Testing for Neglected Nonlinearity in Long Memory Models

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, London E1 4NS, U.K.

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

Abstract

This paper constructs tests for the presence of nonlinearity of unknown form in addition to a fractionally integrated, long memory component in a time series process. The tests are based on artificial neural network approximations and do not restrict the parametric form of the nonlinearity. Some theoretical results for the new tests are obtained and detailed simulation evidence is also presented on the power of the tests. The new methodology is then applied to a wide variety of economic and financial time series.

Key Words: Long Memory, Non-linearity, Artificial Neural Networks, Realized Volatility, Absolute Returns, Real Exchange Rates.

1 Introduction

A considerable amount of recent work in time series econometrics has focused on alternative representations compared with the conventional $I(0)$ and $I(1)$ paradigms. In particular, there have been substantial developments in the modeling of long memory processes, and also in the mainly unrelated area of modeling non-linearity. However, there has been relatively little consideration of the issue of combining, or distinguishing between these types of processes. Notable exceptions are Diebold and Inoue (2001) who show how a process with Markov switching regime changes can be mistaken for a long memory process. Also, Kapetanios and Shin (2002) suggested a formal test for distinguishing between non-stationary long memory and nonlinear geometrically ergodic processes in small samples; while van Dijk, Frances, and Paap (2002) considered a long memory and Exponential Smooth Transition Autoregressive ($ESTAR$) model to represent the US unemployment rate. While the first two articles are concerned about the possibility of confusing non-linearity and long memory, the third paper addresses the possibility that a process may exhibit both long memory dynamics and non-linearity in the short memory dynamics.

This paper focuses on the issue of providing a general, formal testing framework for non-linearity in a time series process which may include a long memory, fractionally integrated component. One motivation for the study is to provide a basis for determining whether an apparent long memory model requires the addition of nonlinear terms. An attractive feature of our procedure is that it does not require specification of the exact parametric form of non-linearity, since a neural network approximation is used, which is combined with the long memory component. Two classes of tests are considered; the first is based on artificial neural network approximations, while the second uses a Taylor series approximation. One test is based on a Wald testing procedure when the terms corresponding to the non linearity approximation are jointly estimated with the long memory parameter. The other tests are based on applying tests for non linearity to fractionally filtered series where a variety of estimates of the long memory parameter are implemented. The power performance of the test statistics are shown to depend on the order of the neural network approximations and also the number of lagged terms being included. The performance of the various test statistics are documented by means of an extensive simulation study with a variety of nonlinear data generating mechanisms. Some of the test statistics perform quite well and give rise to optimism that nonlinear effects can be distinguished *within* a long memory process. The results indicate the desirability of jointly modeling the nonlinear and long memory components of a time series. As noted by Granger and Teräsvirta (1993) the allowance for non-linearity can provide superior forecasts and be relevant for economic theory for short memory processes. The results in this paper indicate how these effects can be tested and possibly incorporated into long memory processes.

The power performance of the tests can be significantly affected in small samples by the use of relatively inefficient initial estimators of the long memory parameter. Hence, this study considers several estimators of the long memory parameter in both the time and frequency domains. The reported simulation evidence is generally favorable to the Local Whittle estimator and also to a time domain approximate MLE where the long memory parameter is estimated jointly with terms from an artificial neural network expansion. These estimators are generally found to be preferable to using the Fox-Taqqu estimator, although it should be noted that the Fox-Taqqu procedure is applied to the estimation of an $ARFIMA$ model, unlike the proposed artificial neural network time domain estimators. Overall, the analysis shows the desirability of taking non-linearity into account when estimating long memory components. The results also highlight the extent to which the Local Whittle and other estimators of the long memory parameter are adversely affected by certain types of non-linearity.

The paper also includes some applications of the above methodology to various economic and financial time series. The results indicate the widespread presence of both nonlinear and long memory components in many macroeconomic time series, including the monthly rate of inflation and real exchange rates. The application of the tests to different measures of volatility in exchange rate returns also yields interesting results. Daily absolute returns on seven major currencies vis a vis the US dollar are found to be well represented by pure long memory for only three series. A series of fifteen years of the daily logged Realized Volatility for the DM-$ appears to only possess marginal non-linearity in addition to long memory. However, the corresponding series for the Yen-$ and the Yen-DM is found to exhibit significant non-linearity.

The structure of the rest of the paper is as follows. Section 2 presents the theoretical framework. Section 3 discusses the various tests, and section 4 their implementation to the problem of testing for neglected non-linearity. Section 5 presents some detailed simulation evidence concerning the performance of the tests, while the next section discusses various empirical examples. Finally, there is short section giving some concluding remarks.

2 Nonlinear long memory models

Long memory, fractionally integrated processes were originally introduced by Granger and Joyeux (1980), Granger (1980) and Hosking (1981) to represent the slow hyperbolic rates of decay associated with the impulse response weights and autocorrelations of a series. See Beran (1994) and Baillie (1996) for detailed surveys of these models and the latter for discussion of the application to economics and finance. A univariate process with fractional integration in its conditional mean can be represented as

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

where L is the lag operator and where u_t is a short memory, $I(0)$ process. Then y_t is said to be a fractionally integrated process of order d, or $I(d)$. In this study an $I(0)$ process is defined according to Davidson and DeJong (2000), as a process whose partial sums converge weakly to Brownian motion. Hence, the parameter d represents the degree of "long memory" behavior for 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 an impulse response function which tends to zero for long horizons, which implies that shocks to the level of the series decay in the long term. If the short memory component can be represented by an $ARMA(p,q)$ process, then equation (1) becomes the $ARFIMA(p,d,q)$ model,

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t \tag{2}$$

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. The Wold decomposition, or infinite order moving average representation of this process is given by

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

while the infinite order autoregressive representation is given by

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

For large lags i, these coefficients decay at very slow hyperbolic rates of $\psi_i(d) \sim c_1 i^{d-1}$ and $\varpi_i(d) \sim c_2 i^{-d-1}$, 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 maybe a nonlinear process rather than a conventional pure $ARMA$

process. For example, the long memory model in (1) can be combined with a short memory $ESTAR$ process,

$$u_t = \alpha_0 + \sum_{i=1}^{p_l} \alpha_i u_{t-i} + \sum_{i=1}^{p_n} \beta_i \left[1 - exp(- \gamma_1 (u_{t-D} - \gamma_0)^2)\right] u_{t-i} + \epsilon_t \tag{5}$$

where D is the delay parameter. However, there is no requirement to restrict attention to this particular form of non-linearity and u_t can be modeled in terms of other nonlinear structures such as threshold autoregressions, other forms of smooth transition autoregressions or bilinear models.

It should be noted that while the theoretical properties of long memory models were originally derived for the conditional mean, recent work has found strong empirical evidence for the presence of long memory in transformations of absolute returns in equity and currency markets and also in realized volatility series; see Ding, Granger, and Engle (1993), Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003) and others. There has also been a corresponding literature on the development of long memory $ARCH$ models, see Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996), and long memory stochastic volatility models, see Breidt, Crato, and de Lima (1998). The methods developed in this paper can be directly applied to either the levels of an economic or financial time series data, or alternatively can be directly applied to absolute returns, or Realized Volatility, or any other metric of volatility in financial market data.

Since the long memory parameter is generally unknown, the practical implementation of the tests require replacing it with a consistent estimate. One approach advocated in this study, is to jointly estimate the long memory parameter d along with the parameters of the terms approximating the nonlinear structure. A further set of tests consist of applying tests for non-linearity to the series u_t, which is obtained by fractionally filtering the original series y_t using some estimate of d. It should be noted that the application of standard linear $ARFIMA$ model estimation is inappropriate in the presence of non-linearity, since it will generally result in an inconsistent estimate of d. Under the alternative hypothesis of neglected non-linearity, the construction of a test for non-linearity, using an $ARFIMA$ estimate of d, is likely to be less powerful than one based on the true value of d. It should be made clear that this issue is related primarily to the power of the test. Under the null hypothesis, d will be estimated consistently through an $ARFIMA$ model and therefore, the test will be correctly sized.

The estimates of the long memory parameter are based on a variety of methods including approximate MLE in the time domain, or alternatively a non-parametric approach in the frequency domain by local Whittle or related techniques. The different effects of these estimators are analyzed in the Monte Carlo study in section 5 of this paper. The tests developed in the paper may be viewed as a first step to a parametric analysis of the neglected non-linearity. Subsequently the investigator may use a model belonging to a class of non-linear models used to investigate weakly dependent stationary processes such as threshold autoregressive (TAR) or smooth transition autoregressive ($STAR$) models. It should also be noted that neural network specifications have been used to test for the presence of threshold type nonlinearity, see e.g., Lee, White, and Granger (1993).

The proposed solution for estimating d in this study is to consider a neural network type model for u_t. Once d is estimated, an estimate of u_t is obtained from fractionally filtering y_t. This estimate of u_t is then tested for non-linearity using standard neural network tests described in the next section.

3 Tests Based on Neural Networks and Taylor Series Approximations

This section considers three different, but related tests for the presence of non linearity within a long memory process. Hence, the conditional mean of the model can be expressed as

$$u_t = (1-L)^d y_t = \sum_{i=1}^{\infty} \pi_i(d) y_{t-i} \tag{6}$$

and

$$u_t = F(u_{t-1} \dots u_{t-p}) + \epsilon_t \tag{7}$$

The short memory part of the process is a possibly nonlinear autoregression involving the last p lags of the variable u_t. One possibility to be considered is the joint estimation of the long memory parameter d with a suitably parameterized non linear function. This one step approach requires the specification of a suitably general process to represent the non linearity and for joint estimation of d and the parameters associated with the non linear terms. The approach consists of testing that the coefficients of these nonlinear terms are zero. Clearly, an important issue which is addressed later, concerns the properties of estimates of d in the presence of short memory $I(0)$ non linear disturbances.

