# Semi Parametric Estimation of Long Memory: Comparisons and Some Attractive Alternatives

Richard T. Baillie*
Departments of Economics and Finance, Michigan State University
Department of Economics, Queen Mary University of London

George Kapetanios
Department of Economics, Queen Mary University of London

August 4, 2009

**Abstract**

The semi parametric Local Whittle ($LW$) estimator of the long memory parameter in a univariate time series is shown to perform poorly in the presence of persistent short memory. The $LW$ estimator also has poor properties for some important practical situations including the estimation of short run parameters and Impulse Response Weights ($IRWs$). These problems are shown to exist even when optimal bandwidths, or local polynomial Whittle methods are used. The paper suggests the use of high order autoregressions to approximate long memory processes from which estimated $IRWs$ can be derived. The autoregressive approximations are also shown to be relevant and perform well for some non-stationary processes.

Key Words: Long Memory, Local Whittle estimator, Non-stationarity, Autoregressive approximations.

JEL Codes: C22, C12.

## 1 Introduction

The work of Granger and Joyeux (1980), Granger (1980) and Hosking (1981) has been very influential for research on long memory time series with hyperbolically decaying autocorrelations. These models have proved extremely relevant for many time series in finance and macroeconomics. However, and perhaps surprisingly, one of the most heavily researched topics in this area, continues to be the issue of semi parametric estimation ($SPE$) of the long memory parameter in the conditional mean of a univariate time series. The Local Whittle ($LW$) estimator

*An earlier version of this paper was presented at the conference in honor of Peter Phillips'sixtieth birthday at Singapore Management University. The authors gratefully acknowledge very helpful comments from the conference participants at the Singapore Management University. Helpful comments were also received from Tim Bollerslev, Luidas Giraitis, Hira Koul, Peter Phillips, Peter Schmidt and Andrew Tremayne. Any remaining errors are our own.

appears to be the most widely used $SPE$ and its properties have been derived by Robinson (1995), Dalla, Giraitis, and Hidalgo (2005) and others.[1]

This paper is concerned with issues that arise in the practical investigation of a time series with long memory characteristics. In general, a researcher will not only be interested in estimating the order of fractional integration of a time series; but typically will also be concerned with (i) constructing a model involving both short and long memory components, (ii) estimating Impulse Response Weights ($IRWs$) for a variety of lags, and (iii) making predictions over short and long horizons. The investigation of the time series will often be an initial stage in a wider context of multivariate modelling.

Our study provides evidence on the performance of the $LW$ estimator of the long memory parameter for different bandwidths and analyzes the effects of using the $LW$ estimator of the long memory parameter to obtain estimates of short run parameters and $IRWs$. The estimation from using $LW$ is found to be generally poor regardless of bandwidth and also compares unfavorably with a benchmark $MLE$ alternative and also with high order autoregressions, that can be interpreted as semiparametric approximations to the unknown full model. The paper also shows that the $AR$ approximation has desirable properties, can be used over non stationary regions of the parameter space and can be particularly useful for $IRW$ analysis. The general findings suggest a more limited role for the $LW$ methodology.

## 2 Preliminaries

A univariate time series process, $y_t$, is said to be fractionally integrated of order $d$, or $I(d)$ if

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

where $L$ is the lag operator and $u_t$ is a short memory, $I(0)$, process. See Granger and Joyeux (1980), Granger (1980) and Hosking (1981) for further definitions and derivations for these processes. Note that the $I(0)$ process is defined as having partial sums that converge weakly to Brownian motion, while $d$ represents the degree of long memory, or persistence in the series. For $-0.5 < d < 0.5$ the process is stationary and invertible; while for $0.5 \leq d < 1$, the process does not have a finite variance, but still has a cumulative impulse response function with a finite limit. The Wold decomposition, or infinite order moving average representation of this process, for $-0.5 < d < 0.5$, is given by

$$y_t = \sum_{i=0}^{\infty} \psi_k \epsilon_{t-k} \tag{2}$$

where $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E(\epsilon_t \epsilon_s) = 0$, $s \neq t$. For $d > 0.5$, this representation is meaningful as long as the process is initiated at a given point in the past. Then, the representation is

[1] The work of Phillips (2007) and Phillips and Shimotsu (2006) is noteworthy in providing extensions and improvements of the $SPEs$ in a number of directions.

given by $y_t = \sum_{i=0}^{t} \psi_k \epsilon_{t-k}$ as discussed in, e.g., Section 2 of Phillips and Shimotsu (2005). For large lags $k$, these coefficients decay at the very slow hyperbolic rates of $\psi_k \sim c_1 k^{d-1}$ and similarly the infinite autoregressive representation coefficients decay at the rate of $c_2 k^{-d-1}$and autocorrelation coefficients at the rate of $c_3 k^{2d-1}$, where $c_1$, $c_2$ and $c_3$ are constants. If the short memory component is represented as a stationary and invertible $ARMA(p,q)$ process, then equation (1) becomes the well known $ARFIMA(p,d,q)$ model,

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t \tag{3}$$

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.

# 3 Parameter Estimation Issues

The $LW$ estimator of $d$, is denoted by $\hat{d}_{LW}$ and is obtained by minimizing the objective function

$$\ln\left[\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d} I(\omega_j)\right] - \frac{2d}{m}\sum_{j=1}^{m}\ln(\omega_j) \tag{4}$$

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 $m$ is the bandwidth. For the $LW$ estimator of $d$, it is known that, for linear processes, $m^{1/2}\left(\hat{d}_{LW} - d_0\right) \to N\{0, 1/4\}$ where $d_0$ denotes the true value of $d$. It is important to note that $m \leq T^{4/5}$, and $m$ is generally chosen in the range of $T^{1/2} \leq m \leq T^{4/5}$. In the usual case of ignorance of the short run dynamics, the bandwidth is generally selected in an ad hoc way and a popular choice is $m = T^{0.5}$. However, when there is substantial persistence in the short run dynamics, the value of $m$ should potentially be reduced so that more weight is placed on ordinates of the periodogram associated with the low frequency components. On denoting the spectral density function (s.d.f.) of $y_t$ as $f(\omega)$, and the s.d.f. of $u_t$ is $f^*(\omega)$, then $f(\omega) = |1-\exp(i\omega)|^{-2d} f^*(\omega)$. The s.d.f. of $y_t$ can be approximated as $\omega \to 0+$, by $f(\omega) = \mathcal{L}(1/\omega)\omega^{-2d}\{1 + c\omega^\beta + o(\omega^\beta)\}$ where $\mathcal{L}(1/\omega)$ is a slowly varying function with $0 < c < \infty$, usually chosen as one, and $\beta \in (0, 2]$. Henry (2001) has found the optimal bandwidth $m^*_{LW}$ for the $LW$ estimator to be $m^*_{LW} = \left(\frac{3}{4\pi}\right)^{4/5}\left|\tau^* + \frac{d}{12}\right|^{-2/5} T^{4/5}$ where $\tau^* = \left[\frac{f^{*''}(0)}{2f^*(0)}\right]_{\omega=0}$ and $\tau^*$ has the interpretation of representing the degree of smoothness of the spectral density of the short memory component $u_t$, as the frequency approaches zero.[2]

