Journal of Uncertain Systems
Vol.6, No.4, pp.289-298, 2012
Online at: www.jus.org.uk

Modeling Long-Memory Time Series with Sparse Autoregressive Processes

Yan Sun*
Department of Mathematics & Statistics, Utah State University, Logan, Utah 84322-3900, U.S.A.

Received 10 June 2012; Revised 2 September 2012

Abstract

Many economic, financial, and engineering time series data exhibit long-term persistence. The autoregressive fractionally integrated moving average (ARFIMA) process is characterized by a slowly decaying autocorrelation function and arises as a popular statistical tool for modeling long memory time series. After years of development on the semiparametric two-stage direct estimation of ARFIMA, recently there has been a considerable interest in the long-order autoregressive (AR) approximation, as it is observed to be simple and effective. This paper proposes a sparse AR approximation to the ARFIMA process based on the penalized conditional likelihood. Simulation study shows that the proposed method leads to better model flexibility and prediction accuracy. Finally, we apply the method to analyze a foreign exchange rate data and the results are very satisfactory.

Keywords: ARFIMA, autoregressive approximation, foreign exchange-rate, long-memory, penalized conditional likelihood, sparse

1 Introduction

There has been a growing body of literature studying the property of long-term dependency in time series data arising from many fields of science. See, for example, [3, 5, 6, 11, 12, 18, 25]. In particular, it has been widely acknowledged that stochastic long memory exists in financial time series, especially in the foreign exchange rates and future prices, although it may not be a typical feature of stock market returns of moderate time intervals (e.g., daily returns). Long memory processes are characterized by a high-order correlation structure, indicating a persistent dependence between distant observations. Short memory processes like autoregressive moving average (ARMA) are unable to capture the dynamics of a long memory series. The autoregressive fractionally integrated moving average (ARFIMA) process by Granger and Joyeux [17], allowing the integration order of a series to take on fractional values, provides a useful tool for modeling and forecasting time series with long-memory properties. Kokoszka and Taqqu [22] extended the ARFIMA model by allowing for an infinite variance.

For decades, researchers have tended to make inferences of the ARFIMA model in two successive steps. In the first step, a semi-parametric method is used to estimate the order of fractional integration. There are various approaches proposed in the literature, for example, the rescaled range analysis (R/S) [21], the modified R/S statistic [25], the de-trended fluctuation analysis (DFA) [27], and the local Whittle estimator [29, 33]). In the second step, one applies the long-memory filter to the process and fits an ARMA model to the resulting short-memory process. Or, alternatively, one can approximate the short-run component by an autoregressive (AR) process [3]. Recently, there has been a trend to approximate both the long-memory and short-memory structure by a high order linear autoregression, and evidences show that it works better than the two-step approach. For instance, Baillie and Kapetanios [4] showed that the high order AR method provides relatively good estimates of the impulse response weights of the ARFIMA model, while the local Whittle two step estimators (LWTSE) generally have quite poor properties in the presence of moderately persistent autocorrelation in the short run component. Poskitt [28] validated their "surprising" findings by proving the autoregressive estimates from long autoregressions to be consistent. Specifically, he showed that

*Corresponding author. Email: yan.sun@usu.edu (Y. Sun).

the L_2 norm of the difference between the vector of the high-order AR approximation coefficients and that of the true autoregressive coefficients vanishes asymptotically.

One problem that arises with the long-order AR approximation method is the choice of the autoregression order. Asymptotically, the autoregression order p is allowed to grow to infinity as the sample size $T \to \infty$. In [28], p is assumed to grow at the rate of $o((T/\log T)^{\frac{1}{2}-d'})$, where $d' = \max\{0, d\}$ and d is the order of fractional integration. In finite sample problems, one has to decide on a specific value of p. Galbraith and Zinde-Walsh [15] gave a rule of thumb, based on $O(\ln T)$, that $p = 8 + 3\ln(T/100)$ when $T > 100$, rounded to the nearest integer. Or equivalently, $p = 3\ln(T) - 6$. This is a very small quantity compared to T.