An alternative two step approach is to obtain an estimate of d, denoted by $\hat{d}$ and to then fractionally filter the y_t process by

$$\hat{u}_t = (1-L)^{\hat{d}} y_t \approx \sum_{i=1}^{t-1} \pi_i(\hat{d}) y_{t-i} \tag{8}$$

and to base tests for non linearity on the filtered series $\hat{u}_t$. One method for doing this is based on the Artificial Neural Network (ANN) test of Lee, White, and Granger (1993), where the null hypothesis of this test is that the conditional mean of u_t given lags of u_t is a linear function of the past information set, so that

$$P\left\{ E(u_t|u_{t-1}\dots u_{t-p}) = \delta_0 + \sum_{i=1}^{p} \delta_i u_{t-p} \right\} = 1 \tag{9}$$

The implementation of the test in this case requires the estimation of d from an auxiliary equation and the fractionally filtered series, $\hat{u}_t$. Note that the effect of truncation in (8) is asymptotically negligible. It should be noted that other tests for non-linearity have been proposed in previous literature. For example, Keenan (1985) and Tsay (1986) have suggested alternative tests based on Volterra expansions and are a different approach to the framework considered in our study. See Li (2004) and Granger and Teräsvirta (1993) for a review of alternative tests.

The implementation of the ANN testing framework, specifies that the nonlinear part of $F(.)$ in (7) is given by $\sum_{j=1}^{q} \beta_j \phi(\sum_{i=1}^{p} \gamma_{ij} \hat{u}_{t-i})$ where $\phi(\lambda)$ is the logistic function, given by $[1+\exp(-\lambda)]^{-1}$. As noted by Lee, White, and Granger (1993), this functional form can approximate arbitrarily well any continuous function.

The coefficients γ_{ij} are randomly generated from a uniform distribution over $[\gamma_l,\ \gamma_h]$. It should be noted that the use of random γ_{ij} has two purposes. First, it bypasses the need for computationally expensive estimation techniques and second, and most importantly, solves the identification problem for γ_{ij} since these parameters are not identified under the null hypothesis of linearity. For a given q, the constructed regressors $\phi(\sum_{i=1}^{p} \gamma_{ij} \hat{u}_{t-i})$, $j = 1, \dots, q$ may suffer from multicollinearity. Following the suggestion of Lee, White, and Granger (1993), the $\tilde{q}$ in this study is taken to be the largest principle components of the constructed regressors excluding the largest one be used as regressors in

$$\hat{u}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{u}_{t-i} + \sum_{j=1}^{\tilde{q}} \beta_j \tilde{\phi}_{j,t} + \epsilon_t \tag{10}$$

where $\tilde{\phi}_{j,t}$ denotes the $(j+1)$-th principal component. A standard LM test is then be performed and Lee, White, and Granger (1993) suggest constructing the test statistic as TR^2, where R^2 is the uncentred squared multiple correlation coefficient of a regression of $\hat{\varepsilon}_t$ on a constant, $\hat{u}_{t-i}$, $i = 1\dots,p$, $\tilde{\phi}_{j,t}$, $j = 1,\dots,\tilde{q}$, where $\hat{\varepsilon}_t$ is the residual of the regression of $\hat{u}_t$ on a constant and $\hat{u}_{t-i}$, $i = 1\dots,p$. Under the null hypothesis, this test statistic has an asymptotic $\chi^2_{\tilde{q}}$ distribution. Under the alternative hypothesis, this test is consistent as discussed in Stinchcombe and White (1998).

An alternative two step approach is to apply the logistic neural network test proposed by Teräsvirta, Lin, and Granger (1993) to the fractionally filtered series. This test approximates the logistic neural network by a Taylor series expansion and tests for the significance of the additional terms when they are subsequently substituted into the model for $\hat{u}_t$. See Blake and Kapetanios (2003) for an alternative interpretation of the logistic neural network test without long memory.

The appropriate order of terms will typically depend on the degree of non linearity in the data. The second order expansion is,

$$\hat{u}_t = \beta_0 + \sum_{i=1}^{p}\beta_i\hat{u}_{t-i} + \sum_{i=1}^{p}\gamma_{0,i,2}\hat{u}^2_{t-i} + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{1,i,j}\hat{u}_{t-i}\hat{u}_{t-j} + \epsilon_t \tag{11}$$

The third order expansion, which is recommended by Teräsvirta, Lin, and Granger (1993) is of the form

$$\hat{u}_t = \beta_0+\sum_{i=1}^{p}\beta_i\hat{u}_{t-i}+\sum_{j=2}^{3}\sum_{i=1}^{p}\gamma_{0,i,j}\hat{u}^j_{t-i}+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{1,i,j}\hat{u}_{t-i}\hat{u}_{t-j}+\sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{2,s,i,j}\hat{u}^{2-s}_{t-i}\hat{u}^{s+1}_{t-j}+\epsilon_t \tag{12}$$

In some of the work reported in this paper, the following fourth order Taylor series expansion is also used:

$$\hat{u}_t = \beta_0 + \sum_{i=1}^{p}\beta_i\hat{u}_{t-i} + \sum_{j=2}^{4}\sum_{i=1}^{p}\gamma_{0,i,j}\hat{u}^j_{t-i} + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{1,i,j}\hat{u}_{t-i}\hat{u}_{t-j} + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{2,i,j}\hat{u}^2_{t-i}\hat{u}^2_{t-j}+$$

$$\sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{3,s,i,j}\hat{u}^{2-s}_{t-i}\hat{u}^{s+1}_{t-j} + \sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p}\gamma_{4,s,i,j}\hat{u}^{3-2s}_{t-i}\hat{u}^{2s+1}_{t-j} + \epsilon_t \tag{13}$$

Clearly, these are all very general approximations with considerable numbers of terms and interactions. For the purposes of this study, it was decided to restrict the number of parameters in the third and fourth order Taylor series expansions, by only considering cross products and powers of up to two lags. Given this restriction, the null hypothesis corresponding to the absence of non linearity is equivalent to the γ coefficients being zero. This null is then tested by means of a Wald statistic and the regressions underlying the tests are denoted by the TLG_i regressions, $i = 2, 3, 4$.

4 Implementation of the tests

An important issue both in terms of the suggested modeling strategy and for the implementation of the tests, concerns the estimation of the long memory parameter d. As previously stated, estimation and testing can either be carried out jointly, or in two separate stages. One approach adopted in this paper is to estimate a model including the long memory parameter and a component based on the logistic neural network approximation. Assuming that such a model is correct under the alternative, then it is shown in the Appendix, that the approximate MLE of d, is consistent and has an asymptotic normal distribution. The numerical implementation of the approximate MLE is equivalent to nonlinear least squares in this case, and is relatively straightforward. However, the approximate MLE for a model including the long memory parameter and a parametric non linear component such as $ESTAR$ is relatively difficult and computationally intensive; e.g. see the discussion in Van Dijk, Terasvirta, and Frances (2002), and is not pursued in this paper. For this reason the recommendation in this study is that joint estimation be carried out through estimating the following models,

$$u_t(d) = \beta_0 + \sum_{i=1}^{p} \beta_i u_{t-i}(d) + \sum_{i=1}^{p} \gamma_{0,i,2} u_{t-i}^2(d) + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i}(d) u_{t-j}(d) + \epsilon_t \tag{14}$$

Or,

$$u_t(d) = \beta_0 + \sum_{i=1}^{p} \beta_i u_{t-i}(d) + \sum_{j=2}^{3} \sum_{i=1}^{p} \gamma_{0,i,j} u_{t-i}^j(d) + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i}(d) u_{t-j}(d) +$$

$$\sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{2,s,i,j} u_{t-i}^{2-s}(d) u_{t-j}^{s+1}(d) + \epsilon_t \tag{15}$$

Or,

$$u_t(d) = \beta_0 + \sum_{i=1}^{p} \beta_i u_{t-i}^j(d) + \sum_{j=2}^{4} \sum_{i=1}^{p} \gamma_{0,i,j} u_{t-j}^j(d) + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i}(d) u_{t-j}(d) +$$

$$\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{2,i,j} u_{t-i}^2(d) u_{t-j}^2(d) + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{3,s,i,j} u_{t-i}^{2-s}(d) u_{t-j}^{s+1}(d) + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{4,s,i,j} u_{t-i}^{3-2s}(d) u_{t-j}^{2s+1}(d) + \epsilon_t \tag{16}$$

and where $u_t(d) = y_t - \sum_{l=0}^{t-p} \pi_l(d) y_{t-l} \approx y_t - \sum_{l=0}^{\infty} \pi_l(d) y_{t-l} = (1-L)^d y_t$.

In each case the joint estimation and testing is carried out by the minimization of the sum of squared residuals $\sum_{t=1}^{T} \hat{\epsilon}_t$. It is convenient to present the regressions again, with $u_t(d)$ replacing $\hat{u}_t(d)$, to distinguish the joint estimation and testing case from the case where testing is carried out using a filtered series using some previously obtained estimate of d.

On assuming Gaussianity of the white noise process ϵ_t, estimation by minimizing the conditional sum of squared residuals is equivalent to approximate MLE in the time domain and is numerically straightforward. It is shown in the appendix that models implied by equations (14) through (16) 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. One issue concerns the effect of the presence of a non zero intercept β_0. The Fox and Taqqu (1986) study considered the $ARFIMA$ model with intercept,

$$(1-L)^d(y_t - \beta_0) = u_t, \; t = 1, \ldots, T \tag{17}$$

and showed that when the model is expressed in deviation form from its mean, the MLE of d and the $ARFIMA$ parameters will all be $T^{1/2}$ consistent and are asymptotically Normally distributed. The proof in Fox and Taqqu (1986) requires the assumption of Gaussian innovations, while Hosoya (1997) shows the result is still valid in the presence of non Gaussian innovations that satisfy mild mixing and moment conditions. Hence all tests based on parametric estimation, considered in this study will depend on $T^{1/2}$ consistent and asymptotically Normally distributed quantities, which implies the χ^2 distribution of the test statistic for linearity. This is summarized in the following theorem:

Theorem 1 *Under assumptions 1-3 of Appendix 1, the null hypothesis of linearity given by (9), a given lag order, p, and for* $d \leq 1/2$, *the asymptotic distribution of the parametric ANN and TLG tests is* χ^2_r *where r is the number of parameters being tested.*

The proof is given in the Appendix. A separate and essentially unrelated issue concerns the asymptotic behavior of the MLE of the intercept in $ARFIMA$ models. In general, the MLE of the intercept will converge at the slower rate of $T^{1/2-d}$; see Dahlhaus (1989). However, this is irrelevant for the purposes of this study, since the tests being considered are based on the structural parameters once the model has been estimated in deviation form around the mean. Cheung and Diebold (1994) discuss the properties of the estimation of the long memory parameter with the $ARFIMA(p, d, q)$ with intercept process, with the intercept replaced by sample median, or other estimators. They find these adjustments have little effect except in very small sample sizes.

