

Department of Economics

Nonlinear Models with Strongly Dependent Processes
and Applications to Forward Premia and Real Exchange Rates

Richard Baillie and George Kapetanios

Working Paper No. 570

September 2006

ISSN 1473-0278



Nonlinear Models with Strongly Dependent Processes and Applications to Forward Premia and Real Exchange Rates

Richard T. Baillie

Departments of Economics and Finance, Michigan State University, East Lansing, MI
48824 (baillie@msu.edu)

Department of Economics, Queen Mary University of London

and

George Kapetanios

Department of Economics, Queen Mary University of London,
London E1 4NS, U.K. (g.kapetanios@qmul.ac.uk)

First version June, 2006

This version: August, 2006

Abstract

This paper considers estimation and inference in some general non linear time series models which are embedded in a strongly dependent, long memory process. Some new results are provided on the properties of a time domain *MLE* for these models. The paper also includes a detailed simulation study which compares the time domain *MLE* with a two step estimator, where the Local Whittle estimator has been initially employed to filter out the long memory component. The time domain *MLE* is found to be generally superior to two step estimation. Further, the simulation study documents the difficulty of precisely estimating the parameter associated with the speed of transition. Finally, the fractionally integrated, nonlinear autoregressive-*ESTAR* model is found to be extremely useful in representing some financial time series such as the forward premium and real exchange rates.

Key Words: Non-linearity, *ESTAR* models, Strong dependence, Forward Premium, Real Exchange Rates.

JEL Codes: C22, C12, F31.

1 Introduction

The literature on non linear time series models has largely developed within the context of stationary $I(0)$ processes, where the nonlinear component is a complement to an *ARMA* model. However, there is now a large body of work on the theoretical properties of strongly dependent, long memory processes in econometrics, which has developed independently of the work on nonlinear time series. There is also a burgeoning literature on the application of long memory processes in applied econometric work. As yet, there has been relatively little research on combining nonlinearity with the literature on long memory processes. A notable exception is the work of van Dijk, Frances, and Paap (2002), who consider an interesting combination of a fractionally integrated model with *LSTAR* errors, which they apply to a series of quarterly Dutch unemployment. A more recent paper by Baillie and Kapetanios (2006), develops various tests for the presence of general forms of nonlinearity that are complementary to a long memory process. These tests are based on logistic approximations and artificial neural networks and are known to be very good approximations of quite general forms of non linearity. Baillie and Kapetanios (2006) also find evidence for the presence of both non linear and long memory components in many economic and particularly financial time series. This current paper is a logical extension to issues arising in the econometric estimation and modeling of combined non linear and long memory processes.

Particular attention is given to the estimation of models with an adjustment mechanism that tends to zero (or any other constant), as the transition variable becomes either extremely large or extremely small. One important model in this class is the Exponential Smooth Transition Autoregression, or *ESTAR* model see Terasvirta (1994) and Granger and Teräsvirta (1993) for more details. It is also possible to estimate a parametric long memory model simultaneously with a generic nonlinear model based on approximation expansions. While this is certainly technically quite feasible, it is not pursued here, since this paper concentrates on the more economically meaningful models with a parsimonious parametrization. The methodology estimates all the parameters simultaneously and for some classes of model is seen to realize parameter estimates that have a limiting Normal distribution. The theoretical results are complemented by a detailed simulation study.

Additional results are presented for a two step estimator, where a semi parametric estimator of the long memory parameter is used to subsequently estimate the parameters associated with the stationary, $I(0)$ linear and non linear parts of the model. The combined nonlinear, long memory models appear to be of considerable use in many areas of financial

econometrics. Detailed examples are given of applications to two controversial areas of international finance; namely forward premia and real exchange rates. The models suggested in this paper turn out to be very appropriate for these time series.

The plan for the rest of the paper is as follows: Section 2 presents the theoretical framework of the various parametric models and approximating structures for nonlinearity. The estimation procedures, including the discussion of the theorem proving consistency and asymptotic normality of the time domain *MLE* are discussed in Section 3. Then, Section 4 presents detailed simulation evidence on the performance of the estimator and its finite sample properties. The *MLE* is compared with a two step estimator which utilizes the Local Whittle estimator in the first step. The time domain *MLE* is found to be generally superior to the two step estimator and also illustrates the difficulty of precisely estimating the parameter associated with the speed of transition. Surprisingly, the simulation appears to be the first study to deal with the performance of the estimators of the parameters in an *ESTAR* model. Section 5 applies the methodology to monthly forward premium since 1979 and finds strong evidence for long memory and nonlinearity, and provides some insight into how these results integrate into the literature on the forward premium anomaly. Then, section 6 applies the new methodology to a series of over two hundred years of real exchange rates. Again the model performs very well and is indicative of slow responses to parity conditions. The paper ends with a short section containing a summary and conclusion.

2 Nonlinear-Long Memory Models

Long memory, fractionally integrated processes are characterized by having slow hyperbolic rates of decay associated with the impulse response weights and autocorrelations. Following Granger and Joyeux (1980), Granger (1980) and Hosking (1981), a univariate time series process with fractional integration in its conditional mean is represented by

$$(1 - L)^d y_t = u_t, \quad t = 1, \dots, T \tag{1}$$

where L is the lag operator and u_t is a short memory, $I(0)$ process. Then y_t is said to be a fractionally integrated process of order d , or $I(d)$. An $I(0)$ process is defined as having partial sums that converge weakly to Brownian motion. The d parameter represents the degree of "long memory", or persistence in the series. For $-0.5 \leq d \leq 0.5$ the process is stationary and invertible; while for $0.5 \leq d \leq 1$, the process does not have a finite variance, but still has a cumulative impulse response function with finite sum. If the short memory component

is represented as an $ARMA(p, q)$ process, then equation (1) becomes the $ARFIMA(p, d, q)$ model,

$$\phi(L)(1 - L)^d y_t = \theta(L)\epsilon_t \quad (2)$$

where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E(\epsilon_t \epsilon_s) = 0$, $s \neq t$, and where $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator of orders p and q respectively, with all their roots lying outside the unit circle. The Wold decomposition, or infinite order moving average representation of this process is given by

$$y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (3)$$

while the infinite order autoregressive representation is given by

$$y_t = \sum_{i=1}^{\infty} \varpi_i y_{t-i} + \epsilon_t \quad (4)$$

For large lags i , these coefficients decay at the very slow hyperbolic rates of $\psi_i \sim c_1 i^{d-1}$ and $\varpi_i \sim c_2 i^{-d-1}$, where c_1 and c_2 are constants. The hyperbolic decay that is generated by such a process is known as the ‘Hurst effect’, after Hurst (1951), who first discovered the phenomenon in hydrological time series data. This paper considers situations where the short memory process u_t is allowed to be a nonlinear process rather than a pure $ARMA$ process.

A separate, but related issue concerns the origin of the long memory characteristic in economic and financial time series. In particular, Granger (1980) developed the theory of how the contemporaneous aggregation of stationary, independent first order autoregressive processes can generate a fractionally integrated process. Parke (1999) has considered a discrete time error duration model where a set of *iid* shocks have different durations or times of being “alive” and has shown that this model generates long memory. This approach is related to that of Cioczek-Georges and Mandelbrot (1995) and shows the relationship between long memory and heavy tailed densities. Furthermore, Diebold and Inoue (2001) have shown that a process with Markov switching regimes can be mistaken for a long memory process. These papers strongly suggest that long memory can arise from forms of some nonlinearity. However, the tests of Baillie and Kapetanios (2006) suggest the existence of *both* nonlinear and long memory components in many economic and financial time series. Hence this paper is directed at the joint modeling of this phenomenon.

It is possible to use techniques which allow for quite general non linear processes, and which can be expressed by the equations

$$u_t = (1 - L)^d y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} \quad (5)$$

and

$$u_t = F(u_{t-1}, \dots, u_{t-p}) + \epsilon_t \quad (6)$$

so that the short memory part of the process is a possibly nonlinear autoregressive involving the last p lags of the variable u_t . The non-linear part of the model could in theory be approximated by an Artificial Neural Network (*ANN*) methodology, such as that underlying Lee, White, and Granger (1993), where the conditional mean of u_t given lags of u_t is a linear function of the past information set, and the nonlinear part of $F(\cdot)$ in (6) is given by $\sum_{j=1}^q \beta_j \phi(\sum_{i=1}^p \gamma_{ij} \hat{u}_{t-i})$ where $\phi(\lambda)$ is some basis function such as the logistic function, given by $[1 + \exp(-\lambda)]^{-1}$.¹ As noted by Lee, White, and Granger (1993) the logistic function can approximate arbitrarily well any continuous function. One possibility would be to jointly estimate the long memory component in (1) with a nonlinear model based on an *ANN* expansion. While it is in principle straightforward to estimate such a system, the route adopted in this study is to concentrate on a simple parametric structure of the nonlinearity. Hence a natural model is where the strongly dependent component is represented by a fractionally integrated process as in (1). While the stationary $I(0)$ component is composed of a linear autogression i.e. $AR(p)$ process with a nonlinear autoregression of order k whose argument is a smooth transition transformation of the stationary $I(0)$ component. For sake of notational convenience, this model is labeled as *FI(d) - NLAR(p, k) - STAR* model and is represented as,

$$(1 - L)^d y_t = u_t, \quad (7)$$

$$u_t = \alpha(L)u_{t-1} + \beta(L)u_{t-1}\phi(u_{t-D}) + \epsilon_t \quad (8)$$

where the polynomials in the lag operator are $\alpha(L) = \alpha_0 + \sum_{i=0}^{p-1} \alpha_i L^i$ and $\beta(L) = \beta_0 + \sum_{i=0}^{k-1} \beta_i L^i$ and $\phi(u_{t-D})$ is the Smooth Transition Autoregression function, i.e. *STAR*. Finally, D is a delay parameter and is set equal to unity in the simulations reported in this paper.

¹The original work of Lee, White, and Granger (1993) and some extensions in Baillie and Kapetanios (2006) were primarily directed at testing. Then various issues concerning the generation of the γ_{ij} coefficients to deal with the identification problem become important. See Lee, White, and Granger (1993) for details of the construction of the regressors $\phi(\sum_{i=1}^p \gamma_{ij} \hat{u}_{t-i})$, $j = 1, \dots, q$ and the formation of an *LM* test. Stinchcombe and White (1998) show the testing procedure to be consistent under the alternative hypothesis.