A further method proposed by Andrews and Sun (2004) is the Local Polynomial Whittle, or $LPW$ method which approximates the logarithm of the spectral density of the short memory

[2]Henry (2001) considers iterating between successive choices of $\hat{d}_{LW}$ and $m^*_{LW}$. However, in this study the optimal bandwidth is used to find $m^*_{LW}$ given knowledge of the underlying short memory process to ensure the maximum advantage of the $LWTSE$ methodology.

component by a polynomial. This leads to the $\hat{d}_{LPW}$ estimator of $d$ which has a reduced asymptotic bias, but higher variance. All the simulations involving $\hat{d}_{LPW}$, in this paper, use the first order approximation as in Nielsen and Frederiksen (2004).

As noted in the introduction, an investigator generally wishes to go beyond simply estimating the long memory parameter and will typically want to estimate other characteristics of the time series; such as its short run dynamics and $IRWs$. If $d$ is known, then the observed $y_t$ series can be fractionally filtered to obtain $u_t = y_t - \sum_{l=1}^{t-p} \pi_l(d) y_{t-l}$ where $(1-L)^d y_t = y_t - \sum_{l=1}^{\infty} \pi_l(d) y_{t-l}$, and $\pi_l(d)$ are the coefficients of the infinite AR representation of $y_t$ in terms of $u_t$, so that $\pi_l(d) = \Gamma(l-d)\Gamma(-d)^{-1}\Gamma(l+1)$. In practice, $d$ is unknown and can be replaced by the $LW$ estimate, $\hat{d}_{LW}$. Then, the Feasible Fractionally Filtered ($FFF$) series based on observable quantities is

$$\hat{u}_t = y_t - \sum_{l=1}^{t-p} \hat{\pi}_l(\hat{d}_{LW}) y_{t-l} \tag{5}$$

where $\hat{\pi}_l(\hat{d}_{LW}) = \Gamma(l-\hat{d}_{LW})\Gamma(-\hat{d}_{LW})^{-1}\Gamma(l+1)$. For concreteness, this paper focuses on the estimation of a univariate $ARFIMA(p,d,q)$ process. The complete parameter vector is denoted by $\vartheta = (d, \beta)'$, where the $(p+q)$ $ARMA$ parameters are in the vector $\beta = (\phi_1, ..\phi_{\mathbf{p}}, \theta_1, ..., \theta_q)'$. The true parameter values are denoted as $\beta_0(d_0)$, and the $LW$ two step estimator ($LWTSE$ ) of $\beta$, based on the feasible fractionally filtered series are $\hat{\beta}_{LWTSE}(\hat{d}_{LW})$. Then the $ARMA(p,q)$ parameters of the original $ARFIMA(p,d,q)$ process in equation (2) are estimated by minimizing the conditional sum of squares, $CSS$, conditional on $\hat{d}_{LW}$. The following result provides consistency and a rate of convergence for the two step estimator of the $ARFIMA(p,d,q)$ model.

**Theorem 1** *Let $y_t$ be given by the $ARFIMA(p,d,q)$ process in (2), 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. Let the disturbance $\epsilon_t$ be i.i.d.$(0,\sigma^2)$, with $E(\epsilon_t^4) < \infty$. Then, $\hat{\beta}_{LWTSE}(\hat{d}_{LW}) - \beta_0(d_0) = O_p(m^{-1/2})$.*

The theorem is proven in the appendix, and is useful for examining the properties of the $LWTSE$. A valid benchmark for comparison is ML Estimation ($MLE$). On assuming the innovations in the $ARFIMA(p,d,q)$ process in equation (3) to be $NID(0,\sigma^2)$, then the Gaussian likelihood is numerically maximized with respect to the complete vector of parameters $\vartheta$. Under these conditions, Fox and Taqqu (1986) have shown the asymptotic distribution of the $MLE$ to be $T^{1/2}\left(\hat{\vartheta} - \vartheta_0\right) \to N\{0, A(\vartheta_0)^{-1}\}$, where $\vartheta_0$ denotes the true value of the vector of parameters, and where $A(\vartheta_0)$ is the information matrix.[3] Hosoya (1997) relaxed the Gaussianity assumption, made by Fox and Taqqu (1986), and shown that $T^{1/2}\left(\hat{\vartheta} - \vartheta_0\right) \to N\{0, A(\vartheta_0)^{-1}B(\vartheta_0)A(\vartheta_0)^{-1}\}$

[3]Dahlhaus (1989) has shown that the same asymptotic properties hold also when the unconditional mean is not known and has to be estimated. In this case the $MLE$ of the parameter estimates will be $T^{1/2}$ consistent. The inclusion of an intercept parameter, will result in a $T^{1/2-d}$ consistent estimator.

| Table 1. Results for estimated $d$ | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bias | | | | | | | | | | | | | | | | |
| | Two-Step (LW) ($m = T^{0.5}$) | | | | Two-Step (LPW) ($m = T^{0.5}$) | | | | Two-Step (LW) (optimal $m$) | | | | Maximum Likelihood | | | |
| | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | |
| $d$/AR | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 |
| 0.2 | 0.229 | 0.706 | 0.078 | 0.496 | -0.008 | 0.456 | -0.024 | 0.237 | 0.161 | 0.431 | 0.048 | 0.080 | 0.014 | -0.001 | 0.009 | -0.002 |
| 0.4 | 0.247 | 0.657 | 0.084 | 0.487 | 0.032 | 0.453 | -0.022 | 0.223 | 0.201 | 0.471 | 0.048 | 0.089 | 0.001 | -0.095 | 0.001 | -0.047 |
| 0.49 | 0.262 | 0.634 | 0.098 | 0.486 | 0.051 | 0.455 | -0.009 | 0.239 | 0.214 | 0.449 | 0.061 | 0.089 | -0.029 | -0.146 | -0.002 | -0.092 |
| Variance | | | | | | | | | | | | | | | | |
| | Two-Step (LW) ($m = T^{0.5}$) | | | | Two-Step (LPW) ($m = T^{0.5}$) | | | | Two-Step (LW) (optimal $m$) | | | | Maximum Likelihood | | | |
| | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | |
| $d$/AR | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 |
| 0.2 | 0.207 | 0.198 | 0.114 | 0.118 | 0.383 | 0.388 | 0.197 | 0.195 | 0.254 | 0.468 | 0.136 | 0.337 | 0.154 | 0.118 | 0.082 | 0.041 |
| 0.4 | 0.201 | 0.191 | 0.110 | 0.119 | 0.385 | 0.359 | 0.186 | 0.193 | 0.248 | 0.426 | 0.138 | 0.349 | 0.166 | 0.154 | 0.080 | 0.065 |
| 0.49 | 0.202 | 0.177 | 0.111 | 0.121 | 0.377 | 0.334 | 0.194 | 0.197 | 0.233 | 0.430 | 0.138 | 0.348 | 0.186 | 0.231 | 0.087 | 0.101 |

