SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-14

When Long Memory Meets the Kalman Filter: A Comparative Study

Stefano Grassi and Paolo Santucci de Magistris

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

When Long Memory Meets the Kalman Filter: A Comparative Study*

Stefano Grassi†
Aarhus University and CREATES

Paolo Santucci de Magistris‡
Aarhus University and CREATES

May 2, 2011

Abstract

The finite sample properties of the state space methods applied to long memory time series are analyzed through Monte Carlo simulations. The state space setup allows to introduce a novel modeling approach in the long memory framework, which directly tackles measurement errors and random level shifts. Missing values and several alternative sources of misspecification are also considered. It emerges that the state space methodology provides a valuable alternative for the estimation of the long memory models, under different data generating processes, which are common in financial and economic series. Two empirical applications highlight the practical usefulness of the proposed state space methods.

Keywords: ARFIMA models, Kalman Filter, Missing Observations, Measurement Error, Level Shifts.

JEL Classification: C10, C22, C80

*The authors are grateful to Massimiliano Caporin, Robinson Kruse, Alessandra Luati, Tommaso Proietti, Eduardo Rossi and Daniel Ventosa Santaularia, for their valuable suggestions. The authors acknowledge the research support of CREATES (funded by the Danish National Research Foundation).

†School of Economics and Management, Aarhus University, Bartholins Alle 10, 1322 University Park, DK-8000 Aarhus C, Denmark; phone: +45 89 421 547; email address: sgrassi@creates.au.dk

‡School of Economics and Management, Aarhus University, Bartholins Alle 10, 1322 University Park, DK-8000 Aarhus C, Denmark; phone: +45 89 421 547; email address:: psantucci@creates.au.dk

1 Introduction

Long-range dependent data arise in a wide variety of scientific disciplines, from hydrology to economics. The subject of long memory time series was brought to prominence by Hurst (1951) on hydrological time series and has subsequently received extensive attention in the literature. The use of fractional processes in economics and econometrics has been introduced by the seminal paper by Granger (1980) and Granger and Joyeux (1980). See the volumes by Beran (1994) and Palma (2007) and the collections of Robinson (2003) and the references therein. The starting point of the econometric literature on autoregressive fractionally integrated moving average (ARFIMA) models has been motivated by the fact that many economic and financial time series show evidence of being neither $I(0)$ nor $I(1)$. Nowadays, a broad range of applications in finance and macroeconomics have shown that fractional integration and long memory are relevant, see among others Diebold et al. (1991) for exchange rate data, Andersen et al. (2001a) and Andersen et al. (2001b) for financial volatility series, and Baillie et al. (1996) for inflation data. Early papers on the estimation of the long-range dependent models are Fox and Taqqu (1986), Dahlhaus (1989), Sowell (1992) and Robinson (1995). See Chan and Palma (2006) for a more recent review.

In an alternative framework, Chan and Palma (1998) proposed a state space approach to compute the maximum likelihood (ML henceforth) estimates for an ARFIMA model. The authors suggest to truncate the infinite MA and AR representation of an ARFIMA model and to write the ARFIMA in state space form. The long memory parameter, d, can be then estimated by means of the Kalman filter (KF henceforth). The estimates obtained by this method are consistent, asymptotically normal and efficient under mild regularity conditions. This methodology, although conceptually simple, was computationally very intensive in the 90's, and, for this reason, not commonly used. With the computational capabilities at hand nowadays, it is possible to estimate such models in few seconds even for large datasets. Several simulation studies, by Rea et al. (2008), Bisaglia and Gugan (1998), Nielsen and Frederiksen (2005) and Haldrup and Nielsen (2007) provide a comparison between different estimation strategies, but without considering the state space alternative. This underlines that, although, great effort has been spent in the estimation of fractional processes with semiparametric and maximum likelihood methods, little has been done to explore the state space alternative.

This paper aims to fill this gap, through an extensive Monte Carlo simulation exercise and two novel contributions regarding measurement errors and level shifts in the long memory framework. The practical usefulness of the state space methods relies on the possibility of a straightforward modeling approach for measurement errors, outliers, level shifts and missing values. It is shown by simulations that the KF provides unbiased and efficient estimates of the model parameters also in these cases. The proposed Monte Carlo simulations are therefore intended as an attempt to explore the potentiality of the state space approach in the long memory framework. With this purpose, we consider several parametric and semiparametric estimation methods for the ARFIMA models, and compare them with the state space alternatives. The methods are contrasted using several data generating processes with alternative short-run dynamics including the possibility of t-distributed innovations, missing values, measurement errors and level shifts. Following Nielsen and Frederiksen (2005) and Haldrup and Nielsen (2007), the bias and the root mean squared error (RMSE) are adopted to measure the finite sample performances of the estimators.

Our results can be summarized as follows: firstly, state space methodology is a valid alternative to the usual estimation procedures and it is robust to non-Gaussian shocks and over-specification. In case of under-specification, the Akaike Information Criterion, AIC, always selects the right model. Second, when the series at hand has missing observations, then the state space estimation method is superior to the traditional ones as it has low bias and RMSE. Third, in case of measurement errors, the KF largely outperforms the traditional estimators as well as the corrected local Whittle of Hurvich et al. (2005). Finally, we provide a novel and promising approach to the joint modeling of long memory and level shifts, by a modified version of the Lu and Perron (2010) filter, that can handle both features at the same time. We show that, differently from usual estimators that are upwardly biased, our methods produces unbiased estimates of the d parameter, even in presence of level shifts. This model also allows for a parametric test of long memory versus level shifts.

The paper is organized as follows. Sections 2 and 3 review the ARFIMA processes and the semiparametric estimators considered in our analysis. Section 4 introduces the approximate ML estimators. State space methodology is presented in Section 5. Section 6 reports the extensive Monte Carlo experiments and the methodological contributions that are the core of this paper. Section 7 provides two empirical applications of the long memory state space models. Section 8 concludes the paper.

2 ARFIMA processes

An ARFIMA(p, d, q) process y_t is defined as:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\eta_t \tag{1}$$

where $\{\eta_t\}$ is a sequence of independent random variables with zero mean and constant variance σ_η^2, the lag operator L is such that $Ly_t = y_{t-1}$; $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is the autoregressive operator, $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$, is the moving average operator, and $(1-L)^d$ is the fractional difference operator. The parameter d determines the long memory of the process. If $d > -1/2$ the process is invertible and possesses a linear representation. If $d < 1/2$ it is covariance stationary. Furthermore, if $d > 0$ the process is said to have long memory since the autocorrelations die out at an hyperbolic rate (and indeed are no longer absolutely summable) in contrast to the much faster exponential rate in the weak dependence case. For $d \in (0, 1/2)$, model (1) is a stationary *long-memory* process with non-summable autocorrelations, $\sum_{k=0}^{\infty} |\rho_k| = \infty$. If $d = 0$ the spectral density is bounded at the origin and the process has only weak dependence (short memory), and it is the well known ARMA process. In most practical cases the parameter $d \in (0, 1/2)$ has been proved to be relevant for many applications. The fractional difference operator $\Delta^d = (1-L)^d$ in equation (1) is defined by its binomial expansion:

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function. Hosking (1981) shows that a stationary ARFIMA(p, d, q) admits infinite MA and AR expansions as

$$y_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j} \tag{3}$$

$$y_t = \sum_{j=1}^{\infty} \pi_j y_{t-j} + \eta_t \tag{4}$$

Hosking (1981) also provides a formula to compute the weights ψ_j and π_j for low order ARFIMA processes. An alternative, although not equivalent, definition of long range dependence can be stated in the frequency domain. In particular, the spectral density of the ARFIMA(p, d, q)

process (1) can be represented as follows:

$$f_\Theta(\lambda) \sim G|\lambda|^{-2d} \quad \text{for} \quad \lambda \to 0 \tag{5}$$

where $\Theta = (d, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$ and G corresponds to the spectral density of an ARMA(p, q) process, see Beran (1994). In Sections 3 and 4 we present the estimation methods that are considered in our study. Section 3 presents the semiparametric approaches, based on the approximation of the frequency domain representation (5), for frequencies close to the origin. Section 4 introduces the parametric models, that are based on likelihood methods. The parametric class presents the advantage, over the semiparametric approaches, that it does not rely on any particular choice of a bandwidth parameter. However, it assumes the correct specification of the short-run dynamics. To have a fair comparison between parametric and semiparametric estimators we present the results for two bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$.

3 Semiparametric estimators

If we were interested only in the estimate of the long memory parameter d, then a natural approach would be to use one of the semiparametric methods. The semiparametric estimators of d require only a local characterization of the spectral density. It is well known that the drawback of the global long memory estimators, such as the Whittle estimator, is that they require a full specification of the spectral density. Instead, a consistent estimate of d can be obtained simply by specifying the shape of the spectral density at the origin, these methods are referred to as local methods. In theory, the semiparametric approach has the advantage, over the parametric ones, that it does not require a full specification of the dynamics of the process. This implies that semiparametric estimation should be more robust to the misspecification of the dynamics. In particular, the semiparametric approaches are based on the characterization of the spectrum as $\lambda \to 0$. According to equation (5), the spectrum of the of a long memory process has a pole as $\lambda \to 0$, that is proportional to λ^{-2d}.

3.1 Local Whittle Estimator

Robinson (1995) studies the properties of the local Whittle estimator (*LW* henceforth) of d, which is the minimand of the following contrast function,

$$Q(G,d) = \frac{1}{b}\sum_{j=1}^{b}\left\{\log\left[G\lambda_j^{-2d}\right] + \frac{\lambda_j^{2d} I_y(\lambda_j)}{G}\right\}, \tag{6}$$

$\lambda_j = \frac{2\pi j}{T}$ are the Fourier frequencies. In particular, concentrating G out, it yields

$$R(d) = \frac{1}{b}\sum_{j=1}^{b}\log\left(\lambda_j^{-2d}\right) + \log\left(\frac{1}{b}\sum_{j=1}^{b}\lambda_j^{2d} I_y(\lambda_j)\right), \tag{7}$$

where I_y is the periodogram of the series under analysis. The local Whittle estimates of d is

$$\hat{d} = \arg\min_{d\in\mathbb{D}} \hat{R}(d) \tag{8}$$

where $\mathbb{D}$ is the admissible set of values for d. Robinson (1995) proves consistency and asymptotic normality of the proposed estimator. Outside the stationary region, it is known that the asymptotic theory for the local Whittle estimator is discontinuous at $d = \frac{3}{4}$ and, it is inconsistent when $d > 1$.

3.2 Exact Local Whittle Estimator

Shimotsu and Phillips (2005) extend the range of application of existing semiparametric methods, studying an exact form of the local Whittle estimator which does not rely on differencing or tapering (see Velasco, 1999) and which seems to offer a good general purpose estimation procedure for the memory parameter that applies throughout the stationary and nonstationary regions of d. The exact Whittle objective function is

$$Q(G,d) = \frac{1}{b}\sum_{j=1}^{b}\left(\log G{\lambda_j}^{-2d} + \frac{1}{G} I_{\Delta^d y}(\lambda_j)\right), \tag{9}$$

where $\lambda_j^{2d} I(\lambda_j)$ is replaced by $I_{\Delta^d y_t}(\lambda_j)$, that is the periodogram of the fractionally differenced series in correspondence of the true parameter. The objective function is then concentrated with respect to G, that is estimated by the averaged periodogram,

$$\hat{G}(d) = \frac{1}{b} \sum_{j=1}^{b} I_{\Delta^d y}(\lambda_j). \tag{10}$$

Shimotsu and Phillips (2005) prove consistency and by means of a simulation exercise, they highlight superiority of the exact local Whittle (ELW henceforth), with respect to alternative semiparametric estimators, in terms of MSE in finite samples.

We also implement the 2-step feasible exact local Whittle estimator ($FELW$ henceforth) of Shimotsu (2010) that extends the ELW estimator and it is robust to the presence of deterministic components in the data.

4 ML estimators

4.1 Beran Method

The approximated likelihood method, proposed by Beran (1994), is based on the best linear prediction of y_t, given all its past values as in equation (4), see Beran (1994). In particular, replacing the infinite sum in (4) by a finite number of past values

$$\tilde{\eta}_t = y_t - \sum_{j=1}^{t-1} \pi_j y_{t-j} \quad t = 2, \ldots, T \tag{11}$$

is equivalent to assume that $y_t = 0$ for $t \leq 0$. An approximated log likelihood function, is then given by

$$\log \mathrm{L}(y_t, \Theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2} \sum_{t=1}^{T} \frac{\tilde{\eta}_t^2}{\sigma_\eta^2}, \tag{12}$$

under the hypothesis that $\tilde{\eta} \sim iid\mathrm{N}(0, \sigma_\eta^2)$, where $\Theta = (d, \phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q, \sigma_\eta^2)$. The QMLE estimate of Θ is then provided by the maximization of equation (12). As has been noted by Beran (1994), the infinite autoregressive representation is not restricted to the case where y_t is stationary, but it can be extended to any $d > -\frac{1}{2}$. The approximate maximum likelihood estimator is therefore defined for any stationary and non stationary fractional processes. Moreover, Rossi and Santucci de Magistris (2009) show that this method can be adapted to non Gaussian innovations and extended to the multivariate case using copul*ae*.