One of the most widely used models in the nonlinear literature is the *ESTAR* model, where the *STAR* part of the model is represented by $\phi(u_{t-D}) = \exp(-\gamma(u_{t-D} - c)^2)$. For the empirical implementation of the *ESTAR* model it is usual to set the transition function to $1 - \exp(-\gamma(u_{t-D} - c)^2)$, i.e., to $1 - \phi(u_{t-D})$. Hence the *NLAR*(p, k) – *ESTAR* part of the model can also be represented as

$$u_t = (\alpha_0 + \beta_0 [1 - \exp(-\gamma(u_{t-D} - c)^2)]) + \sum_{i=1}^p \alpha_i u_{t-i} + \sum_{i=1}^k \beta_i u_{t-i} [1 - \exp(-\gamma(u_{t-D} - c)^2)] + \epsilon_t \quad (9)$$

This is the form of the model that will be used in the Monte Carlo and empirical sections of the paper.

Since the *ESTAR* model is so widespread in applied work for stationary processes, it has also been implemented in this study. Hence both the simulation study and empirical results are built around the case of a long memory process with a stationary component which has both linear and non linear autoregressive parts and where the nonlinear transformation is of the form of *ESTAR*. It should be noted that there are many other possible parameterizations, such as the Gaussian transformation, or the Logistic Smooth Transition Autoregressive *LSTAR* process

$$u_t = (\alpha_0 + \beta_0 [1 + \exp(-\lambda_1(u_{t-D} - \lambda_0))]^{-1}) + \sum_{i=1}^p \alpha_i u_{t-i} + \sum_{i=1}^k \beta_i u_{t-i} [1 + \exp(-\lambda_1(u_{t-D} - \lambda_0))]^{-1} + \epsilon_t \quad (10)$$

which has been used by van Dijk, Frances, and Paap (2002). This is clearly one of many possible interesting parametric forms to represent the adjustment process; and has been used by Michael, Nobay, and Peel (1997) and Sarantis (1999) among others in empirical work. Evidence is presented in sections 5 and 6 of this paper, for the use of the *ESTAR* formulation in conjunction with a long memory process.

3 Properties of Time Domain MLE

This section now considers the properties of the time domain *MLE* and proves consistency and asymptotic normality for a useful class of process. While a quite general formulation is captured by the *FI*(d) – *NLAR*(p, k) – *ESTAR* model, the results on the properties of the class of models for which the time domain *MLE* is valid is considerably more extensive. In

particular, attention is confined to the $FI(d) - NLAR(p, k) - STAR$ process, where $STAR$ refers to a general smooth transition process that satisfies the various assumptions listed below. To precede the theorem, it is necessary to state the following assumptions:

Assumption 1 y_t is a zero mean process.

Assumption 2 The range of $\phi(\cdot)$ is the unit interval, while the $\lim_{y \rightarrow \pm\infty} \phi(y)y = 0$ and $\phi(\cdot)$ belongs to C^3 .

Assumption 3 The disturbance ϵ_t is *i.i.d.* $(0, \sigma^2)$, with $E(\epsilon_t^4) < \infty$ The density of ϵ_t , denoted by φ , is positive everywhere, continuous and thrice differentiable.

Assumption 4 The roots of the lag polynomials $\alpha(L) = \alpha_0 - \alpha_1 L - \dots - \alpha_p L^p$ are outside the unit circle.

It should be noted that the last assumption covers the behavior of the lag polynomial that is relevant for the outer regime. The behavior of $\alpha(L) + \beta(L)$, the polynomial in the lag operator for the inner regime, is not relevant for the global stationarity of the system.

Theorem 1 For the following $FI(d) - NLAR(p, k) - STAR$ process

$$(1 - L)^d y_t = u_t \tag{11}$$

$$u_t = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i} + \sum_{j=0}^k \beta_j \phi(u_{t-1}, \dots, u_{t-l}; \gamma, c) u_{t-j} + \epsilon_t \tag{12}$$

where without loss of generality it is assumed that $l \leq \max(p, k)$. On denoting $\theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_0, \beta_1, \dots, \beta_p, \gamma, c, d)'$, and under assumptions 1 through 4, the MLE of the parameter vector θ is $T^{1/2}$ consistent and has a limiting normal distribution.

The proof of this theorem is given in the Appendix. It should again be emphasized that Theorem 1 allows for a very wide class of nonlinear functions $\phi(\cdot)$, which are considerably more general than the $ESTAR$ type of specification. The behavior of these functions is not specified at all points on the real line, but only at the tails. In this sense, the theorem takes on a semi-parametric interpretation. The need to specify the tails of the function via assumption 2, motivates the use for a general model in the theoretical discussion. This model when specialized to the $ESTAR$ case becomes an alternative representation to the formulation used in the empirical sections of this paper. See the details in the discussion

above (9). It is shown in the proof of the theorem in the appendix that models implied by equations (1) and (12) satisfy the conditions of Fox and Taqqu (1986) for the consistency and asymptotic normality of the *MLE* estimates of all the parameters, under the null hypothesis. As in Fox and Taqqu (1986) the proof is in terms of a demeaned process.

On assuming Gaussianity of the white noise process ϵ_t , then estimation by minimizing the conditional sum of squared residuals is equivalent to approximate *MLE* in the time domain and is numerically straightforward. Hence joint estimation and testing is carried out by the minimization of the sum of squared residuals $\sum_{t=1}^T \hat{\epsilon}_t^2$. The approximate *MLE* for a model including the long memory parameter and a parametric non linear component such as *ESTAR* is relatively computationally intensive, but is fortunately quite feasible given the computational power that is available today. This approach has previously been successfully applied to other models with a long memory component, such as *ARFIMA* with *GARCH*; see Baillie, Chung, and Tieslau (1996). Although the method does not take into account starting values as considered by Sowell (1992), it has been shown in several studies to perform well in sample sizes of 100 observations or more: see Cheung (1993), Cheung and Diebold (1994) and Taqqu and Teverovsky (1998). Various standard information criteria, such as *AIC* are used for model selection and are found to generally lead to quite adequate *FI(d) – NLAR(p, k) – ESTAR* models being selected and estimated.

An alternative to the one step *MLE* is to use a two step estimation procedure, where an estimate of d is initially found from a semi parametric procedure, such as the Local Whittle, which is then used to fractionally filter the observed series y_t from the relationship $u_t = y_t - \sum_{l=0}^{t-p} \pi_l y_{t-l} - \sum_{l=0}^{\infty} \pi_l y_{t-l} = (1-L)^d y_t$. The filter with estimated parameters generates \hat{u}_t ; and the next stage is to then to apply *MLE* to estimate a *NLAR(p, k) – ESTAR* model to the \hat{u}_t series. For the purposes of this study it was decided to use the Local Whittle semi-parametric estimator for d , which is obtained by minimizing the objective function

$$\ln \left[\frac{1}{m} \sum_{j=1}^m \omega_j^{2d} I(\omega_j) \right] - \frac{2d}{m} \sum_{j=1}^m \log(\omega_j) \quad (13)$$

with respect to d , where $I(\omega_j)$ is the periodogram given by

$$I(\omega_j) = \frac{1}{2\pi T} \left| \sum_{j=1}^T y_t e^{i\omega_j t} \right|^2$$

and the usual bandwidth used in both the simulation and empirical work is to choose $m = [T^{0.5}]$. As a robustness check, other choices of bandwidth, such as $m = [T^{0.75}]$ were also

considered in this study. However, in terms of the results reported in this study, the semi-parametric estimate of d , is always based on the choice of $m = [T^{0.50}]$. Full details of the simulation and empirical results for other bandwidths are omitted for reasons of space, but are available from the authors on request. More generally, it might be of interest, although beyond the scope of the present paper, to consider data dependent methods for setting the bandwidth such as, e.g., those discussed in Andrews and Sun (2004) or Henry and Robinson (1996). Details on the theoretical properties of the Local Whittle estimator may be found in Robinson (1995). See also Dalla, Giraitis, and Hidalgo (2005) for a discussion of properties of the Local Whittle estimator for general nonlinear processes.

The lag order of the models, p , may be determined by an information criterion or chosen a priori. While there is no formal justification for applying information criteria in this particular context, it should be noted that available results in the literature strongly suggest the standard asymptotic properties of the various information criteria to hold in this context. The relevant properties include consistent model selection of the Schwartz (Schwarz (1978)) and Hannan-Quinn (Hannan and Quinn (1979)) information criteria; while the Akaike information criterion is known to be inconsistent. In particular, Sin and White (1996) and Kapetanios (2001) have shown that these properties extend to nonlinear models for weakly dependent processes; while Hidalgo (2002) has shown that similar results are valid for regressions involving long memory regressors. Finally, recent work by Poskitt (2005) has extended the optimality results of Shibata (1980) to stationary long memory processes.

4 Simulation Study

This section reports the results from a detailed Monte Carlo study to investigate the performance of the various estimation procedures applied to data generated from various specifications of the $FI(d) - NLAR(p, k) - ESTAR$ model. There are nine different designs which allow for the specification of the basic bell shaped $ESTAR$ process, and also allow for various degrees of persistence in the linear autoregressive and non linear autoregressive coefficients. Also, two values of $d = 0.2$ and $d = 0.4$ were selected for the long memory parameter for each of the nine basic experiments. Hence, there are eighteen experiments, i.e. nine with $d = 0.2$ and nine with $d = 0.4$. The numerical values for the parameters in each design were

- Experiment 1 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.2, \beta_1 = -0.1, \gamma = 0.2$

- Experiment 2 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.4, \beta_1 = -0.2, \gamma = 0.2$
- Experiment 3 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.8, \beta_1 = -0.4, \gamma = 0.2$
- Experiment 4 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.2, \beta_1 = -0.1, \gamma = 0.4$
- Experiment 5 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.4, \beta_1 = -0.2, \gamma = 0.4$
- Experiment 6 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.8, \beta_1 = -0.4, \gamma = 0.4$
- Experiment 7 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.2, \beta_1 = -0.1, \gamma = 1$
- Experiment 8 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.4, \beta_1 = -0.2, \gamma = 1$
- Experiment 9 $\alpha_0 = 0, \beta_0 = 0, \alpha_1 = 0.8, \beta_1 = -0.4, \gamma = 1$

The degree of peakedness in the shape of the *ESTAR* function is determined by the value of the γ parameter, with experiments 1 through 3 being very tightly concentrated and very peaked since $\gamma = 0.2$, while designs 4 through 6 with $\gamma = 0.4$ are smoother and more of the classic bell shape. Finally, experiments 7 through 9 with $\gamma = 1.0$ have an *ESTAR* function which is relatively flat and spread out. The main purpose of this simulation was to assess the performance of the *MLE* and two step Local Whittle estimator for all eight parameters in the *FI(d) – NLAR(1, 1) – ESTAR* model, including the long memory parameter. However, rather surprisingly, this simulation analysis appears to be the first Monte Carlo which has investigated the properties of parameter estimators for *STAR* models.