As discussed above, an approximate MLE in the time domain of the unknown parameters can be achieved by means of the conditional sum of squares (CSS) method, which has previously been successfully applied to models 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 (1992a), 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).

The two step estimation and corresponding testing procedures are carried out by first obtaining an estimate of d and then applying the nonlinearity test to the fractionally filtered series $\hat{u}_t(d)$. The two step procedure can be implemented either parametrically, or semi-parametrically. The first parametric case considered uses the d estimates from the TLG_i regressions and applies the ANN test to the filtered series. The second parametric case is to use the methodology of Fox and Taqqu (1986)to obtain a frequency domain MLE of the linear $ARFIMA$ model, where the estimation is carried out over all frequencies. It should be noted that the Fox Taqqu estimator again assumes Gaussianity of the disturbances. For the semi-parametric case the Local Whittle (LW) estimator is implemented, and is chosen since it is probably the most widely used semi-parametric estimator of the parameters of an $ARFIMA$ model. Once a semi-parametric estimate of d has been obtained either the TLG or the ANN tests are applied.

It should be noted that estimation in either the time or the frequency domain by assuming a linear $ARFIMA$ model is problematic under the alternative hypothesis of non-linearity, and it is expected that the tests would accordingly suffer from low power. Clearly, it seems desirable to consider a model which approximates the nonlinear structure of the series when estimating d. Following the above discussion the TLG_i models for $i = 2, 3, 4$ are motivated as approximations to the nonlinear component of the processes and furthermore are straightforward to estimate jointly with the long memory parameter by approximate MLE in the time domain.

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.

Under the alternative hypothesis of nonlinearity, a formal power result can be established for the ANN test following the work of Stinchcombe and White (1998) and Bierens (1990). The TLG testing framework approximates a neural network but no formal results on its potential universal approximation properties seem available in the literature. The class of functions for the ANN test are defined as:

$$\mathcal{H}_{ANN} = \left\{ h(u_1, \ldots, u_p) = \sum_{i=1}^{m} \zeta_i \left(1 - e^{-\sum_{j=1}^{p} \gamma_{ij} u_j}\right)^{-1}, \quad \zeta_i, \gamma_{ij} \in \mathbb{R}, j = 1, \ldots, p, i = 1, \ldots, m \right\} \tag{18}$$

where $e_t = F(\ u_{t-1}, ..., u_{t-p}) + \epsilon_t$ by definition.

Theorem 2 *Let $F(\ u_{t-1}, ..., u_{t-p})$ be L_p-bounded for some $p \geq 1$. Then there exists a function $h \in \mathcal{H}_{ANN}$ such that $E(h(u_{t-1}, \ldots, u_{t-p})e_t) \neq 0$.*

In words, this theorem states that there exists some ANN specification that can capture the nonlinearity due to the presence of $F(.)$, which therefore implies that the ANN test is consistent. Again the proof of this theorem is given in the appendix.

In order to clarify the nature of the tests being proposed, it is useful to consider the $p = 1$ case for illustrative purposes. As previously explained, the Monte Carlo analysis considers both parametric and semi-parametric estimators for d. The approximate time domain MLE for d are obtained from the models

$$u_t(d) = \beta_0 + \beta_1 u_{t-1}(d) + \epsilon_t$$

$$u_t(d) = \beta_0 + \beta_1 u_{t-1}(d) + \gamma_{0,1,2} u_{t-1}^2(d) + \epsilon_t$$

$$u_t(d) = \beta_0 + \beta_1 u_{t-1}(d) + \gamma_{0,1,2} u_{t-1}^2(d) + \gamma_{0,1,3} u_{t-1}^3(d) + \epsilon_t$$

$$u_t(d) = \beta_0 + \beta_1 u_{t-1}(d) + \gamma_{0,1,2} u_{t-1}^2(d) + \gamma_{0,1,3} u_{t-1}^3(d) + \gamma_{0,1,4} u_{t-1}^4(d) + \epsilon_t$$

The joint estimation and testing procedure has a byproduct of an estimate of d and the estimated covariance matrix for the γ parameters, which provide a test for linearity. The alternative two step procedures are based on the filtered series $\hat{u}_t$, which is obtained from an estimate of d, which is either obtained from the above equations, or the Fox Taqqu frequency domain parametric estimator or from the semiparametric Local Whittle estimator. Then, the ANN non-linearity test consists of testing that $\beta_i = 0$, $i = 1, \ldots, \tilde{q}$ in the model

$$\hat{u}_t = \alpha_0 + \alpha_1 \hat{u}_{t-1} + \sum_{j=1}^{\tilde{q}} \beta_j \tilde{\phi}_{j,t} + \epsilon_t$$

for the various estimators of d. In the case of the Fox Taqqu estimator and the semi-parametric estimator of d, the TLG tests were implemented by testing that $\gamma_{0,1,j} = 0$, for $j = 2, 3, 4$ in the following regressions

$$\hat{u}_t = \beta_0 + \beta_1 \hat{u}_{t-1} + \gamma_{0,1,2} \hat{u}_{t-1}^2 + \epsilon_t$$

$$\hat{u}_t = \beta_0 + \beta_1 \hat{u}_{t-1} + \gamma_{0,1,2} \hat{u}_{t-1}^2 + \gamma_{0,1,3} \hat{u}_{t-1}^3 + \epsilon_t$$

$$\hat{u}_t = \beta_0 + \beta_1 \hat{u}_{t-1} + \gamma_{0,1,2} \hat{u}_{t-1}^2 + \gamma_{0,1,3} \hat{u}_{t-1}^3 + \gamma_{0,1,4} \hat{u}_{t-1}^4 + \epsilon_t$$

All the above procedures were implemented in the following simulation study to investigate their small sample properties.

5 Simulation Study

This section reports the results obtained from a Monte Carlo study to investigate the size and power properties of the proposed new tests. The simulation experiment considers neglected non-linearity of the $ESTAR$ form and is consistent with the type of non-linearity investigated by van Dijk, Frances, and Paap (2002) in their analysis of US unemployment data. Tables 1 through 6 detail the results of three sets of experiments concerning the size and power of the tests. In the set of Experiments A, the parameters are set as $d = 0.2$ and $p = 1$; and in the set of Experiments B, as $d = 0.2$ and $p = 2$. Finally, in the set of Experiments C, the parameters are set as $d = 0.4$ and $p = 1$. It should be noted that experiments B are of additional interest, since the lag order used to generate the data is 1 and therefore different from that used in the testing procedures.

The simulation then considers two size experiments and eight power experiments for every set of experiments, where the null hypothesis is an $ARFIMA$ model and the alternative nonlinear hypothesis is an $ARFIMA(1, d, 0)$-$ESTAR$ model with $d = 0.2$ and $d = 0.4$. In the case of size experiments the model is $ARFIMA(0, d, 0)$ for experiment 1 and $ARFIMA(1, d, 0)$ with AR coefficient set equal to 0.8 for experiment 2. Experiments 3 to 10 are power experiments. For Experiments 3 through 10 inclusive, the disturbance u_t follows an $ESTAR$ model. The precise specification of the $ESTAR$ models for each experiment, referring back to (5), are:

- Experiment 3 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1.5$ $\gamma_1 = 0.01$
- Experiment 4 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1$ $\gamma_1 = 0.01$
- Experiment 5 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1.5$ $\gamma_1 = 0.05$
- Experiment 6 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1$ $\gamma_1 = 0.05$
- Experiment 7 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1.5$ $\gamma_1 = 0.01$
- Experiment 8 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1$ $\gamma_1 = 0.01$
- Experiment 9 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1.5$ $\gamma_1 = 0.05$
- Experiment 10 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1$ $\gamma_1 = 0.05$

The chosen parameter values for the nonlinear specifications is similar to those in other Monte Carlo studies of ESTAR models such as, e.g., Kapetanios, Shin, and Snell (2003). All the experiments, or designs, represent geometrically ergodic processes for u_t. The last four experiments allow for the corridor regime of the nonlinear process, (i.e. the regime closer to the mean of the process), to be locally explosive as the polynomial of the autoregressive part of the specification at the corridor regime has a root which is inside the unit circle. Such processes have been found to be of particular use for modeling certain macroeconomic series, such as US GDP by Kapetanios (2003). 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 are also β-mixing and hence α-mixing by (Davidson, 1994, Ch. 14), with sufficiently rapidly decaying mixing coefficients. An important implication of this finding is that the processes are hence I(0).