In this paper, we propose a sparse long-order AR approximation to ARFIMA models based on the penalized conditional likelihood. Being widely applied to independent data models, the penalty-based regularization techniques effectively achieve concise and parsimonious parameter estimation. See, for example, [10, 13, 14, 20, 36]. Especially, when the penalty function is appropriately chosen, these techniques leads to a sparse parameter estimation, and thereby simultaneously accomplishes model selection. Sun and Lin [35] extended these methodologies to the time series context and developed the penalized conditional maximum likelihood estimation (PCMLE) for a large class of stationary multivariate time series models. Sang and Sun [31] further studied the PCMLE for AR processes, allowing heavy-tailed innovations. The sparse feature of the penalized conditional maximum likelihood estimation is particularly useful in the long-order AR approximation in that it enables automatic selections of significant lags within the order p. This facilitates the order selection in the sense that redundant parameters are automatically eliminated from the model and increasing p does not bring unnecessary noise into the estimation. We show in the simulation study that our method allows for a much larger value of p compared to the aforementioned literature, and it performs consistently well over a wide range of different p values. We further demonstrate the advantages of the proposed method through the analysis of a foreign exchange data, which exhibits long-memory property.

The rest of the paper is organized as follows: in Section 2, we formally introduce the ARFIMA model and its two-stage estimation procedure. We also give a detailed review of the AR approximation methodology. Section 3 represents the penalty-based order selection framework for the autoregression order. The simulation study is reported in Section 4, followed by a foreign exchange data analysis in Section 5. We give concluding remarks in Section 6.

2 Review of the ARFIMA Model and Existing Estimation Methods

2.1 The ARFIMA Model

An ARFIMA time series $Y = Y(k), k \in \mathbb{Z}$ with white noise innovations is defined as the stationary solution to the back-shift operator equation:

$$\Phi_p(B)Y(k) = \Theta_q(B)(1-B)^{-d}Z(k), \ k \in \mathbb{Z}, \tag{1}$$

where the innovations $Z(k)$ are iid random variables with variance σ^2, B is the back-shift operator such that $BY(k) = Y(k-1)$, and $\Phi_p(z)$, $\Theta_q(z)$ are real polynomials of degrees p, q, respectively, i.e.,

$$\begin{aligned} \Phi_p(z) &= 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p, \\ \Theta_q(z) &= 1 - \Theta_1 z - \Theta_2 z^2 - \cdots - \Theta_q z^q, \end{aligned} \tag{2}$$

with roots outside the unit disk. These conditions guarantee the existence and uniqueness of the solution Y to (1). In addition, the fractional difference parameter d satisfies $0 < d < 1/2$. This can be generalized to $-1/2 < d < 1$, which includes both stationary and non-stationary ARFIMA processes. If $d \geq 1/2$, then Y(k) has infinite variance. See [17].

It has been shown that when $-1 < d < 1/2$, $d \neq 0$, the autocorrelation function $\rho(\cdot)$ for the ARFIMA model (1) is given by:

$$\rho(h) = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(h+1-d)}. \tag{3}$$

For large h, the above quantity is well approximated by Ch^{2d-1} for some constant C. Namely,

$$\rho(h) \propto h^{2d-1}, \tag{4}$$

asymptotically. Whereas for a stationary ARMA model, $\rho(h)$ is asymptotically proportional to θ^h for some θ such that $|\theta| < 1$, which decays to zero much faster compared to (4). This illustrates the "long-memory" aspect of the ARFIMA model. See, e.g., [30] for a comprehensive review of the literature on long-memory time series.

2.2 The Two-Stage Semi-Parametric Estimation

There have been substantial contributions in literature to the estimation of the ARFIMA models, which are mainly categorized into two ways. One considers a joint estimation of the fractional integration parameter d and the ARMA parameters for the short-run dynamics [34, 37]. The other estimates d alone in the first stage and leaves the ARMA parameters for a second stage. Either way, the estimation of the fractional integration parameter d is required and crucial. Its value is also viewed as an important criterion for testing the long-memory existence. A few most notable contributions on the estimation of d include [16, 19, 21, 24, 25, 29]. Below we give the formula of three popularly used estimators of the fractional integration parameter d: rescaled range analysis (R/S), modified rescaled range analysis (MRS), and de-trended fluctuation analysis (DFA).