The chosen parameter values for the nonlinear specifications is similar to those in other Monte Carlo studies using *ESTAR* models such as Kapetanios, Shin, and Snell (2003). All the experiments, or designs, represent geometrically ergodic processes for u_t . The geometric ergodicity of the above processes has been proven for *STAR* models by Kapetanios, Shin, and Snell (2003) using the drift condition by Tweedie (1975). Since these processes are geometrically ergodic, they possess sufficiently rapid decay in their coefficients to satisfy β -mixing and hence α -mixing; see Davidson (1994, Ch. 14) for further details. An important corollary is that the processes are $I(0)$.

The results of the mean and standard deviation of all the various parameter estimates, from sample sizes of $T = 400, T = 600$ and $T = 1,000$ are reported in Tables 1 through 8. Table 9 reports some additional experiments for the much larger sample size of $T = 5,000$. All the simulation results for each separate experiment were averaged over 500 replications.

Tables 1 through 4 are for the long memory parameter being set as $d = 0.2$. In particular, tables 1 and 2 report the average values of the various parameter estimates, while tables 3 and 4 report their corresponding standard deviations. Tables 5 through 8 follow exactly the same format, only for $d = 0.4$. The following facts emerge from the tables:

(i) The constants in the linear and non linear autoregressive polynomials are estimated very accurately across all designs and different estimation methods.

(ii) There appears to be an interesting trade off between bias and variance when comparing estimation methods for d . Comparisons of the last columns of tables 1 and 2 and also tables 5 and 6 reveal that the Local Whittle estimator is superior in terms of bias to the *MLE* across all experiments. However, comparisons of the last columns of tables 3 and 4 and tables 7 and 8 indicate that the Local Whittle estimator has considerably higher variance than the *MLE* again across all experiments. This finding is consistent with the known slower rate of convergence and $m^{1/2}$ consistency, where m is the number of bandwidths used in applying the Local Whittle estimator. Note however, that this theoretical result is usually derived under some linearity assumption (see, e.g., Dalla, Giraitis, and Hidalgo (2005)).

(iii) While the Local Whittle estimator arguably compares very favorably with *MLE* in terms of just estimating d , the one step *MLE* is generally superior in terms of both bias and variance for all the other parameters. Hence an investigator may reasonably choose the Local Whittle estimator for simply estimating the long memory parameter. However, the simulation results caution against using the Local Whittle estimator to fractionally filter the series before estimating the remaining parameters in a second step. The alternative one step approximate *MLE* appears to produce superior estimates of all the remaining parameters not including the long memory parameter.

(iv) The estimates of γ are generally poor for all designs. This is not surprising given the nature of the parameter measuring slope or curvature. The quality of the estimator improves both with increasing degrees of persistence in the autoregressive processes and also with increasing sample sizes. Some further aspects of the issue of estimating the γ parameter can be seen from Table 9 where the average values and standard deviation of the parameter estimates are investigated for Experiments 3, 6 and 9 for the sample size of $T = 5,000$. It should be noted that the bias in the *MLE* of γ is now very small for all three designs, although its standard deviation remains quite large for Experiment 9.

(v) The parameter, α_1 , which is associated with the degree of persistence in the linear autoregressive processes is generally estimated with a downward bias in all cases. This is in

contrast to the parameter, β_1 , which represents with the degree of persistence in the non-linear autoregressive processes and is estimated with an upward bias across experiments. In general a measure of total persistence, which is the sum of the estimated parameters $\alpha_1 + \beta_1$ is extremely well estimated.

(vi) The value of d does not appear to be related to the quality of any of the *MLE* parameter estimates, including the long memory parameter itself.

5 Analysis of the Forward Premium

There has been considerable previous literature on the most appropriate time series representation for the forward premium in currency markets. The possibility of the forward premium being a strongly dependent, fractionally integrated process has key implications for the theory, and as shown below, the possibility of the forward premium being a nonlinear process is also highly relevant. First it is necessary to provide some background information. From the theory of Uncovered Interest Parity, (*UIP*), the expected future rate of appreciation, (depreciation) on a currency is equal to the current forward premium (discount), or equivalently the interest rate differential. Hence,

$$E_t(\Delta s_{t+1}) = (f_t - s_t) = (i_t - i_t^*) \quad (14)$$

where $E_t(\cdot)$ denotes the conditional expectation based on a sigma field of all relevant information at time t . The variable s_t is the logarithm of the spot exchange rate and is measured in terms of the number of dollars in terms of a unit of foreign currency at time t ; while i_t and i_t^* denote the common maturity nominal interest rates available on similar domestic and foreign assets respectively. A common test for *UIP* is to estimate the regression equation

$$\Delta s_{t+1} = \alpha + \beta(f_t - s_t) + \epsilon_{t+1}, \quad (15)$$

where under the null hypothesis of *UIP*, $\alpha = 0$, $\beta = 1$ and ϵ_{t+1} is a serially uncorrelated disturbance. The forward premium anomaly is the widespread empirical finding of a negative slope coefficient in the above regression, so that the rate of appreciation of the spot exchange rate is negatively correlated with the lagged forward premium. This phenomenon has been consistently found for most freely floating currencies in the current float and appears robust to the choice of numeraire currency. Hence the forward premium anomaly implies that the country with the highest interest rate will have an appreciating currency, and not a depreciating currency, as implied by the theory of uncovered interest rate parity. Several

previous papers have questioned the econometric specification of (15), since it appears to be unbalanced, with spot returns which are almost serially uncorrelated being regressed on the highly persistent forward premium. The time series properties of the forward premium are closely related to the issue of cointegration between spot and forward rates. While there is large body of evidence which finds that both spot and forward rates are well represented by $I(1)$ processes and are cointegrated with a coefficient of unity; Baillie and Bollerslev (1994) and Maynard and Phillips (2001) show that the forward premium is well approximated as a long memory process, which suggests a form of fractional cointegration. Hence 15 is then unbalanced and the theory of UIP is immediately seen to be inappropriately expressed in terms of the equation (15).

For the purposes of illustrating the methodology developed earlier in this paper, monthly data on spot exchange rates, and one month forward exchange rates were used for the currencies of the Belgian Franc (BF), Canadian Dollar (CD), French Franc (FF), Italian Lira (IL), Japanese Yen (JY), and the UK Pound (UK) vis a vis the US Dollar. The data are provided by the Bank of International Settlements and are end of month mid rates, from December 1978 through December 1998 for the Eurozone currencies of the BF , FF , and IL ; and are from December 1978 through January 2002 for the other currencies of CD , JY and UK . Table 10 presents some basic summary statistics of these time series together with estimates of the $FI(d) - NLAR(p, k) - ESTAR$ model.

In order to motivate the use of a nonlinear long memory model, tests for neglected nonlinearity are presented for each series. In particular, the tests developed in Baillie and Kapetanios (2006) which consider the presence of linear long memory under the null hypothesis and long memory with nonlinearity in the short memory component under the alternative hypothesis, are applied to each series. The tests are based on Taylor approximations and artificial neural network approximations. They are denoted as TLG and ANN respectively. More details can be found in Baillie and Kapetanios (2006). Probability values for these tests are reported in Table 10. It should be noted that to the best of our knowledge, these are the only tests that have been developed explicitly for long memory processes, unlike tests for $ESTAR$ nonlinearity which are designed for short memory processes. The chosen orders, p and k of the models that optimized the information criteria vary across the currencies. The most appropriate order of the linear autoregressive component varied between 10 for JY and 15 for BF , while the optimum order for the nonlinear autoregressive component varied between one for BL , FF , IL and UK , to 2 for the CD and 3 for JY . The MLE

of the long memory parameter d was relatively close to the Local Whittle estimate for most currencies and was only in the stationary region of $0.0 < d < 0.5$ for the *IL*, and was in the region of $0.50 < d < 1.00$ for the others, which implies non stationarity of the process, but nevertheless existence of its cumulative impulse response weights. The implied non stationarity of the forward premium is also consistent with the findings of previous authors such as Baillie and Bollerslev (1994) and Maynard and Phillips (2001) who used different methods. Hence for all eight currencies, the *FI - NLAR(p, k) - ESTAR* model is found to work extremely well and includes as few as 16 parameters for the UK and as many as 21 parameters for Belgium. For two of the estimated models on the forward premium series the estimated models have the first nonlinear autoregressive coefficient β_1 being small and not significantly different from zero. However, the effect of nonlinearity enters through the statistically significant constant term β_0 .

The diagnostic analysis of the estimated models can also be based on examining the roots of the estimated lag polynomial $\alpha(L) + \beta(L)$, which is the relevant operator for the outer regime. Note that this is different from the theoretical analysis of Section 3 since there the transition function is $\phi(u_{t-D})$ rather than $1 - \phi(u_{t-D})$ where $\phi(u_{t-D}) = \exp(-\gamma(u_{t-D} - c)^2)$ (cf. assumption 4). As seen in Table 10, all the eigenvalues of the determinantal polynomial, (i.e. the inverse of the lag polynomial), lie inside the unit circle, which is consistent with the process being stationary. It should be noted that the possible existence of an explosive root in the lag polynomial $\alpha(L)$ is only of interest in terms of the behavior inside the regime and is not relevant for the issue of global stationarity.² Since the estimated long memory parameters exceeded 0.5 in most cases, the models were also estimated after having first differenced the data. It was found that this did not change the results after having added unity to the estimate of the long memory parameter.

Table 10 also presents the Ljung-Box statistic of the residuals, denoted by *LBR*. While the asymptotic distribution of this test is strictly speaking only correct for i.i.d. processes under the null, it is known that if the process is simply uncorrelated, then the assumed asymptotic distribution can be arbitrarily misleading and lead to spurious over rejection of the null hypothesis; see Romano and Thombs (1996). However, for all the estimated models in Table 10, this is not an issue as the test never rejects the null hypothesis of no serial

²It did not prove possible to find a satisfactory model for the German Deutschmark, since the estimate of γ was outside the range of $0.0 < \gamma < 3.0$, and both the Local Whittle and *MLE* methods were unable to reject a unit root of $d = 1$. Full details of these results are not reported here, but are available from the authors on request.

correlation. Note that this test is a portmanteau type test and has power against cases where neglected nonlinearity, arising for example from mis-specification due to an inappropriate parametric nonlinear model, gives rise to serial correlation in the residuals. Note that most conditional mean nonlinear model used in time series analysis, give rise to serial correlation. Further useful information on the estimated models can be obtained from plots of the transition functions in Figure 1. The first six panels present the transition functions for the forward premia whereas the last panel presents the transition function for the Real Exchange Rate data discussed in Section 6. As we can see there is considerable variability across series both with respect to the speed of transition and the minimum value for the transition function.