The time domain parametric versions of the TLG testing methodology are one step procedures since the long memory parameter d is estimated simultaneously with the coefficients associated with the non linear terms. The TLG test statistics are denoted by $TLG^{ML}_{j,i}$, where the tests are based on the time domain MLE of d being used. Similarly, TLG^{FT} denotes the TLG statistic which is based on the frequency domain estimator of Fox and Taqqu (1986) of d, which then is used to obtain $\hat{u}_t(d)$, before computing the TLG^{FT} test statistic. Analogously, TLG^{LW} corresponds to the TLG test statistic using the Local Whittle estimator of d. The symbol i, for $i = 2, 3, 4$, in the parametric time domain TLG test refers to the order of the TLG model being used to estimate the d parameter. The subscript j refers to the order of the Taylor expansion used to test the null hypothesis of linearity. As mentioned above, all the nonlinear specifications used have a joint estimation and testing procedure, so that $j = i$. Further, for completeness we consider two tests where it is assumed that either d is known and used to obtain the filtered series $\hat{u}_t(d)$, or is estimated in the time domain via an $ARFIMA(p, d, 0)$ model. These two tests are respectively denoted by $TLG^{ML}_{j,t}$ (subscript t for true d) and $TLG^{ML}_{j,\ell}$ (subscript ℓ for the linear ARFIMA model used to estimate d). For these cases, $j = 3$, following Teräsvirta, Lin, and Granger (1993). A third order Taylor expansion is also used for the TLG^{FT} and TLG^{LW} tests. Analogous notations for the ANN tests denoted by ANN^{ML}_i for $i = 2, 3, 4$, ANN^{FT}, ANN^{LW}, ANN^{ML}_t and ANN^{ML}_ℓ are also used throughout.

The Local Whittle semi-parametric estimator for d is obtained by minimizing the objective function

$$\ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d}I(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\log(\omega_j) \tag{19}$$

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 bandwidth is set as $m = [T^{0.5}]$ throughout. 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 test statistics being analyzed and the estimates of d, the choice of $m = [T^{0.50}]$ was generally found to be preferred to the alternatives that were considered. Full details of the simulation 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 ANN tests require the choice of the number of principal components to be included and following Lee, White, and Granger (1993), the setting of $q = 10$ is selected. The simulation also imposes $\tilde{q} = 2$, $\gamma_h = 2$ and $\gamma_l = -2$. The error term ϵ_t is generated from a $NID(0, 1)$ process for all replications; the results are presented for samples of size $T = 200$ and $T = 400$. The number of replications is set to 10,000. The nominal significance level is set to 5%. Rejection probabilities, the average estimates of d and $RMSE$s for these estimates are reported in the tables. The results are quite revealing. First, it can be seen from Tables 2, 4 and 6, that all the time domain based estimates of d, perform relatively well in terms of bias and $RMSE$ for most experiments, especially for the third and fourth order expansions. It is interesting to note that these findings appear robust to all the stationary and invertible $ARFIMA(p, d, 0)$ data generating processes. However, the frequency domain Fox-Taqqu estimator can be seen to have substantial upward biases in a number of cases for all experiments as expected. The Fox-Taqqu estimator performs well for some cases for a sample size of $T = 400$, but the substantial upward bias is again evident for a sample size of $T = 200$.

A further interesting feature is the behavior of the semi-parametric Local Whittle estimator, which is expected to work well despite the presence of non-linearity in the short run dynamics. The estimator performs well for experiments 1 through 6. However, in experiments 7 through 10, where the nonlinear process is still geometrically ergodic and also $I(0)$ but highly persistent this estimator is severely upward biased. In contrast the time domain approximate MLE based on the various neural network specifications works extremely well.

However, the most interesting results concern the rejection frequencies of the tests. Inspection of the size properties of the tests from experiment 1 indicates that the overall performance is satisfactory. There is a slight degree of over rejection for the ANN_3^{ML}, ANN_4^{ML}, $TLG_{3,3}^{ML}$ and $TLG_{4,4}^{ML}$ tests, especially in the case of experiments B where $p = 2$, for samples of 200 observations which is significantly reduced for samples of 400 observations.

The power properties indicate that the tests based on estimates of d from using either the linear models or the semi-parametric estimator are uniformly less powerful that the tests based on estimates of d from models approximating the nonlinear models. In some cases the advantage can reach 20%. More interestingly the proposed tests have comparable power to tests based on the true value of d. It is intuitive to expect that tests using the true value of d possess the property of reaching the upper bound in terms of power given specific linearity tests. Therefore, it can be seen that tests based on an estimate of d obtained from a neural network type model, clearly has better power properties than tests based on an estimate of d obtained from $ARFIMA$ models.

A final point worth noting, is that increasing the lag order p from 1 to 2 seems to introduce a deterioration of the performance of the tests, in particular with respect to the rejection probabilities in the case of a linear $ARFIMA(1, 0.2, 0)$ model. This is to be expected given the fact that estimation of d becomes more difficult. Consequently, it suggests that choosing relatively low lag orders to detect nonlinearity may be more appropriate, even in the case where the true nonlinear process is of a higher order.

6 Empirical Applications

There is now a considerable body of literature reporting finding evidence for the presence of long memory in economic and financial time series. For example many studies have found evidence of long memory in real GNP, see Diebold and Rudebusch (1989) and Sowell (1992b). There is also a wealth of evidence for the presence of both long memory and non-linearity in inflation rates, which is one of the many macroeconomic series that could reasonably be expected to be regime dependent with consequent non linearities. An even more substantial body of literature now exists concerning the very strong evidence for the existence of long memory in various measures of volatility in financial markets, including absolute returns and realized volatility. This section considers some relevant applications of the previously developed techniques.

Throughout this section, the long memory parameter, d,is estimated within an $ARFIMA(p, d, 0)$ model by means of the approximate time domain MLE and by using a neural network approximation where a third order Taylor series expansion is used. In both cases the lag order, p, is chosen using the Bayesian information criterion with maximum lag order 4. The same maximum lag order is used throughout the empirical section for quarterly data whereas $p = 12$ is used for monthly data and $p = 4$ is used for daily data. Both the ANN and TLG tests are implemented using the estimate of d obtained both from the linear and nonlinear models. The ANN test follows Lee, White, and Granger (1993) by setting $q = 10$, $\tilde{q} = 2$, $\gamma_h = 2$ and $\gamma_l = -2$. The TLG test is implemented by means of a third order Taylor series expansion in all the cases below.

6.1 Real Exchange Rates

Many previous studies, e.g., Adler and Lehman (1983), 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 nonstationarity using standard unit root tests.

This study considers bilateral real exchange rates against the DM; so that for the *i-th* currency at time t, the real exchange rate is $(q_{i,t})$ as $q_{i,t} = s_{i,t} + p_{j,t} - p^*_{i,t}$, where j represents the German DM and $s_{i,t}$ is the corresponding nominal exchange rate in terms of (*i-th* currency per DM). Also, $p_{j,t}$ the price level in Germany and $p^*_{i,t}$ the price level of the *i-th* country. Thus, a rise in $q_{i,t}$ implies a real numeraire country currency appreciation against the *i-th* currency.

The price levels are all Consumer Price Indices and all variables are in natural logs, where the data are taken from the International Monetary Fund's *International Financial Statistics* in CD-ROM and are not seasonally adjusted. The data are quarterly, spanning 1973Q1-1998Q4 for seventeen countries, and details of the results are presented in Table 7. In general the TLG and ANN tests give rise to very similar conclusions, with evidence of non-linearity for nine countries at the .10 level and for four countries at the .05 level.

6.2 Monthly Inflation

The properties of monthly rates of inflation has been extensively investigated in the economics and econometrics literature. A central issue in much of this research, has been the degree of persistence of the shocks, and is related to the controversy concerning the possible existence of a unit root in inflation. In particular, Ball and Cecchetti (1990), Brunner and Hess (1998)), Barsky (1987) and Nelson and Schwert (1977), have argued that US inflation contains a unit root so that shocks to inflation are completely persistent. Alternatively, Hassler and Wolters (1995), Baillie, Chung, and Tieslau (1996), Baum, Barkoulas, and Caglayan (1999), and others, have all found evidence that inflation is fractionally integrated. The above articles provide quite consistent evidence across countries and time periods that inflation is fractionally integrated with a differencing parameter which is significantly different from zero and unity. However, Brunner and Hess (1998) use non-linear methods, particularly switching regime models to represent inflation. The time series properties of inflation are important from a finance perspective, since as originally noted by Rose (1988), most asset pricing models require ex post real rates of interest to be stationary. For a long memory inflation process, this implies the existence of a non-standard form of cointegration between inflation and nominal interest rates. The type and properties of non-linearities are also important in terms of the real rate of interest.

The inflation series analyzed in this paper are monthly inflation defined as $y_t = \Delta ln(CPI_t)$, where CPI_t is monthly Consumer Price Index. The countries considered are US, Canada, France, Germany, Italy, Japan, UK, Argentina and Brazil. The series extend from January 1957 to April 1990 and is the same data set as analyzed by Baillie, Chung, and Tieslau (1996). Results are presented in Table 8 and provide overwhelming evidence for neglected non-linearity especially when d is estimated using a TLG approximation. A notable result is obtained for France where a negative estimate for d obtained using an $ARFIMA(p, d, 0)$ is reversed when a neural network model is used.

6.3 Volatility of Exchange Rate Returns

Many different methods have been proposed for measuring volatility in asset market returns. In particular, Ding, Granger, and Engle (1993) noted the widespread finding that absolute returns in speculative auction markets have pronounced long memory features, resulting in the very slow hyperbolic decay of their autocorrelations. Ding, Granger, and Engle (1993) considered power transformations of absolute returns and the top panel of Table 9 reports the results of our testing procedures to the absolute returns from daily exchange rate returns for the seven freely floating exchange rates of Belgium, Canada, France, Germany, Italy, Japan and the UK vis a vis the US dollar, from March 1980 through June 1998; a total of 4,950 observations. Four of the absolute returns are found to have substantial non-linearity in addition to the long memory component. The presence of non linearity as well as long memory suggests the possibility of the formulation of comparable GARCH models, such as FIGARCH with non linear terms.