4.2 Whittle approximation

Alternatively to the time domain approximated ML estimators, a fast numerical approach has been proposed by Fox and Taqqu (1986) and Dahlhaus (1989) in the frequency domain. In particular, the Whittle estimator is a ML estimate based on the calculation of the periodogram by means of the Fast Fourier Transform (FFT). Given that the log-likelihood of a Gaussian process, $Y_t = \{y_1, \ldots, y_T\}$ can be written as

$$\log \mathrm{L}(Y_t, \Theta) = -\frac{1}{2T} \log |\Gamma_\Theta| - \frac{1}{2T} Y_t' \Gamma_\Theta^{-1} Y_t \tag{13}$$

where Γ_Θ is the covariance matrix of Y_t, and $\Theta = (d, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$. The first term in (13) is approximated as

$$\frac{1}{2T} \log |\Gamma_\Theta| \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_\Theta(\lambda) \mathrm{d}\lambda \tag{14}$$

and the second as

$$\frac{1}{2T} Y_t' \Gamma_\Theta^{-1} Y_t \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_\Theta(\lambda)} \mathrm{d}\lambda \tag{15}$$

The discrete version of the estimator approximates the log-likelihood function as

$$\log \mathrm{L}(Y_t, \Theta) = -\frac{1}{2T} \left[\sum_{j=1}^{T} \log f_\Theta(\lambda_j) + \sum_{j=1}^{T} \frac{I(\lambda_j)}{f_\Theta(\lambda_j)} \right] \tag{16}$$

The asymptotic distribution of this estimator is stated in Fox and Taqqu (1986).

5 State Space Form for Long Memory Processes

5.1 General introduction

Although a long-range dependent process has an infinite-dimensional state space representation, Chan and Palma (1998) proved that the likelihood of an ARFIMA process can be computed, by means of the KF, in a finite number of steps. For instance, the ARFIMA(p, d, q) model has linear MA or AR representations given by formulae (3) and (4). To make the KF recursions feasible, Chan and Palma (1998) consider an approximation of equations (3) and (4) based on a truncation up to lag m. In this study, both kinds of representation will be considered. Finally, Chan and Palma (1998) provide the asymptotic properties of these approximate maximum likelihood estimates. The interesting results is that under mild regularity conditions the maximum likelihood estimators are both consistent and efficient. It is worth noticing that Palma (2007) suggest to exploit the MA approximation in first differences, which guarantees computational advantages. However, the first difference introduces an MA unit root term in the measurement equation when considering the possible presence of measurement errors, see Section 6.3. Therefore, the analysis of the MA in first differences is excluded from the present analysis and a solution of the problem outlined above is left for future research.

5.2 State Space Form

The state space representation consists of two equations. The first is the *measurement equation*, which relates the time series y_t to the state vector:

$$y_t = \mathrm{Z}\alpha_t + \mathrm{D}\varepsilon_t, \qquad t = 1, 2, \ldots, T, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2), \tag{17}$$

where Z is a $1 \times m$ matrix.The second term is the *transition equation*, that defines the evolution of the state vector α_t as a first order vector autoregression:

$$\alpha_{t+1} = \mathrm{T}\alpha_t + \mathrm{H}\eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \mathrm{Q}), \tag{18}$$

where T is $m \times m$ and H is $m \times g$ selection matrix, and η_t is a $g \times 1$ disturbance vector. Q is an $m \times m$ matrix of zeros, whose first element is equal to σ_η^2.

Chan and Palma (1998) and Palma (2007) show that there are basically two ways to rewrite a long memory model in state space form. The first is based on the AR(∞) representation and the second is based on the MA(∞) representation. Selecting a truncation lag m that is large enough, it is sufficient for the evaluation of the likelihood. This leads to an AR(m) and an MA(m), that can be casted in a state space form and estimated using the KF recursions, see Harvey (1989) and Harvey and Proietti (2005) for an introduction.

The AR(m) approximation, (SS-AR henceforth), can be written as follows:

$$\begin{aligned} &\mathrm{Z} = [1, 0, \ldots, 0], \qquad \mathrm{D} = 0, \qquad t = 1, 2, \ldots, T, \\ &\mathrm{T} = \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_m \\ 0 & & I_{m-1} & \end{pmatrix}, \qquad \mathrm{H} = \begin{pmatrix} 1 & 0 & \ldots & 0 \end{pmatrix}. \end{aligned} \tag{19}$$

where π_j for $j = 1, ..., m$ comes from (4), see Hosking (1981).

The MA(m) approximation, (SS-MA henceforth), can be written as follows:

$$\begin{aligned}
&\mathrm{Z} = [1, 0, \ldots, 0], \quad \mathrm{D} = 0, \quad t = 1, 2, \ldots, T, \\
&\mathrm{T} = \begin{pmatrix} 0 & I_{m-1} \\ 0 & 0 \end{pmatrix}, \quad \mathrm{H} = \begin{pmatrix} 1 & \psi_1 & \psi_2 & \ldots & \psi_m \end{pmatrix}.
\end{aligned} \tag{20}$$

where ψ_j for $j = 1, ..., m$ comes from (3), see Hosking (1981).

The approximated representation of a causal ARFIMA(p, d, q) has computational advantages over the exact one of Sowell (1992). The order of the MLE algorithm is reduced from $O(n^3)$ to $O(n)$.[1]

5.3 Kalman Filter

The KF (see Harvey, 1989, and Durbin and Koopman, 2001) is a fundamental algorithm for the statistical treatment of a state space model. Under the Gaussian assumption it produces the minimum mean square estimator of the state vector along with its mean square error matrix, conditional on past information; this is used to build the one-step-ahead predictor of y_t and its mean square error matrix. Due to the independence of the one-step-ahead prediction errors, the likelihood can be evaluated via the prediction error decomposition, see Schweppe (1965).

The simple case corresponds to the equations (17)-(18), with $\alpha_1 \sim \mathrm{N}(\alpha_{1|0}, \mathrm{P}_{1|0})$, where $\alpha_{1|0}$ and $\mathrm{P}_{1|0}$ need to be specified, see below. Then, defining $Y_t = \{y_1, y_2, \ldots, y_t\}$ and the conditional mean and variance of the state vector, $\tilde{\alpha}_{t|t-1} = \mathrm{E}(\alpha_t|Y_{t-1})$, $\mathrm{Var}(\alpha_t|Y_{t-1}) = \mathrm{P}_{t|t-1}$, the KF is given by the following recursive formulae and definitions for $t = 1, \ldots, T$:

$$\begin{aligned}
&\nu_t = y_t - \mathrm{Z}\tilde{\alpha}_{t|t-1}, && \mathrm{F}_t = \mathrm{Z}\mathrm{P}_{t|t-1}\mathrm{Z}' + \sigma_\varepsilon^2 \mathrm{D}\mathrm{D}', \\
&\mathrm{K}_t = \mathrm{T}\mathrm{P}_{t|t-1}\mathrm{Z}'\mathrm{F}_t^{-1}, && \mathrm{L}_t = \mathrm{T} - \mathrm{K}_t\mathrm{Z} \\
&\tilde{\alpha}_{t+1|t} = \mathrm{T}\tilde{\alpha}_{t|t-1} + \mathrm{K}_t\nu_t, && \mathrm{P}_{t+1|t} = \mathrm{T}\mathrm{P}_{t|t-1}\mathrm{L}_t' + \mathrm{H}\mathrm{Q}_t\mathrm{H}'.
\end{aligned} \tag{21}$$

The matrix K_t is the so called *Kalman gain*, and $\nu_t = y_t - \mathrm{E}(y_t|Y_{t-1})$ are the innovations or one-step-ahead prediction errors, with variance matrix F_t. The KF is thus an algorithm that receives as an input the observations and the system matrices, and returns the innovations, the one-step-ahead forecasts of the states, along with their conditional covariance matrix. Recalling that the conditional expectation is the minimum mean square estimator of the state vector,

[1]Due to computational burden, the exact maximum likelihood estimator of Sowell (1992) is not considered in the simulations.

when the Gaussianity assumption is removed the KF still yields the minimum mean square linear estimator, see section 6 and section 7. In the case when $\alpha_{1|0}$ and $P_{1|0}$ are known and finite, the KF enables the likelihood function to be written in prediction error decomposition, that excluding constant terms, is given by:

$$\log \mathrm{L}(\Theta) = \frac{1}{2}\left\{\sum_{t=1}^{T} \log |\mathrm{F}_t| + \sum_{t=1}^{T} \frac{\nu_t^2}{\mathrm{F}_t}\right\} \tag{22}$$

where $\Theta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d, \sigma_\eta^2)$ is the parameter vector associated with the ARFIMA(p, d, q) representation (1). The Kalman filter estimator, $\hat{\Theta}_{KF}$, is the parameter vector which maximizes (22). In order to evaluate the log-likelihood function, the initial conditions are chosen as $\alpha_{1|0} = \mathrm{E}[\alpha_1] = 0$ and $\mathrm{P}_{1|0} = \mathrm{E}[\alpha_1 \alpha_1^{'}]$. For the autoregressive approximation, the $\mathrm{P}_{1|0}$ matrix is set equal to the empirical autocovariances up to lag m, and it has a Toeplitz structure. For the moving average representation, following Chan and Palma (1998) and Palma (2007), the initialization of $\mathrm{P}_{1|0}$ is $\mathrm{P}_{1|0} = |\omega(i,j)|_{i,j=1,2,\ldots,m}$ where $\omega(i,j) = \sum_{k=0}^{\infty} \psi_{i+k}\psi_{j+k}$.
Chan and Palma (1998) show the following results:

- **Consistency**: Assume that $m = T^k$ with $k > 0$, then as $T \rightarrow \infty$, $\hat{\Theta}_{KF} \rightarrow \Theta_0$ in probability.
- **Asymptotic Normality**: Assume that $m = T^k$ with $k > 1/2$, then as $T \rightarrow \infty$, $\sqrt{T}(\hat{\Theta}_{KF} - \Theta_0) \rightarrow N(0, \Sigma(\Theta_0))$.
- **Efficiency**: Assume that $m = T^k$ with $k > 1/2$, then $\hat{\Theta}_{KF}$ is an efficient estimator.

It next section, the finite sample performance of the state space approach will be investigated by means of several Monte Carlo simulations.

6 Monte Carlo Simulations

In this section a finite sample comparison of the estimation methods outlined above is carried out by means of Monte Carlo simulations. The objective of this exercise is to shed some light on which estimator is the most accurate in practical applications, with realistic sample sizes. The Monte Carlo experiment is based on 1,000 artificial time series with 250, 500 and 1000 observations. The simulations are performed as follows: first a vector of iid innovations, η_t, is simulated from a standard Normal distribution. A long memory series of order $d > 0$,

$z_t = (1-L)^{-d}\eta_t$, is generated by means of the MA expansion in equation (3). Finally, an ARMA structure is added to z_t. Haldrup and Nielsen (2007) show that there are no differences among the semiparametric and the parametric estimators when the data are generated with the simulation scheme used here, or unconditionally, by the Cholesky decomposition of the $T \times T$ autocovariance matrix. We do not report all the tables; the missing tables are available from the authors upon request. The sample size is selected to reflect realistic empirical samples of macroeconomic or financial datasets. Although financial samples may be larger than the sample sizes considered here, empirical analysis is frequently based on some aggregated measures such as monthly realized volatility, so that the sample sizes considered here are relevant. We study four different data generating processes (DGPs) in our Monte Carlo simulations. The first one is the simple ARFIMA $(0, d, 0)$ model

$$(1-L)^d(y_t - \mu) = \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{23}$$

where the parameters values $\mu = 0$ and $\sigma^2 = 1$ are chosen for the simulations.[2] For the parameter d, we consider the values $\{0, 0.4\}$. The case $d = 0$ corresponds to estimating d to the case where the process is not long memory. The other three models considered are the ARFIMA$(1, d, 0)$, ARFIMA$(0, d, 1)$ and an ARFIMA$(1, d, 1)$ given by:

$$(1-\phi L)(1-L)^d(y_t - \mu) = \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{24}$$

$$(1-L)^d(y_t - \mu) = (1+\theta L)\eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{25}$$

$$(1-\phi L)(1-L)^d(y_t - \mu) = (1+\theta L)\eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{26}$$

with $\phi = -0.5$ and $\theta = -0.4$ in all cases.[3]

An useful preliminary assessment is related to the choice of the truncation lag, which can have an influence on the parameter estimates. Bondon and Palma (2005) show that the quality of the AR truncation depends asymptotically only on d and not on the short memory components. For different choices of d, ϕ and θ the authors find that the convergence is good, even for small values of m. Unfortunately, their result does not consider the fact that, in finite samples, increasing

[2]A non zero mean could also be added to model (23). In that case, the proposed parametric methods should be computed on the series in difference from the sample mean, say on $x_t = y_t - \bar{y}_t$, with $\bar{y}_t = \frac{1}{T}\sum_{t=1}^{T} y_t$, see also Chan and Palma (1998).

[3]Simulation with $\phi = 0.5$ and $\theta = 0.4$ are also provided. They are available upon request to the authors.