Further insights into how the empirical results interact with the theory of uncovered interest rate parity (*UIP*) can be derived from the Euler equation for a risk averse investor or trader considering being long or short in the forward market. Then

$$E_t \frac{(S_{t+1} - F_t) U'(C_{t+1})}{P_t U'(C_t)} = 0 \quad (16)$$

where and a Taylor series approximation gives

$$E_t \Delta s_{t+1} = (f_t - s_t) + 0.5 \text{Var}_t[s_{t+1}] + \text{Cov}_t[s_{t+1}p_{t+1}] + \text{Cov}_t[s_{t+1}q_{t+1}] \quad (17)$$

where the last term is derived from the Lucas Breeden asset pricing model, and on denoting $\rho_{t+1} = \text{Cov}_t[s_{t+1}q_{t+1}]$, is the time dependent risk premium. Hence the relevant equation for estimation is

$$\Delta s_{t+1} = (f_t - s_t) + 0.5 \text{Var}_t[s_{t+1}] + \text{Cov}_t[s_{t+1}p_{t+1}] + \rho_{t+1} + \epsilon_{t+1} \quad (18)$$

Given a long memory forward premium, the terms involving the time dependent risk premium and possibly the Jensen inequality term $\text{Var}_t[s_{t+1}]$ will have to enter into the fractionally cointegrating relationship to maintain balance in (15). The issue of nonlinearity and long memory in the forward premium is related to the specification of the *UIP* equation and whether is appropriate. For example, Baillie and Kilic (2006) have estimated a model of the form

$$\Delta s_{t+1} = [\alpha_1 + \beta_1(f_{t,1} - s_t)] + [\alpha_2 + \beta_2(f_{t,1} - s_t)]F(z_t, \gamma, c) + \epsilon_{t+1}, \quad (19)$$

where ϵ_{t+1} is again the disturbance term, and $F(\cdot)$ is a logistic the transition function

$$F(z_t; \gamma, c) = (1 + \exp(-\gamma(z_t - c)/\sigma_{z_t}))^{-1} \quad \text{with } \gamma > 0, \quad (20)$$

where z_t is the transition variable, σ_{z_t} is the standard deviation of z_t , while γ is a slope parameter and c is a location parameter. The parameter restriction $\gamma > 0$ is an identifying restriction. The logistic function (20), is bounded between 0 and 1, and depends on the transition variable z_t so that $F(z_t; \gamma, c) \rightarrow 0$ as $z_t \rightarrow -\infty$, $F(z_t; \gamma, c) = 0.5$ for $z_t = c$, and $F(z_t; \gamma, c) \rightarrow 1$ as $z_t \rightarrow +\infty$. When $\gamma \rightarrow \infty$, $F(z_t; \gamma, c)$ becomes a step function, such that the LSTR model becomes effectively a threshold model. Therefore, the LSTR model nests a two-regime threshold model. For $\gamma = 0$, $F(z_t; \gamma, c) = 0.5$ for all z_t , in which case the model reduces to a linear regression model with parameters $\alpha = \alpha_1 + 0.5\alpha_2$, and $\beta = \beta_1 + 0.5\beta_2$. The exponent in (20) is normalized by dividing by σ_{z_t} , which allows the parameter γ to be approximately scale-free. This is particularly useful for the initial estimates for the nonlinear optimization used to estimate the parameters in (19). The values taken by the transition variable and the transition parameter γ will determine the speed of reversion to *UIP*. For any given value of z_t , the transition parameter γ determines the slope of the transition function and hence the speed of transition between extreme regimes, with low values of γ implying slower transitions. The parameter c can be interpreted as the *threshold* between the two regimes corresponding to $F(z_t; \gamma, c) = 0$ and $F(z_t; \gamma, c) = 1$, in the sense that the logistic function changes monotonically from 0 to 1 as z_t increases, while $F(c; \gamma, c) = 0.5$. The model implies an inner regime when $z_t = c$ and $F(z_t = 0; \gamma, c) = \frac{1}{2}$ and equation (19) becomes a standard linear *UIP* regression

$$\Delta s_{t+1} = [\alpha_1 + \beta_1(f_{t,1} - s_t)] + u_{t+1}, \quad (21)$$

and is consistent with a region where the forward premium anomaly holds. There is also an outer regime when $\lim_{z_t \rightarrow +\infty} F(z_t; \gamma, c)$ where (19) becomes

$$\Delta s_{t+1} = [(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)(f_{t,1} - s_t)] + u_{t+1}. \quad (22)$$

and $\alpha_1 + \alpha_2 = 0$ and $\beta_1 + \beta_2 = 1$, which corresponds to an outer regime where the theory of *UIP* has a high probability of holding. Baillie and Kilic (2006) find that transition variables associated with higher US money growth differential, high volatility of US money growth and low US interest rates, and other variables implied by the fundamentals and risk premium variables are associated with observations in the outer regime where *UIP* has a higher probability of holding. An interesting area for future research would be to combine the forms of nonlinearity present in the forward premium series with the structural changes in *UIP* and the regimes where risk premium are relatively important.

6 Analysis of the Dollar-Pound Real Exchange Rate

Many previous studies, e.g., Diebold, Husted, and Rush (1991), Papell (1997) and Cheung and Lai (2001) have considered whether the real exchange rate series exhibit mean reversion, and whether there is evidence of long run absolute Purchasing Power Parity. Some of these studies have also tried to measure the magnitude and duration of shocks. The evidence has generally been mixed with less evidence of stationarity in the post Bretton Woods regime. One of the motivations of previous studies has been to explain the puzzling inability to reject the null hypothesis of unit root non-stationarity using standard unit root tests. This section of the paper considers an interesting historical series of the *US* dollar vis a vis the *UK* pound. The data are annual and stretch from 1791 through 1994 and were compiled and analyzed by Lothian and Taylor (1996). Clearly such a long historical series covers a number of regimes of floating and fixed exchange rates. However, many previous researchers have argued for the validity of testing the law of one price over such a long span of data. The last column of table 10 duly presents the estimates of an $FI(d) - NLAR(15, 6) - ESTAR$ for the logarithm of the *US - UK* real exchange rate. As with the forward premium series, all the eigenvalues of the relevant determinantal polynomial lie inside the unit circle, implying that the estimated u_t component is indeed $I(0)$. Interestingly, the estimated long memory parameter is 0.403 and has a two sided confidence interval that is still within the stationarity region of being < 0.50 , in contrast to the Local Whittle estimate of 0.554, which has a two sided confidence interval that includes 0.50. Hence the estimated $FI(d) - NLAR(p, k) - ESTAR$ model provides evidence of slow, mean reversion of this historical Dollar-Pound real exchange rate.

7 Conclusion

This paper has examined estimation and model specification of quite general fractionally integrated, nonlinear autoregressions with smooth transition regimes; i.e. so called $FI(d) - NLAR(p, k) - STR$ models. Hence the nonlinear times series structure is embedded in a strongly dependent, long memory process, unlike most previous studies which have considered nonlinearity in a purely stationary environment. The paper has provided some new theoretical results on the properties of a time domain MLE for these models. A detailed simulation study found that the time domain MLE overall performed quite well; although the parameter representing the speed of transition in the $ESTAR$ model is relatively poorly estimated in small samples. However, the MLE of all the parameters performs very well in

a sample of 5,000 which is quite relevant for many applications in financial econometrics.

The simulation study also considered the method of first employing a semi parametric, Local Whittle estimator to estimate the long memory parameter before fractionally filtering the series and then estimating the remaining parameters by *MLE* on the filtered series. The Local Whittle estimator has less bias, but increased *MSE* compared with the time domain *MLE* of the long memory component. However, one step *MLE* is generally superior in terms of both bias and *MSE* for all other parameters in the model. Hence, while an investigator may possibly choose the Local Whittle estimator for simply estimating the long memory parameter, the results of the simulation suggest using the full *MLE* is preferable for estimation of the complete model. The estimated *FI(d) – NLAR(p, k) – ESTAR* models are is found to be extremely successful in representing the nonlinear structures and strong dependencies within forward premium and real exchange rates. The use of these models appears promising for future modeling in these areas.

8 Appendix

Fox and Taqqu (1986), Dahlhaus (1989) and Hosoya (1997) have shown that under certain conditions the *MLE* of parametric stationary long memory models produce \sqrt{T} consistent and asymptotically normal parameter estimates. The rest of the proof will verify the sufficient conditions for the case in this study. For the purposes of the analysis in this paper it is sufficient to show the following two conditions:

Condition (C.1): y_t is a zero-mean stationary long memory $I(d)$ process.

Condition (C.2): The spectral density of $u_t \equiv u_t(d)$, is denoted by $f_u(x, \theta)$ and is continuous and belongs to C^3 . It should be noted that the spectral density of y_t , denoted $f_y(x, \theta)$ can be written as

$$f_y(x, \theta) = |e^{-ix} - 1|^{-2d} f_u(x, \theta) \quad (23)$$

which further illustrates the decomposition of the nonlinear short memory component and the long memory component. This emphasizes the similarity between the class of nonlinear models analysed in this paper compared with the standard *ARFIMA* model. Equation (23) together with conditions (C.1) and (C.2) above, imply the conditions (B.1) through (B.4) of Fox and Taqqu (1986), via the second part of Theorem 3 of Fox and Taqqu (1986), which implies conditions (A.1) through (A.6) of the same paper. The conditions (A.1) through (A.6) relate to the continuity and differentiability of the spectral density of y_t . The non parametric nature of these conditions allows for the forms of non linearity considered in the present paper. It is also important to note that Fox and Taqqu (1986) and subsequent authors such as Dahlhaus (1989) and this paper consider the process y_t to have been demeaned prior to the application of the *MLE*.