However, the concept of Realized Volatility (RV) has now generally become accepted as a more desirable measure of daily volatility when high frequency financial market data are available; see Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen, Bollerslev, Diebold, and Labys (2003). The RV series have the attraction of being a pure, model free measure of volatility and does not depend on the assumption of an $ARCH$, or stochastic volatility, or other model formulation. Andersen, Bollerslev, Diebold, and Labys (2003) and Andersen, Bollerslev, Diebold, and Labys (2001) show that under the assumption that logarithmic asset prices follow a univariate diffusion, the volatility is naturally measured by the associated quadratic variation process. The observed realized volatility is then calculated at the daily level by using the high frequency squared returns aggregated over the day. The lower panel of Table 9 of this study reports the results of the tests for non-linearity applied to the logarithm of the RV series of the DM-\$, Yen-\$ and DM-Yen; where the RV series are computed from 3,045 days from December 1, 1986 through June 30, 1999. Our estimates of

d from the time domain MLE are extremely close to those of Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen, Bollerslev, Diebold, and Labys (2003) with the estimates of d being 0.385, 0.433 and 0.440 for the DM-\$, Yen-\$ and DM-Yen respectively, compared with the respective estimates of 0.387, 0.413 and 0.430 reported by Andersen, Bollerslev, Diebold, and Labys (2003) from semi-parametric estimation. They also use a multivariate semi-parametric estimator for the combined series and report an estimate of d of 0.401. The tests for non-linearity are interesting. While the RV series are often considered to be virtually pure long memory, the findings in Table 9 of this paper indicate a failure to reject linearity at the .05 level for the DM-\$, but strong rejections at that level for the Yen-\$ and Yen-DM. It may be that the non-linearity is due to jumps and continuous price adjustments as suggested by Maheu and McCurdy (2002) in this context. Hence the approach of estimating switching regime models and/or threshold, or $STAR$ type models may give rise to further improvements in modeling and forecasting RV in currency markets.

7 Conclusion

This paper has suggested a new modeling strategy and testing framework for non-linearity in a time series process with a fractionally integrated component. Our suggested procedure does not require specification of the exact parametric form of non-linearity and is based on both artificial neural network and Taylor series approximations. It is found that using a linear model to estimate the long memory parameter prior to applying the tests for linearity leads to a significant loss of power. The paper suggests a procedure that estimates the long memory parameter within a model comprising an approximate neural network model, which is capable of picking up arbitrary forms of non-linearity. It is found that this strategy entails no loss of power compared to the case of the long memory parameter being known and is thus our recommended approach. The test statistics generally perform quite well and indicate that non-linear effects can be distinguished within a long memory process.

The performance is documented for different estimators of the long memory parameter. In the applications section, the results indicate the widespread presence of both non-linear and long memory components in monthly inflation rates and also in various definitions of real exchange rates. There also seems to be evidence of non-linearity in addition to long memory for daily absolute returns on some exchange rates against the US dollar; while there is also some evidence of non-linearity in the Realized Volatility for three major currencies.

Acknowledgements

The authors gratefully acknowledge the very helpful comments of an Associate Editor and two anonymous referees and also those of the seminar participants at the NSF-NBER Time Series Conference at Southern Methodist University, Dallas, TX, Iowa State University, Michigan State University and Wayne State University. The authors also acknowledge the assistance of Yoryos Chortareas, Florence Hubert and the Federal Reserve Bank of Cleveland for supplying some of the data used in this study.

Appendix

Proof of Theorem 1

The initial part of the theorem concentrates on the ANN test. The proof of the theorem requires showing that the MLE of the model

$$(1-L)^d y_t = u_t(d) \tag{20}$$

where $-0.5 < d < 0.5$,

$$u_t(d) = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}(d) + \sum_{j=1}^{\tilde{q}} \beta_j \phi(\sum_{i=1}^{p} \gamma_{ij} u_{t-i}(d)) + \epsilon_t \tag{21}$$

produces $\sqrt{T}$ consistent and asymptotically normal estimates for $\theta = (d, \alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_{\tilde{q}}$, $\gamma_{11}, \ldots, \gamma_{p\tilde{q}})'$ where $\phi(\lambda)$ is the logistic function, given by $[1+\exp(-\lambda)]^{-1}$. This will be shown to be true for a class of functions given by the assumption below:

Assumption 1 *$\phi(.)$ is a bounded function belonging to C^3.*

Assumption 2 *$E(\epsilon_t^s) < \infty$ for $s = 16$. The density of ϵ_t, denoted by φ, is positive everywhere, continuous and thrice differentiable.*

Assumption 3 *The roots of the lag polynomial $\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p$ are outside the unit circle.*

It should be noted that the need for 16-th moments in assumption 2 is motivated by the polynomial expansions used for the TLG test. For example, if only third order Taylor expansions were analyzed, then only $s = 12$ would be required in Assumption 2. 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. While the main emphasis in the literature has been for the pure $ARFIMA$ case, the purpose of the theorem in this paper is to extend the class of models to the long memory and non linear model considered in this study. The rest of the proof will verify the sufficient conditions that are relevant for this case. 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 stationary long memory $I(d)$ Gaussian process with $-05. < d < 0.5$.

Condition (C.2): The spectral density of $u_t(d^0)$, denoted $f_u(x, \theta)$ is continuous and belongs to C^3.

The spectral density of y_t, denoted $f_y(x, \theta)$ can be written as

$$f_y(x, \theta) = |e^{-ix} - 1|^{-2d} f_u(x, \theta) \tag{22}$$

which further illustrates the decomposition of the nonlinear short memory component and the long memory component. This emphasizes the similarity between the class of long memory nonlinear models analyzed in this paper compared with the standard $ARFIMA$ model. Equation (22) together with the above conditions (C.1) and (C.2) imply the conditions (B.1) to (B.4) of Fox and Taqqu (1986), which establish conditions for the second part of Theorem 3 of Fox and Taqqu (1986). This implies conditions (A.1)-(A.6) of the same paper and are necessary for the proof of convergence of the MLE.

It should be noted that under assumptions 1-3 it is shown below that the process u_t in (6) satisfies assumptions A (i), (ii) and (iii) of Hosoya (1997) which are a form of mixing condition on u_t. The non parametric nature of this assumption is sufficient to allow for the general forms of non linearity implicity considered earlier in this present paper. It is also important to note that Fox and Taqqu (1986), Dahlhaus (1989) and this present paper consider the process y_t to having been demeaned prior to the application of the MLE.

It is therefore necessary to now establish the conditions (C.1) and (C.2). The requirement for (C.1) is to show Gaussianity of y_t which in turn follows from the Gaussianity of u_t under the null. It can be noted that Gaussianity is only needed for asymptotic normality of the maximum likelihood estimator and not for consistency. For the rest of the proof of this theorem for the ANN test, the null hypothesis is not used. The Gaussianity assumption is in fact relaxed in Hosoya (1997) and replaced with a mixing assumption for u_t and a finite fourth moment assumption for y_t which is satisfied in our case for all models considered if Assumption 2 above holds. This follows from the finiteness of the fourth moment of u_t which follows from assumptions 1 and 2, (3) and square summability of $\psi_i(d)$ since by the Marcinkiewicz-Zygmund inequality (see, e.g., equation (25) of Park (2002)) for some constant c

$$E|y_t|^4 \leq c \left(\sum_{i=1}^{\infty} \psi_i(d)^2 \right) E|u_t|^4 < \infty, \tag{23}$$

It is then clear that y_t is an $I(d)$ process. This is immediate if it can be shown that u_t is a mixing process that satisfies a functional central limit theorem and is therefore $I(0)$ (see, e.g., Davidson and DeJong (2000)). The process u_t is β-mixing if it is geometrically ergodic, so that it is then necessary to show geometric ergodicity of u_t.

Geometric ergodicity for u_t is easily established under assumptions 1-3 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, u_t is geometrically ergodic if there exists constants $\delta < 1$, $B, L < \infty$, and a small set C such that

$$E\left[\|u_t\| \mid u_{t-1} = u\right] \leq \delta \|u\| + L, \quad \forall y \notin C, \tag{24}$$

$$E\left[\|u_t\| \mid u_{t-1} = u\right] \leq B, \quad \forall y \in C, \tag{25}$$

where $\|\cdot\|$ is the Euclidean norm. This easily follows if the autoregressive polynomial of the linear part of (21) has no roots outside the unit circle since the nonlinear part is bounded by assumption 1. Hence this establishes the condition (C.1).

To prove our condition (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 (21) implies that the closed form of the autocovariance function is not trivial to obtain. It is necessary to start by establishing 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 by geometric ergodicity of u_t, and $h(u)$ is the solution to the following integral equation

$$h(y) = \int \varphi\left(y - \alpha_0 + \sum_{i=1}^{p} \alpha_i x + \sum_{j=1}^{\tilde{q}} \beta_j \phi\left(\sum_{i=1}^{p} \gamma_{ij} x\right)\right) h(x)dx = \int g(y, x; \theta) h(x)dx \tag{26}$$

where $\varphi(.)$ denotes the density of ϵ_t. This is a special case of a nonlinear Volterra integral equation as discussed in, e.g., 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 thrice-differentiability properties of ϕ and φ are inherited by h.

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 thrice-differentiable with respect to the parameters. Hence the theorem is proved for the ANN test. It is important to note here that under the null hypothesis of linearity the parameters, γ_{ij}, are not identified. However, the β_j are identified each having a true value of zero. Hence, their estimates are consistent and asymptotically normal leading to a valid test statistic.

It is now possible to move on to the TLG test, and it is necessary to consider the second order expansion model given by

$$u_t(d) = \beta_0 + \sum_{i=1}^{p} \beta_i u_{t-i}(d) + \sum_{i=1}^{p} \gamma_{0,i,2} u_{t-i}^2(d) + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i}(d) u_{t-j}(d) + \epsilon_t \tag{27}$$

A similar analysis can be carried out for the models with higher order expansions. First note that it is not possible to establish geometric ergodicity, via the drift condition proposed by Tweedie (1975), for this model if any of $\gamma_{0,i,2} \neq 0$. As a result it is not possible to establish C.1 for the TLG models. However, under the null hypothesis the nonlinear part of the model is not active and as a result the process under the null is a linear $ARFIMA(p, d, 0)$ model which is stationary and Gaussian. A careful examination of the proof of Theorem 2 of Fox and Taqqu (1986) reveals that all relevant quantities such as the derivatives of the variance of the long memory process are evaluated for the true parameter value $\theta^0 = (d^0, \beta_0^0, \beta_1^0, ..., \beta_p^0, 0, ...0)'$, which implies that under the null, that the MLE of (27) are consistent and asymptotically $\sqrt{T}$-normal.

