10 \times \hat{\sigma}_b$	5	1	12	3

Notes. Half lives are calculated by Monte Carlo integration as discussed in the text; $\hat{\sigma}_b$ denotes the sample standard deviation of the basis.

about one week for very small shocks of one standard deviation. The FTSE 100 basis displays much higher persistence, with half lives ranging from three days for a ten standard deviations shock to just less than two weeks for a one standard deviations shock.

These results seem to shed some light on the importance of nonlinearities in basis dynamics. For small shocks occurring when the basis is near its equilibrium level the nonlinear models consistently yield relatively long half lives, presumably because transactions costs prevent profitable arbitrage opportunities. Large shocks imply, however, faster mean reversion in the basis and fairly plausible half lives, albeit perhaps far longer than believers in a no-arbitrage world would expect. These results contrast with the microstructure literature focusing on intraday data, which typically suggests that arbitrage opportunities are washed out within a day or so. Of course, one possibility is that futures index markets are characterized by heterogeneous traders' populations with different horizons of arbitrage. Nevertheless, although the approach taken in this paper does not allow to distinguish between different plausible explanations of slow or gradual mean reversion in the futures basis, the data clearly suggest that mean reversion in futures markets is, in general, puzzlingly slow at the daily frequency.

7. CONCLUSION

This article illustrates how, in a world characterized by nonzero transactions costs, the resulting correction to the futures price and the basis may yield a process for the futures basis which exhibits nonlinearly mean-reverting behavior. Employing a nonlinear empirical model for the

basis designed to capture the implications of plausible theoretical considerations provided strong confirmation that the bases of both the S&P 500 and the FTSE 100 indices are well characterized by nonlinearly mean reverting processes over the post-crash period since 1988. The crucial estimated parameters, the transition parameters, were of the correct signs and plausible magnitudes and were shown to be strongly statistically significantly different from zero, allowing for a unit root process under the null hypothesis and calculating their empirical significance levels by Monte Carlo methods. The estimated models imply an equilibrium level of the basis in the neighborhood of which the behavior of the basis is nonstationary, becoming increasingly mean reverting with the absolute size of the deviation from equilibrium, consistent with theoretical arguments in the spirit of nonzero transactions costs.

Impulse response functions were calculated by Monte Carlo integration. Because of the nonlinearity, the half-lives of shocks to the basis vary both with the size of the shock and with the initial conditions. By taking account of statistically significant nonlinearities, the speed of adjustment of the basis towards its equilibrium value is found to be an increasing function of the size of the shock (deviation from equilibrium). However, the half lives recorded in this paper also suggest that deviations from equilibrium appear to be quite persistent in the futures markets considered.

Although these results aid the profession's understanding of basis behavior, they should be viewed as a tentatively adequate characterization of the data that appears to be consistent with both the underlying pricing theory and the view held by a number of academics and practitioners that "arbitrage is like gravity." Although the nonlinear model proposed appears superior to linear basis modeling in a number of respects and highlights important features of the dynamics that characterizes the futures basis of the major stock index futures markets examined, it is of course capable of improvement. In particular, one may gain further insights into the adjustment process by developing nonlinear equilibrium correcting systems of equations involving spot prices, futures prices and other economic and financial variables capable of affecting both the equilibrium level of the basis and the dynamic adjustment of the basis towards equilibrium. Also, it would be interesting to use the nonlinear model proposed in this paper or variants of it to investigate the forecasting performance of this nonlinear framework relative to conventional linear and nonparametric methods used for modeling and forecasting purposes in stock index futures markets. These challenges remain on the agenda for future research.

BIBLIOGRAPHY

- Blank, S. C. (1991). "Chaos" in futures markets? A nonlinear dynamical analysis. *Journal of Futures Markets*, 11, 711-728.
- Brennan, M. J., & Schwartz, E. S. (1990). Arbitrage in stock index futures. *Journal of Business*, 63, 7-31.
- Brenner, R. J., & Kroner, K. F. (1995). Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets. *Journal of Financial and Quantitative Analysis*, 30, 23-42.
- Chan, K. (1992). A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies*, 5, 123-152.
- Dubofsky, D. A. (1992). *Options and financial futures: Valuation and uses*. New York: McGraw Hill.
- Dumas, B. (1992). Dynamic equilibrium and the real exchange rate in spatially separated world. *Review of Financial Studies*, 5, 153-180.
- Dumas, B. (1994). Partial equilibrium versus general equilibrium models of the international capital market. In F. Van Der Ploeg (Ed.), *The handbook of international macroeconomics* (Chapter 10). Oxford: Blackwell.
- Dwyer, G. P., Locke, P., & Yu, W. (1996). Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash. *Review of Financial Studies*, 9, 301-332.
- Eitrheim, O., & Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74, 59-75.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 55, 987-1007.
- Figlewski, S. (1984). Hedging performance and basis risk in stock index futures. *Journal of Finance*, 39, 657-669.
- Fujihara, R. A., & Mougoue, M. (1997a). Linear dependence, nonlinear dependence and petroleum futures market efficiency. *Journal of Futures Markets*, 17, 75-99.
- Fujihara, R. A., & Mougoue, M. (1997b). An examination of linear and nonlinear causal relationships between price variability and volume in petroleum futures markets. *Journal of Futures Markets*, 17, 385-416.
- Fuller, W. A. (1976). *Introduction to statistical time series*. New York: Wiley & Sons.
- Gallant, A. R., Rossi, P. E., & Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, 61, 871-908.
- Gao, A. H., & Wang, G. H. K. (1999). Modeling nonlinear dynamics of daily futures price changes. *Journal of Futures Markets*, 19, 325-351.
- Garbade, K. D., & Silber, W. L. (1983). Price movements and price discovery in futures and cash markets. *Review of Economics and Statistics*, 65, 289-297.
- Granger, C. W. J., & Lee, T. H. (1999). The effect of aggregation on nonlinearity. *Econometric Reviews*, 18, 259-269.
- Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford: Oxford University Press.

- Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.
- Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, 2, 1–14.
- Jansen, E. S., & Teräsvirta, T. (1996). Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, 58, 735–763.
- Jarque, C. M., & Bera, A. (1987). Efficient tests for normality, heteroscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255–259.
- Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, 12, 231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–1580.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated VAR models*. Oxford: Oxford University Press.
- Karatzas, I., & Shreve, S. E. (1998). *Methods of mathematical finance*. New York: Springer.
- Kawaller, I. G. (1991). Determining the relevant fair value(s) of S&P 500 futures: A case study approach. *Journal of Futures Markets*, 11, 453–460.
- Kawaller, I. G., Koch, P. D., & Koch, T.-W. (1987). The temporal price relationship between S&P 500 futures and the S and P 500 Index. *Journal of Finance*, 42, 1309–1329.
- Klemkosky, R. C., & Lee, J.-H. (1991). The intraday ex post and ex ante profitability of index arbitrage. *Journal of Futures Markets*, 11, 291–311.
- Koop, G., Pesaran, M. H., & Potter, S. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74, 119–147.
- Lekkos, I., & Milas, C. (2001). The predictability of excess returns on UK bonds: A non-linear approach. Bank of England and University of Sheffield, Discussion Paper.
- Liu, J., & Longstaff, F. A. (2000). Losing money in arbitrage. *Review of Financial Studies*, forthcoming.
- Lo, A. W., & MacKinlay, A. C. (1999). *A non-random walk down Wall Street*. Princeton, NJ: Princeton University Press.
- Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75, 491–499.
- MacKinlay, A. C., & Ramaswamy, K. (1988). Index-futures arbitrage and the behavior of stock index futures prices. *Review of Financial Studies*, 1, 137–158.
- MacKinnon, J. (1991). Critical values for cointegration tests. In R. F. Engle & C. W. J. Granger (Eds.), *Long-run economic relationships* (pp. 267–276). Oxford: Oxford University Press.
- Martens, M., Kofman, P., & Vorst, T. C. F. (1998). A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics*, 13, 245–263.

- Michael, P., Nobay, A. R., & Peel, D. A. (1997). Transactions costs and non-linear adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy*, 105, 862–879.
- Miller, M. H., Muthuswamy, J., & Whaley, R. E. (1994). Mean reversion of Standard & Poor's 500 Index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance*, 49, 479–513.
- Modest, D. M., & Sundaresan, M. (1983). The relationship between spot and futures prices in stock index futures markets: Some preliminary evidence. *Journal of Futures Markets*, 3, 15–41.
- Neal, R. (1996). Direct tests of index arbitrage models. *Journal of Financial and Quantitative Analysis*, 31, 541–562.
- Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54, 461–472.
- Parhizgari, A. M., & de Boyrie, M. E. (1997). Predicting spot exchange rates in a nonlinear estimation framework using futures prices. *Journal of Futures Markets*, 17, 935–956.
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75, 335–346.
- Ramsey, J. B. (1969). Tests for specification errors in classical least-squares regression analysis. *Journal of the Royal Statistical Analysis, Series B*, 31, 350–371.
- Sarno, L., & Valente, G. (2000). The cost of carry model and regime shifts in stock index futures markets: An empirical investigation. *Journal of Futures Markets*, 20, 603–624.
- Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52, 923–973.
- Shiller, R. J., & Perron, P. (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters*, 18, 381–386.
- Skalin, J. (1998). Testing linearity against smooth transition autoregression using a parametric bootstrap. Department of Economic Statistics, Stockholm School of Economics, Working Paper Series in Economics and Finance No. 276.
- Sofianos, G. (1993). Index arbitrage profitability. *Journal of Derivatives*, 1, 6–20.
- Stoll, H. R., & Whaley, R. E. (1990). The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative Analysis* 25, 441–468.
- Taylor, N., van Dijk, D., Franses, P. H., & Lucas, A. (2000). SETS, arbitrage activity, and stock price dynamics. *Journal of Banking and Finance*, 24, 1289–1306.
- Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89, 208–218.
- Teräsvirta, T. (1998). Modelling economic relationships with smooth transition regressions. In D. E. A. Giles & A. Ullah (Eds.), *Handbook of applied economic statistics* (pp. 502–552). New York: Marcel Dekker.

- Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Processes and Their Applications*, 21, 251–273.
- Tong, H. (1990). *Nonlinear time series: A dynamical system approach*. Oxford: Clarendon Press.
- Yadav, P. K., Pope, P. F., & Paudyal, K. (1994). Threshold autoregressive modeling in finance: The price differences of equivalent assets. *Mathematical Finance*, 4, 205–221.
- Yang, S. R., & Brorsen, B. W. (1993). Nonlinear dynamics of daily futures prices: Conditional heteroskedasticity or chaos? *Journal of Futures Markets*, 13, 175–191.
- Yang, S. R., & Brorsen, B. W. (1994). Daily futures price changes and nonlinear dynamics. *Structural Change and Economic Dynamics*, 5, 111–132.

Copyright of Journal of Futures Markets is the property of John Wiley & Sons, Inc. / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.