It is therefore necessary to establish conditions (C.1) and (C.2). The results of Fox and Taqqu (1986) require the additional assumption of the Gaussianity of y_t which follows from the Gaussianity of u_t under the null. It should be noted that Gaussianity is only needed for asymptotic normality of the *MLE* and not for consistency of the estimator. The assumption of Gaussianity is relaxed by Hosoya (1997) and is replaced with a mixing assumption for u_t and a finite fourth moment assumption for y_t which is satisfied in our case if Assumption 3 above holds³. This follows from the finiteness of the fourth moment of u_t which follows from assumptions 2 and 3 and the Wold decomposition in (3) and the square summability of the

³The main assumptions required for proving normality for the QMLE in Hosoya (1997) are mixing, a Lindeberg type and Lipschitz continuity condition (Condition A of Hosoya (1997)), and differentiability and continuity conditions for the spectral density of y_t , (Conditions C and D). Both are satisfied if the conditions

ψ_i coefficients. The latter is obtained from the Marcinkiewicz-Zygmund inequality (see, e.g., equation (25) of Park (2002)) for some constant c

$$E|y_t|^4 \leq c \left(\sum_{i=1}^{\infty} \psi_i^2 \right) E|u_t|^4 < \infty, \quad (24)$$

It is then clear that y_t is $I(d)$. This property is immediate if it can be shown that u_t is a mixing process that satisfies a functional central limit theorem and is therefore an $I(0)$ process. For example, see Davidson and DeJong (2000). Since the process u_t is β -mixing it follows that it is then geometrically ergodic.

The next step is to show geometric ergodicity for u_t , which is easily established under assumptions 2-4 using the drift condition proposed by Tweedie (1975). This condition states that a process is ergodic under the regularity condition that disturbances have positive densities everywhere if the process tends towards the center of its state space at each point in time. More specifically, a multivariate Markov process v_t is geometrically ergodic if there exists constants $\delta < 1$, $B, L < \infty$, and a small set C such that

$$E [\|v_t\| \mid v_{t-1} = v] \leq \delta \|v\| + L, \quad \forall v \notin C, \quad (25)$$

$$E [\|v_t\| \mid v_{t-1} = v] \leq B, \quad \forall v \in C, \quad (26)$$

where $\|\cdot\|$ is the Euclidean norm. To see how this criterion implies geometric ergodicity for the case in this paper, it is necessary to define $v_t = (u_t(d), \dots, u_{t-\max(p,q)}(d))$. By the boundedness of $\phi(\cdot)$, it easily follows that (26) holds for any compact set C . Hence it is necessary to show that there exists a compact set C such that (25) holds. Let $E^l[\cdot]$ denote the conditional expectation operator for the linear model given by

$$u_t(d) = \alpha_0 + \sum_{i=1}^p \alpha_i u_{t-i}(d) + \epsilon_t \quad (27)$$

Then, for any $\varepsilon > 0$, by virtue of the fact that $\lim_{y \rightarrow \pm\infty} \phi(y)y = 0$, there exists a compact set C_ε such that

$$E [\|v_t\| \mid v_{t-1} = v] \leq E^l [\|v_t\| \mid v_{t-1} = v] + \varepsilon \quad (28)$$

(A.1) through (A.6) of Fox and Taqqu (1986) are satisfied without assuming Gaussianity (see Hosoya (1997, pp. 106)).

But, by assumption 4, for any compact set C there exists $\delta_1 < 1$ and $L_1 < \infty$ such that⁴

$$E^l [\|v_t\| \mid v_{t-1} = v] \leq \delta_1 \|v\| + L_1, \quad \forall v \notin C, \quad (29)$$

Then, (25) follows immediately from (28) and (29) since ε can be as small as desired. The above then establishes our condition (C.1).

To prove (C.2), it is necessary to show that the continuity and thrice-differentiability parts of (C.2) holds for the autocovariance function of u_t . The nonlinearity of (12) implies that the closed form of the autocovariance function is non trivial to obtain. However, it is possible to establish continuity and thrice-differentiability for the stationary density function of u_t , which is denoted by $h(u)$. It is known that $h(u)$ exists from the geometric ergodicity of u_t . For the case of $p = 1$, the function $h(u)$ is the solution to the following integral equation

$$h(y) = \int \varphi \left(y - \alpha_0 - \alpha_1 x - \sum_{j=1}^{\bar{q}} \beta_j \phi(x, \gamma) x \right) h(x) dx = \int g(y, x; \theta) h(x) dx \quad (30)$$

where $\varphi(\cdot)$ denotes the density of ϵ_t . The solution takes on a more complicated but similar form for $p > 1$. The above equation (30) is a special case of a nonlinear Volterra integral equation as discussed in Corduneanu (1991). It is easy to see that continuity and thrice-differentiability of ϕ and φ implies continuity and thrice-differentiability for h . To see this note that a standard method for proving existence theorems for this integral equation is via the method of successive approximations which simply involve taking successive integrals containing ϕ and φ (see, e.g., (Corduneanu, 1991, Sec. 1.3)). As a result the continuity and three times differentiability properties of ϕ and φ are inherited by $h(\cdot)$.

Having established the above for the probability density of u_t it is straightforward to see that the autocovariance function of u_t will be continuous and three times differentiable with respect to the parameters. Hence the result is proven.

Acknowledgements

This paper is prepared for presentation at the International Conference on Econometrics, Finance and Risk at the University of Western Australia, Perth on June 29 through July 1, 2006.

⁴In fact, any δ_1 which exceeds the maximum eigenvalue, in absolute value, of the companion matrix $\begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_p \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 \end{pmatrix}$ is admissible.

References

- ANDREWS, D. W. K., AND Y. SUN (2004): “Adaptive Local Polynomial Whittle Estimation of Long Range Dependence,” *Econometrica*, 72(2), 569–614.
- BAILLIE, R. T., AND T. BOLLERSLEV (1994): “The long memory of the forward premium,” *Journal of International Money and Finance*, 13, 565–571.
- BAILLIE, R. T., C. F. CHUNG, AND M. A. TIESLAU (1996): “Analysing Inflation by the Fractionally Integrated ARFIMA-GARCH Model,” *Journal of Applied Econometrics*, 11, 23–40.
- BAILLIE, R. T., AND G. KAPETANIOS (2006): “Testing for neglected nonlinearity in long memory models,” *Journal of Business and Economic Statistics*, forthcoming.
- BAILLIE, R. T., AND R. KILIC (2006): “Do asymmetric and nonlinear adjustments explain the forward premium anomaly?,” *Journal of International Money and Finance*, 22, 25–47.
- CHEUNG, Y. W. (1993): “Tests for Fractional Integration: A Monte Carlo Investigation,” *Journal of Time Series Analysis*, 14, 331–345.
- CHEUNG, Y. W., AND F. X. DIEBOLD (1994): “On Maximum Likelihood Estimation of the Differencing Parameter of Fractionally Integrated Noise with Unknown Mean,” *Journal of Econometrics*, 62, 301–316.
- CHEUNG, Y. W., AND K. S. LAI (2001): “Long Memory and Nonlinear Mean Reversion in Japanese Yen-Based Real Exchange Rates,” *Journal of International Money and Finance*, 20, 115–132.
- CIOCZEK-GEORGES, R., AND B. B. MANDELBROT (1995): “A class of micropulses and antipersistent fractional Brownian motion,” *Stochastic Processes and their Applications*, 60, 1–18.
- CORDUNEANU, C. (1991): *Integral Equations and Applications*. Cambridge University Press.
- DAHLHAUS, R. (1989): “Efficient Parameter Estimation for Self-Similar Processes,” *Annals of Statistics*, 17(4), 1749–1766.

- DALLA, V., L. GIRAITIS, AND J. HIDALGO (2005): “Consistent Estimation of the Long Memory Parameter for Nonlinear Time Series,” Unpublished Manuscript, London School of Economics.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*, Advanced Tests in Econometrics. Oxford University Press.
- DAVIDSON, J., AND R. DEJONG (2000): “The Functional Central Limit Theorem and Convergence to Stochastic Integrals II: Fractionally Integrated Processes,” *Econometric Theory*, 16, 621–642.
- DIEBOLD, F. X., S. HUSTED, AND M. RUSH (1991): “Real Exchange Rates Under the Gold Standard,” *Journal of Political Economy*, 99, 1252–1271.
- DIEBOLD, F. X., AND A. INOUE (2001): “Long Memory and Regime Switching,” *Journal of Econometrics*, 105, 131–159.
- FOX, R., AND M. S. TAQQU (1986): “Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Processes,” *Annals of Statistics*, 14, 517–532.
- GRANGER, C. W. J. (1980): “Long Memory Relationships and the Aggregation of Dynamic Models,” *Journal of Econometrics*, 14, 227–238.
- GRANGER, C. W. J., AND R. JOYEUX (1980): “An Introduction to Long Memory Time Series Models and Fractional Differencing,” *Journal of Time Series Analysis*, 1, 15–39.
- GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press.
- HANNAN, E. J., AND B. G. QUINN (1979): “The Determination of the Order of an Autoregression,” *Journal of the Royal Statistical Society (Series B)*, 41, 190–195.
- HENRY, M., AND P. M. ROBINSON (1996): “Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence,” in *Athens Conference on Applied Probability and Time Series, Volume II: Time Series in Memory of E. J. Hannan*, ed. by P. M. Robinson, and M. Rosenblatt. Springer-Verlag.
- HIDALGO, J. (2002): “Consistent order selection with strongly dependent data and its application to efficient estimation,” *Journal of Econometrics*, 110, 213–239.

- HOSKING, J. R. M. (1981): “Fractional Differencing,” *Biometrika*, 65, 165–176.
- HOSOYA, Y. (1997): “A Limit Theory for Long Range Dependence and Statistical Inference on Related Models,” *Annals of Statistics*, 25, 105–137.
- HURST, H. E. (1951): “Long Term Storage Capacity of Reservoirs,” *Transactions of the American Society of Civil Engineers*, 116, 770–799.
- KAPETANIOS, G. (2001): “Model Selection in Threshold Models,” *Journal of Time Series Analysis*, 22, 733–754.
- KAPETANIOS, G., Y. SHIN, AND A. SNELL (2003): “Testing for a Unit Root in the Non-linear STAR Framework,” *Journal of Econometrics*, 112(2), 359–379.
- LEE, T. H., H. WHITE, AND C. W. J. GRANGER (1993): “Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests,” *Journal of Econometrics*, 56, 269–290.
- LOTHIAN, J. R., AND M. P. TAYLOR (1996): “Real exchange rate behavior: the recent float from the perspective of the past two centuries,” *Journal of Political Economy*, 104, 488–509.
- MAYNARD, A., AND P. PHILLIPS (2001): “Rethinking an old empirical puzzle; the forward premium paradox,” *Journal of Applied Econometrics*, ??, 215–254.
- MICHAEL, P., A. R. NOBAY, AND D. A. PEEL (1997): “Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation,” *Journal of Political Economy*, 105, 862–879.
- PAPELL, D. H. (1997): “Searching for Stationarity: Purchasing Power Parity under the Current Float,” *Journal of International Economics*, 43, 313–332.
- PARK, J. Y. (2002): “An Invariance Principle for Sieve Bootstrap in Time Series,” *Econometric Theory*, 18, 469–490.
- PARKE, W. R. (1999): “What is fractional integration?,” *Review of Economics and Statistics*, 81, 632–638.
- POSKITT, D. (2005): “Autoregressive Approximation in Nonstandard Situations: The Non-Invertible and Fractionally Integrated Case,” *Monash University Working Paper*.