where $B(\vartheta_0)$ is the outer product gradient, evaluated at the true parameter values $\vartheta_0$. This result is valid when the innovations in (3) merely satisfy some mild moment and mixing conditions. Recently, Baillie and Kapetanios (2008) have extended this analysis to include the estimation of long memory models with nonlinear autoregressive structures. The computation of such models is achieved by the minimization of the conditional sum of squares ($CSS$). This estimator has been shown by Robinson (2006) to be a consistent and asymptotically normal and to be approximately equivalent to the Gaussian $MLE$.

We now provide simulation evidence on the performance of the estimators given above for the estimation of both $d$ and the short memory parameters of an $ARFIMA$ model. A previous study by Nielsen and Frederiksen (2004) has provided Monte Carlo evidence on the small sample properties of $\hat{d}_{LW}$. The simulation evidence in this paper extends their results by also considering optimal bandwidths and also the effects of estimating the short memory parameters, which are essential when estimating $IRWs$. In order to provide a valid benchmark, corresponding results from $MLE$ are also reported. This benchmark is of course parametric, and is therefore expected to outperform the $SPE$. However, the extent of the improvement in performance is informative when considering the the practical usefulness of the $SPE$.

Realizations of $ARFIMA(1, d, 0)$ processes were generated for two different sample sizes of $T = 200$ and $T = 1,000$, and for 6 different simulation designs of the $AR$ coefficient $\phi = 0.80$ and $\phi = 0.95$ and for $d = 0.20$, $d = 0.40$ and $d = 0.49$. The biases and variances of the $\hat{d}_{LW}$ from the original observed series are presented in Table 1. Similar results for $\hat{\phi}_{LWTSE}$ which is estimated from the $FFF$ series are presented in Table 2. Results are shown for three different versions of the $LW$; (i) $LW$ using a bandwidth of $m = T^{0.5}$, (ii) $LPW$ and (iii) $LW$ using the optimal bandwidth $m^*_{LW}$. Corresponding analysis for the $MLE$ of $d$ and $\phi$, namely $\hat{d}_{MLE}$ and $\hat{\phi}_{MLE}$ is also provided across the different designs and sample sizes. Note that using the optimal bandwidth $m^*_{LW}$, presents the $LW$ in the most favorable circumstances, since an investigator would not have access to this information in practice. The consideration of this estimator mitigates to an extent the mismatch in our comparison of a parametric and a $SPE$ such as $LW$ and $MLE$. For the $ARFIMA(1, d, 0)$ process, it can be shown that $\tau^*$ in the equation for $m^*_{LW}$

| Table 2. Results for estimated AR coefficient | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bias | | | | | | | | | | | | | | | | |
| | Two-Step (LW) ($m = T^{0.5}$) | | | | Two-Step (LPW) ($m = T^{0.5}$) | | | | Two-Step (LW) (optimal $m$) | | | | Maximum Likelihood | | | |
| | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | |
| $d$/AR | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 |
| 0.2 | -0.230 | -0.673 | -0.075 | -0.412 | -0.089 | -0.438 | -0.016 | -0.171 | -0.181 | -0.432 | -0.053 | -0.119 | -0.032 | -0.021 | -0.014 | -0.003 |
| 0.4 | -0.246 | -0.615 | -0.079 | -0.403 | -0.118 | -0.432 | -0.016 | -0.162 | -0.216 | -0.450 | -0.053 | -0.129 | -0.032 | -0.012 | -0.009 | 0.004 |
| 0.49 | -0.259 | -0.585 | -0.088 | -0.401 | -0.129 | -0.425 | -0.024 | -0.175 | -0.221 | -0.433 | -0.064 | -0.131 | -0.024 | -0.036 | -0.008 | 0.009 |
| Variance | | | | | | | | | | | | | | | | |
| | Two-Step (LW) ($m = T^{0.5}$) | | | | Two-Step (LPW) ($m = T^{0.5}$) | | | | Two-Step (LW) (optimal $m$) | | | | Maximum Likelihood | | | |
| | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | | T=200 | | T=1000 | |
| $d$/AR | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 | 0.8 | 0.95 |
| 0.2 | 0.185 | 0.220 | 0.094 | 0.142 | 0.238 | 0.339 | 0.129 | 0.153 | 0.129 | 0.078 | 0.069 | 0.017 | 0.211 | 0.373 | 0.102 | 0.199 |
| 0.4 | 0.185 | 0.217 | 0.091 | 0.143 | 0.261 | 0.317 | 0.124 | 0.151 | 0.143 | 0.118 | 0.065 | 0.018 | 0.209 | 0.364 | 0.105 | 0.213 |
| 0.49 | 0.187 | 0.200 | 0.092 | 0.142 | 0.255 | 0.294 | 0.130 | 0.159 | 0.162 | 0.211 | 0.071 | 0.047 | 0.204 | 0.355 | 0.107 | 0.205 |

reduces to $\tau^* = -\phi(1-\phi)^{-2}$, and the optimal bandwidth $m^*_{LW}$ can be determined for each design in the parameter space.

The most striking aspect of tables 1 and 2 are the relatively poor properties of $\hat{d}_{LW}$ and $\hat{\phi}_{LWTSE}$; and also how the performance deteriorates with increasing persistence of the short memory process. The results for the $m = T^{1/2}$ bandwidth, indicate very substantial upward bias in the estimate $\hat{d}_{LW}$ and a corresponding severe downward bias in $\hat{\phi}_{LWTSE}$. For example, when the true data generating process is an $ARFIMA(1, d, 0)$ process with $d = 0.40$ and $\phi = 0.95$, the average estimated values are $\hat{d}_{LW} = 0.88$ and $\hat{\phi}_{LWTSE} = 0.37$. In general, the $LW$ methodology appears unable to discriminate between the short run $I(0)$ autoregressive dynamics and long memory $I(d)$ dynamics.