In the R/S analysis proposed by Hurst [21], the time period is divided into m sub-periods, each with length n, and an R/S value is calculated for each sub-period by:

$$R/S = \frac{\max_{0\le k\le n} \sum_{t=1}^{k}(Y_t - \bar{Y}_n) - \min_{0\le k\le n} \sum_{t=1}^{k}(Y_t - \bar{Y}_n)}{S(n)}, \tag{5}$$

where $S(n)$ is the sample standard deviation for the sub-period. The R/S value for this division R/S(n) is the average of R/S's for the m sub-periods. Repeat the above calculation for a range of different integers n, and the Hurst exponent H is the OLS slope of the regression:

$$\log(R/S(n)) = a + H\log(n). \tag{6}$$

A value of H between 0 and 1 indicates long-memory of the time series. Finally, the fractional integration order d is given by $d = H - 0.5$.

The R/S analysis was modified by Lo [25], resulting in an MRS analysis. It replaces the sample standard deviation in (5) by the root of a modified sample variance, augmented by a weighted sample autocovariance up to lag q:

$$\sigma_n^2(q) = \sigma_Y^2(q) + \frac{2}{n}\sum_{j=1}^{q} w_j(q)[\sum_{t=j+1}^{n}(Y_t - \bar{Y}_n)(Y_{t-j} - \bar{Y}_n)] \tag{7}$$

with the weights

$$w_j(q) = 1 - \frac{j}{q+1},\ q < n. \tag{8}$$

The rest of the procedure is the same as R/S. The MRS statistic was later found to be too strong to indicate a true long-memory process.

Peng et al. [27] proposed the de-trended fluctuation analysis (DFA), which aims at detecting the long-range correlation embedded in a seemingly non-stationary time series. The algorithm first integrated the times series in the following way:

$$X(k) = \sum_{t=1}^{k}[Y(k) - \bar{Y}]. \tag{9}$$

Then the integrated time series $\{X(k)\}$ is divided into m non-overlapping sub-series of length n, and an OLS line is fitted to each sub-series. This is the so-called "local trend" of the sub-series. Denoted by $X_n(k)$ the y coordinate of the line segments in all sub-series, the detrended and integrated time series is defined as:

$$\tilde{X}(k) = X(k) - X_n(k). \tag{10}$$

Next, the root-mean-square fluctuation of $\tilde{X}(k)$ is calculated by:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\tilde{X}^2(k)}, \tag{11}$$

where $N = m \times n$ is the length of the time series. Repeat the above calculations for all time scales (different values of integer n) and a characteristic value of the fluctuations is reflected by the scaling exponent α, which is the slope of the regression line:

$$\log F(n) = \alpha \log n + \beta. \tag{12}$$

An α value in (0.5,1) indicates persistent long-range correlations. In contrast, $0 < \alpha < 0.5$ indicates a different type of long-term correlations that may be called "long-range negative dependence", such that large and small values of the time series are more likely to alternate. Similar to the R/S, $d = \alpha - 0.5$.

2.3 Joint ARFIMA Parameters Estimations

For modeling long-memory processes in practice, researchers have tended more often to use the two-stage estimator. This is probably because of the convenience to apply a traditional ARMA filter to the short-run component in the second stage. An alternative strategy in the literature is to jointly estimate all the parameters d, Φ, and Θ. The ML estimator [34] finds the maximizer of the exact likelihood function:

$$f(Y_T, \Sigma) = (2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} Y_t^{'} \Sigma^{-1} Y_T^{'}\right\}, \tag{13}$$

where $Y_T = [y_1, y_2, \ldots, y_T]' \sim N_T(0, \Sigma)$, and Σ is determined exclusively by the parameters d, Φ, and Θ. The computational cost of the ML method, which involves iterative inversions of the $T \times T$ covariance matrix, prohibits the use of even mildly large samples.