- ROBINSON, P. M. (1995): “Gaussian Semiparametric Estimation of Long Range Dependence,” *Annals of Statistics*, 23, 1630–1661.
- ROMANO, J. P., AND L. A. THOMBS (1996): “Inference for Autocorrelations under Weak Assumptions,” *Journal of the American Statistical Association*, 91, 590–600.
- SARANTIS, N. (1999): “Modelling Nonlinearities in Real Effective Exchange Rates,” *Journal of International Money and Finance*, 18, 27–45.
- SCHWARZ, G. (1978): “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461–464.
- SHIBATA, R. (1980): “Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process,” *Annals of Statistics*, 8, 147–164.
- SIN, C. Y., AND H. WHITE (1996): “Information Criteria for Selecting Possibly Misspecified Parametric Models,” *Journal of Econometrics*, 71(1–2), 207–225.
- SOWELL, F. (1992): “Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models,” *Journal of Econometrics*, 53, 165–188.
- STINCHCOMBE, M. B., AND H. WHITE (1998): “Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative,” *Econometric Theory*, 14, 295–325.
- TAQQU, M. S., AND V. TEVEROVSKY (1998): “On estimating the Intensity of Long Range Dependence in Finite and Infinite Variance Time Series,” in *A Practical Guide to Long Memory Processes*, pp. 177–217. Santa Barbara, CA.
- TERASVIRTA, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208–218.
- TWEEDIE, R. L. (1975): “Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space,” *Stochastic Processes Appl.*, 3, 385–403.
- VAN DIJK, D., P. H. FRANCES, AND R. PAAP (2002): “A Nonlinear Long Memory Model with Application to US Unemployment,” *Journal of Econometrics*, 110(2), 135–165.

Figure 1: Transition functions for forward premia and real exchange rate data.

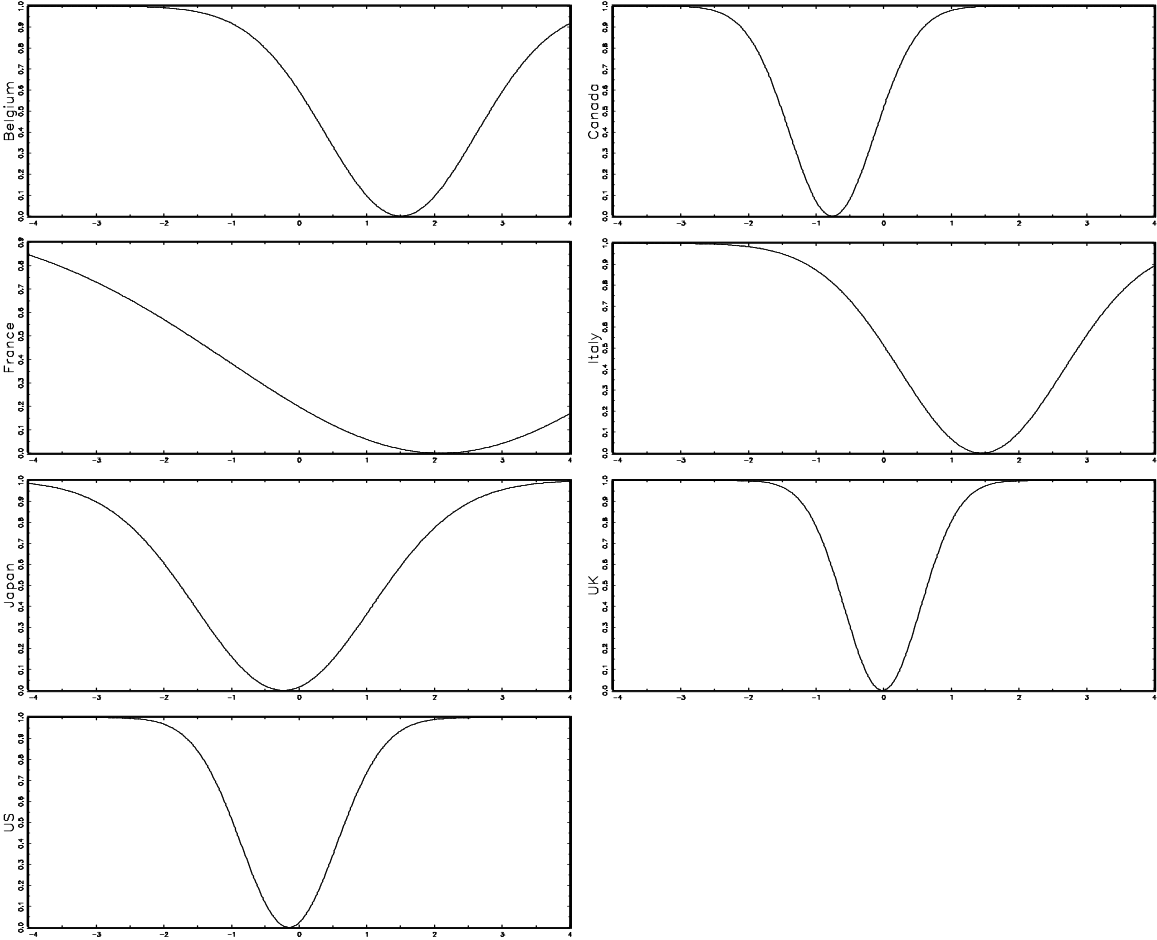


Table 1. Average of the One Step *MLE* for $d=0.2$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	-0.006	0.135	-0.003	-0.029	1.144	0.068	0.169
	600	0.012	0.166	-0.015	-0.087	1.124	-0.016	0.177
	1000	0.004	0.143	-0.001	-0.060	1.057	0.030	0.186
Exp. 2	400	-0.015	0.294	0.005	-0.085	1.063	-0.032	0.140
	600	0.000	0.351	-0.012	-0.152	0.988	-0.009	0.163
	1000	0.001	0.306	0.007	-0.131	0.889	0.026	0.179
Exp. 3	400	-0.001	0.632	0.010	-0.316	0.552	-0.014	0.174
	600	0.045	0.683	-0.072	-0.344	0.529	0.069	0.176
	1000	0.029	0.704	-0.054	-0.354	0.391	0.052	0.184
Exp. 4	400	0.012	0.130	-0.002	-0.050	1.273	0.005	0.168
	600	0.054	0.124	-0.066	-0.051	1.222	0.031	0.180
	1000	-0.003	0.117	0.020	-0.042	1.142	-0.022	0.186
Exp. 5	400	-0.015	0.304	0.026	-0.133	1.041	-0.036	0.157
	600	-0.048	0.273	0.053	-0.109	1.050	-0.017	0.167
	1000	-0.008	0.321	0.002	-0.161	0.994	0.014	0.181
Exp. 6	400	0.016	0.739	-0.023	-0.371	0.797	0.036	0.166
	600	0.028	0.741	-0.028	-0.379	0.713	0.042	0.174
	1000	0.003	0.759	-0.018	-0.392	0.600	-0.013	0.186
Exp. 7	400	0.006	0.175	-0.000	-0.080	1.458	-0.006	0.166
	600	-0.033	0.203	0.030	-0.120	1.520	-0.091	0.181
	1000	0.008	0.189	0.002	-0.112	1.412	0.021	0.186
Exp. 8	400	0.000	0.350	0.008	-0.172	1.485	0.025	0.156
	600	-0.022	0.357	-0.015	-0.181	1.488	0.011	0.173
	1000	-0.015	0.364	0.034	-0.195	1.476	-0.016	0.181
Exp. 9	400	-0.018	0.821	0.009	-0.428	1.372	0.057	0.160
	600	-0.018	0.822	0.006	-0.438	1.435	0.006	0.175
	1000	0.001	0.834	-0.002	-0.446	1.370	-0.019	0.185

Table 2. Average of the Two Step Estimator for $d=0.2$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.010	0.082	-0.030	0.019	1.233	0.072	0.181
	600	-0.024	0.156	0.014	-0.084	1.206	-0.032	0.190
	1000	0.016	0.134	-0.026	-0.055	1.128	0.042	0.194
Exp. 2	400	-0.007	0.200	0.005	-0.020	1.145	-0.040	0.191
	600	-0.027	0.285	0.030	-0.120	1.062	-0.017	0.196
	1000	-0.013	0.262	0.024	-0.095	0.947	-0.004	0.195
Exp. 3	400	0.027	0.510	-0.044	-0.238	0.587	-0.028	0.198
	600	0.049	0.563	-0.051	-0.264	0.575	0.056	0.210
	1000	0.018	0.614	-0.012	-0.291	0.377	0.053	0.202
Exp. 4	400	-0.006	0.107	0.001	-0.020	1.350	0.027	0.180
	600	0.030	0.097	-0.046	-0.023	1.237	0.024	0.199
	1000	-0.034	0.103	0.043	-0.026	1.225	-0.019	0.202
Exp. 5	400	-0.001	0.228	-0.011	-0.101	1.153	-0.048	0.206
	600	-0.032	0.213	0.041	-0.083	1.100	0.005	0.197
	1000	-0.024	0.272	0.020	-0.110	1.088	-0.001	0.196
Exp. 6	400	-0.001	0.609	0.003	-0.287	0.824	0.041	0.206
	600	0.009	0.624	-0.021	-0.312	0.739	0.029	0.205
	1000	-0.004	0.670	-0.018	-0.322	0.609	-0.010	0.200
Exp. 7	400	0.028	0.175	-0.010	-0.090	1.514	-0.012	0.189
	600	-0.023	0.158	0.016	-0.078	1.590	-0.106	0.188
	1000	0.032	0.165	-0.022	-0.085	1.494	0.038	0.199
Exp. 8	400	-0.001	0.288	-0.004	-0.136	1.507	0.082	0.183
	600	-0.031	0.301	-0.016	-0.152	1.522	0.025	0.193
	1000	0.002	0.328	0.008	-0.163	1.508	0.014	0.188
Exp. 9	400	-0.017	0.659	0.008	-0.305	1.432	0.020	0.186
	600	-0.017	0.688	0.017	-0.336	1.402	0.012	0.195
	1000	-0.002	0.740	0.001	-0.363	1.388	0.002	0.193