With the use of optimal bandwidths, the trade off in terms of bias of the estimates of $d$ and $\phi$ when $m = T^{0.5}$ is no longer present. Although $\hat{d}_{LW}$ has reduced bias, there is a corresponding increase in parameter estimation variability. This is due to the considerably smaller value of $m$, in this case. This increased variability is inherited by the $FFF$ series $\hat{u}_t$ and hence second stage estimation of the $\phi$ parameter is not generally improved since $\hat{\phi}_{LWTSE}$ has a reduced bias compared to before, but much larger variability. The use of the local polynomial Whittle, $\hat{d}_{LPW}$ does not generally lead to any great improvement. In contrast, and noting the obvious caveat that $MLE$ assumes the correct parametric model, the benchmark $MLE$ estimator appears to perform extremely well in terms of bias and variability even compared to the infeasible estimator that uses the unknown optimal bandwidth. The only further caveat to this appears for $d = 0.49$ and $\phi = 0.95$, when there is a simultaneous move to non stationarity from both the autoregressive and long memory parameters. In this case $MLE$ does not perform as well as for the other experiments.

# 4 A Semiparametric Autoregressive Approximation Method

A straightforward approach of modeling long memory processes, is to implicitly ignore the presence of long memory and to simply estimate a high order $AR(P)$ model. In particular, the

$ARFIMA(p,d,q)$ model can be represented by an infinite autoregressive expansion of the form

$$y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + v_t \tag{6}$$

A possible method is to directly estimate by $OLS$ the truncated autoregressive, $AR(P)$, expansion

$$y_t = \sum_{j=1}^{P} \pi_j y_{t-j} + \tilde{v}_t \tag{7}$$

where the order $P$, is obtained by some information criterion. This approach has recently been theoretically analyzed by Poskitt (2005). On denoting the least squares estimates of $\pi_j$, obtained by fitting an $AR(P)$ model to the data, by $\hat{\pi}(j)^{(P)}$; and on further denoting the coefficients that solve the Yule-Walker equations for an $AR(P)$ model by $\pi(j)^{(P)}$, then theorem 5.1 of Poskitt (2005) states that $\sum_{j=1}^{P} \left|\hat{\pi}(j)^{(P)} - \pi(j)^{(P)}\right|^2 = o_p(1)$ for all $P$ such that $P \to \infty$ and $P = o(T^\alpha)$ for all $\alpha > 0$. For example, an acceptable sequence for $P$ is $(lnT)^\alpha$ for some $\alpha > 1$. Further, by the extension of Baxter's inequality proven in Theorem 4.1 of Inoue and Kasahara (2006) it follows that

$$\sum_{j=1}^{P} \left|\pi(j)^{(P)} - \pi(j)\right| = o(1), \tag{8}$$

as long as $P \to \infty$. Then, overall,

$$\sum_{j=1}^{P} \left|\hat{\pi}(j)^{(P)} - \pi(j)\right|^2 = o_p(1) \tag{9}$$

which implies that the $IRWs$ can be consistently estimated by fitting an approximating autoregressive model to the long memory time series realization. On the important issue of choosing $P$, it has been shown by Poskitt (2005), via his Theorem 5.3, that selecting $P$ by information criteria such as the $AIC$ or $BIC$ is asymptotically efficient in the sense of Shibata (1980). In the Monte Carlo study in this paper the value of $P$ is fixed at $(lnT)^2$. Of course, the validity of this approximation extends well beyond finite order ARFIMA models. It can be validly used when the short memory component has an infinite AR representation. As a result the AR approximation takes on the interpretation of a semiparametric model of long memory.

A further attraction of using an infinite $AR(P)$ approximation for a long memory process is that the approach extends to non stationary processes. Non-stationary long memory processes are still amenable to impulse response analysis, since when $0.5 \le d \le 1$ the process does not have a finite variance, but still has finite cumulative $IRWs$. Of course, even for higher values of $d$, there may be interest on the path of the $IRW$ even though this path does not converge to zero. The appendix provides a probabilistic bound for the estimated coefficients of the autoregressive approximation similar to that given in (9). It should be noted that there are some existing results concerning the theoretical properties of $LW$ and $MLE$ when applied to non-stationary long memory processes. In particular, Velasco (1999) and Phillips and Shimotsu (2006) have shown that the $LW$ estimator of $d$ is consistent when $0.5 < d < 1$, and asymptotically normally distributed when $d < 3/4$, but not otherwise. Phillips and Shimotsu (2005) provide an exact version of the $LW$ estimator that produces a consistent and asymptotically normal estimator of $d$ for any value of the true $d$, under certain regularity conditions. Some results on parametric estimation are available in Tanaka (1999), Ling and Li (2001) and Johansen and Nielsen (2008). The study by Johansen and Nielsen (2008) is especially relevant since it proves consistency of $MLE$ for a parametric autoregressive non-stationary long memory model. The model is related but not equivalent to a non-stationary $ARFIMA$ model. Note there is no theoretical analysis of any semi parametric model of non-stationary long memory which models both the short and long memory component and, therefore, the analysis of the appendix is of some relevance. Overall, given that the $AR(P)$ approximation models both short and long memory, it is particularly attractive for practical purposes.