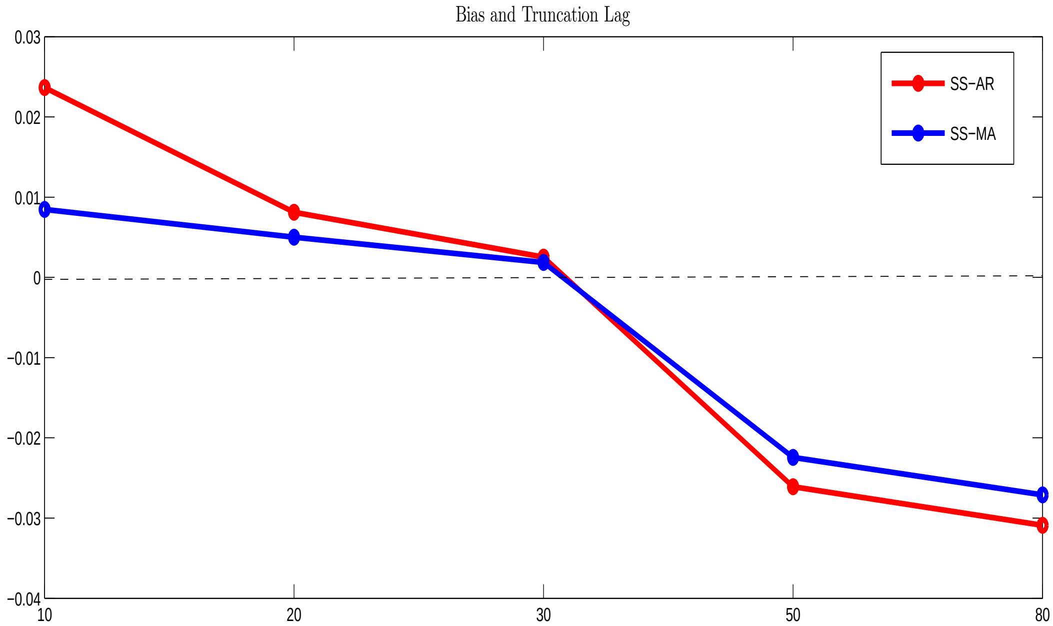
m leads to discard the first m observations. Figure 1(a) plots the estimation bias for d as a function of the truncation lag, m, for the $SS-MA$ and $SS-AR$, based on an ARFIMA$(0, d, 0)$ with $d = 0.4$ and $T = 500$. It is clear that, in both cases, increasing the truncation lag leads to better parameter estimates. However, choosing $m > 30$ when $T = 500$ leads to an increase of the bias, since too many initial observations are lost due to the truncation. The same evidence is obtained when the sample size is 1000, see Figure 1(b). In particular, the minimum bias is reached when $T = 50$, while using 80 lags introduces a larger downward bias. In light of this evidence, the truncation lag used in the paper is 30 for $T \leq 500$, and 50 for $T > 500$.

Tables 1 and 2 report the Monte Carlo results in the case of Gaussian innovations and correct specification of the model. It is evident that both semiparametric and parametric estimators generally provide correct estimates of the long memory parameter for all specifications. There is a slight downward bias in the semiparametric estimates for the ARFIMA$(1, d, 1)$ when $T = 250$ and $T = 500$. This is due to the fact that, in small samples, the estimates of the spectral density are highly affected by the ARMA parameters. The SS-AR and SS-MA provide estimates of d that are well centered on the true value. The same holds true for the Beran estimator, which has generally the lowest bias and RMSE. In particular, the RMSE of the parametric models are generally lower than the corresponding semiparametric estimates. This results could be easily anticipated provided that we are assuming that the short run dynamics are correctly specified when estimating the ARFIMA with the parametric models.

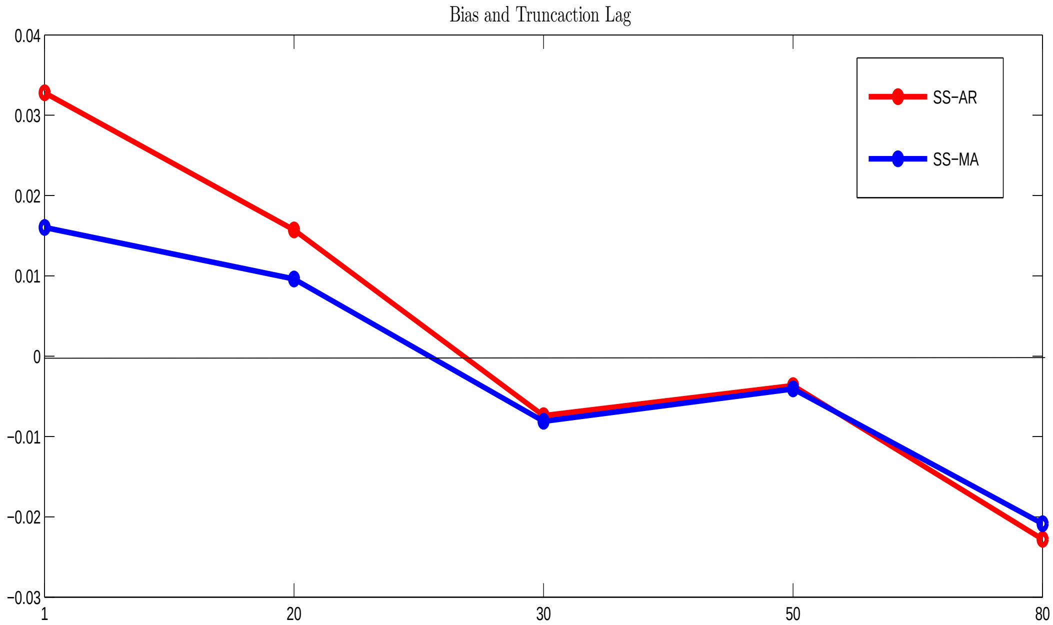
Therefore, in subsection 6.1 we investigate the effect of model misspecification on the estimates of the d parameter; non-Gaussian shocks are considered as well as a misspecification in the ARMA component of the process. The problem of missing values is considered in subsection 6.2. Subsection 6.3 and 6.4 consider the case of measurement errors and random level shifts, respectively.

6.1 Model Misspecification

The purpose of this section is to show how far, from the true parameter value, the parametric estimates of d are located, when the model is not correctly specified. It is clear that if the ARMA terms of model (1) are not correctly identified, then the parametric estimators are not able to provide reliable estimates of d. Three sources of misspecification are therefore investigated here: non-Gaussian errors, under-specification and over-specification of the short run dynamics of the model.



(a) Bias as a function of m. Model is an ARFIMA$(0, d, 0)$ with $d = 0.4$ and $T = 500$.



(b) Bias as a function of m. Model is an ARFIMA$(0, d, 0)$ with $d = 0.4$ and $T = 1000$.

For what concerns the robustness of the different estimators in case of errors that are non-Gaussian, we consider models (23)-(26) with Student's t innovations with three degrees of freedom. Table 4 reports the estimation results for an ARFIMA(p, d, q), with $d = 0.4$ and $T = \{250, 500\}$. As already found in Haldrup and Nielsen (2007) and Nielsen and Frederiksen (2005), the semiparametric estimators are robust to the misspecification of the error term, such as neglected GARCH effects as well as non-Gaussian errors. On the other hand, despite the KF is not the optimal estimator with non-Gaussian errors (see Durbin and Koopman (2001)), it turns out to be reliable also in this case. In particular the bias and the RMSE are of the same order as those reported in table 2. This holds true also for the Beran and the Whittle estimators. Further improvements can be achieved with the non linear filter such as the extended KF, but this is beyond the scope of this paper.

Table 5 shows the consequences of the over-specification in the short run dynamics. In case of over-specification, the true DGP is an ARFIMA$(0, d, 0)$, but an ARFIMA$(1, d, 1)$ is estimated on the simulated series. Obviously, estimating ϕ or θ, or both, when they are not present (i.e. overfitting the model) may reduce the efficiency of the estimates of the parameter d. This clearly emerges from the Table. In particular, all the parametric estimators, including SS-AR and SS-MA, have small bias but large RMSE when compared to the semiparametric ones.

On the other hand, under-specification means that the true DGP is an ARFIMA$(1, d, 1)$, but an ARFIMA$(0, d, 0)$ is estimated, completely neglecting the ARMA components. This is a strong form of misspecification, and it can cause large biases in the parametric estimators. The estimates of the parameter d are largely downward biased for the parametric estimators, while, as expected, the semiparametric ones are robust also in this case. However, as suggested in Iglesias et al. (2006), when adopting one of the usual likelihood based information criteria, such as the Akaike's, allows to select the right model in all cases. Therefore, in practical applications, when the true DGP is unknown, it could be useful to adopt one of the information criteria for the selection of the short run dynamics.

6.2 Missing Values

Missing data are a common feature in economic and financial time series. A considerable advantage of the state space approach is the ease with which the missing observations can be dealt with. The KF is able to calculate the prediction errors even in presence of missing values, this is based on a skipping approach, see Durbin and Koopman (2001).

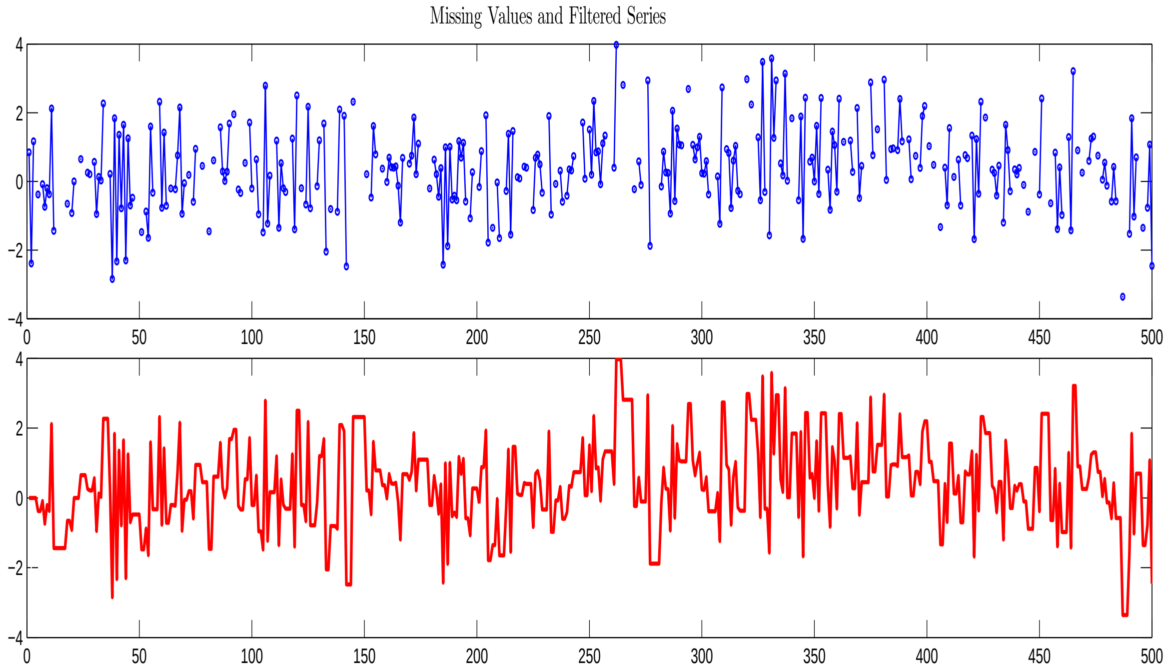


Figure 1: Upper graph: ARFIMA process with missing. Bottom graph: The smoothed estimate of the ARFIMA process.

Theoretical derivation of the prediction error variance for state space long memory in case of missing values has been provided by Palma and del Pino (1999). The authors show that the information is lost during the block of missing observations at a much slower rate than it is gained after that gap. However, they do not provide an extensive Monte Carlo exercise to check the behavior of state space model for long memory in presence of missing data and short run dynamics. In particular, a straightforward extension of the equation (21), allows to use the state space approach in the case of missing observations,

$$
\begin{aligned}
&\nu_t = y_t - \mathrm{Z}\tilde{\alpha}_{t|t-1}, && \mathrm{F}_t = \mathrm{ZP}_{t|t-1}\mathrm{Z}' + \sigma_\varepsilon^2 \mathrm{DD}', \\
&\mathrm{K}_t = \mathrm{TP}_{t|t-1}\mathrm{Z}'\mathrm{F}_t^{-1}, && \mathrm{L}_t = \mathrm{T} - \mathrm{K}_t\mathrm{Z} \\
&\tilde{\alpha}_{t+1|t} = \begin{cases} \mathrm{T}\tilde{\alpha}_{t|t-1} + \mathrm{K}_t\nu_t, & \text{No Missing Values} \\ \mathrm{T}\tilde{\alpha}_{t|t-1} & \text{Missing Values} \end{cases} \\
&\mathrm{P}_{t+1|t} = \begin{cases} \mathrm{TP}_{t|t-1}\mathrm{L}_t' + \mathrm{HQ}_t\mathrm{H}' & \text{No Missing Values} \\ \mathrm{TP}_{t|t-1}\mathrm{T}' + \mathrm{HQ}_t\mathrm{H}' & \text{Missing Values} \end{cases}
\end{aligned}
\tag{27}
$$

so that the latent long memory process can be easily tracked by means of this KF recursion.

As example, we report in Figure 1 a simulated long memory process with 40% percent of missing values, blue line. The figure highlights the fact that the KF, red line, is able to provide a good tracking of the latent long memory process. On the other hand, the alternative estimators considered in this paper cannot directly handle this kind of data irregularity, so that the missing values are replaced with the last available observation.

Table 6 reports the results models of formula (26) with $T = \{250, 500\}$. The number of missing values considered are 10%, 20%, 30% and 40%. Those results are clearly in favor of the state space methods. The bias of the SS-AR and the SS-MA estimators generally range between 0.03 and 0.05, while other estimators are extremely downward biased. The RMSE are smaller for the state space approaches. It is noteworthy the fact that the bias of SS-AR and SS-MA does not increase as the percentage of missing values increases, so that the relative performance of the state space estimators of long memory improves, compared to the standard estimators.