Table 3. Standard Deviation of the One Step *MLE* for $d=0.2$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.941	0.662	1.079	0.701	1.107	1.212	0.089
	600	0.758	0.566	0.900	0.601	1.126	1.130	0.069
	1000	0.625	0.495	0.780	0.503	1.105	1.057	0.047
Exp. 2	400	1.003	0.720	1.166	0.718	1.103	1.277	0.128
	600	0.748	0.623	0.963	0.648	1.064	1.197	0.098
	1000	0.717	0.466	0.932	0.526	0.963	1.241	0.064
Exp. 3	400	0.874	0.640	1.180	0.639	0.695	1.329	0.095
	600	0.718	0.538	0.969	0.544	0.682	1.211	0.078
	1000	0.555	0.356	0.870	0.432	0.484	1.074	0.059
Exp. 4	400	0.970	0.724	1.111	0.747	1.192	1.204	0.078
	600	0.824	0.586	0.947	0.604	1.183	1.177	0.064
	1000	0.725	0.541	0.841	0.573	1.095	1.139	0.045
Exp. 5	400	0.893	0.713	1.085	0.711	1.022	1.265	0.112
	600	0.766	0.635	0.928	0.611	1.056	1.156	0.087
	1000	0.656	0.464	0.785	0.494	0.981	1.107	0.064
Exp. 6	400	0.674	0.570	0.870	0.578	0.813	1.011	0.095
	600	0.588	0.482	0.824	0.485	0.721	0.921	0.076
	1000	0.345	0.318	0.557	0.335	0.508	0.753	0.056
Exp. 7	400	0.831	0.673	0.945	0.660	1.218	1.139	0.082
	600	0.728	0.591	0.839	0.609	1.211	1.111	0.059
	1000	0.589	0.469	0.705	0.486	1.150	1.004	0.046
Exp. 8	400	0.871	0.714	1.029	0.682	1.230	1.125	0.102
	600	0.781	0.629	0.871	0.602	1.162	1.012	0.078
	1000	0.558	0.487	0.708	0.480	1.172	0.953	0.051
Exp. 9	400	0.552	0.586	0.676	0.576	1.067	0.796	0.105
	600	0.459	0.485	0.577	0.469	1.074	0.738	0.088
	1000	0.259	0.338	0.333	0.324	0.933	0.565	0.063

Table 4. Standard Deviation of the Two Step Estimator for $d=0.2$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.973	0.728	1.115	0.720	1.206	1.332	0.192
	600	0.850	0.577	1.026	0.599	1.191	1.255	0.145
	1000	0.658	0.504	0.826	0.533	1.139	1.138	0.143
Exp. 2	400	1.003	0.762	1.176	0.744	1.142	1.477	0.189
	600	0.747	0.637	0.973	0.658	1.111	1.300	0.148
	1000	0.793	0.542	1.062	0.571	1.021	1.385	0.135
Exp. 3	400	0.946	0.677	1.198	0.620	0.762	1.475	0.190
	600	0.775	0.587	1.021	0.546	0.721	1.383	0.138
	1000	0.714	0.404	1.051	0.414	0.446	1.312	0.135
Exp. 4	400	0.990	0.795	1.161	0.768	1.244	1.357	0.188
	600	0.967	0.635	1.097	0.630	1.196	1.317	0.136
	1000	0.743	0.586	0.856	0.594	1.187	1.258	0.140
Exp. 5	400	0.976	0.769	1.171	0.784	1.129	1.431	0.186
	600	0.870	0.686	1.026	0.682	1.094	1.299	0.149
	1000	0.691	0.528	0.866	0.535	1.079	1.219	0.143
Exp. 6	400	0.792	0.604	0.931	0.613	0.888	1.203	0.188
	600	0.707	0.538	0.963	0.530	0.781	1.133	0.143
	1000	0.428	0.356	0.645	0.350	0.602	0.963	0.146
Exp. 7	400	0.859	0.801	0.968	0.737	1.246	1.323	0.180
	600	0.830	0.707	0.944	0.698	1.297	1.292	0.148
	1000	0.588	0.516	0.683	0.502	1.216	1.117	0.141
Exp. 8	400	0.890	0.793	1.011	0.739	1.249	1.325	0.186
	600	0.790	0.595	0.899	0.622	1.210	1.123	0.143
	1000	0.618	0.557	0.722	0.548	1.186	1.112	0.137
Exp. 9	400	0.677	0.651	0.768	0.629	1.133	1.060	0.182
	600	0.502	0.520	0.680	0.517	1.052	0.867	0.146
	1000	0.349	0.403	0.437	0.395	1.050	0.739	0.137

Table 5. Average of the One Step *MLE* for $d=0.4$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.017	0.167	-0.009	-0.067	1.131	0.035	0.357
	600	0.016	0.150	-0.039	-0.070	1.045	0.009	0.372
	1000	-0.022	0.143	0.023	-0.056	1.021	-0.008	0.388
Exp. 2	400	0.010	0.303	-0.027	-0.110	0.974	0.008	0.330
	600	-0.030	0.299	0.045	-0.113	0.943	-0.065	0.353
	1000	-0.029	0.318	0.032	-0.131	0.875	-0.057	0.381
Exp. 3	400	0.030	0.665	-0.050	-0.298	0.571	0.041	0.356
	600	0.027	0.688	-0.006	-0.338	0.489	0.030	0.366
	1000	0.053	0.676	-0.061	-0.354	0.390	0.092	0.384
Exp. 4	400	-0.007	0.163	-0.018	-0.083	1.233	0.054	0.361
	600	-0.017	0.135	0.013	-0.077	1.231	0.017	0.381
	1000	0.015	0.156	-0.003	-0.104	1.126	0.016	0.387
Exp. 5	400	-0.021	0.307	0.013	-0.136	1.114	-0.015	0.343
	600	-0.080	0.313	0.079	-0.147	1.020	-0.113	0.363
	1000	0.023	0.326	-0.028	-0.167	0.990	0.066	0.380
Exp. 6	400	-0.024	0.736	0.016	-0.343	0.787	-0.012	0.354
	600	-0.010	0.746	0.019	-0.373	0.740	0.011	0.370
	1000	-0.011	0.746	0.032	-0.374	0.612	-0.035	0.384
Exp. 7	400	0.006	0.160	0.013	-0.083	1.486	-0.059	0.366
	600	-0.033	0.184	0.056	-0.111	1.470	-0.040	0.379
	1000	0.027	0.147	-0.036	-0.081	1.485	0.006	0.390
Exp. 8	400	-0.010	0.363	-0.009	-0.182	1.432	0.017	0.349
	600	0.037	0.348	-0.066	-0.171	1.511	0.018	0.369
	1000	-0.026	0.353	0.023	-0.188	1.404	0.002	0.383
Exp. 9	400	-0.027	0.842	0.023	-0.448	1.405	-0.047	0.351
	600	0.010	0.821	-0.014	-0.422	1.398	0.041	0.359
	1000	0.015	0.843	-0.023	-0.445	1.366	0.009	0.377

Table 6. Average of the Two Step Estimator for $d=0.4$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	-0.040	0.125	0.022	-0.039	1.302	0.030	0.402
	600	0.034	0.098	-0.053	-0.030	1.156	0.008	0.406
	1000	-0.028	0.130	0.035	-0.042	1.155	-0.032	0.402
Exp. 2	400	-0.026	0.235	0.024	-0.086	1.058	0.035	0.389
	600	-0.042	0.226	0.043	-0.088	1.011	-0.084	0.405
	1000	-0.047	0.258	0.025	-0.097	0.977	-0.047	0.399
Exp. 3	400	-0.009	0.502	-0.000	-0.192	0.640	0.026	0.396
	600	-0.013	0.574	0.031	-0.281	0.531	-0.005	0.415
	1000	0.046	0.595	-0.057	-0.306	0.407	0.086	0.408
Exp. 4	400	0.032	0.090	-0.031	-0.030	1.339	0.063	0.391
	600	-0.021	0.131	0.012	-0.076	1.295	-0.021	0.402
	1000	-0.000	0.149	0.016	-0.095	1.186	0.026	0.397
Exp. 5	400	0.021	0.216	-0.017	-0.097	1.185	-0.021	0.411
	600	-0.060	0.264	0.067	-0.129	1.087	-0.104	0.400
	1000	0.024	0.260	-0.028	-0.110	1.049	0.064	0.404
Exp. 6	400	-0.051	0.582	0.040	-0.242	0.864	-0.007	0.403
	600	-0.003	0.609	-0.012	-0.292	0.741	0.043	0.412
	1000	-0.000	0.658	0.016	-0.331	0.618	-0.047	0.401
Exp. 7	400	0.015	0.104	-0.021	-0.054	1.543	-0.079	0.398
	600	-0.014	0.140	0.035	-0.086	1.504	-0.035	0.396
	1000	0.049	0.125	-0.041	-0.076	1.564	-0.007	0.401
Exp. 8	400	-0.011	0.258	0.006	-0.109	1.499	-0.006	0.399
	600	0.029	0.262	-0.034	-0.115	1.543	0.010	0.406
	1000	-0.014	0.295	-0.012	-0.147	1.440	0.002	0.410
Exp. 9	400	-0.040	0.653	0.045	-0.307	1.442	-0.039	0.405
	600	-0.008	0.689	0.006	-0.326	1.394	0.043	0.405
	1000	0.007	0.715	-0.004	-0.353	1.349	0.008	0.397

Table 7. Standard Deviation of the One Step *MLE* for $d=0.4$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.876	0.662	1.102	0.692	1.148	1.155	0.099
	600	0.778	0.607	0.944	0.663	1.101	1.161	0.067
	1000	0.624	0.499	0.791	0.538	1.073	1.105	0.045
Exp. 2	400	0.960	0.730	1.191	0.783	1.063	1.302	0.127
	600	0.871	0.615	1.085	0.663	1.037	1.276	0.100
	1000	0.724	0.508	0.872	0.523	0.932	1.218	0.060
Exp. 3	400	0.912	0.594	1.163	0.631	0.754	1.299	0.080
	600	0.728	0.534	0.962	0.490	0.638	1.196	0.070
	1000	0.699	0.433	0.931	0.413	0.501	1.135	0.059
Exp. 4	400	0.873	0.670	1.019	0.687	1.146	1.148	0.092
	600	0.808	0.706	0.985	0.703	1.167	1.156	0.058
	1000	0.664	0.560	0.797	0.575	1.102	1.099	0.047
Exp. 5	400	0.896	0.759	1.073	0.737	1.142	1.261	0.113
	600	0.746	0.560	0.915	0.593	1.019	1.146	0.085
	1000	0.600	0.452	0.779	0.478	0.972	1.091	0.058
Exp. 6	400	0.731	0.589	0.905	0.606	0.782	0.994	0.087
	600	0.573	0.507	0.770	0.516	0.743	0.915	0.074
	1000	0.400	0.356	0.585	0.344	0.531	0.765	0.056
Exp. 7	400	0.826	0.689	0.920	0.689	1.218	1.108	0.080
	600	0.708	0.608	0.840	0.596	1.185	1.109	0.064
	1000	0.585	0.488	0.715	0.486	1.230	1.016	0.044
Exp. 8	400	0.799	0.704	0.918	0.703	1.161	1.037	0.110
	600	0.692	0.632	0.823	0.635	1.196	1.034	0.078
	1000	0.560	0.468	0.725	0.458	1.097	0.941	0.053
Exp. 9	400	0.574	0.583	0.696	0.565	1.037	0.804	0.088
	600	0.508	0.509	0.592	0.507	1.025	0.733	0.084
	1000	0.274	0.336	0.420	0.323	0.940	0.547	0.062