# 5 Estimation of Impulse Response Weights

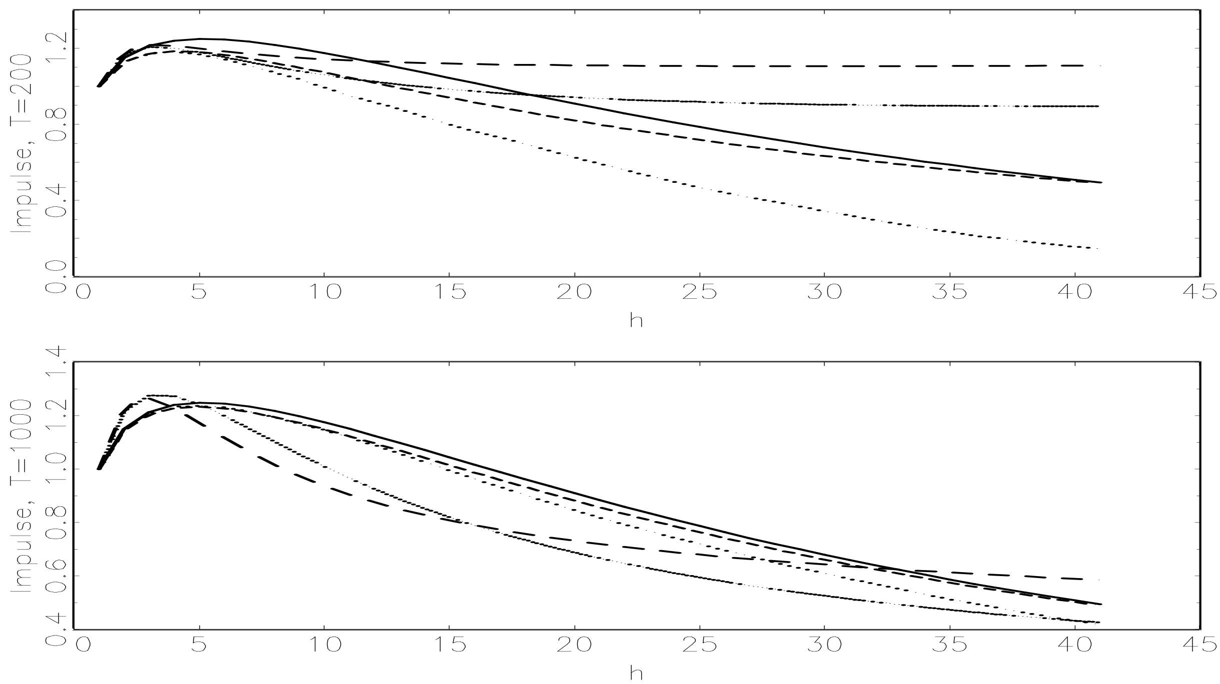
Many applications of long memory models have been concerned with $IRW$ analysis. For example, Diebold and Rudebusch (1989) and Jensen (2007) when analyzing real $GNP$ growth and money growth, Diebold, Husted, and Rush (1991) for persistence in real exchange rates, Baillie, Chung and Tielsau (1996) for inflation and the Friedman hypothesis, and Andersen et al (2001) and Andersen et al (2003) for the realized volatility of exchange rates. Given an $ARFIMA$ data generating process the implied $IRWs$, denoted by $\psi_k$ for $k = 1, , 2, ....$ are generated from

$$\psi(L) = \theta(L)(1 - L)^{-d}\phi(L)^{-1} \tag{10}$$

where $\psi(L) = \sum_{k=1}^{\infty} \psi_k L^k$ and the underlying "true" model is the $ARFIMA(p, d, q)$ process in equation (3). The estimated $IRWs$ are obtained by replacing the true theoretical parameters with their corresponding estimates. For large lag $k$, these Wold decomposition coefficients decay at the approximate rate of $\psi_k \sim c_1 k^{d-1}$. However, the presence of a relatively persistent $AR(1)$ component process can considerably alter the appearance of the $IRWs$ for short to moderate impulse response horizons.

Figures 1 through 4 report results for different $IRWs$ for horizons $k = 1, 2, ....40$ for $ARFIMA(1, d, 0)$ models. The values 0.2, 0.4, 0.6 and 0.8 are considered for $d$ and the value 0.95 for $\phi$. Results

Figure 1: Impulse Responses: d=0.2, ar=0.95

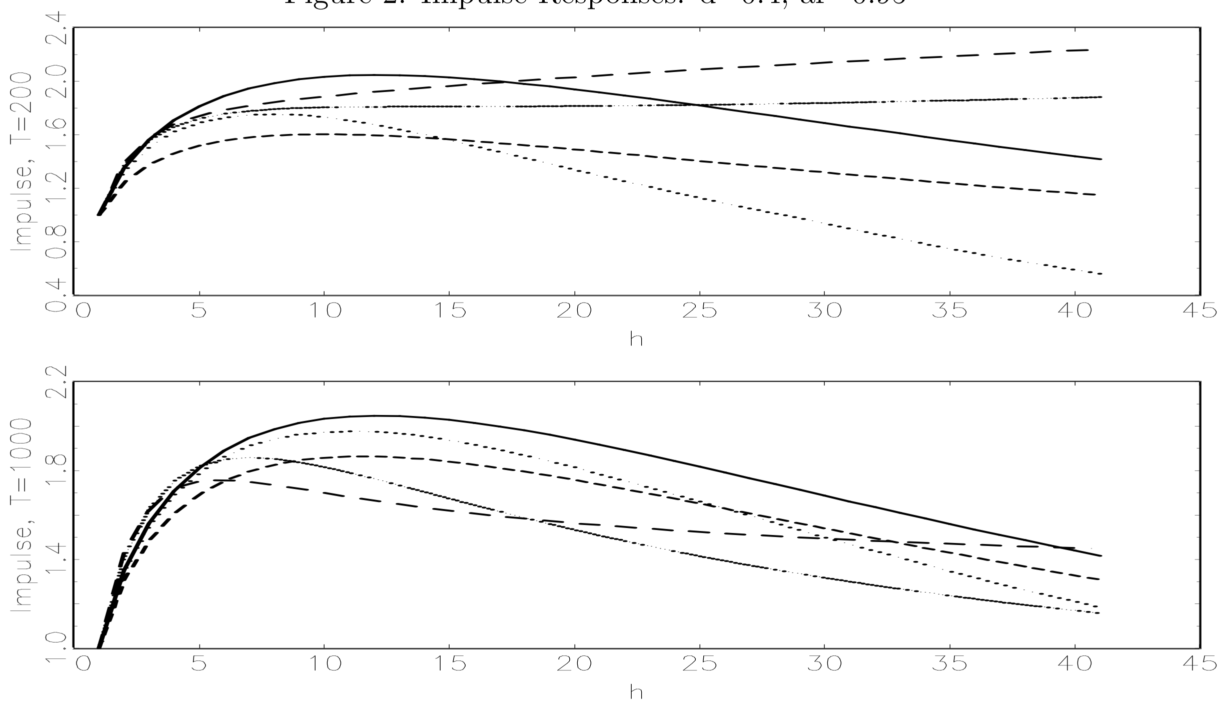



Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE; Dense Dotted Line (...): Two-Step LPW.

were also obtained for the cases of $\phi = 0.5$ and $\phi = 0.8$, but are omitted due to space constraints and are available upon request. The $IRW$ are estimated from $AR$ approximations, $MLE$, $LWTSE$ for all the cases, and also from using $LPW$, rather than $LW$, for the $LWTSE$ estimation, in the stationary cases. The estimated $IRWs$ from using the $LW$ and $LPW$ methods are constructed using a bandwidth of $m = \ T^{0.5}$ . For models with $d = 0.2, 0.4$ and quite persistent short memory, Figures 1 and 2 indicate that $IRWs$ estimated from the $LWTSE$ approach perform poorly in comparison with corresponding estimates from $MLE$. The $IRWs$ estimated from $MLE$ with $d$ in the stationary region dominates alternative methods; however $MLE$ estimated $IRWs$ are poor for $d = 0.6$, or $d = 0.8$ when there is persistent autocorrelation of $\phi = 0.95$. In this case the $AR(P)$ approximation does surprisingly well and is the preferred method.