Another alternative is to fit a high-order AR(p) process to the ARFIMA process, and obtain estimation of the ARFIMA parameters from the estimated AR coefficients. For any linear process $\{Y_t\}$ that is invertible of $\{Z_t\}$, there exists constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}, \tag{14}$$

for all t. Based on this observation, there has been a long history on the use of long-order AR processes, which dates back to the early work of Yule [38] and later Akaike [1, 2] and Parzen [26]. For invertible processes, the AR approximation works very well, provided that the autoregression order is allowed to go to infinity as the sample size grows. Therefore, it has become a standard procedure to analyze empirical time series using AR models. Poskitt [28] extended the theory of AR approximation to non-standard situations when the process is fractionally integrated and non-invertible. His work verified a conjecture of Beran [7] concerning the predictive optimality of AIC due to Shibata [32] to fractionally integrated and non-invertible processes. His major results show an L_2 norm convergence of the difference vector between the least squares and the true autoregressive coefficients, assuming the autoregression order

$$p_T = o\left\{ (\frac{T}{\ln T})^{0.5-d'} \right\}, \tag{15}$$

where $d^{'} = \max\{0, d\}$, and d is the fractional integration order. Empirical studies on the AR approximation to long-memory processes include [8, 9, 23], among others.

The estimators for ARFIMA parameters based on autoregressive approximation [15] minimizes the distance:

$$D = (\tilde{\mathbf{a}} - \delta(d, \Phi, \Theta))^{'} \Omega (\tilde{\mathbf{a}} - \delta(d, \Phi, \Theta)), \tag{16}$$

where $\tilde{\mathbf{a}}$ is a preliminary estimates for the coefficients of the $AR(\infty)$ representation up to order k, and the weight matrix Ω is the inverse of the estimated covariance matrix of $\tilde{\mathbf{a}}$. The theoretical values of the AR coefficients are represented explicitly by the ARFIMA parameters d, Φ, and Θ:

$$\delta_k = -b_k + \sum_{i=1}^{q} \theta_i a_{k-i} + \sum_{i=1}^{p} \phi_i b_{k-i}, \tag{17}$$

where the b_k's are the coefficients satisfying $\sum_{j=0}^{\infty} b_j L^j = (1-L)^d$. More precisely, $b_0 = 1$, $b_1 = -d$, $b_2 = d(d-1)/2$, $b_j = b_{j-1}(j-1-d)/j$, $j \geq 3$. The order k is chosen to be $3\ln(T) - 6$, rounded to the nearest integer.

3 Sparse AR Approximation to ARFIMA Based on Penalized Conditional Likelihood

Theoretically, the more lags one includes in the approximating AR model, the better the result would be. Including too many parameters in the model, unfortunately, brings unnecessary noise and harms the estimation accuracy. A consevative way is not to include many lags and a rule of thumb in the literature is to set the AR order $p = 3\ln(T) - 6$, rounded to the nearest integer, where T is sample size. We propose to approximate ARFIMA models by sparse AR models, whose coefficient profiles have sparse structures, and to set the order p to be much larger than $O(\ln(T))$. The sparse structure of the AR coefficients is realized by the penalized conditional maximum likelihood estimation [31], which excludes insignificant lags from the model and removes unnecessary noise in a data-driven way.

One question arises naturally as to why the estimation for the ARFIMA parameters is needed at all, if our goal is to predict the future based on past observations. Apparently, a more straightforward way is to carry out prediction directly based on the estimated high-order AR(p) representation, which can be viewed as a truncation of the infinite order autoregressive expansion of the ARFIMA model in (1):

$$Y_t = \sum_{j=1}^{\infty} \Phi_j Y_{t-j} + \epsilon_t, \tag{18}$$

where $\Phi_j = O(j^{-d-1})$ as $j \to \infty$ [28]. Therefore, we also propose a direct forecasting of ARFIMA models based on the estimated AR process.

4 Simulation

In this section, we compare the performances of the estimators for the ARFIMA(p,d,q) parameters based on autoregressive approximation, sparse autoregressive approximation, and exact likelihood approaches.

4.1 An ARFIMA(0,d,0) Model

We compare the three methods for an ARFIMA(0,d,0) model where $d = \pm 0.2, \pm 0.3, \pm 0.4$. As discussed above, the exact ML method only allows for data with up to 100 observations. So we first simulate 100 independent ARFIMA time series, each having 100 observations. Inferences are drawn based on all the three methods and the results are summarized in Table 1. For a small sample like this, the performances of the three methods are comparable. The difference between the AR-based and the sparse AR-based estimators are not distinct either. We continue the study with larger samples for which the exact ML method does not work anymore. Apparently, a larger sample size brings in drastic improvements on the performances of the estimators based on AR approximations, and this is unattainable by the small sample ML estimate. In addition, the sparse AR-based estimator performs consistently better than the AR-based estimator.