Table 8. Standard Deviation of the Two Step Estimator for $d=0.4$

	T	α_0	α_1	β_0	β_1	γ	c	d
Exp. 1	400	0.980	0.754	1.171	0.757	1.248	1.366	0.185
	600	0.910	0.704	1.026	0.721	1.179	1.292	0.142
	1000	0.696	0.582	0.828	0.581	1.192	1.216	0.139
Exp. 2	400	0.972	0.786	1.237	0.775	1.137	1.468	0.192
	600	0.898	0.662	1.122	0.676	1.097	1.384	0.142
	1000	0.800	0.558	0.988	0.584	1.054	1.319	0.143
Exp. 3	400	1.008	0.666	1.269	0.685	0.850	1.490	0.188
	600	0.774	0.596	1.049	0.567	0.703	1.353	0.143
	1000	0.663	0.473	1.027	0.455	0.528	1.257	0.140
Exp. 4	400	0.990	0.740	1.174	0.748	1.222	1.288	0.185
	600	0.862	0.666	1.055	0.660	1.220	1.289	0.150
	1000	0.670	0.582	0.848	0.587	1.150	1.186	0.144
Exp. 5	400	0.988	0.783	1.120	0.809	1.168	1.418	0.186
	600	0.823	0.598	1.046	0.619	1.079	1.283	0.144
	1000	0.663	0.508	0.798	0.465	1.052	1.230	0.142
Exp. 6	400	0.796	0.657	1.005	0.642	0.904	1.213	0.185
	600	0.722	0.518	0.888	0.535	0.758	1.100	0.142
	1000	0.446	0.419	0.663	0.407	0.589	0.916	0.141
Exp. 7	400	0.890	0.806	1.022	0.803	1.297	1.297	0.186
	600	0.735	0.624	0.900	0.607	1.228	1.249	0.143
	1000	0.631	0.530	0.724	0.553	1.277	1.129	0.137
Exp. 8	400	0.888	0.804	0.998	0.765	1.205	1.250	0.190
	600	0.738	0.677	0.862	0.644	1.222	1.172	0.145
	1000	0.590	0.555	0.776	0.556	1.143	1.104	0.143
Exp. 9	400	0.697	0.686	0.855	0.649	1.130	1.085	0.183
	600	0.545	0.590	0.636	0.580	1.070	0.896	0.146
	1000	0.361	0.389	0.514	0.393	0.978	0.744	0.135

Table 9. Average and St. Deviation of the One Step *MLE* for $d=0.2$ and 5000 observations

Average							
	α_0	α_1	β_0	β_1	γ	c	d
Exp. 3	0.004	0.785	-0.036	-0.406	0.219	0.042	0.197
Exp. 6	0.000	0.803	-0.004	-0.408	0.431	0.003	0.199
Exp. 9	0.001	0.826	-0.002	-0.422	1.106	0.001	0.194
Standard Deviation							
Exp. 3	0.048	0.066	0.272	0.075	0.089	0.451	0.025
Exp. 6	0.024	0.055	0.092	0.049	0.130	0.219	0.026
Exp. 9	0.028	0.091	0.042	0.087	0.392	0.139	0.024

Table 10: Empirical Results

Table 10: Empirical Results							
	Forward Premium						RER
Country	Belgium	Canada	France	Italy	Japan	UK	US
ACF1	0.780	0.916	0.698	0.841	0.940	0.939	0.8909
ACF2	0.666	0.803	0.626	0.663	0.854	0.842	0.7695
ACF3	0.595	0.714	0.584	0.521	0.769	0.758	0.6643
ACF4	0.525	0.660	0.445	0.394	0.691	0.694	0.5799
LB	863.36	1987.29	720.12	688.19	1544.32	1666.69	1214.83
LW	0.489	0.619	0.734	0.649	0.619	0.558	0.554
Nonlinearity Tests							
TLG	0.000	0.133	0.012	0.000	0.041	0.111	0.022
ANN	0.000	0.124	0.006	0.000	0.038	0.100	0.027
Estimation of FI-ESTAR Model							
p	15	12	12	12	10	10	12
k	1	2	1	1	3	1	6
Linear AR Parameters							
α_0	2.0084 (0.8984)	-0.5103 (0.0140)	1.1464 (1.8215)	-0.0089 (0.4168)	-0.0143 (0.1457)	-0.2949 (0.0578)	-0.0002 (0.0458)
α_1	-0.5914 (0.4613)	-0.4114 (0.1578)	-0.3566 (0.4579)	1.0155 (0.3201)	0.7971 (0.4005)	0.7175 (0.4611)	0.8326 (0.1755)
α_2	-0.2014 (0.1234)	0.1390 (0.1290)	-0.0051 (0.1392)	-0.0598 (0.0783)	0.2258 (0.2012)	-1.3513 (0.1873)	-0.1358 (0.1264)
α_3	0.0784 (0.1426)	-0.0573 (0.0792)	0.1657 (0.3288)	0.1323 (0.0796)	0.2088 (0.2016)	0.3525 (0.2156)	0.2999 (0.1249)
α_4	0.1587 (0.1468)	-0.2338 (0.0806)	-0.0080 (0.4203)	-0.2253 (0.0781)	-0.0494 (0.0746)	-0.0722 (0.2212)	0.0027 (0.0734)
α_5	0.0867 (0.1424)	0.1056 (0.0821)	0.0863 (0.3360)	-0.0659 (0.0800)	-0.1294 (0.0723)	0.1940 (0.2199)	-0.3569 (0.1039)
α_6	-0.1329 (0.1545)	-0.2097 (0.0749)	0.0016 (0.2295)	-0.0276 (0.0773)	0.1484 (0.0708)	-0.2386 (0.2197)	0.1433 (0.1044)
α_7	-0.2982 (0.1515)	0.0801 (0.0794)	0.1191 (0.3660)	0.1806 (0.0762)	-0.2179 (0.0731)	-0.0913 (0.2211)	0.1014 (0.0784)
α_8	0.1043 (0.1441)	-0.0392 (0.0753)	-0.0813 (0.3123)	0.0132 (0.0768)	0.1532 (0.0722)	0.5569 (0.2103)	0.0808 (0.0767)
α_9	0.0702 (0.1524)	0.2174 (0.0737)	0.0736 (0.4015)	-0.1318 (0.0758)	0.2441 (0.0739)	-0.6044 (0.1633)	-0.1162 (0.0770)
α_{10}	0.1280 (0.1422)	-0.2166 (0.0734)	0.0886 (0.2948)	-0.0093 (0.0786)	-0.2747 (0.0579)	0.2162 (0.0789)	0.0308 (0.0753)
α_{11}	0.0249 (0.1502)	0.0248 (0.0739)	0.1962 (0.3285)	0.0121 (0.0801)			0.0175 (0.0747)
α_{12}	-0.0037 (0.1836)	-0.0767 (0.0708)	0.1651 (0.2953)	0.1586 (0.0643)			0.0412 (0.0609)
α_{13}	-0.3544 (0.1534)						
α_{14}	-0.0597 (0.1317)						
α_{15}	0.2155 (0.1444)						

Table 10 (cont.): Empirical Results							
	Forward Premium						RER
Country	Belgium	Canada	France	Italy	Japan	UK	US
Non-linear AR Parameters							
β_0	-3.3514 (0.8990)	0.9872 (0.0750)	-5.4067 (1.8141)	0.0569 (0.4331)	0.0747 (0.1756)	0.3045 (0.0587)	0.0588 (0.1680)
β_1	-0.0101 (0.4603)	0.5086 (0.1947)	-0.0124 (0.6346)	-0.5379 (0.3340)	0.0642 (0.3963)	1.3189 (0.4701)	-0.2149 (0.2602)
β_2		-0.5072 (0.2377)			-0.3227 (0.2195)		0.3413 (0.2707)
β_3					-0.5033 (0.2176)		-1.1910 (0.5498)
β_4							0.0054 (0.2788)
β_5							0.8325 (0.4075)
β_6							-0.6630 (0.3109)
θ	0.4004 (0.3036)	1.2650 (0.0728)	0.0506 (0.1607)	0.3431 (0.7259)	0.2986 (0.2524)	1.5778 (0.0177)	1.0091 (1.0580)
c	1.5006 (0.1188)	-0.7599 (0.0471)	2.0852 (0.4066)	1.4468 (0.4109)	-0.2351 (0.1056)	-0.0194 (0.1854)	-0.1530 (0.0941)
d	0.6837 (0.1013)	0.7556 (0.0959)	0.5777 (0.0472)	0.3148 (0.0633)	0.5273 (0.0278)	0.2938 (0.0275)	0.4032 (0.0389)
Maximum Eigenvalue of Lag Polynomial							
Linear	0.9764	1.0334	0.9401	0.9964	1.0955	1.1526	0.9715
Linear + Nonlinear	0.9762	0.9065	0.9391	0.9056	0.9897	0.9896	0.9313
LBR	11.21	22.520	10.89	14.38	18.60	25.69	14.76

Key: The standard errors of the parameter estimates are in parentheses below their corresponding parameter estimates; ACF_j denotes the autocorrelation coefficient of the original series at lag j , and LB is the Ljung-Box statistic of the first 20 of these autocorrelations. The symbol LW denotes the Local Whittle estimate of the long memory parameter. The symbols ANN and TLG denote the test statistics for nonlinearity, and LBR is the Ljung-Box statistic of the first 20 of the autocorrelations of the residuals from the estimated model.

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2006 Richard Baillie and George Kapetanios
All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**