Proof of Theorem 2

First note that any element of $\mathcal{H}_{ANN}$ is L_q-bounded for all $q \geq 0$. Since $F(\ u_{t-1}, ..., u_{t-p})$ is L_p-bounded for some $p \geq 1$ it follows that it is possible to find some q such that $1/q + 1/p = 1$. The next stage is to apply the technical apparatus of Stinchcombe and White (1998). It is required to show that for some continuous non-polynomial function, G, with $\Sigma(G, \mathbb{R}^m) \subset \mathcal{H}_{ANN}$ where:

$$\Sigma(G, \mathbb{R}^m) = \left\{ h(x_1, \ldots, x_{k_x}) = \sum_{i=1}^{m} \zeta_i G \left(\sum_{j=1}^{k_x} a_{ij} x_j \right), \quad \zeta_i, a_{ij} \in \mathbb{R}, j = 1, \ldots, k_x, i = 1, \ldots, m \right\} \tag{28}$$

This is obvious from the definition of $\mathcal{H}_{ANN}$ and the required result follows from Lemma 3.5 of Stinchcombe and White (1998).

References

Adler, M., and B. Lehman (1983): "Deviations from purchasing power parity in the long run," *Journal of Finance*, 38, 1471–1487.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96(2), 42–55.

——— (2003): "Modelling and Forecasting Realised Volatility," *Econometrica*, 71(2), 579–625.

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. (1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

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.

Ball, L., and S. G. Cecchetti (1990): "Inflation and Uncertainty at Short and Long Horizons," *Brookings Papers on Economic Activity*, pp. 215–254.

Barsky, R. B. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3–24.

Baum, C. F., J. T. Barkoulas, and M. Caglayan (1999): "Persistence in International Inflation Rates," *Southern Economics Journal*, 65, 900–913.

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

Bierens, H. B. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

Blake, A. P., and G. Kapetanios (2003): "A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity," *Forthcoming in Econometrics Journal.*

Bollerslev, T., and H. O. Mikkelsen (1996): "Modelling and Pricing Long Memory in Stock Price Volatility," *Journal of Econometrics*, 73, 151–184.

Breidt, J. F., M. Crato, and P. J. F. de Lima (1998): "On the Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325–348.

BRUNNER, A. D., AND G. D. HESS (1998): "Are Higher Levels of Inflation Less Predictable? A State-Dependent Conditional Heteroskedasticity Approach," *Journal of Business and Economic Statistics*, 11, 187–197.

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.

CORDUNEANU, C. (1991): *Integral Equations and Applications*. Cambridge University Press.

DAHLHAUS, R. (1989): "Effcient 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.

DIEBOLD, F. X., AND G. D. RUDEBUSCH (1989): "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24, 189–209.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

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.

HASSLER, U., AND J. WOLTERS (1995): "Long Memory in Inflation Rates: International Evidence," *Journal of Business and Economic Statistics*, 13, 37–45.

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.

——— (2003): "Threshold Models for Trended Time Series," *Empirical Economics*, 28, 687–708.

KAPETANIOS, G., AND Y. SHIN (2002): "Testing for Nonstationary Long Memory Against Nonlinear Ergodic Models," Unpublished Manuscript, University of Edinburgh.

Kapetanios, G., Y. Shin, and A. Snell (2003): "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics*, 112(2), 359–379.

Keenan, D. M. (1985): "A Tukey Non Additivity Type test for Time Series Nonlinearity," *Biometrika*, 72, 39–44.

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.

Li, W. K. (2004): *Diagnostic Testing*. Chapman and Hall.

Maheu, J. M., and T. H. McCurdy (2002): "Nonlinear features of realized FX volatility," *Review of Economics and Statistics*, 84, 668–681.

Nelson, C. R., and G. W. Schwert (1977): "Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Rate of Interest is Constant," *American Economic Review*, 67, 478–486.

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.

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.

Rose, A. K. (1988): "Is the Real Interest Rate Stable?," *Journal of Finance*, 43, 1095–1112.

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. (1992a): "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.

Sowell, F. B. (1992b): "Modelling Long Run Behaviour with the Fractional ARIMA Model," *Journal of Monetary Economics*, 29, 277–302.

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.

Teräsvirta, T., C. F. Lin, and C. W. J. Granger (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 209–220.

Tsay, R. S. (1986): "Nonlinearity Tests for Time Series," *Biometrika*, 72, 461–466.

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.

Van Dijk, D., T. Terasvirta, and P. H. Frances (2002): "Smooth transition autoregressive models - a survey of recent developments," *Econometric Reviews*, 21, 1–47.

Table 1. Experiments A. Test Rejection Probabilities for Models with $p = 1$ and d=0.2

ANN tests							
	T=200						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.046	0.047	0.046	0.058	0.062	0.063	0.051
Exp 2	0.043	0.050	0.042	0.065	0.085	0.084	0.043
Exp 3	0.136	0.183	0.147	0.169	0.241	0.231	0.131
Exp 4	0.104	0.131	0.110	0.130	0.186	0.182	0.091
Exp 5	0.639	0.732	0.557	0.599	0.773	0.759	0.605
Exp 6	0.415	0.511	0.380	0.423	0.568	0.555	0.389
Exp 7	0.235	0.910	0.754	0.877	0.919	0.919	0.238
Exp 8	0.144	0.623	0.478	0.639	0.661	0.657	0.143
Exp 9	0.382	0.997	0.905	0.934	0.998	0.997	0.835
Exp 10	0.183	0.995	0.894	0.938	0.995	0.996	0.511
	T=400						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.050	0.049	0.050	0.052	0.056	0.057	0.053
Exp 2	0.041	0.044	0.038	0.054	0.072	0.071	0.040
Exp 3	0.285	0.337	0.281	0.301	0.388	0.383	0.264
Exp 4	0.167	0.200	0.177	0.190	0.258	0.249	0.159
Exp 5	0.919	0.960	0.857	0.871	0.965	0.965	0.886
Exp 6	0.730	0.799	0.681	0.695	0.825	0.818	0.677
Exp 7	0.342	0.996	0.859	0.918	0.993	0.994	0.333
Exp 8	0.266	0.921	0.846	0.915	0.925	0.922	0.251
Exp 9	0.459	0.998	0.965	0.975	0.998	0.999	0.963
Exp 10	0.219	1.000	0.969	0.980	0.999	0.999	0.767
TLG tests							
	T=200						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.047	0.047	0.045	0.064	0.064	0.077	0.052
Exp 2	0.041	0.048	0.042	0.052	0.089	0.086	0.044
Exp 3	0.138	0.182	0.149	0.049	0.243	0.215	0.132
Exp 4	0.106	0.130	0.111	0.049	0.188	0.166	0.096
Exp 5	0.647	0.738	0.566	0.139	0.778	0.731	0.610
Exp 6	0.417	0.515	0.385	0.095	0.577	0.530	0.396
Exp 7	0.232	0.912	0.756	0.789	0.919	0.896	0.239
Exp 8	0.142	0.621	0.477	0.706	0.658	0.617	0.142
Exp 9	0.382	1.000	0.913	0.268	1.000	1.000	0.839
Exp 10	0.183	0.998	0.904	0.328	0.998	0.997	0.513
	T=400						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.049	0.048	0.049	0.052	0.056	0.060	0.053
Exp 2	0.042	0.044	0.037	0.053	0.073	0.069	0.041
Exp 3	0.291	0.340	0.286	0.048	0.394	0.342	0.267
Exp 4	0.175	0.202	0.177	0.044	0.259	0.227	0.163
Exp 5	0.923	0.967	0.865	0.140	0.972	0.957	0.895
Exp 6	0.738	0.809	0.688	0.089	0.834	0.791	0.686
Exp 7	0.345	0.996	0.863	0.721	0.994	0.992	0.332
Exp 8	0.265	0.920	0.845	0.911	0.925	0.901	0.251
Exp 9	0.463	1.000	0.972	0.210	1.000	1.000	0.963
Exp 10	0.222	1.000	0.975	0.256	1.000	1.000	0.769

Notes: Numbers reported are estimated rejection probabilities in 10,000 replications.
For ANN_t^{ML} and $TLG_{3,t}^{ML}$ the true value of d has been used.
For ANN_2^{ML}, ANN_3^{ML}, ANN_4^{ML} and $TLG_{2,2}^{ML}$, $TLG_{3,3}^{ML}$, $TLG_{4,4}^{ML}$, d has been estimated using (14), (15) and (16) respectively.
For ANN_ℓ^{ML}, and $TLG_{3,\ell}^{ML}$, d has been estimated using an $ARFIMA(p, d, 0)$.
For ANN^{FT}, ANN^{LW} and TLG^{FT}, TLG^{LW}, d has been estimated using Fox-Taqqu and Local Whittle.