6.3 Measurement Error

Consider a generic ARFIMA(p, d, q) process, y_t. In case of measurement errors, y_t is a latent process which cannot be observed due to the contamination in the data. What we observe instead is the noisy series ζ_t, defined as follows:

$$\zeta_t = y_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2). \tag{28}$$

A large literature, see Deo and Hurvich (2001), Hurvich et al. (2005) and Haldrup and Nielsen (2007), discusses the properties of the semiparametric long memory estimators, such as the log-periodogram regression and the local Whittle estimator, when the long memory signal is contaminated by a noise term. Deo and Hurvich (2001) show that the Geweke and Porter-Hudak (1984) estimator (GPH) is biased by a constant factor that depends on the variance of the noise term. Sun and Phillips (2003) suggest to introduce an additional term in the log-periodogram regression, $\beta\lambda^{2d}$ to account for the effect of the additive noise term, that is allowed to be weakly dependent. Arteche (2004) suggests that an optimal choice of the bandwidth, b, is important to minimize the influence of the added noise term, since the variance of the measurement error heavily restricts the allowable bandwidth in finite samples. With a larger variance of the noise with respect to the signal, only frequencies very close to the origin contain a valuable information. We compute here also the corrected local Whittle estimator (CLW henceforth) by Hurvich and Ray (2003) and Hurvich et al. (2005), which is designed to account for the extra constant term that enters in the spectral density when the long memory signal is contaminated by an *iid* noise. They propose to modify the local Whittle objective function as

$$Q(G,d,\beta)=\frac{1}{b}\sum_{j=1}^{b}\left\{\log\left[G\lambda_j^{-2d}(1+\beta\lambda_j^{2d})\right]+\frac{\lambda_j^{2d}I_\zeta(\lambda_j)}{G(1+\beta\lambda_j^{2d})}\right\}, \tag{29}$$

Concentrating G out, it yields

$$R(d,\beta)=\frac{1}{b}\sum_{j=1}^{b}\log\left(\lambda_j^{-2d}(1+\beta\lambda_j^{2d})\right)+\log\left(\frac{1}{b}\sum_{j=1}^{b}\frac{\lambda_j^{2d}I_\zeta(\lambda_j)}{(1+\beta\lambda_j^{2d})}\right), \tag{30}$$

where the local Whittle estimator is obtained setting $\beta=0$ in the minimization of R. The local Whittle estimates of d and β are

$$(\hat{d}_c,\hat{\beta})=\arg\min_{(d,\beta)\in\mathbb{D}\times\mathbb{B}}\hat{R}(d,\beta) \tag{31}$$

where $\mathbb{D}$ and $\mathbb{B}$ are the admissible sets of d and β.

The KF with the smoothing recursions (see Durbin and Koopman (2001)) is able to estimate the latent process and to provide a smoothed estimate of the long memory signal, $\hat{y}_t$, that is disentangled from the estimated noise component $\hat{\varepsilon}_t$. This case corresponds to the well-known AR(m) or MA(m) plus noise model that has the state space representations presented in equations (19) and (20). The only difference is that the D matrix must be set equal to 1. The log-likelihood has to be maximized with respect to the following parameter vector $\theta=(\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,d,\sigma_\eta^2,\sigma_\varepsilon^2)$. In such a case, a computational efficient algorithm is based on the concentrated likelihood, with respect to σ_ε^2, see Durbin and Koopman (2001). In this context, a key quantity is the so called *noise-to-signal ratio*, *nsr* henceforth, which is the ratio between the variance of the noise over the variance of the signal. Differently from Haldrup and Nielsen (2007), who set $nsr=\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}$, our Monte Carlo simulations are carried out by setting the nsr proportional to the underlying long memory signal. In particular, setting $nsr=1$ and $nsr=2$ means that the noise term has a variance that is 1 and 2 times larger than the variance of the signal, respectively. Table 7 reports the estimation results for the case when the long memory signal is contaminated by an error term. As it is emerges from the tables, the state space estimators clearly outperform other estimators in term of bias and RMSE. As found in previous studies, the usual semiparametric and the parametric estimators are extremely downward biased, since the contamination due to the measurement error hides the long memory signal, resulting in

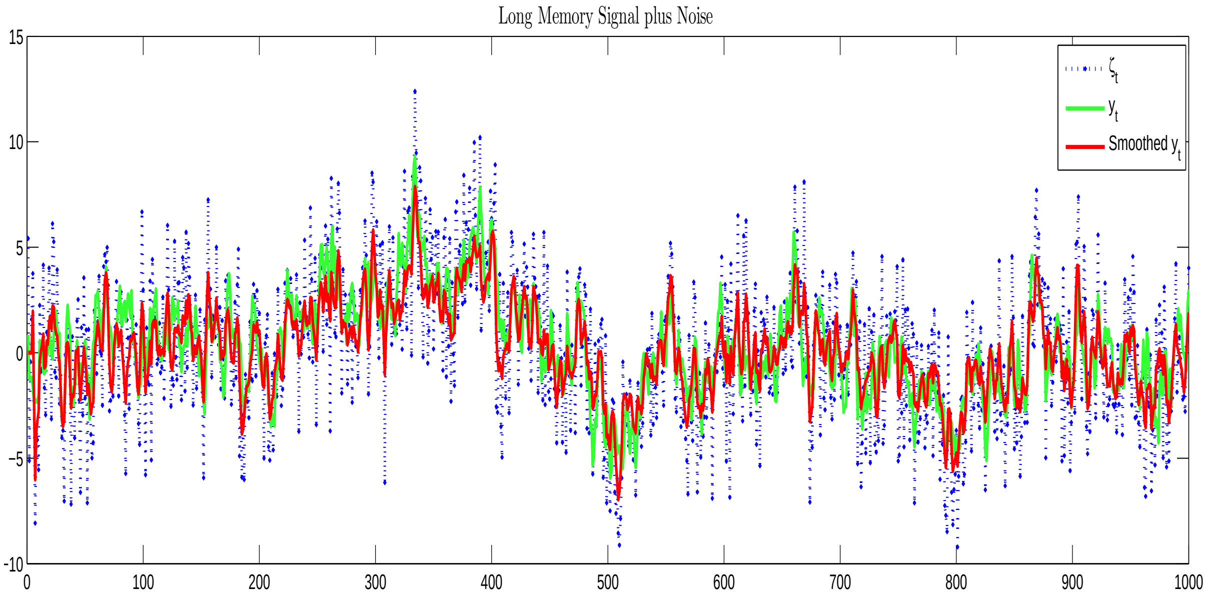


Figure 2: Dotted blue line: Series ζ_t contaminated by noise. Green line: True ARFIMA$(0.5, 0.4, 0)$. Red line: Smoothed estimate of the true ARFIMA$(0.5, 0.4, 0)$

lower estimates. Modifying the bandwidth parameter, b, for the semiparametric estimators does not change the results substantially. On the other hand, the CLW estimator, which incorporates an estimate of the nsr, slightly overestimates d and it has a large $RMSE$. This is probably due to the fact that the sample size is too small for this semiparametric estimator. For example, Hurvich et al. (2005) show, by means of Monte Carlo simulations, that the CLW has good properties for large sample sizes. On the other hand, the state space approach provides better estimates of the long memory parameter. In particular, with the exception of the ARFIMA$(0, d, 1)$, the estimates of d have small biased and smaller $RMSE$ than other estimators, especially when $nsr = 2$. In the case of ARFIMA$(0, d, 1)$, the state space methods slightly underestimate the true parameter. This is perhaps due to the fact that a negative MA term, $\theta = -0.4$, reduces the first order autocorrelation thus resulting in a less clear evidence of long memory, so that the signal is more hardly detected.

Figure 2 reports a realization of an ARFIMA $(0.5, 0.4, 0)$ plus a noise process, blue line, with $nsr = 1$. The true long memory signal, in green, is plotted together with the estimated smoothed signal, red line. It is evident that, based on the parameter estimates, the KF is able to provide a good tracking of the underlying long memory series. These results highlights the usefulness of the state space approach, so that filtering techniques can be exploited when a forecast of the underlying signal is needed.

6.4 Level Shifts

A recent debate in literature focuses on the possible confusion between truly long memory processes and short memory processes with structural changes in levels, see among others Perron and Qu (2010). It is well known that, when a stationary process is contaminated by level shifts, the fractional integration order is positively biased, so that long memory may just be the spurious effect of the presence of one or more level shifts. Granger and Hyung (2004) show that the long memory property of volatility could be induced spuriously by the presence of structural breaks in the mean, and, when removing them, it results in a weaker evidence of long memory. Several testing procedures have been carried out in order to assess the presence of true or spurious long memory. A recent paper by Ohanissian et al. (2008) exploits the self-similarity condition of truly long memory to test whether the long run dependence in the data is due to a truly long memory process. Dolado et al. (2005) illustrate how the slow hyperbolic decay of the autocorrelations, which is typical of the long memory series, could be generated by a short memory process plus level shifts. They provide an extension of the well known Dickey-Fuller statistic to test for the null hypothesis of true long memory against the alternative of a $I(0)$ process plus structural breaks. Shimotsu (2006) provides a test based on the sample splitting, where, under the null of true long memory, each subsample has the same degree of fractional integration. More recently, Perron and Qu (2010) propose a testing procedure which is based on the evaluation at different bandwidths of the GPH estimator.

In a state space setting, Lu and Perron (2010) present an univariate random shifts model for stock volatilities, exploiting a procedure to estimate the probability and the magnitude of the shifts, and showing that the random occurrence of shifts induces spurious long memory. The model proposed by Lu and Perron (2010) allows for a random level shift process plus AR dynamics. We extend their methodology, assuming that the data generating process is given by the sum of a random level shift and an ARFIMA process. This model setup allows to correctly estimate the long memory parameter, when a long memory process is contaminated by level shifts. We assume the following DGP

$$x_t = \tau_t + y_t, \tag{32}$$

where τ_t is the random shift component,

$$\tau_t = \tau_{t-1} + \gamma_t \kappa_t, \tag{33}$$

with $\kappa_t \sim N(0, \sigma_\kappa^2)$ and $\delta_t = \gamma_t \kappa_t$ where γ_t is a binomial variable that takes value 1 with probability α, so that if $\gamma_t = 1$ a level shift occurs. y_t is an ARFIMA(p, d, q) process, which is defined by its autoregressive representation (4). Taking the first differences as

$$\Delta x_t \quad = \quad y_t - y_{t-1} + \delta_t \tag{34}$$

so that the corresponding state-space form is

$$\Delta y_t \quad = \quad \mathrm{Z}\alpha_t + \delta_t \tag{35}$$

$$\alpha_t \quad = \quad \mathrm{T}\alpha_{t-1} + \mathrm{H}\eta_t, \tag{36}$$

where $\mathrm{Z} = [1, -1, 0, \ldots, 0]$ is a $1 \times m$ vector, which selects the elements of α_t. $\alpha_t = [y_t, y_{t-1}, \ldots, y_{t-m}]'$ is an $m \times 1$ vector and η_t is a $m \times 1$ vector of innovations with a diagonal variance covariance matrix Q. Finally $\mathrm{H} = [1, 0, \ldots, 0]'$ is a selection matrix. The $m \times m$ transition matrix T, is defined as

$$\mathrm{T} = \begin{bmatrix} \pi_1 & \pi_1 & \ldots & \pi_m \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{37}$$

where the first row contains the parameters of the truncated AR expansion of an ARFIMA(p, d, q), see (4). The state space formulation (36) admits the KF recursions with the modification proposed by Lu and Perron (2010). We call it *MSS-AR*.[4] The procedure provides the estimates of the shifts parameters α and σ_κ^2, and the estimates of the ARFIMA parameters, allowing to disentangle the long memory components from the level shifts process. It is interesting to highlight the fact that the modified KF allows to test for the presence of long memory or level shifts, or both, by testing the nullity of the parameters α, σ_κ^2 and d.

Figure 3, reports a simulated trajectory of process (32), for 500 observations. It is clear

[4]Due to constraints on the number of pages of the present paper, we do not report the details on the modified KF technique. A detailed description of this method can be found in Lu and Perron (2010).

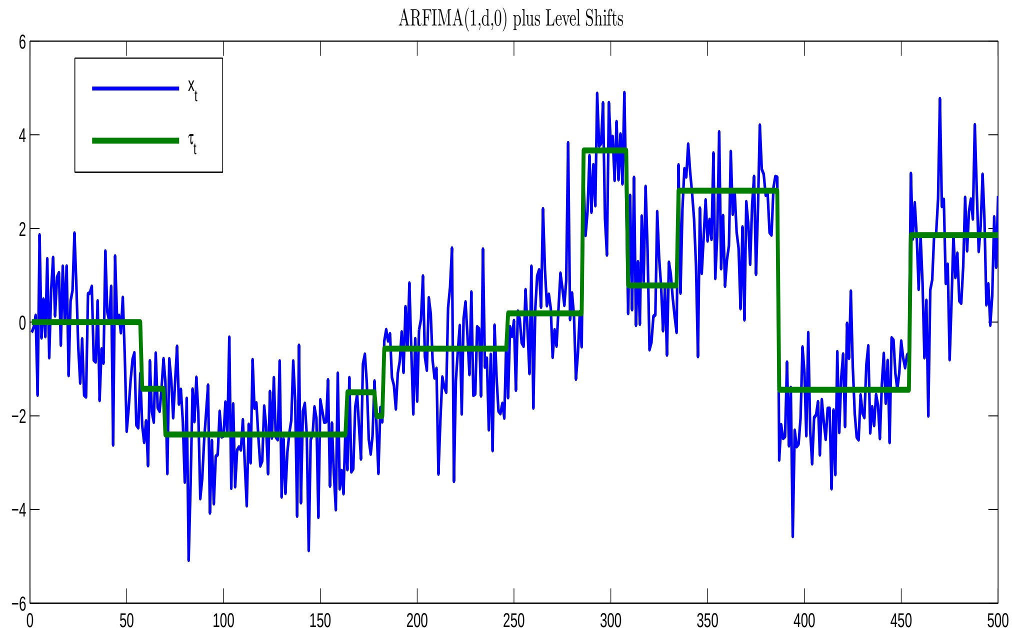


Figure 3: A simulated trajectory of process (32), with $\sigma_\kappa = 4$, $\alpha = 0.015$, $d = 0.4$ and $\phi = -0.3$.

that the level shift process, green line, is added to a long memory term, resulting into an highly persistent time series, blue line, where it is accentuated the presence of long periods above and below the unconditional mean. For example, this is likely to be the case of the realized volatilities. As noted by Raggi and Bordignon (2010) in a Bayesian framework, after removing the significant level shifts from the realized volatility series, there is still presence of long memory. However, the degree of (fractional) integration of a long memory plus level shift needs a careful theoretical investigation and it is beyond the scope of the present paper. As expected, the simulation results on the level shifts process highlight the fact that usual long memory estimators are highly upward biased, when the shifts are present. On the other hand, the modified KF technique is able to provide corrected estimates of the long memory parameter as well as of the shift parameters (not reported). It is also interesting to note that the proposed MSS-AR is unbiased also in the case of a pure long memory process, setting $\sigma_k^2 = 0$ and $\alpha = 0$, or a pure level shift process, setting $d = 0$, see bottom lines of Table 8. Given these results, we can figure out a parametric testing procedure, such as a LR test, for truly long memory versus level shifts based on the state space form representation. This aspect will be investigated in future research.