It is worth commenting more extensively on the performance of the $AR$ approximation. For the large sample size of $T = 1,000$ and for designs of ($d = 0.6$, $\phi = 0.5$) and ($d = 0.8$, $\phi = 0.5$) (not reported to save space but available upon request), the $MLE$ performs extremely well, with the high order $AR$ approximation generally being slightly superior to the $LWTSE$. For the designs of ($d = 0.6$, $\phi = 0.95$) and ($d = 0.8$, $\phi = 0.95$) in figures 3 and 4 respectively, the high order $AR$ approximation performs outstandingly well, with the $MLE$ a poor third compared

Figure 2: Impulse Responses: d=0.4, ar=0.95



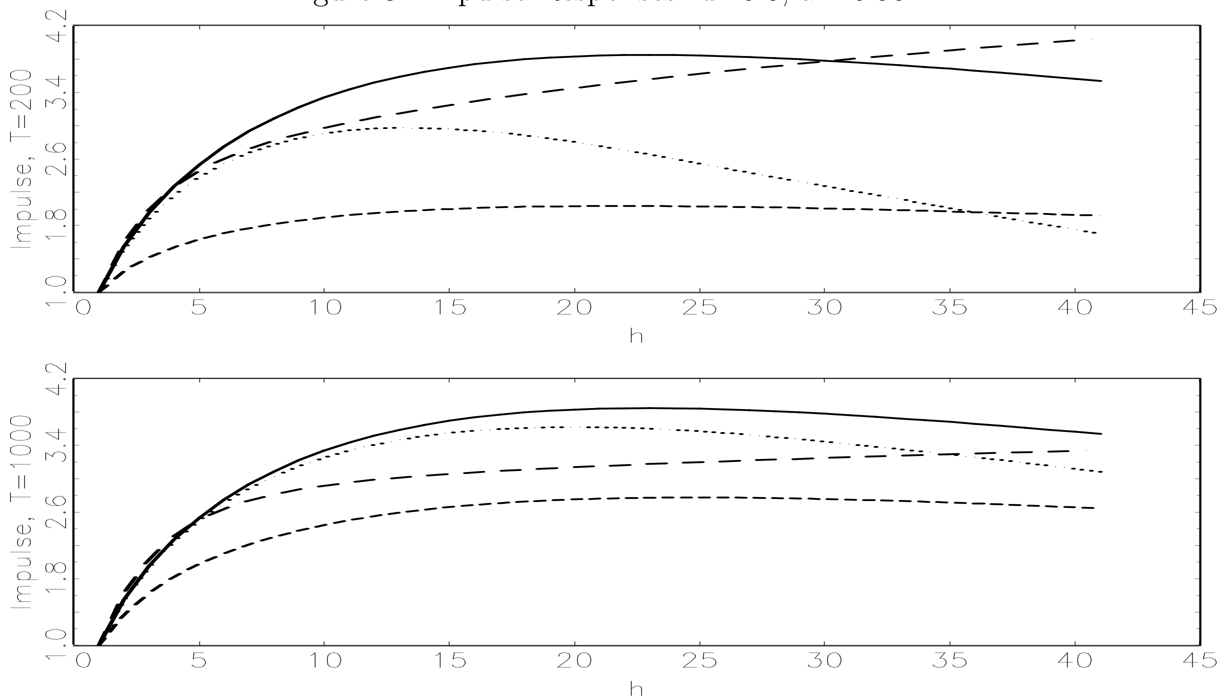

Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE; Dense Dotted Line (...): Two-Step LPW.

with the $LWTSE$. Hence, there seems some evidence that $MLE$ works well for non stationary long memory processes provided that there is only moderate degree of persistence in the short run dynamics. However, when a non stationary long memory process has a very persistent short run component, the high order $AR$ approximation method is extraordinarily accurate compared with $MLE$ and the $LWTSE$. The excellent performance of the high order $AR(P)$ method raises important issues as to whether it is worth an investigator being concerned with the presence of long memory if the investigator's main interest is to only to assess the impact of shocks or innovations on a series.

# 6 Conclusions

This paper has shown that the semi parametric Local Whittle ($LW$) estimator of the long memory parameter in a univariate time series performs poorly in the presence of persistent short memory. The $LW$ estimator also has poor properties for some important practical situations including the estimation of short run parameters and Impulse Response Weights ($IRWs$). These problems are shown to exist even when optimal bandwidths, or local polynomial Whittle methods are used. The results are compared with benchmark $MLEs$ and the paper also suggests

Figure 3: Impulse Responses: d=0.6, ar=0.95




Notes: Solid Line (——): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE.

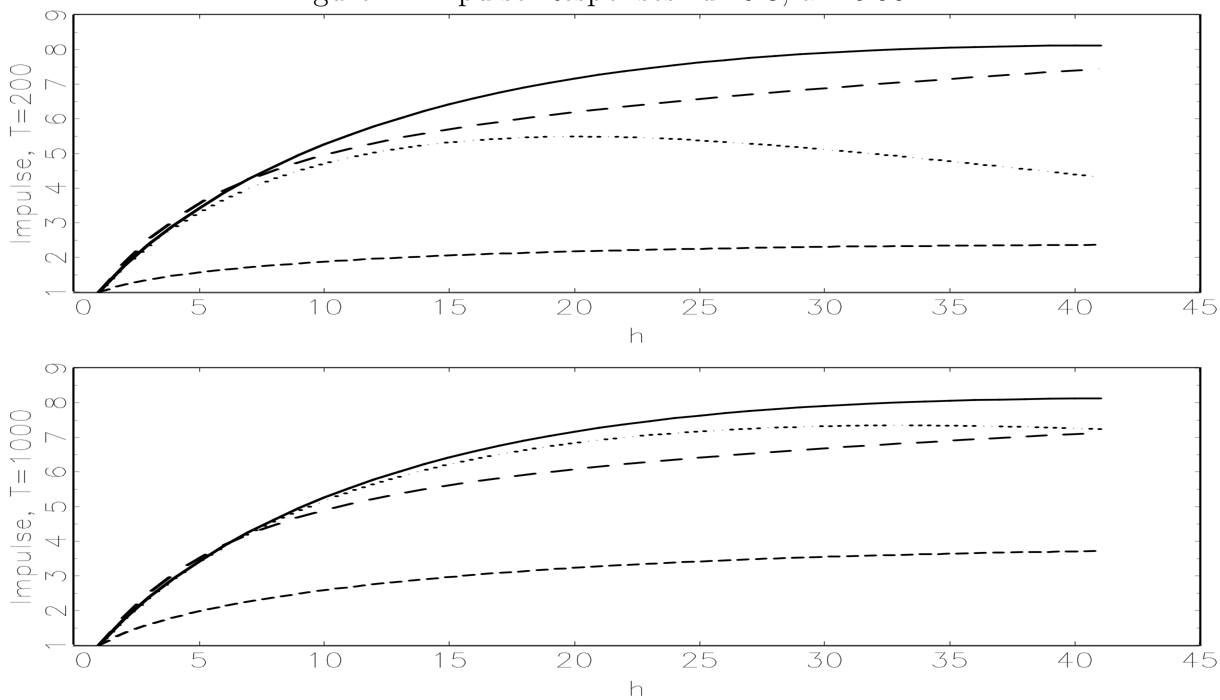
the use of, and provides results on the application of, high order autoregressions to approximate long memory processes, from which estimated $IRWs$ can be derived. The autoregressive approximations are shown to perform very well. Further, they are found to be relevant and perform relatively well for some non-stationary processes.