Table 2. Experiments A. Properties of the Estimated Long Memory Parameter for $p = 1$ and d=0.2

	Average estimated d over replications											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.196	0.129	0.127	0.126	0.128	0.188	0.200	0.165	0.165	0.165	0.165	0.189
Exp 2	0.353	0.174	0.187	0.190	0.200	0.269	0.269	0.180	0.188	0.191	0.196	0.227
Exp 3	0.271	0.126	0.139	0.143	0.152	0.227	0.220	0.141	0.150	0.160	0.165	0.211
Exp 4	0.286	0.142	0.157	0.158	0.169	0.231	0.234	0.157	0.165	0.166	0.171	0.211
Exp 5	0.196	0.046	0.069	0.134	0.139	0.202	0.193	0.087	0.101	0.166	0.166	0.201
Exp 6	0.212	0.068	0.089	0.126	0.133	0.205	0.190	0.098	0.111	0.152	0.154	0.192
Exp 7	1.035	0.047	0.135	0.178	0.181	1.088	0.970	0.023	0.095	0.193	0.194	1.070
Exp 8	1.076	0.083	0.157	0.159	0.165	1.097	0.988	0.100	0.160	0.180	0.182	1.079
Exp 9	0.803	0.021	0.081	0.186	0.189	0.415	0.811	0.007	0.035	0.188	0.190	0.317
Exp 10	0.947	0.028	0.084	0.185	0.190	0.657	0.935	0.015	0.049	0.186	0.189	0.518
	RMSE of estimate of d											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.089	0.153	0.155	0.162	0.162	0.233	0.056	0.086	0.087	0.088	0.089	0.184
Exp 2	0.265	0.132	0.136	0.143	0.148	0.248	0.171	0.107	0.110	0.115	0.120	0.186
Exp 3	0.195	0.157	0.156	0.149	0.151	0.242	0.133	0.133	0.134	0.119	0.122	0.184
Exp 4	0.203	0.148	0.151	0.149	0.151	0.239	0.137	0.124	0.125	0.118	0.122	0.189
Exp 5	0.179	0.244	0.232	0.149	0.150	0.237	0.140	0.204	0.194	0.092	0.095	0.186
Exp 6	0.178	0.213	0.204	0.157	0.158	0.243	0.137	0.178	0.173	0.111	0.114	0.187
Exp 7	0.836	0.234	0.159	0.093	0.098	0.895	0.771	0.258	0.194	0.059	0.055	0.877
Exp 8	0.877	0.204	0.136	0.134	0.132	0.900	0.789	0.189	0.119	0.082	0.083	0.882
Exp 9	0.635	0.242	0.215	0.047	0.047	0.328	0.628	0.240	0.215	0.029	0.029	0.229
Exp 10	0.752	0.214	0.178	0.051	0.052	0.516	0.737	0.213	0.187	0.032	0.030	0.375

Notes: FT: Fox-Taqqu Frequency Domain ARFIMA Estimator, ML_ℓ: ARFIMA Time Domain approximate MLE, ML_i, $i = 2, 3, 4$: Time Domain Approximate MLE of (14), (15) and (16) respectively,
LW: Local Whittle Semiparametric Frequency Domain Estimator.

Table 3. Experiments B. Test Rejection Probabilities for Models with $p = 2$ and d=0.2

ANN tests							
	T=200						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.048	0.051	0.045	0.073	0.075	0.069	0.049
Exp 2	0.050	0.047	0.043	0.083	0.103	0.084	0.049
Exp 3	0.060	0.070	0.053	0.097	0.121	0.101	0.066
Exp 4	0.053	0.062	0.050	0.095	0.112	0.099	0.055
Exp 5	0.144	0.161	0.123	0.175	0.187	0.182	0.147
Exp 6	0.103	0.129	0.091	0.138	0.167	0.146	0.112
Exp 7	0.151	0.743	0.595	0.713	0.756	0.743	0.150
Exp 8	0.116	0.582	0.477	0.587	0.600	0.585	0.104
Exp 9	0.306	0.519	0.368	0.448	0.526	0.521	0.348
Exp 10	0.155	0.584	0.424	0.530	0.594	0.600	0.226
	T=400						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.055	0.053	0.056	0.064	0.063	0.068	0.050
Exp 2	0.048	0.054	0.049	0.083	0.103	0.085	0.049
Exp 3	0.084	0.095	0.067	0.111	0.134	0.127	0.084
Exp 4	0.071	0.077	0.060	0.103	0.129	0.117	0.074
Exp 5	0.215	0.241	0.189	0.233	0.247	0.255	0.219
Exp 6	0.155	0.170	0.138	0.169	0.198	0.189	0.158
Exp 7	0.266	0.896	0.728	0.828	0.896	0.886	0.227
Exp 8	0.195	0.757	0.669	0.764	0.755	0.757	0.173
Exp 9	0.413	0.610	0.498	0.535	0.606	0.605	0.504
Exp 10	0.183	0.683	0.585	0.619	0.693	0.684	0.382
TLG tests							
	T=200						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.045	0.041	0.043	0.076	0.076	0.086	0.047
Exp 2	0.051	0.055	0.051	0.096	0.138	0.121	0.049
Exp 3	0.097	0.124	0.094	0.091	0.227	0.201	0.106
Exp 4	0.076	0.089	0.074	0.085	0.183	0.165	0.081
Exp 5	0.487	0.549	0.450	0.135	0.634	0.551	0.481
Exp 6	0.286	0.330	0.255	0.106	0.447	0.372	0.281
Exp 7	0.147	0.868	0.736	0.744	0.889	0.849	0.145
Exp 8	0.100	0.601	0.512	0.674	0.657	0.594	0.091
Exp 9	0.651	0.996	0.927	0.286	0.996	0.990	0.792
Exp 10	0.262	0.985	0.846	0.329	0.989	0.976	0.408
	T=400						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.051	0.055	0.051	0.062	0.068	0.071	0.053
Exp 2	0.051	0.052	0.053	0.090	0.134	0.120	0.053
Exp 3	0.180	0.212	0.166	0.078	0.305	0.265	0.187
Exp 4	0.121	0.139	0.108	0.081	0.240	0.208	0.127
Exp 5	0.845	0.877	0.806	0.140	0.896	0.846	0.824
Exp 6	0.568	0.626	0.524	0.097	0.689	0.598	0.554
Exp 7	0.297	0.985	0.932	0.802	0.987	0.977	0.254
Exp 8	0.185	0.808	0.745	0.878	0.830	0.758	0.166
Exp 9	0.823	1.000	0.994	0.275	1.000	1.000	0.965
Exp 10	0.291	1.000	0.974	0.305	1.000	1.000	0.690
See notes in Table 1.							

Table 4. Experiments B. Properties of the Estimated Long Memory Parameter for $p = 2$ and d=0.2

	Average estimated d over replications											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.194	0.083	0.089	0.087	0.101	0.196	0.197	0.139	0.137	0.136	0.137	0.188
Exp 2	0.252	-0.017	0.045	0.070	0.115	0.271	0.208	0.031	0.060	0.068	0.099	0.238
Exp 3	0.185	-0.051	0.016	0.061	0.102	0.240	0.174	0.004	0.034	0.081	0.096	0.217
Exp 4	0.199	-0.044	0.024	0.070	0.111	0.246	0.179	0.000	0.033	0.071	0.093	0.218
Exp 5	0.150	-0.025	0.023	0.102	0.116	0.208	0.174	0.037	0.055	0.135	0.134	0.198
Exp 6	0.156	-0.039	0.012	0.083	0.105	0.216	0.168	0.021	0.048	0.116	0.122	0.205
Exp 7	1.014	0.053	0.103	0.181	0.184	1.096	0.961	0.035	0.089	0.188	0.192	1.066
Exp 8	1.058	0.068	0.121	0.161	0.176	1.118	1.002	0.097	0.142	0.172	0.178	1.080
Exp 9	0.497	0.048	0.116	0.189	0.187	0.407	0.403	0.056	0.095	0.194	0.195	0.309
Exp 10	0.800	0.066	0.130	0.188	0.189	0.653	0.799	0.053	0.095	0.193	0.193	0.518
	RMSE of estimate of d											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.133	0.235	0.223	0.235	0.228	0.229	0.083	0.143	0.143	0.152	0.157	0.187
Exp 2	0.271	0.339	0.302	0.300	0.278	0.249	0.202	0.287	0.268	0.279	0.260	0.190
Exp 3	0.273	0.365	0.317	0.285	0.260	0.235	0.207	0.311	0.286	0.244	0.238	0.184
Exp 4	0.270	0.362	0.314	0.286	0.263	0.238	0.208	0.316	0.287	0.259	0.248	0.184
Exp 5	0.258	0.348	0.308	0.226	0.221	0.236	0.177	0.287	0.266	0.160	0.168	0.182
Exp 6	0.267	0.358	0.317	0.248	0.236	0.236	0.198	0.302	0.276	0.194	0.193	0.185
Exp 7	0.815	0.237	0.225	0.102	0.100	0.905	0.762	0.257	0.222	0.075	0.068	0.872
Exp 8	0.860	0.231	0.195	0.140	0.127	0.924	0.805	0.205	0.154	0.102	0.105	0.883
Exp 9	0.428	0.250	0.218	0.068	0.081	0.324	0.367	0.210	0.193	0.036	0.039	0.223
Exp 10	0.644	0.224	0.197	0.074	0.082	0.513	0.639	0.202	0.177	0.035	0.039	0.373

See notes in Table 2.

Table 5. Experiments C. Test Rejection Probabilities for Models with $p = 1$ and d=0.4