Table 1: Average bias and RMSE of the exact ML estimates, the AR-based and sparse AR-based estimates for the ARFIMA (0,d,0) parameter from 100 independent replications of 100 observations

	bias			**RMSE**		
d	d_{ML}	d_{AR}	d_{spsAR}	d_{ML}	d_{AR}	d_{spsAR}
0.2	0.063	0.0673	0.0669	0.0921	0.09	0.0897
-0.2	0.0636	0.0624	0.0688	0.0789	0.0841	0.0869
0.3	0.0629	0.0864	0.0888	0.0813	0.1032	0.1025
-0.3	0.0674	0.0723	0.0987	0.0814	0.0898	0.1054
0.4	0.0686	0.0917	0.0827	0.0885	0.1118	0.108
-0.4	0.0717	0.0684	0.0673	0.09	0.0882	0.0832

Table 2: Average bias and RMSE of the AR-based and sparse AR-based estimates for the ARFIMA (0,d,0) parameter from 100 independent replications of 1000 observations

	bias		**RMSE**	
d	d_{AR}	d_{spsAR}	d_{AR}	d_{spsAR}
0.2	0.0212	0.0207	0.0265	0.0258
-0.2	0.0225	0.0216	0.0279	0.0276
0.3	0.0293	0.0263	0.0377	0.0337
-0.3	0.0249	0.0243	0.0306	0.0293
0.4	0.0407	0.0347	0.0539	0.0464
-0.4	0.0346	0.0304	0.0407	0.0372

4.2 An ARFIMA(1,d,1) Model

We proceed to compare the three methods for an ARFIMA(1,d,1) model, where $|d| = 0.2$, $|\phi_1| = 0.7$, and $|\theta_1| = 0.3$. We consider all the combinations of the parameter signs. Obviously large samples are more advantageous than small ones. So we only compare the AR and sparse AR based estimators for a larger sample size $T = 1000$. Table 3 and Table 4.2 report the average bias and RMSE of 100 independent replications, respectively. Notice that both the average bias and RMSE reduce significantly by including more observations in the sample and, consequently, more legs in the approximating AR model. What is more important, the sparse AR-based estimator now shows consistently better estimation bias and efficiency over the regular AR-based estimator.

Table 3: Average bias of the AR-based and sparse AR-based estimators for the ARFIMA (1,d,1) parameters from 100 replications of 1000 observations each method

d	ϕ^1	θ^1	d_{AR}	d_{sps}	ϕ^1_{AR}	ϕ^1_{sps}	θ^1_{AR}	θ^1_{sps}
0.2	0.7	0.3	0.151	0.136	0.115	0.105	0.081	0.077
-0.2	0.7	0.3	0.158	0.122	0.143	0.123	0.116	0.010
0.2	-0.7	0.3	0.066	0.062	0.021	0.021	0.080	0.078
-0.2	-0.7	0.3	0.085	0.062	0.023	0.022	0.095	0.075
0.2	0.7	-0.3	0.119	0.109	0.104	0.094	0.034	0.035
-0.2	0.7	-0.3	0.111	0.102	0.090	0.082	0.041	0.040
0.2	-0.7	-0.3	0.040	0.038	0.042	0.042	0.072	0.070
-0.2	-0.7	-0.3	0.032	0.032	0.041	0.040	0.068	0.069

5 Analysis of a Foreign Exchange Rate Data

There has been evidence of long-memory in exchange rate data (see, e.g., [11]), and therefore, ARFIMA models are considered the appropriate model to analyze such data. In this section we provide an analysis of a weekend exchange rate data of Japanese yen versus US dollar using ARFIMA models. Historical prices are quoted for the period from 1/2/1990 to 10/1/2010. Thursday prices are used when Friday quotations are not available. The raw data consists of 1086 observations in total, and the input logarithm return data consists of 1085 observations. The AR order p and MA order q are set to be $p = 1, 2$ and $q = 1, 2$ based on empirical evidence, and the 4 combinations are all included in the analysis.