7 Empirical analysis

The purpose of this section is to highlight some interesting financial and economic applications of the state space approach in the long memory framework. Two examples are provided in order to show the advantages, with respect to the usual estimators, of using the KF methods in the empirical analysis. The first example is a simple version of the long memory stochastic volatility model, $LMSV$ henceforth, which is investigated by Breidt et al. (1998) and Hurvich et al. (2005), among many others. In particular, the $LMSV$ model is based on the following assumptions on the log-returns, r_t,

$$r_t = \sigma \cdot \exp(y_t/2) \cdot z_t \tag{38}$$

where $\sigma_t = \sigma \cdot \exp(y_t/2)$ is the volatility of r_t, and y_t is the long memory component of the stochastic volatility. z_t is a iid standard Gaussian innovation, independent of y_t. Taking $\zeta_t = \log r_t^2$ yields

$$\zeta_t = \mu + y_t + \varepsilon_t \tag{39}$$

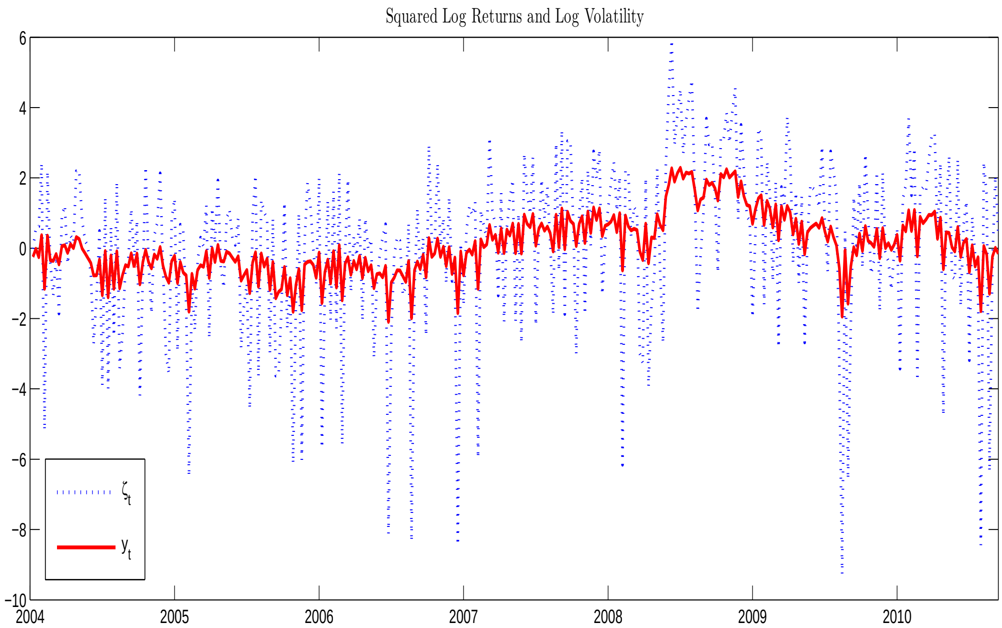
where $\mu = \log(\sigma^2) + \mathrm{E}(\log(z_t^2))$ is the unconditional mean, and $\varepsilon_t = \log(z_t^2) - \mathrm{E}(\log(z_t^2))$. Therefore, ε_t is a zero mean iid sequence, and ζ_t is a long memory process plus an iid noise. As in Harvey et al. (1994), we treat ε_t as it were $iidN(0, \sigma_\varepsilon^2)$, where $\sigma_\varepsilon^2 = \frac{\pi^2}{2}$, is treated as an additional unknown parameter in the state space representation. The rationale behind this choice is to verify *ex-post* whether the Kalman Filter, despite not being the minimum mean square estimator in this case, is able to provide an estimate of σ_ε^2 that is reasonably close to $\frac{\pi^2}{2}$. This could be taken as an evidence that the Kalman filter is able to disentangle the long memory signal from the noise. Assume now that, as in Comte and Renault (1998) and Casas and Gao (2008), the logarithm of the stochastic volatility in continuous time follows a fractional Ornstein-Uhlenbeck process:

$$\mathrm{d}\log y(t) = -k \log y(t)\mathrm{d}t + \gamma \mathrm{d}W_d(t) \tag{40}$$

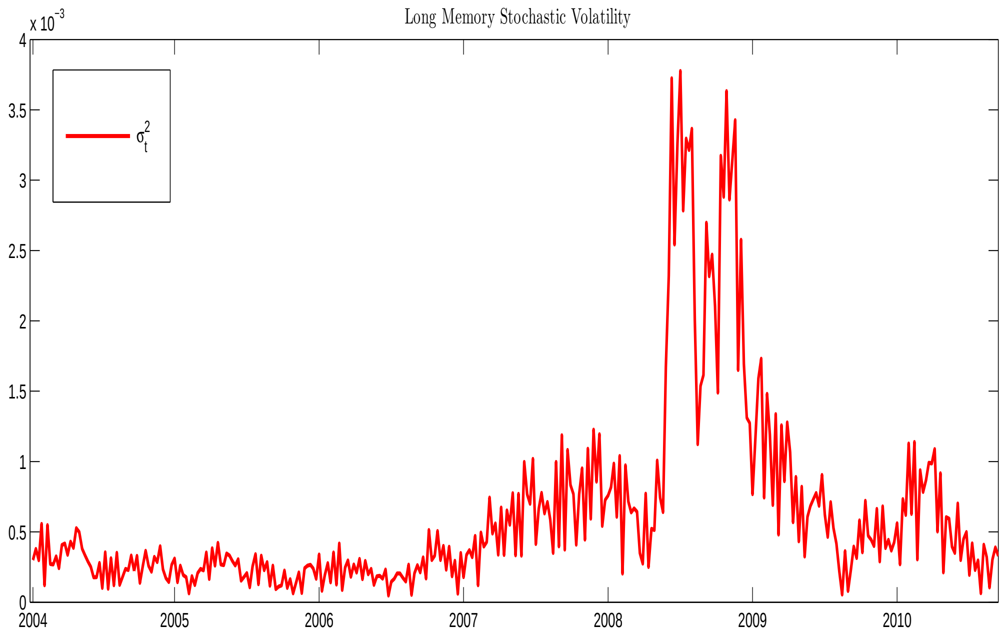
where $k > 0$ is the drift parameter and $W_d(t)$ is a fractional Brownian motion. It is therefore reasonable to assume that the discrete time counterpart of model (40) follows an ARFIMA$(1, d, 0)$ process. We therefore propose an ARFIMA$(1, d, 0)$ state space representation for y_t, allowing for the presence of a measurement error. Table 9 reports the results of the estimation of the long memory parameter for the LMSV model computed on the S&P with 415 weekly returns from January 1, 2003 to December 31, 2010. The estimates obtained with the traditional estimators of long memory are all below 0.3. On the other hand, CLW is larger than 0.8, with an estimated noise-to-signal ratio, $\hat{\beta}$, close to 80. It is evident that CLW over-estimates the degree of long memory in the LMSV, since we know from the theory that the variance of the measurement error, σ_ε^2, must be equal to $\frac{\pi^2}{2}$. This evidence confirms that, in small samples, the CLW is a weak estimator.[5] State space approaches provide instead an estimate of the long memory parameter close to 0.45, that is in line with the results provided in the financial literature, see among others Casas and Gao (2008). The state space proves to be reliable since the estimated σ_ε^2 is not far from the true value, both for AR and MA approximations. This means that the KF recursion is able to well disentangle the long memory signal from the noise, and to correctly estimate the parameters. Figure 4 plots the signal long memory series and the observed noisy series ζ_t. It is clear, from the Panel 4(a), that the signal has been extracted from an highly noisy series, blue line, and that the extracted variance series, Panel Panel 4(b), presents the long memory characteristic.

As second example, we apply the MSS-AR to model the monthly inflation series of US. A growing literature, dating back to the paper by Hassler and Wolters (1995), investigates the degree of long memory in the inflation series, see among others the recent papers by Sun and Phillips (2005) and Sibbertsen and Kruse (2009). As shown in Gadea and Mayoral (2006), following the argument by Zaffaroni (2004), long memory in the inflation series is consistent with an *la* Calvo (1983) price generating process. On the other hand, Hsu (2005) suggests that the dynamics of the inflation series could be characterized by level shifts. The author shows that inflation is a long memory stationary series, after the removal of the level shifts. However, Hsu (2005) methodology only allows for an unique break date. Table 10 reports the estimates of the MSS-AR on the seasonally adjusted U.S. City Average core consumer price index, P_t, of the Bureau of Labor Statistics (BLS), which excludes the direct effect of price changes for food and energy. The sample covers the period January 1965-February 2011. The series is sampled at monthly frequency, for a total of 516 observation. Inflation, is computed as $\pi_t = \log(P_t) - \log(P_{t-1})$. Figure 5 plots the monthly US inflation series, where it is clear the persistence in the dynamics, so that long periods of high levels of inflation rates during the 70's are followed by a long period of inflation in the second part of the sample.

[5] This result is obtained with any choices of the bandwidth, b.



(a) $\zeta_t = log(r_t^2)$, blue line, and Smoothed KF estimate of y_t, red line.



(b) Smoothed estimate of σ_t^2.

Figure 4: Smoothed volatility series

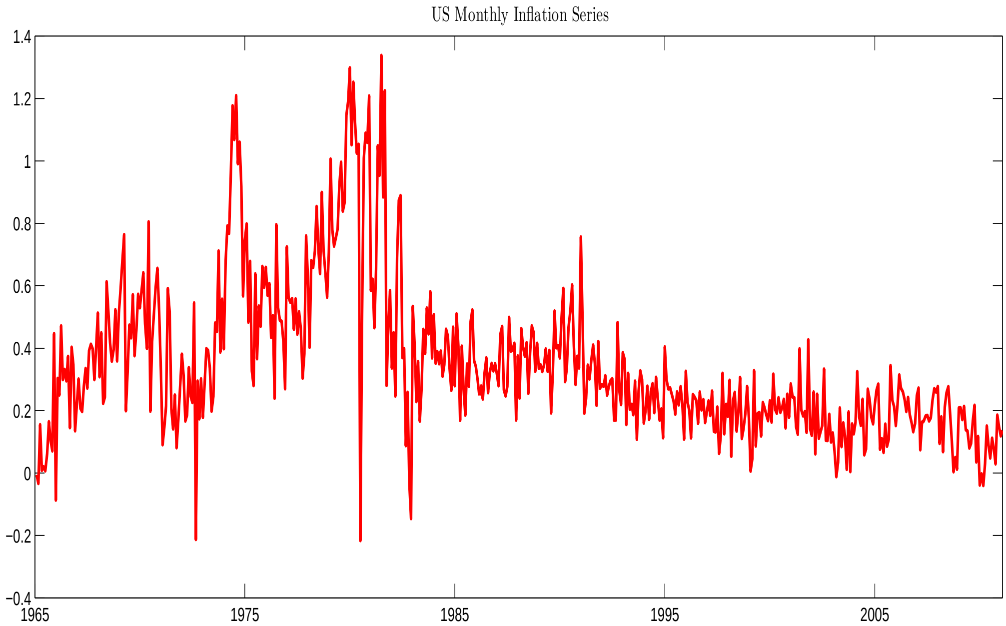


Figure 5: US inflation series

Table 10 reports the estimation results of long memory in inflation. The degree of persistence of π_t, obtained with the semiparametric estimators, is approximately 0.7, pointing toward non-stationarity in inflation. For what concerns the parametric methods, an ARFIMA$(1, d, 0)$ is estimated.[6] All the parametric estimators, including SS-AR and SS-MA find similar degrees of long memory, which are close to 0.55, confirming the results obtained with the semiparametric estimators. On the other hand, when considering the possible presence of level shifts, then the conclusions drastically change. In particular, the model selected by the Akaike information criterion is an ARFIMA$(0, d, 0)$. Moreover, the value of the log-likelihood of the $MSS - AR$ is significantly higher than that obtained with the SS-AR on an ARFIMA$(1, d, 0)$ without shifts, thus suggesting that level shifts cannot be neglected. In this case, the estimated degree of long memory is not statistically different from zero, while there is a positive probability of level shifts, with a significantly positive variance, σ^2_κ. The estimated parameter α is 0.06, meaning that, on average, there is a shift every 17 months, so that shifts in inflation are likely to happen very often. Their size depends on σ^2_κ that is equal to 0.17, so that, with 95% of probability the shift size is in the range between $[-0.80\%, +0.80\%]$. The estimated high value of α could be caused by the fact that the variance of π_t is much higher in the first part of the sample. Therefore, model (32) can be easily modified to account for a break in the variance of η_t. We set the break

[6]The selection of the model has been done according to the Akaike information criterion.

date on January 1984, that is widely known to be the starting date of the *Great Moderation*, see among others Stock and Watson (2007). The last column of table 10 reports the estimates for this model, that is named $MSS - AR_{\sigma^2_\eta}$. The results highlight that, when accounting for break in the innovation variance, then there is a lower, but significant, probability of shifts, equal to 0.0293, with shift size variance of the same order as before. This result suggests that the persistence in the inflation series is due to the role of possible large level shifts in the mean, which excludes the possibility of truly long memory (the parameter d is not statistically different from zero in both cases). It is therefore reasonable to think at the inflation series as resulting from the interaction between a step function and a short memory process. As a consequence, the CPI index is a trend stationary series, where the trend is Note that, model 32 excludes the presence of measurement error, which is a relevant features of the inflation series when focusing on the so called *core inflation*. The long memory state space form (32) can be adapted to incorporate also this feature. The investigation of a state space estimation method, that is jointly robust both to measurement errors and level shifts, is left for future research.