ANN tests							
	T=200						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.044	0.041	0.042	0.052	0.067	0.069	0.049
Exp 2	0.040	0.048	0.043	0.058	0.081	0.079	0.039
Exp 3	0.134	0.185	0.152	0.171	0.232	0.229	0.137
Exp 4	0.089	0.131	0.108	0.124	0.179	0.176	0.097
Exp 5	0.626	0.734	0.545	0.593	0.769	0.758	0.602
Exp 6	0.405	0.500	0.382	0.412	0.553	0.541	0.383
Exp 7	0.233	0.913	0.740	0.843	0.891	0.890	0.243
Exp 8	0.156	0.633	0.493	0.597	0.614	0.613	0.160
Exp 9	0.361	0.997	0.885	0.919	0.998	0.997	0.831
Exp 10	0.170	0.996	0.884	0.925	0.994	0.995	0.479
	T=400						
Exp	ANN^{FT}	ANN_t^{ML}	ANN_ℓ^{ML}	ANN_2^{ML}	ANN_3^{ML}	ANN_4^{ML}	ANN^{LW}
Exp 1	0.048	0.050	0.051	0.054	0.059	0.059	0.052
Exp 2	0.041	0.045	0.044	0.050	0.068	0.067	0.038
Exp 3	0.278	0.331	0.286	0.299	0.380	0.372	0.263
Exp 4	0.169	0.203	0.171	0.189	0.248	0.249	0.154
Exp 5	0.916	0.955	0.840	0.856	0.955	0.953	0.874
Exp 6	0.727	0.803	0.680	0.702	0.824	0.823	0.695
Exp 7	0.358	0.998	0.820	0.886	0.986	0.988	0.348
Exp 8	0.299	0.921	0.830	0.872	0.888	0.888	0.292
Exp 9	0.463	0.999	0.955	0.965	0.999	1.000	0.964
Exp 10	0.224	0.999	0.961	0.976	0.998	0.998	0.776
TLG tests							
	T=200						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.043	0.042	0.042	0.063	0.068	0.078	0.050
Exp 2	0.040	0.048	0.042	0.052	0.082	0.082	0.038
Exp 3	0.133	0.187	0.152	0.050	0.238	0.215	0.138
Exp 4	0.091	0.132	0.107	0.042	0.180	0.166	0.096
Exp 5	0.634	0.742	0.553	0.130	0.775	0.731	0.603
Exp 6	0.408	0.507	0.387	0.084	0.560	0.507	0.384
Exp 7	0.235	0.912	0.742	0.742	0.890	0.868	0.244
Exp 8	0.155	0.633	0.493	0.660	0.613	0.578	0.160
Exp 9	0.362	1.000	0.896	0.269	0.999	0.999	0.836
Exp 10	0.171	0.998	0.892	0.322	0.997	0.997	0.481
	T=400						
Exp	TLG^{FT}	$TLG_{3,t}^{ML}$	$TLG_{3,\ell}^{ML}$	$TLG_{2,2}^{ML}$	$TLG_{3,3}^{ML}$	$TLG_{4,4}^{ML}$	TLG^{LW}
Exp 1	0.048	0.050	0.049	0.059	0.058	0.058	0.053
Exp 2	0.041	0.044	0.044	0.050	0.068	0.066	0.037
Exp 3	0.284	0.336	0.290	0.044	0.387	0.338	0.268
Exp 4	0.172	0.207	0.177	0.039	0.253	0.211	0.159
Exp 5	0.923	0.961	0.846	0.140	0.962	0.945	0.881
Exp 6	0.735	0.812	0.693	0.088	0.833	0.786	0.702
Exp 7	0.361	0.998	0.826	0.672	0.986	0.986	0.350
Exp 8	0.298	0.921	0.831	0.875	0.887	0.859	0.293
Exp 9	0.466	1.000	0.961	0.213	1.000	1.000	0.966
Exp 10	0.228	1.000	0.966	0.265	1.000	1.000	0.778
See notes in Table 1.							

Table 6. Experiments C. Properties of the Estimated Long Memory Parameter for $p = 1$ and d=0.4

	Average estimated d over replications											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.409	0.312	0.311	0.309	0.310	0.401	0.409	0.363	0.362	0.362	0.362	0.400
Exp 2	0.592	0.380	0.393	0.393	0.400	0.475	0.484	0.385	0.391	0.393	0.398	0.436
Exp 3	0.497	0.330	0.341	0.344	0.353	0.429	0.433	0.343	0.348	0.355	0.358	0.411
Exp 4	0.515	0.353	0.362	0.362	0.370	0.433	0.441	0.360	0.367	0.366	0.370	0.410
Exp 5	0.412	0.243	0.269	0.332	0.337	0.398	0.403	0.279	0.294	0.360	0.361	0.398
Exp 6	0.435	0.280	0.295	0.328	0.335	0.405	0.410	0.298	0.308	0.353	0.355	0.406
Exp 7	1.165	0.307	0.373	0.419	0.419	1.189	1.120	0.248	0.318	0.420	0.419	1.191
Exp 8	1.160	0.397	0.428	0.434	0.434	1.145	1.102	0.390	0.420	0.438	0.438	1.126
Exp 9	1.015	0.253	0.305	0.389	0.393	0.621	1.012	0.225	0.256	0.388	0.390	0.524
Exp 10	1.135	0.252	0.303	0.388	0.393	0.859	1.126	0.227	0.261	0.386	0.389	0.718
	RMSE of estimate of d											
	T=200						T=400					
Exp	FT	ML_ℓ	ML_2	ML_3	ML_4	LW	FT	ML_ℓ	ML_2	ML_3	ML_4	LW
Exp 1	0.086	0.188	0.188	0.193	0.194	0.235	0.057	0.090	0.094	0.095	0.095	0.185
Exp 2	0.313	0.159	0.164	0.168	0.172	0.252	0.196	0.121	0.123	0.127	0.130	0.191
Exp 3	0.223	0.177	0.176	0.165	0.168	0.239	0.149	0.143	0.143	0.128	0.130	0.187
Exp 4	0.235	0.164	0.166	0.162	0.165	0.243	0.153	0.135	0.137	0.130	0.132	0.185
Exp 5	0.185	0.256	0.240	0.155	0.155	0.240	0.141	0.210	0.200	0.096	0.099	0.189
Exp 6	0.189	0.218	0.211	0.162	0.165	0.237	0.142	0.184	0.179	0.114	0.116	0.184
Exp 7	0.767	0.245	0.196	0.139	0.140	0.800	0.722	0.268	0.220	0.111	0.106	0.801
Exp 8	0.761	0.211	0.180	0.172	0.174	0.749	0.703	0.203	0.169	0.144	0.144	0.733
Exp 9	0.638	0.267	0.239	0.056	0.058	0.328	0.626	0.254	0.235	0.032	0.032	0.231
Exp 10	0.740	0.236	0.203	0.062	0.064	0.513	0.728	0.223	0.199	0.034	0.034	0.371

See notes in Table 2.

Table 7. Tests for Non-Linearity on DM-$ Real Exchange Rates.

Country	ANN_3^{ML}	$TLG_{3,3}^{ML}$	d_{ML_3}	ANN_ℓ^{ML}	$TLG_{3,\ell}^{ML}$	d_{ML_ℓ}
Austria	0.287	0.384	0.441	0.411	0.436	0.275
Canada	0.349	0.375	0.758	0.622	0.651	0.570
Denmark	0.316	0.337	0.119	0.357	0.365	0.023
France	0.239	0.237	0.053	0.214	0.236	0.055
Japan	0.661	0.682	0.205	0.725	0.742	0.289
Korea	0.074	0.072	0.431	0.464	0.855	0.262
Mexico	0.041	0.044	0.639	0.769	0.759	0.158
Malaysia	0.076	0.082	0.321	0.186	0.187	0.474
New Zealand	0.071	0.074	0.669	0.417	0.439	0.419
Norway	0.052	0.074	0.623	0.819	0.568	0.307
Portugal	0.050	0.048	0.583	0.289	0.219	0.390
S. Africa	0.079	0.106	0.497	0.090	0.091	0.402
Spain	0.022	0.022	0.703	0.225	0.234	0.315
Sweden	0.027	0.028	0.217	0.027	0.027	0.215
Switzerland	0.056	0.061	0.391	0.132	0.137	0.267
UK	0.612	0.639	0.670	0.775	0.864	-0.303
US	0.582	0.600	0.264	0.527	0.542	0.302
No. of Rejections	9	9		2	2	

Notes: ANN_ℓ^{ML}, ANN_3^{ML}, $TLG_{3,\ell}^{ML}$ and $TLG_{3,3}^{ML}$ are nonlinearity test probability values. For ANN_ℓ^{ML}, ANN_3^{ML} and $TLG_{3,\ell}^{ML}$, $TLG_{3,3}^{ML}$, d has been estimated using an $ARFIMA(p,d,0)$ model and (15) respectively. The third column presents an estimate of d, denoted d_{ML_3}, using the TLG approximation whereas the sixth column presents an estimate of d, denoted d_{ML_ℓ} using an $ARFIMA(p,d,0)$ model. Number of rejections reported at 10% significance level.

Table 8. Tests for Non-Linearity on Inflation rates.

Country	ANN_3^{ML}	$TLG_{3,3}^{ML}$	d_{ML_3}	ANN_ℓ^{ML}	$TLG_{3,\ell}^{ML}$	d_{ML_ℓ}
US	0.009	0.009	0.508	0.012	0.011	0.487
Canada	0.072	0.070	0.368	0.781	0.165	0.550
France	0.007	0.007	0.610	0.617	0.006	-0.217
Germany	0.433	0.438	0.292	0.441	0.438	0.284
Italy	0.025	0.031	0.314	0.145	0.158	0.422
Japan	0.002	0.003	0.170	0.967	0.077	0.367
UK	0.416	0.014	0.225	0.519	0.172	0.353
Argentina	0.000	0.000	0.224	0.005	0.000	0.270
Brazil	0.040	0.000	0.142	0.001	0.001	0.373
No. of Rejections	7	8		3	5	

See notes in Table 7.

Table 9. Tests for Non-Linearity on the Volatility of Exchange Rate Returns.

Daily Absolute Returns						
Country	ANN_3^{ML}	$TLG_{3,3}^{ML}$	d_{ML_3}	ANN_ℓ^{ML}	$TLG_{3,\ell}^{ML}$	d_{ML_ℓ}
Belgium	0.117	0.119	0.158	0.000	0.000	0.223
Canada	0.151	0.152	0.212	0.203	0.204	0.223
France	0.030	0.029	0.176	0.035	0.001	0.228
Germany	0.005	0.005	0.167	0.000	0.000	0.227
Italy	0.000	0.000	0.223	0.108	0.079	0.224
Japan	0.000	0.000	0.196	0.000	0.000	0.215
UK	0.216	0.226	0.208	0.263	0.231	0.211
No. of Rejections	4	4		4	5	
Log Realized Volatility of Returns						
Exch. Rate	ANN_3^{ML}	$TLG_{3,3}^{ML}$	d_{ML_3}	ANN_ℓ^{ML}	$TLG_{3,\ell}^{ML}$	d_{ML_ℓ}
DM-$	0.084	0.084	0.385	0.877	0.111	0.458
Yen-$	0.039	0.039	0.433	0.043	0.043	0.442
Yen-DM	0.012	0.012	0.440	0.507	0.161	0.250
No. of Rejections	3	3		1	1	
See notes in Table 7.						