Considering the large sample size, the exact ML method is not used. We only use the AR-based and sparse AR-based approximations to estimate the ARFIMA parameters. The first 1000 observations of the log return data are used for in-sample estimation, and the last 85 are used for out-of-sample forecast. The estimated

Table 4: RMSE's of the AR-based and sparse AR-based estimators for the ARFIMA (1,d,1) parameters from 100 replications of 1000 observations each method

d	ϕ^1	θ^1	d_{AR}	d_{sps}	ϕ^1_{AR}	ϕ^1_{sps}	θ^1_{AR}	θ^1_{sps}
0.2	0.7	0.3	0.193	0.175	0.155	0.141	0.120	0.112
-0.2	0.7	0.3	0.206	0.162	0.247	0.228	0.189	0.171
0.2	-0.7	0.3	0.091	0.091	0.027	0.027	0.111	0.111
-0.2	-0.7	0.3	0.102	0.075	0.028	0.027	0.117	0.089
0.2	0.7	-0.3	0.144	0.131	0.128	0.117	0.045	0.045
-0.2	0.7	-0.3	0.135	0.125	0.107	0.101	0.053	0.051
0.2	-0.7	-0.3	0.049	0.045	0.058	0.057	0.095	0.093
-0.2	-0.7	-0.3	0.039	0.039	0.051	0.051	0.085	0.085

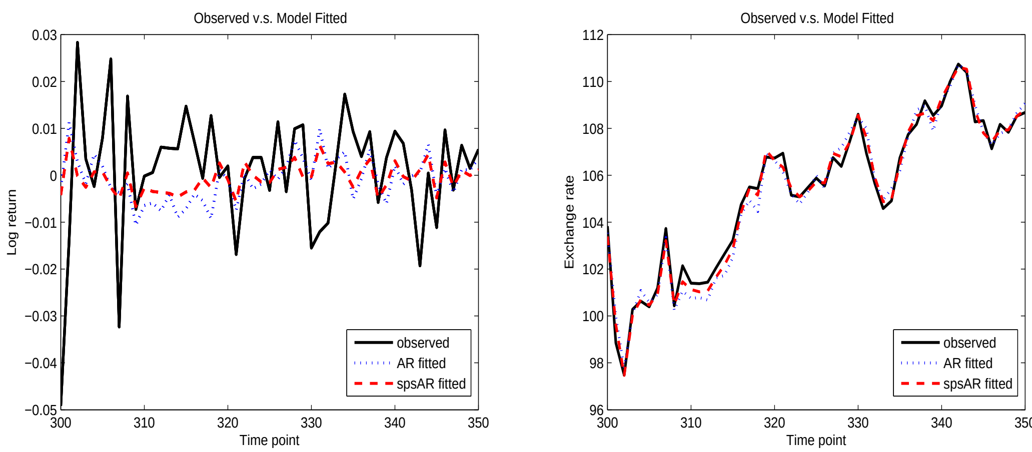

Figure 1: Comparison of direct model fittings from the AR and sparse AR approximations

parameter values for the 4 models are reported in Table 5. Forecasts for the log return series are generated from the truncated AR(∞) representation. In particular, at time t, the k-step-ahead forecast is given by

$$\hat{Y}_{t+k} = \sum_{j=1}^{k-1} \psi_j \hat{Y}_{t+k-j} + \sum_{j=k}^{100} \psi_j Y_{t+k-j}, \tag{19}$$

where $(1-\psi_1 L - \psi_2 L^2 - \cdots) = \Theta^{-1}(L)\Phi(L)(1-L)^d$. That is, the AR($\infty$) representation is truncated at order 100. Forecasts for the original exchange rate series are then constructed and evaluated by mean absolute error (MAE) and root mean squared error (RMSE). The results are summarized in Table 6 for forecast horizons $1, 10, 20$. The differences among the last three models are insignificant. Each one is considered as the optimal model. For all of the four models considered, forecasts based on sparse AR approximation are consistently better than those based on AR approximation. Finally, we compare model fittings from the estimated AR processes directly. Figure 1 shows plots of the fitted values against the observed values for the log return process (left) and the original exchange rate process (right), respectively, for a randomly chosen time period. The fitted values are calculated as the 1-step-ahead predictions from the estimated AR models. The dotted line denotes predictions from the fitted AR process, the dashed line denotes those from the fitted sparse AR process, while the observed values are displayed by the solid line. It is clearly seen from these plots that predictions from the sparse AR approximation capture more precisely the dynamics, while those from the AR approximation show larger biases.