8 Conclusion

The finite sample properties of a wide range of long memory estimators are studied through Monte Carlo simulations. Simulations clearly highlight the usefulness of the state space methods as a valid alternative to the traditional semiparameteric and parametric estimators. The bias and RMSE of the state space methods are of the same order of the other parametric estimators considered in the study, confirming the results found in Haldrup and Nielsen (2007). A potential pitfall of all the parametric methods, including state space, is that they rely on the correct specification of the short run dynamics of the process. The Monte Carlo simulations show that the usual information criteria are able to provide the right guidance in selecting the true model. The potentiality of the state space methods is then exploited when additional features are added to the data. In case of Student's t distribution the state space approach proves to be robust even thought the KF is not optimal, at least in theory. Moreover, when the series at hand are affected by missing values, the KF technique clearly outperform the traditional estimators especially when the percentage of missing observations increases. A typical problem, faced in practical applications, is that many economic and financial series are contaminated by measurement errors. In that case, it is shown that, as expected, traditional long memory estimators with the exception of CLW, are extremely downward bias. The state space methodology allows to directly tackle this issue, providing corrected estimates and filtering the latent long memory signal. This could be exploited for forecasting purposes.

Finally, we provide a novel and promising approach to the estimation of the ARFIMA processes, when they are contaminated by level shifts. A modified version of the Lu and Perron (2010) filter is proposed and its finite sample properties are analyzed with Monte Carlo simulations. It turns out that in the case of shifts, the semiparametric and the parametric estimators provide upward biased estimates of the parameter d. The proposed estimator is able to disentangle the long memory signal from the random level shift process providing corrected estimates of the probability of the shifts, of their size as well as of the ARFIMA parameters. Under the assumption of no shifts the estimates of d parameter are unbiased. The same holds true also under the assumption that only level shifts are present, but not long memory. The results highlight that the state space methods are highly reliable, even in small samples, and are an useful tool to analyze long memory processes under alternative potential misspecification.

(a) $d = 0.0,\ \phi = 0,\ \theta = 0$

T		$b = T^{0.5}$ LW	ELW	FLW	$b = T^{0.65}$ LW	ELW	FLW	ML Estimators Beran	Whittle	SS-AR	SS-MA
250	Bias	-0.0160	-0.0174	-0.0088	-0.0114	-0.0024	**-0.0021**	-0.0022	-0.0048	-0.0036	-0.0185
	RMSE	0.1852	0.1892	0.2109	0.0994	0.0991	0.0989	**0.0499**	0.0505	0.0561	0.0679
500	Bias	-0.0142	-0.0142	-0.0080	-0.0070	-0.0025	-0.0021	**-0.0007**	-0.0022	-0.0010	-0.0066
	RMSE	0.1373	0.1375	0.1573	0.0769	0.0770	0.0766	**0.0339**	0.0343	0.0353	0.0378
1000	Bias	0.0059	-0.0071	-0.0004	0.0020	0.0020	0.0036	**0.0000**	-0.0006	-0.0000	-0.0028
	RMSE	0.1124	0.1121	0.1195	0.0611	0.0615	0.0615	**0.0246**	0.0259	0.0255	0.0266

(b) $d = 0.0,\ \phi = -0.5,\ \theta = 0$

T		$b = T^{0.5}$ LW	ELW	FLW	$b = T^{0.65}$ LW	ELW	FLW	ML Estimators Beran	Whittle	SS-AR	SS-MA
250	Bias	-0.0218	-0.0286	-0.0092	-0.0383	-0.0317	-0.0316	-0.0068	-0.0108	-0.0087	**0.0019**
	RMSE	0.1858	0.1878	0.2353	0.1058	0.1039	0.1037	**0.0601**	0.0618	0.0645	0.0692
500	Bias	-0.0168	-0.0196	-0.0029	-0.0222	-0.0191	-0.0187	-0.0035	-0.0055	-0.0037	**-0.0013**
	RMSE	0.1376	0.1369	0.1900	0.0799	0.0793	0.0787	**0.0417**	0.0424	0.0428	0.0454
1000	Bias	-0.0073	-0.0098	-0.0061	-0.0094	-0.0078	-0.0078	-0.0014	0.0445	-0.0013	**0.0000**
	RMSE	0.1125	0.1114	0.1308	0.0618	0.0616	0.0616	**0.0298**	0.0618	0.0310	0.0337

(c) $d = 0.0,\ \phi = 0,\ \theta = -0.4$

T		$b = T^{0.5}$ LW	ELW	FLW	$b = T^{0.65}$ LW	ELW	FLW	ML Estimators Beran	Whittle	SS-AR	SS-MA
250	Bias	-0.0424	-0.0488	-0.0339	-0.1077	-0.1008	-0.1013	**-0.0018**	-0.0163	0.0035	**0.0018**
	RMSE	0.1892	0.1912	0.2275	0.1462	0.1422	0.1418	**0.1495**	0.1531	0.1621	0.1612
500	Bias	-0.0274	-0.0304	-0.0121	-0.0829	-0.0835	-0.0834	-0.0041	-0.0104	0.0021	**0.0009**
	RMSE	0.1394	0.1387	0.1972	0.1134	0.1138	0.1134	**0.0979**	0.1004	0.1100	0.1085
1000	Bias	-0.0123	-0.0151	-0.0108	-0.0396	-0.0379	-0.0380	**-0.0009**	-0.0119	-0.0038	0.0025
	RMSE	0.1131	0.1123	0.1349	0.0728	0.0720	0.0721	0.0722	0.1298	**0.0700**	0.0730

(d) $d = 0.0,\ \phi = -0.5,\ \theta = -0.4$

T		$b = T^{0.5}$ LW	ELW	FLW	$b = T^{0.65}$ LW	ELW	FLW	ML Estimators Beran	Whittle	SS-AR	SS-MA
250	Bias	-0.0480	-0.0690	-0.0471	-0.1331	-0.1323	-0.1332	0.0063	0.0134	0.0102	**0.0012**
	RMSE	0.1902	0.1947	0.2498	0.1660	0.1661	0.1662	**0.1669**	0.2058	0.1719	0.1924
500	Bias	-0.0297	-0.0417	-0.0157	-0.0168	-0.0196	-0.0029	0.0047	0.0294	0.0059	**0.0007**
	RMSE	0.1398	0.1390	0.2122	0.1376	0.1369	0.1900	**0.1112**	0.0618	0.1159	0.1231
1000	Bias	-0.0137	-0.0212	-0.0143	-0.0484	-0.0492	-0.0493	0.0003	0.0228	0.0041	**0.0001**
	RMSE	0.1131	0.1109	0.1507	0.0780	0.0780	0.0782	**0.0667**	0.1525	0.0720	0.0771

Table 1: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA model, with $d = 0$ and different combinations of ϕ and θ. Sample size is $T = \{250, 500, 1000\}$. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, $SS\text{-}AR$ and $SS\text{-}MA$. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.4,\ \phi = 0,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0104	-0.0174	0.0027	-0.0100	-0.0025	0.0070	-0.0022	0.0047	0.0036	**-0.0012**
	RMSE	0.1890	0.1892	0.1920	0.1006	0.0992	0.1071	**0.0499**	0.0526	0.0550	0.0567
500	Bias	-0.0092	-0.0142	-0.0020	-0.0050	0.0025	0.0026	**-0.0008**	0.0042	0.0069	0.0017
	RMSE	0.1391	0.1375	0.1453	0.07890	0.0771	0.0828	**0.0340**	0.0355	0.0385	0.0388
1000	Bias	-0.0051	-0.0088	-0.0035	0.0005	0.0020	0.0036	**-0.0007**	0.0023	0.0071	0.0037
	RMSE	0.1134	0.1138	0.1163	0.0623	0.0617	0.0632	**0.0251**	0.0259	0.0291	0.0296

(b) $d = 0.4,\ \phi = -0.5,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0163	-0.0286	-0.0139	-0.0371	-0.0317	-0.0243	-0.0069	0.0392	**0.0005**	-0.0033
	RMSE	0.1893	0.1878	0.1830	0.1067	0.1039	0.1072	**0.0602**	0.0965	0.0646	0.0681
500	Bias	-0.0112	-0.0196	-0.0096	-0.0203	-0.0191	-0.0143	-0.0035	0.0023	0.0066	**0.0015**
	RMSE	0.1394	0.1369	0.1404	0.0814	0.0793	0.0833	**0.0417**	0.0433	0.0464	0.0479
1000	Bias	-0.0025	-0.0098	0.0031	-0.0077	-0.0078	-0.0051	**-0.0014**	0.0445	0.0089	0.0054
	RMSE	0.1125	0.1114	0.1140	0.0625	0.0616	0.0643	**0.0299**	0.0618	0.0348	0.0357

(c) $d = 0.4,\ \phi = 0,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0369	-0.0487	-0.0336	-0.1064	-0.1008	-0.0958	**-0.0020**	-0.1268	0.0334	-0.0025
	RMSE	0.1921	0.1912	0.1875	0.1463	0.1422	0.1390	**0.1488**	0.1464	0.1860	0.1576
500	Bias	-0.0258	-0.0336	-0.0222	-0.0677	-0.0660	-0.0632	**-0.0064**	-0.1200	0.0350	-0.0076
	RMSE	0.1440	0.1477	0.1460	0.1021	0.1006	0.0992	0.0990	0.1305	0.1389	**0.0988**
1000	Bias	-0.0225	-0.0304	-0.0213	-0.0662	-0.0650	-0.0620	**-0.0007**	-0.0115	0.0397	0.0012
	RMSE	0.1408	0.1387	0.1413	0.1032	0.1009	0.1003	**0.0676**	.1210	0.1049	0.0685

(d) $d = 0.4,\ \phi = 0.5,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0434	-0.0685	0.0573	-0.1334	-0.1323	-0.1282	0.0242	0.0600	0.0365	**-0.0002**
	RMSE	0.1935	0.1941	0.1869	0.1671	0.1661	0.1622	0.1916	0.2368	0.1956	**0.1648**
500	Bias	-0.0254	-0.0417	-0.0346	-0.0815	-0.0835	-0.0802	0.0112	0.0434	0.0440	**0.0007**
	RMSE	0.1414	0.1390	0.1388	0.1138	0.1138	0.1128	0.1324	0.1847	0.1530	**0.1055**
1000	Bias	-0.0088	-0.0212	-0.0156	-0.0467	-0.0492	-0.0473	**0.0006**	0.02121	0.0603	0.0077
	RMSE	0.1128	0.1109	0.1127	0.0777	0.0780	0.0777	**0.0710**	0.1142	0.1396	0.0767

Table 2: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA model, with $d = 0.4$, and different combinations of ϕ and θ. Sample size is $T = \{250, 500, 1000\}$. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, $SS\text{-}AR$ and $SS\text{-}MA$. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.6,\ \phi = 0,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	0.0013	-0.0174	0.0060	-0.002	-0.003	0.012	-0.002	0.014	0.004	-0.032
	RMSE	0.1920	0.1892	0.1779	0.1021	0.0991	0.0940	0.0499	0.0570	0.0556	0.0499
500	Bias	0.0020	-0.0142	0.0047	0.0036	-0.0025	0.0101	-0.0008	0.0116	0.0071	-0.0211
	RMSE	0.1403	0.1375	0.1323	0.0808	0.0771	0.0732	0.0340	0.0393	0.0383	0.0464
1000	Bias	0.0090	-0.0072	0.0060	0.0086	0.0020	0.0103	0.0000	0.0086	0.0095	0.0352
	RMSE	0.1136	0.1121	0.1061	0.0649	0.0615	0.0593	0.0246	0.0287	0.0326	0.0928

(b) $d = 0.6,\ \phi = -0.5,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.0045	-0.0286	-0.0123	-0.0285	-0.0317	-0.0218	-0.0069	0.0125	0.0017	-0.0273
	RMSE	0.1919	0.1878	0.1810	0.1060	0.1039	0.0967	0.0602	0.0682	0.0644	0.0678
500	Bias	-0.0008	-0.0196	-0.0067	-0.0116	-0.0191	-0.0110	-0.0036	0.0114	0.0069	-0.0148
	RMSE	0.1403	0.1369	0.1338	0.0818	0.0793	0.0737	0.0417	0.0475	0.0462	0.0535
1000	Bias	0.0077	-0.0110	-0.0049	-0.0004	-0.0078	-0.0051	-0.0014	0.0091	0.0107	-0.0107
	RMSE	0.1136	0.1114	0.1055	0.0644	0.0616	0.0571	0.0298	0.0342	0.0364	0.0436

(c) $d = 0.6,\ \phi = 0,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.0246	-0.0488	-0.0319	-0.0969	-0.1008	-0.0802	-0.0009	0.0340	0.0262	-0.0650
	RMSE	0.19334	0.1912	0.1848	0.1413	0.1422	0.1300	0.1512	0.1782	0.1643	0.1251
500	Bias	-0.0111	-0.0304	-0.0184	-0.0570	-0.0650	-0.0501	-0.0029	0.0286	0.0322	-0.0491
	RMSE	0.1409	0.1387	0.1372	0.0994	0.1009	0.0930	0.1011	0.1263	0.1258	0.0968
1000	Bias	0.0028	-0.0151	-0.0118	-0.0302	-0.0379	-0.0301	-0.0002	0.0234	0.0372	-0.0373
	RMSE	0.1134	0.1123	0.1066	0.0715	0.0722	0.0792	0.0703	0.0867	0.0978	0.0747