# Appendix

*Proof of Theorem 1*

Since all the roots of the polynomials in the lag operator $\phi(L)$ and $\theta(L)$ lie outside the unit circle, it follows that $\sum_{k=0}^{\infty}\pi_k^2 < \infty$ and hence that $\sum_{k=1}^{t-1}\pi_k y_{t-k} = O_p(1)$. The Local Whittle estimator $\hat{d}_{LW}$ will generate the FFF series $\hat{u}_t = (1-L)^{\hat{d}_{LW}} y_t = y_t - \sum_{l=1}^{t-p}\hat{\pi}_l(\hat{d}_{LW})y_{t-l}$, where $\hat{\pi}_l(\hat{d}_{LW}) = \Gamma(l-\hat{d}_{LW})\Gamma(-\hat{d}_{LW})^{-1}\Gamma(l+1)$. Since $\hat{u}_t = (1-L)^{\hat{d}_{LW}} y_t$, then $(\hat{u}_t - u_t) = \sum_{j=1}^{\infty}\pi_j(\hat{d}_{LW} - d_0)u_{t-j}$. Since $(\hat{d}_{LW} - d_0) = O_p(m^{-1/2})$ and $u_t = (1-L)^d y_t$, then following the same approach as Wright (1995), $T^{-1}\sum_{j=1}^{\infty}(\hat{u}_t - u_t)^2 = T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1}\pi_j(\hat{d}_{LW} - d_0)u_{t-j}\right)^2$. Then, using the mean value theorem we have that $\pi_j(d) = dX_j^1 + d^2X_j^2$ , where $X_j^1$ denotes the first derivative and $X_j^2$ the second derivative of $\pi_j(.)$. Then, $\sum_{k=1}^{t-1}\pi_k u_{t-k} = d\sum_{k=1}^{t-1}X_j^1 u_{t-j} + d^2\sum_{k=1}^{t-1}X_j^2 u_{t-j}$, and following the same arguments as in Wright (1995), $(\hat{d}_{LW} - d_0)\sum_{k=1}^{t-1}X_j^1 u_{t-j} =$

Figure 4: Impulse Responses: d=0.8, ar=0.95




Notes: Solid Line (—–): True IRF; Long Dashed Line (– – –): Two-Step LW; Dotted Line (. . .): AR Approximation; Short Dashed Line (- - -): MLE.

$O_p(m^{-1/2})$, and $T^{-1}(\hat{d}_{LW} - d_0)\sum_{k=1}^{t-1} X_j^2 u_{t-j} = O_p(m^{-1/2})$, and hence

$$T^{-1}\sum_{t=1}^{T-k} \hat{u}_t \hat{u}_{t+k} = T^{-1}\sum_{t=1}^{T-k} u_t u_{t+k} + O_p(m^{-1/2}) \tag{11}$$

This suffices to prove the Theorem result for an $ARFIMA(p, d, 0)$ model. For the general case of an $ARFIMA(p, d, q)$ model we have that for the second step $ARMA$ estimation, the conditional $MLE$ needs to be numerically maximized. Let us denote the likelihood function by $L(\beta)$. The form of the likelihood may be found in, e.g., (5.6.3) of Hamilton (1994). It is then clear that the likelihood function is differentiable and as long as (11) holds we have that $L\left(\hat{\beta}_{LWTSE}(\hat{d}_{LW})\right) - L\left(\beta_0(d_0)\right) = O_p(m^{-1/2})$. But, by an application of the mean value theorem we have that $L\left(\hat{\beta}(\hat{d}_{LW})\right) = L\left(\beta_0(d_0)\right) + \frac{\partial L}{\partial \beta}\Big|_{\beta=\bar{\beta}} \left(\hat{\beta}(\hat{d}_{LW}) - \beta_0(d_0)\right)$. Hence, the result of the Theorem holds for $ARFIMA(p, d, q)$ models completing the proof.

*A probabilistic bound for the coefficients of the autoregressive approximation in the non-stationary case*

Let the norm of a matrix $A$, $\|A\|$, be given by $tr(A'A)^{1/2}$; and it is assumed that $y_t$ follows (6) with $d > 1/2$ and $v_t$ i.i.d. with finite second moments. First, note that the autoregressive expansion underlying (6) is valid for all values of $d$, even if the resulting process is non-stationary, as discussed in Section 2 of Phillips and Shimotsu (2005). This is the case as long as the process is specified to start at a given point in the past. We wish to derive a bound for $\|\hat{\pi} - \pi\|$ where $\pi = (\pi_1, ..., \pi_P)$ and $\hat{\pi} = (\hat{\pi}_1, ..., \hat{\pi}_P)$. We have that $\|\hat{\pi} - \pi\| = \left\|(X'X)^{-1} X'\tilde{v}\right\|$ where $X = (y_1, ..., y_P)$, $y_i = (y_{P-i+1}, ..., y_{T-i})'$ and $\tilde{v} = (\tilde{v}_{P+1}, .., \tilde{v}_T)'$. Then,

$$\left\|(X'X)^{-1} X'\tilde{v}\right\| \leq \left\|(X'X)^{-1}\right\| \|X'\tilde{v}\| \tag{12}$$