Table 5: Estimated ARFIMA parameters based on AR and sparse AR approximations

ARFIMA Model		Estimated Parameters d	ϕ_1	ϕ_2	θ_1	θ_2
(1,d,1)	AR	0.0885	-0.3177	-	-0.1384	-
	sps	0.0713	-0.3379	-	-0.1745	-
(2,d,1)	AR	0.0317	0.1856	0.115	0.3093	-
	sps	0.0091	0.2377	0.1159	0.3398	-
(1,d,2)	AR	0.0221	0.1986	-	0.3146	-0.122
	sps	0.0087	0.2386	-	0.3404	-0.1153
(2,d,2)	AR	0.0211	0.2082	-0.1011	0.3195	-0.2172
	sps	0.008	0.2533	-0.0585	0.3526	-0.1714

6 Conclusion

We have introduced a new estimation method for ARFIMA processes based on the sparse AR approximation. It has been shown, by simulations and real data applications, to be simpler and more effective than the existing methods: 1) it corrects the bias of the two-stage semiparametric estimation; 2) it overcomes the disadvantage of small sample size of the exact likelihood; 3) it improves on the estimation and prediction accuracy of the (non-sparse) AR approximation. The sparse AR coefficients are realized by the penalized conditional maximum likelihood estimation proposed by Sang and Sun [31]. It accounts for both Gaussian and non-Gaussian cases, allowing the innovation process to be heavy-tailed. Therefore, our proposed method is especially general and flexible in that it can accommodate both long-memory and heavy-tail simultaneously, which are two common characteristics of time series data.

References

[1] Akaike, H., Statistical predictor identification, *Annals of the Institute of Statistical Mathematics*, vol.22, pp.203–217, 1970.

[2] Akaike, H., Maximum likelihood identification of Gaussian autoregressive moving average model, *Biometrica*, vol.60, pp.255–265, 1973.

[3] Anderson, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, Modeling and forecasting realized volatility, *Econometrica*, vol.71, pp.529–626, 2003.

[4] Baillie, R.T., and G. Kapetanios, Nonlinear models for strongly dependent processes with financial applications, *Journal of Econometrics*, vol.147, no.1, pp.60–71, 2008.

[5] Barkoulas, J.T., and C.F. Baum, Long term dependence in stock returns, *Economic Letters*, vol.53, no.3, pp.253–259, 1996.

[6] Barkoulas, J.T., W.C. Labys, and J. Onochie, Long memory in futures prices, *Financial Review*, vol.34, no.1, pp.91–100, 1996.

[7] Beran, J., Fitting long-memory models by generalized linear regression, *Biometrika*, vol.80, no.4, pp.817–822, 1992.

[8] Berk, K.N., Consistent autoregressive spectral estimates, *The Annals of Statistics*, vol.2, pp.489–502, 1974.

[9] Bhansali, R.J., Linear prediction by autoregressive model fitting in the time domain, *The Annals of Statistics*, vol.6, pp.224–231, 1978.

[10] Bickel, P., and E. Levina, Regularized estimation of large covariance matrices, *Annals of Statistics*, vol.36, no.1, pp.199–227, 2008.

Table 6: MAE and RMSE for the 8 combinations of models and methods

		Forecast Evaluation					
Model		1-stp MAE	10-stp MAE	20-stp MAE	1-stp RMSE	10-stp RMSE	20-stp RMSE
(1,d,1)	AR	0.047	0.044	0.033	0.439	0.412	0.307
	sps	0.044	0.042	0.032	0.412	0.385	0.283
(2,d,1)	AR	0.039	0.036	0.025	0.357	0.331	0.235
	sps	0.036	0.033	0.023	0.334	0.308	0.215
(1,d,2)	AR	0.037	0.035	0.024	0.346	0.320	0.225
	sps	0.036	0.033	0.023	0.334	0.308	0.215
(2,d,2)	AR	0.037	0.035	0.024	0.346	0.320	0.225
	sps	0.036	0.033	0.023	0.333	0.307	0.215

[11] Cheung, Y.W., Long memory in foreign-exchange rates, *Journal of Business and Economic Statistics*, vol.11, pp.93–101, 1996.