(d) $d = 0.6,\ \phi = -0.5,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.1237	-0.1323	-0.1132	-0.1237	-0.1323	-0.1132	0.0017	0.077	0.0238	-0.0615
	RMSE	0.1611	0.1661	0.1547	0.1611	0.1661	0.1547	0.1590	0.2516	0.1636	0.1303
500	Bias	-0.0138	-0.0417	-0.0407	-0.0722	-0.0835	-0.0696	-0.0034	0.0736	0.0320	-0.0415
	RMSE	0.1411	0.1390	0.1344	0.1090	0.1138	0.1048	0.0976	0.2184	0.1231	0.0940
1000	Bias	0.0015	-0.0212	-0.0293	-0.0405	-0.0503	0.0421	0.0008	0.0694	0.0377	0.0312
	RMSE	0.1134	0.1109	0.1063	0.0758	0.0785	0.0791	0.0729	0.1978	0.0936	0.0734

Table 3: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA model, with $d = 0.6$, and different combinations of ϕ and θ. Sample size is $T = \{250, 500, 1000\}$. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, SS-AR and SS-MA. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.4,\ \phi = 0,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0231	-0.0296	-0.0104	-0.0125	-0.0041	0.0033	-0.0045	0.0021	**0.0008**	-0.0044
	RMSE	0.1940	0.2004	0.1970	0.1017	0.1011	0.1063	**0.0506**	0.0533	0.0562	0.0581
500	Bias	-0.0124	-0.0186	-0.0062	-0.0066	-0.0040	**0.0001**	-0.0011	0.0038	0.0067	0.0013
	RMSE	0.1423	0.1457	0.1468	0.0763	0.0756	0.0787	**0.0352**	0.0367	0.0390	0.0399
1000	Bias	-0.0051	-0.0088	-0.0035	0.0005	0.0020	0.0036	**-0.0007**	0.0023	0.0071	0.0037
	RMSE	0.1134	0.1138	0.1163	0.0623	0.0617	0.0632	**0.0251**	0.0259	0.0291	0.0296

(b) $d = 0.4,\ \phi = -0.5,\ \theta = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0289	-0.0380	-0.0199	-0.0398	-0.0329	-0.0277	-0.0100	-0.0288	**-0.0033**	-0.0076
	RMSE	0.1945	0.2003	0.1946	0.1084	0.1060	0.1067	**0.0612**	0.0827	0.0663	0.0697
500	Bias	-0.0153	-0.0230	-0.0116	-0.0219	-0.0201	-0.0166	-0.0031	-0.0184	0.0070	**0.0016**
	RMSE	0.1426	0.1453	0.1443	0.0791	0.0782	0.0794	**0.0425**	0.0829	0.0474	0.0493
1000	Bias	-0.0064	-0.0112	-0.0062	-0.0086	-0.0077	-0.0063	**-0.0019**	-0.0181	0.0084	0.0048
	RMSE	0.1135	0.1137	0.1156	0.0629	0.0621	0.0631	**0.0307**	0.0431	0.0354	0.0361

(c) $d = 0.4,\ \phi = 0,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0496	-0.0587	-0.0413	-0.1080	-0.1019	-0.0984	**-0.0037**	-0.0571	0.0349	0.0014
	RMSE	0.1983	0.2049	0.1995	0.1464	0.1417	0.1395	**0.1450**	0.1719	0.1840	0.1574
500	Bias	-0.0258	-0.0336	-0.0222	-0.0677	-0.0660	-0.0632	**-0.0064**	-0.0618	0.0350	-0.0076
	RMSE	0.1440	0.1477	0.1460	0.1021	0.1006	0.0992	0.0990	0.1348	0.1389	**0.0988**
1000	Bias	-0.0113	-0.0163	-0.0115	-0.0387	-0.0379	-0.0367	**0.0008**	-0.0647	0.0384	0.0010
	RMSE	0.1139	0.1143	0.1157	0.0734	0.0725	0.0723	0.0694	0.0998	0.1044	**0.0689**

(d) $d = 0.4,\ \phi = -0.5,\ \theta = -0.4$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0556	-0.0744	-0.0576	-0.1366	-0.1336	-0.1296	0.0119	0.0496	0.0305	**-0.0066**
	RMSE	0.1997	0.2073	0.1995	0.1702	0.1690	0.1647	0.1859	0.2319	0.1980	**0.1669**
500	Bias	-0.0287	-0.0418	-0.0323	-0.0823	-0.0836	-0.0804	0.0097	0.0335	0.0386	**-0.0035**
	RMSE	0.1445	0.1469	0.1439	0.1129	0.1131	0.1119	0.1371	0.1735	0.1476	**0.1050**
1000	Bias	-0.0127	-0.0217	-0.0177	-0.0477	-0.0491	-0.0478	0.0032	0.0236	0.0374	**0.0031**
	RMSE	0.1140	0.1142	0.1141	0.0786	0.0788	0.0784	0.0786	0.1180	0.1035	**0.0708**

Table 4: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA model, with $d = 0.4$, and different combinations of ϕ and θ. Sample size is $T = \{250, 500, 1000\}$. Innovations are distributed as a Student's t with 3 degrees of freedom, zero mean and unit variance. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, *SS-AR* and *SS-MA*. In bold character the lowest bias and RMSE for each line.

(a) **Over-Specification**

T		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
		LW	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	-0.0104	-0.0174	0.0027	-0.0101	**-0.0025**	0.0070	-0.0392	-0.0442	-0.0043	-0.0800
	RMSE	0.1890	0.1892	0.1920	0.1006	**0.0991**	0.1071	0.2036	0.2556	0.2120	0.2422
500	Bias	-0.0092	-0.0142	**-0.0020**	-0.0050	0.0025	0.0026	-0.0187	-0.0134	0.0250	-0.0490
	RMSE	0.1391	0.1375	0.1453	0.0789	**0.0771**	0.0828	0.1341	0.1583	0.1416	0.1552

(b) **Under-Specification**

T		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
		LW	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
250	Bias	**-0.0434**	-0.0685	0.0573	-0.1334	-0.1323	-0.1282	-0.4874	-0.4741	-0.4806	-0.4709
	RMSE	0.1935	0.1941	0.1869	0.1671	0.1661	**0.1622**	0.4920	0.4788	0.4901	0.4620
500	Bias	**-0.0254**	-0.0417	-0.0346	-0.0815	-0.0835	-0.0802	-0.4580	-0.4496	-0.4638	-0.4561
	RMSE	0.1414	0.1390	0.1388	0.1138	0.1138	**0.1128**	0.4613	0.4525	0.4624	0.4728

Table 5: Panel (a) reports the estimated long memory parameter for an ARFIMA$(1, d, 1)$ model for $T = \{250, 500\}$, when the true model is ARFIMA$(0, d, 0)$ with $d = 0.4$. Panel (b) reports the estimated long memory parameter for an ARFIMA$(0, d, 0)$ model, for $T = \{250, 500\}$, when the true model is ARFIMA$(1, d, 1)$ with $d = 0.4$, $\phi = -0.5$ and $\theta = -0.4$. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, SS-AR and SS-MA. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.4$, $\phi = -0.5$, $\theta = -0.4$ with 10% missing

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.1568	-0.1676	-0.1440	-0.2150	-0.2108	-0.2080	-0.1465	-0.1266	0.0344	**0.0146**
	RMSE	0.2450	0.2514	0.2629	0.2382	0.2346	0.2317	0.2129	0.2311	0.2013	**0.1910**
500	Bias	-0.1252	-0.1349	-0.1117	-0.1816	-0.1800	-0.1783	-0.1598	-0.1483	0.0479	**0.0109**
	RMSE	0.1879	0.1919	0.2133	0.1990	0.1971	0.1956	0.1882	0.1849	0.1649	**0.1221**

(b) $d = 0.4$, $\phi = -0.5$, $\theta = -0.4$ with 20% missing

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.2218	-0.2325	-0.2206	-0.2465	-0.2404	-0.2376	-0.1995	-0.1885	0.0420	**0.0287**
	RMSE	0.2890	0.3004	0.2942	0.2669	0.2617	0.2583	0.2307	0.2565	0.2222	**0.2180**
500	Bias	-0.1860	-0.1915	-0.1787	-0.2207	-0.2182	-0.2170	-0.2084	-0.1992	0.0574	**0.0245**
	RMSE	0.2385	0.2409	0.2457	0.2338	0.2311	0.2300	0.2207	0.2183	0.1886	**0.1524**

(c) $d = 0.4$, $\phi = -0.5$, $\theta = -0.4$ with 30% missing

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.2558	-0.2576	-0.2504	-0.2549	-0.2489	-0.2474	-0.2226	-0.2473	0.0414	**0.0289**
	RMSE	0.3148	0.3140	0.3212	0.2742	0.2688	0.2670	0.2742	0.3406	0.2383	**0.2346**
500	Bias	-0.2183	-0.2221	-0.2136	-0.2452	-0.2426	-0.2414	-0.2053	-0.2046	0.0656	**0.0372**
	RMSE	0.2639	0.2655	0.2672	0.2581	0.2557	0.2545	0.2201	0.2324	0.2074	**0.1735**

(d) $d = 0.4$, $\phi = -0.5$, $\theta = -0.4$ with 40% missing

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators			
T		LW	ELW	FLW	LW	ELW	FLW	$Beran$	$Whittle$	SS-AR	SS-MA
250	Bias	-0.2806	-0.2808	-0.2786	-0.2400	-0.2351	-0.2335	-0.2875	-0.3424	0.0307	**0.0270**
	RMSE	0.3377	0.3374	0.3371	0.2645	0.2610	0.2585	0.3606	0.4390	**0.2590**	0.2704
500	Bias	-0.2481	-0.2501	-0.2472	-0.2483	-0.2449	-0.2442	-0.2476	-0.2869	0.0676	**0.0539**
	RMSE	0.2856	0.2858	0.2876	0.2607	0.2573	0.2567	0.2817	0.3317	0.2225	**0.2049**

Table 6: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA$(1, d, 1)$ model with missing values. The ARFIMA parameters are $d = 0.4$, $\phi = -0.5$ and $\theta = -0.5$ in all cases. The sample size are $T = \{250, 500\}$. The percentage of missing observations are 10%, 20%, 30% and 40%. Table reports the semiparametric estimates of LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, SS-AR and SS-MA. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.4$, $\phi = 0$, $\theta = 0$

		$b = T^{0.65}$				ML Estimators			
nsr		*LW*	*ELW*	*FLW*	*CLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
1	Bias	-0.1347	-0.1320	-0.1297	0.0333	-0.1913	-0.1883	0.0342	**0.0038**
	RMSE	0.1557	0.1523	0.1505	0.1914	0.1950	0.1926	0.1241	**0.1013**
2	Bias	-0.1931	-0.1898	-0.1878	0.0827	-0.2450	-0.2435	**0.0149**	-0.0203
	RMSE	0.2084	0.2046	0.2030	0.3340	0.2481	0.2278	0.1172	**0.1007**

(b) $d = 0.4$, $\phi = -0.5$, $\theta = 0$

		$b = T^{0.65}$				ML Estimators			
nsr		*LW*	*ELW*	*FLW*	*CLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
1	Bias	-0.1785	-0.1757	-0.1736	0.0803	-0.16589	-0.1624	0.0345	**0.0077**
	RMSE	0.1949	0.1915	0.1898	0.3033	0.1733	0.1707	0.1386	**0.1045**
2	Bias	-0.2757	-0.2713	-0.2703	0.0655	-0.5194	-0.2568	**0.0034**	-0.0280
	RMSE	0.2870	0.2825	0.2815	0.4155	0.2432	0.2301	0.1507	**0.1291**

(c) $d = 0.4$, $\phi = 0$, $\theta = -0.4$

		$b = T^{0.65}$				ML Estimators			
nsr		*LW*	*ELW*	*FLW*	*CLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
1	Bias	-0.2757	-0.2713	-0.2702	0.0655	-0.5193	-0.2568	**-0.0211**	-0.0884
	RMSE	0.2870	0.28245	0.2815	0.4155	0.5224	**0.2697**	0.3169	0.3077
2	Bias	-0.3213	-0.3158	-0.3157	**0.0093**	-0.4803	-0.3080	-0.0211	-0.0984
	RMSE	0.3311	0.3256	0.3256	0.5442	0.4841	0.3161	0.3368	**0.3077**

(d) $d = 0.4$, $\phi = -0.5$, $\theta = -0.4$

		$b = T^{0.65}$				ML Estimators			
nsr		*LW*	*ELW*	*FLW*	*CLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
1	Bias	-0.1588	-0.1559	-0.1537	0.0805	-0.3262	-0.0866	0.0436	**-0.0226**
	RMSE	0.1772	0.1736	0.1718	0.3196	0.5892	0.1578	0.2503	**0.2085**
2	Bias	-0.1848	-0.1829	-0.1907	0.0942	-0.4217	-0.2817	**0.0236**	-0.0126
	RMSE	0.1772	0.1736	0.1718	0.3196	0.6042	0.3578	0.3150	**0.3085**

Table 7: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA$(1, d, 1)$ plus noise model, with $d = 0.4$, $\phi = -0.5$ and $\theta = -0.5$, for $T = \{500\}$. The measurement errors are distributed as a Gaussian random variable, zero mean and variance that depends on the signal to noise ratio. The *nsr* is set equal to 1 and 2. Table reports the semiparametric estimates of LW, ELW and FLW with bandwidth, $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, SS-AR and SS-MA. In bold character the lowest bias and RMSE for each line.