We examine each term of the right hand side of (12) in turn. We first need to determine the normalising factor for $X'X$. By Lemma 3.2 of Davidson and De jong (2000) we know that $\sum_{t=P+1}^{T} y_t^2 = O_p(T^{2d})$ and so $T^{2d}$ is the normalising constant. Then, we need to determine the order of magnitude of each diagonal term of $\left(\frac{X'X}{T^{2d}}\right)^{-1}$. From Kantorovich's inequality for a square $m \times m$ matrix $A$, we have that for the $i,i$-th element of $A^{-1}$ the following holds: $(A^{-1})_{ii} \leq \frac{1}{4a_{ii}} \left(\frac{\alpha}{\beta} + \frac{\beta}{\alpha} + 2\right)$, where $a_{ii}$ is the $i,i$-th element of $A$, $\alpha = \min_i \lambda_i(A)$, $\beta = \max_i \lambda_i(A)$ and $\lambda_i$ denotes the $i$-th eigenvalue of $A$. Since we assume that $\frac{X'X}{T^{2d}}$ has an inverse we know that $\alpha > 0$ and therefore $\beta > 0$. We need to determine a bound for $\frac{\beta}{\alpha}$ and therefore for $\beta$. $\beta$ is bounded from above by the column sum norm of $\frac{X'X}{T^{2d}}$. But, from the above it easily follows that this norm is $O_p(P)$ and therefore that $\frac{\beta}{\alpha}$ for $\frac{X'X}{T^{2d}}$ is $O_p(P)$. Thus, $\left|((X'X)^{-1})_{ii}\right| = O_p(PT^{-2d})$. Further, by the positive-definiteness of $X'X$, $\left|((X'X)^{-1})_{ij}\right| \leq \left|((X'X)^{-1})_{ii}\right|$, when $i \neq j$. Then, it follows that $\left\|(X'X)^{-1}\right\| = O_p(P^2T^{-2d})$. Moving on to $\|X'\tilde{v}\|$ we need to determine the behavior of $\sum_{t=P+1}^{T} y_{t-i}\tilde{v}_t$. We have

$$\sum_{t=P+1}^{T} y_{t-i}\tilde{v}_t = \sum_{t=P+1}^{T} y_{t-i} \left(\sum_{j=P+1}^{t} \pi_j y_{t-j} + v_t\right) \leq \sum_{j=P+1}^{T} \pi_j \left(\sum_{t=P+1}^{T} y_{t-i}y_{t-j}\right) + \sum_{t=P+1}^{T} y_{t-i}v_t$$

Then, it is necessary to show that

$$\sum_{t=P+1}^{T} \frac{y_{t-i}}{T^{d-1/2}} v_t = O_p\left(T\right) \tag{13}$$

But, by the independence of $v_t$ it follows that $\frac{y_{t-i}}{T^{d-1/2}} v_t$ is a martingale difference sequence with finite variance and so (13) follows by the martingale difference law of large numbers. By the fact that $\sum_{t=P+1}^{T} y_t^2 = O_p(T^{2d})$, it follows that $\sum_{t=P+1}^{T} y_{t-i}y_{t-j} = O_p(T^{2d})$. Since, $\pi_j = O(j^{-d-1})$, it follows that $\sum_{t=P+1}^{T} y_{t-i}\tilde{v}_t = O_p\left(P^{-d}T^{2d}\right) + O_p\left(T^{d+1/2}\right)$ and $\|X'\tilde{v}\| = O_p\left(P^{-d+1/2}T^{2d}\right) + O_p\left(T^{d+1/2}P^{1/2}\right)$. Thus, overall

$$\left\|(X'X)^{-1}\right\| \|X'\tilde{v}\| = O_p\left(P^{-d+5/2}\right) + O_p\left(T^{1/2-d}P^{5/2}\right)$$

This probabilistic bound suggests that $\|\hat{\pi} - \pi\| = o_p(P)$ for $d > 3/2$.

# References

ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F. X. AND LABYS (2001): The distribution of realized exchange rate volatilty, *Journal of the American Statistical Association*, 96, 42-55.

ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F. X. AND LABYS (2003): Modelling and forecasting realised volatilty, *Econometrica*, 71, 579-625.

ANDREWS, D.W.K. AND Y. SUN (2004): Adaptive local polynomial Whittle estimation of long range dependence, *Econometrica*, 72, 569-614.

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 (2008): "Nonlinear models for strongly dependent processes and financial applications, " *Journal of Econometrics*, 147, 60-71.

BERAN, J. (1970): *Statistics for Long Memory Processes*, Chapman-Hall.

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, " *Journal of Time Series Analysis*, 27, 211-251.

DAVIDSON, J., AND R. M. DE JONG (2000): "The functional central limit theorem and weak convergence to stochastic integrals II: Fractionally integrated processes, " *Econometric Thoery*, 16, 643-666.

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 G. RUDEBUSCH (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, 24, 189-209.

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.

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

HENRY, M. (2001): "Robust automatic bandwidth for long memory," *Journal of Time Series Analysis*, 22, 292-316.

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.

INOUE, A. AND Y. KASAHARA (2006): "Explicit representation of finite predictor coefficients and its applications," *Annals of Statistics*, 34, 973–993.

Jensen, M. (2007): "The long-run Fisher effect: can it be tested?," *Journal of Money, Credit and Banking*, forthcoming.

Johansen, S. and M. O. Nielsen (2008): "Likelihood inference for a nonstationary fractional autoregressive model" *Mimeo*.

Ling, S. and W. K. Li (2001): "Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models" *Econometric Theory*, 17, 738-765.

Nielsen, M. O. and P. H. Frederiksen (2004): "Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration" *Econometric Reviews*, 24, 405-443.

Phillips, P.C.B. (2007): "Unit root periodogram," *Journal of Econometrics,* 138, 104-124.

Phillips, P.C.B. and K. Shimotsu. (2005): "Exact local whittle estimation of fractional integration," *Annals of Statistics,* 33, 1890-1933.

Phillips, P.C.B. and K. Shimotsu (2006): "Local Whittle estimation in non stationary and unit root cases," *Annals of Statistics,* 32, 656-692.

Poskitt, D. (2005): "Autoregressive approximation in nonstandard situations: The non-invertible and fractionally integrated case," *Annals of the Institute of Statistical Mathematics,* 59, 697–725.

Robinson, P. M. (1995): "Gaussian semi-parametric estimation of long range dependence," *Annals of Statistics*, 23, 1630–1661.

Robinson, P. M. (2006): "Conditional - sum - of - squares estimation of models for stationary time series with long memory," *London School of Economics working paper.*

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.

Sowell, F. B. (1992): "Maximum likelihood estimation of stationary univariate fractionally integrated time series models," *Journal of Econometrics*, 53, 165-188.

Sowell, F. B. (1992): "Modeling long run behavior with the fractional ARIMA model" *Journal of Monetary Economics*, 29, 277-302.

Tanaka, K. (1999): "The Nonstationary fractional unit root" *Econometric Theory*, 15, 549–583.

Velasco, C. (1999): "Gaussian semi parametric estimation of non stationary time series," *Journal of Time Series Analysis*, 20, 87-127.

Wright, J H (1995): "Stochastic orders of magnitude associated with two-stage estimators of fractional ARMA systems," *Journal of Time Series Analysis*, 16, 119-125