[12] Cheung, Y.W., and K.S. Lai, Do gold market returns have long memory?, *The Financial Review*, vol.28, no.2, pp.181–202, 1993.

[13] Fan, J., and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association, Theory and Methods*, vol.96, no.456, pp.1348–1360, 2001.

[14] Frank, I., and J. Friedman, A Statistical view of some chemometrics regression tools, *Technometrics*, vol.35, no.2, pp.109–148, 1993.

[15] Galbraith, J., and V. Zinde-Walsh, Autoregression-based estimators for ARFIMA models, CIRANO working paper No. 2001s-11, 2001.

[16] Geweke, J., and S. Porter-Hudak, The estimation and application of long memory time series models, *Journal of Time Series Analysis*, vol.4, pp.221–238, 1983.

[17] Granger, C.W.J., and R. Joyeux, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, vol.1, pp.15–39, 1980.

[18] Greene, M.T., and B.D. Fielitz, Long-term dependence in common stock returns, *Journal of Financial Economics*, vol.4, no.3, pp.339–349, 1977

[19] Hosking, J.R.M., Fractional differencing, *Biometrika*, vol.68, pp.165–176, 1981.

[20] Huang, J., S. Ma, H. Xie, and C. Zhang, A group bridge approach for variable selection, *Biometrika*, vol.96, no,2, pp.339–355, 2009.

[21] Hurst, H.R., Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, vol.116, pp.770–799, 1951.

[22] Kokoszka, P.S., and M.S. Taqqu, Infinite variance stable moving averages with log memory, *Journal of Econometrics*, vol.73, pp.79–99, 1996.

[23] Lewis, R., and G.C. Reinsel, Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis*, vol.16, pp.393–411, 1985.

[24] Li, W.K., and A.I. McLeod, Fractional time series modeling, *Biometrika*, vol.73, pp.217–221, 1986.

[25] Lo, A.W., Long-term memory in stock market prices, *Econometrica*, vol.59, pp.1279–1313, 1991.

[26] Parzen, E., Some recent advances in time series modeling, *IEEE Transactions on Automatic Control*, vol.19, pp.723–730, 1974.ieee transactions on automatic control

[27] Peng, C.K., S. Havlin, H.E. Stanley, and A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, *Chaos*, vol.5, no.1, pp.82–87, 1995.

[28] Poskitt, D.S., Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible cases, *Annals of the Institute of Statistical Mathematics*, vol.59, no.4, pp.697–725, 2007.

[29] Robinson, P.M., Gaussian semiparametric estimation of long range dependence, *The Annals of Statistics*, vol.23, no.5, pp.1630–1661, 1995.

[30] Robinson, P.M., *Time Series with Long Memory*, Oxford University Press Inc., New York, 2003.

[31] Sang, H., and Y. Sun, Simultaneous sparse model selection and coefficient estimation for heavy-tailed autoregressive processes, arXiv:1112.2682, 2012.

[32] Shibata, R., Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *The Annals of Statistics*, vol.8, no.1, pp.147–164, 1980.

[33] Shimotsu, K., and P.C.B. Phillips, Local whittle estimation of fractional integration and some of its variants, *Journal of Econometrics*, vol.130, no.2, pp.209–233, 2006.

[34] Sowell, F.B., Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, vol.53, pp.165–188, 1992.

[35] Sun, Y., and X. Lin, Regularization for stationary multivariate time series, *Quantitiative Finance*, vol.12, pp.573–586, 2012.

[36] Tibshirani, R., Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society*, series B (statistical methodology), vol.58, no.1, pp.267–288, 1996.

[37] Tieslau, M.A., P. Schmidt, and R.T. Baillie, A minimum-distance estimator for long-memory processes, *Journal of Econometrics*, vol.71, pp.249–264, 1996.

[38] Yule, G.U., On the time-correlation problem with especial reference to the variate-difference correlation method, *Journal of the Royal Statistical Society*, vol.84, no.4, pp.497–537, 1921.