(a) $d = 0.4$, $\phi = 0$, $\sigma_\kappa^2 = 4$ and $\alpha = 7.5/T$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators		
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*MSS-AR*
500	Bias	0.3544	0.2985	0.3110	0.2994	0.2486	0.2608	0.1551	0.1907	**-0.0155**
	RMSE	0.3947	0.3406	0.3476	0.3287	0.2761	0.2842	0.1682	0.2088	**0.0588**
1500	Bias	0.3354	0.2805	0.2908	0.2671	0.2170	0.2278	0.1075	0.1316	**-0.0119**
	RMSE	0.3640	0.3091	0.3155	0.2902	0.2372	0.2448	0.1179	0.1459	**0.0296**

(b) $d = 0.4$, $\phi = -0.3$, $\sigma_\kappa^2 = 4$ and $\alpha = 7.5/T$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators		
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*MSS-AR*
500	Bias	0.4061	0.3509	0.3632	0.2994	0.2486	0.2608	0.2346	0.2808	**-0.0136**
	RMSE	0.4398	0.3874	0.3950	0.3287	0.2761	0.2842	0.2478	0.2989	**0.0596**
1500	Bias	0.3902	0.3355	0.3439	0.3197	0.2669	0.2752	0.1683	0.2014	**-0.0173**
	RMSE	0.4136	0.3598	0.3659	0.3398	0.2851	0.2909	0.1790	0.2162	**0.0490**

(c) $d = 0.4$, $\phi = -0.3$, $\sigma_\kappa^2 = 0$ and $\alpha = 0$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators		
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*MSS-AR*
500	Bias	-0.0149	-0.0210	-0.0163	-0.0275	-0.0290	-0.0264	**-0.0052**	0.0014	-0.0386
	RMSE	0.1448	0.1438	0.1412	0.1124	0.1138	0.1120	**0.0474**	0.0489	0.0827
1500	Bias	0.0022	-0.0011	-0.0006	0.0006	**-0.0002**	0.0004	0.0039	0.0061	-0.0187
	RMSE	0.1039	0.1012	0.1005	0.0469	0.0455	0.0446	**0.0263**	0.0272	0.0550

(d) $d = 0.0$, $\phi = 0$, $\sigma_\kappa^2 = 4$ and $\alpha = 7.5/T$

		$b = T^{0.5}$			$b = T^{0.65}$			ML Estimators		
T		*LW*	*ELW*	*FLW*	*LW*	*ELW*	*FLW*	*Beran*	*Whittle*	*MSS-AR*
500	Bias	0.6046	0.5389	0.5601	0.7361	0.6781	0.6921	0.4263	0.4645	**-0.0146**
	RMSE	0.6900	0.6195	0.6412	0.7515	0.6928	0.7041	0.3806	0.4112	**0.0628**
1500	Bias	0.8748	0.8228	0.8299	0.7517	0.6940	0.7004	0.4886	0.5217	**-0.0022**
	RMSE	0.8844	0.8333	0.8401	0.7630	0.7042	0.7102	0.4956	0.5312	**0.0298**

Table 8: Panels (a), (b), (c) and (d) report the estimated long memory parameter for an ARFIMA model with level shifts. ARFIMA is generated with with $d = 0.4$ and $\phi = 0, -0.3$ and $\theta = 0$. The sample size is $T = \{500, 1500\}$. The level shifts process process is simulated according to formula (33) with $\sigma_\kappa^2 = 4$ and $\alpha = 7.5/T$. Table reports the semiparametric estimates for LW, ELW and FLW with different bandwidth choices, $b = T^{0.5}$ and $b = T^{0.65}$. Table also reports the approximated maximum likelihood estimates, ML: Beran, Whittle, MSS-AR. In bold character the lowest bias and RMSE for each line.

LMSV model with ARFIMA$(1, d, 0)$

		$b = T^{0.65}$				ML Estimators			
T		*LW*	*ELW*	*FLW*	*CLW*	*Beran*	*Whittle*	*SS-AR*	*SS-MA*
415	$\hat{d}$	0.2690	0.2717	0.2717	0.8217	0.2184	0.2190	0.4476	0.4779
	$\hat{\phi}$	–	–	–	–	-0.1757	-0.1760	-0.3996	-0.3650
	$\hat{\sigma}^2_\eta$	–	–	–	–	5.4055	5.4228	1.0752	1.3688
	$\hat{\sigma}^2_\epsilon$	–	–	–	–	0	0	3.9596	3.8598
	$\hat{\beta}$	–	–	–	76.1550	–	–	–	–

Table 9: Table reports the estimated parameters of the $LMSV$ model, with an ARFIMA(1, d ,0) for the latent stochastic volatility process. Table reports the semiparametric estimates for LW, ELW, FLW and CLW with bandwidth $b = T^{0.65}$. Table also reports the estimates of the parameters ϕ, σ^2_η and σ^2_ϵ for the approximated maximum likelihood estimates. $\hat{\beta} = \widehat{\left(\frac{\sigma^2_\epsilon}{2\pi G}\right)}$ is the estimate of the noise-to-signal ratio, obtained as the additional parameter in the corrected Whittle objective function.

Long Memory in Inflation

T		$b = T^{0.65}$				ML Estimators					
		LW	ELW	FLW	CLW	$Beran$	$Whittle$	SS-AR	SS-MA	MSS-AR	$MSS-AR_{\sigma^2_\eta}$
553	d	0.7111^a	0.6939^a	0.7534^a	0.7556^a	0.5417^a	0.5476^a	0.5381^a	0.5286^a	0.0266	0.1319
	ϕ	–	–	–	–	-0.1336	-0.1334	-0.1166	-0.1018	–	–
	α	–	–	–	–	–	–	–	–	0.0609^a	0.0293^b
	$\hat{\sigma}_\kappa$	–	–	–	–	–	–	–	–	0.1701^b	0.2072^b
	$\hat{\sigma}_\eta$	–	–	–	–	0.0228^a	0.0231^a	0.0231^a	0.0225^a	0.0106^b	–
	$\hat{\sigma}^2_{\eta_1}$	–	–	–	–	–	–	–	–	–	0.0304^a
	$\hat{\sigma}^2_{\eta_2}$	–	–	–	–	–	–	–	–	–	0.0073^a

Table 10: Table reports the estimated parameters of the long memory estimates on the US inflation series. LW, ELW, FLW and CLW are the semiparametric estimates, with bandwidth $b = T^{0.65}$. An ARFIMA$(1, d, 0)$ is estimated with the parametric estimators, including SS-AR and SS-MA. Table also reports the estimates of the MSS-AR on an ARFIMA(0,d,0) plus random level shifts. Table reports the parameter estimates for ϕ, σ^2_η, α and σ^2_κ. $\hat{\sigma}^2_{\eta_1}$ and $\hat{\sigma}^2_{\eta_2}$ are the innovation variances for the periods *Jan:1965-Dec:1983* and *Jan:1984-Mar:2011,* respectively. a, b and c stand for 1%, 5% and 10% significance levels, respectively.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of stock return volatility. *Journal of Financial Economics*, 61:43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.

Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, 119(1):131–154.

Baillie, R. T., Chung, C. F., and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated arfima-garch model. *Journal of Applied Econometrics*, Vol. 11:pp. 23–40.

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

Bisaglia, L. and Gugan, D. (1998). A comparison of techniques of estimation in long-memory processes. *Computational Statistics and Data Analysis*, Vol. 27:pp. 61–81.

Bondon, P. and Palma, W. (2005). Asymptotics for linear predictors of strongly dependent time series. In *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*, pages 847 –852.

Breidt, F. J., Crato, N., and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1-2):325 – 348.

Calvo, G. A. (1983). Staggered prices in a utility maximizing framework. *Journal of Monetary Economics*, 12:383–398.

Casas, I. and Gao, J. (2008). Econometric estimation in long-range dependent volatility models: Theory and practice. *Journal of Econometrics*, 147(1):72–83.

Chan, N. and Palma, W. (2006). Estimation of long-memory time series models: A survey of different likelihood-based methods. *Advances in Econometrics*, Vol. 20:pp. 89–121.

Chan, N. H. and Palma, W. (1998). State space modeling of long-memory processes. *Annals of Statistics*, Vol. 26, No. 2:pp. 719–740.

Comte, F. and Renault, E. (1998). Long-memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8:291–323.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, Vol. 17:pp. 1749–1766.

Deo, R. S. and Hurvich, C. M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, 17(04):686–710.

Diebold, F. X., Husted, S., and Rush, M. (1991). Real exchange rates under the gold standard. *Journal of Political Economy*, Vol. 99:pp. 1252–1271.

Dolado, J. J., Gonzalo, J., and Mayoral, L. (2005). What is what?: A simple time-domain test of long-memory vs. structural breaks. Economics Working Papers 954, Department of Economics and Business, Universitat Pompeu Fabra.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian series. *Annals of Statistics*, Vol. 14:pp. 517–532.

Gadea, M. D. and Mayoral, L. (2006). The persistence of inflation in oecd countries: A fractionally integrated approach. *International Journal of Central Banking*, 2(1).

Geweke, J. and Porter-Hudak, S. (1984). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4:221–238.

Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with application to the s&p 500 absolute stock returns. *Journal of Empirical Finance*, 11:399–421.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 4:221–238.

Haldrup, N. and Nielsen, M. . (2007). Estimation of fractional integration in the presence of data noise. *Computational Statistics and Data Analysis*, Vol. 51:pp. 3100–3114.

Harvey, A. (1989). *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge University Press, Cambridge, UK.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–64.

Harvey, A. C. and Proietti, T. (2005). *Readings in Unobserved Components Models*. Advanced Texts in Econometrics. Oxford University Press, Oxford, UK.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, 13(1):37–45.

Hosking, J. (1981). Fractional differencing. *Biometrika*, 68:165–76.

Hsu, C.-C. (2005). Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters*, 88:289294.

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of The American Society of Civil Engineers*, 1:519–543.

Hurvich, C. M., Moulines, E., and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica*, 73:1283–1328.

Hurvich, C. M. and Ray, B. K. (2003). The local whittle estimator of long-memory stochastic volatility. *Journal of Financial Econometrics*, 1(3):445–470.

Iglesias, P., Jorquera, H., and Palma, W. (2006). Data analysis using regression models with missing observations and long-memory: an application study. *Computational Statistics and Data Analysis*, 50:2028 – 2043.

Lu, Y. K. and Perron, P. (2010). Modeling and forecasting stock return volatility using a random level shift model. 17:138–156.

Nielsen, M. and Frederiksen, P. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, Vol. 24:pp. 405–443.

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2008). True or spurious long memory? a new test. *Journal of Business & Economic Statistics*, 26:161–175.

Palma, W. (2007). *Long-Memory Time Series*. Wiley-Interscience, a John Wiley and Sons, Inc.

Palma, W. and del Pino, G. (1999). Statistical analysis of incomplete long-range dependent data. *Biometrika*, Vol. 86:pp. 165–172.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics*, 28(2):275–290.

Raggi, D. and Bordignon, S. (2010). Long memory and nonlinearities in realized volatility: A markov switching approach. *Computational Statistics & Data Analysis*, In press.

Rea, W., Oxley, L., M., R., and J., B. (2008). The empirical properties of some popular estimators of long memory processes. Working Papers 13, University of Canterbury, Christchurch.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23:1048–1072.

Robinson, P. M. (2003). *Long Memory Time Series*. Oxford University Press.

Rossi, E. and Santucci de Magistris, P. (2009). A no arbitrage fractional cointegration analysis of the range based volatility. Technical report, CREATES, Research Paper 31-2009.

Schweppe, F. (1965). Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11:61–70.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Technical report, Working paper, Queens University.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(02):501–540.

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

Sibbertsen, P. and Kruse, R. (2009). Testing for a break in persistence under long-range dependencies. *Journal of Time Series Analysis*, 30(3):263–285.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53(1-3):165–188.

Stock, J. H. and Watson, M. (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Banking and Credit*, 39:3–33.

Sun, Y. and Phillips, P. C. B. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, 115(2):355–389.

Sun, Y. and Phillips, P. C. B. (2005). Understanding the fisher equation. *Journal of Applied Econometrics*, 89:869886.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20-1:87–127.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120:75 102.

Research Papers 2011

2010-76: Peter R. Hansen, Asger Lunde and James M. Nason: The Model Confidence Set

2011-01: Cristina Amado and Timo Teräsvirta: Modelling Volatility by Variance Decomposition

2011-02: Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity

2011-03: Roxana Halbleib and Valeri Voev: Forecasting Covariance Matrices: A Mixed Frequency Approach

2011-04: Mark Podolskij and Mathieu Rosenbaum: Testing the local volatility assumption: a statistical approach

2011-05: Michael Sørensen: Prediction-based estimating functions: review and new developments

2011-06: Søren Johansen: An extension of cointegration to fractional autoregressive processes

2011-07: Tom Engsted and Stig V. Møller: Cross-sectional consumption-based asset pricing: The importance of consumption timing and the inclusion of severe crises

2011-08: Tommaso Proietti and Stefano Grassi: Bayesian stochastic model specification search for seasonal and calendar effects

2011-09: Matt P. Dziubinski: Option valuation with the simplified component GARCH model

2011-10: Tim A. Kroencke, Felix Schindler and Andreas Schrimpf: International Diversification Benefits with Foreign Exchange Investment Styles

2011-11: Eduardo Rossi and Paolo Santucci de Magistris: Estimation of long memory in integrated variance

2011-12: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Generalized Jackknife Estimators of Weighted Average Derivatives

2011-13: Dennis Kristensen: Nonparametric Detection and Estimation of Structural Change

2011-14: Stefano Grassi and Paolo Santucci de Magistris: When Long Memory Meets the Kalman Filter: A Comparative Study