Pak. J. Statist.
2019 Vol. 35(3), 249-260

MODELING OF ECONOMIC DATA USING ESTIMATING FIRST-ORDER FUNCTIONAL AUTOREGRESSIVE MODELS BASED ON SEMI-PARAMETRIC APPROACH

J. Amiri[1], **R. Farnoosh**[2§] and **M.H. Behzadi**[1]
[1] Department of Statistics, Science and Research Branch
Islamic Azad University, Tehran, Iran
[2] School of Mathematics, Iran University of Science
and Technology, Tehran 16846, Iran
[§] Corresponding author Email: rfarnoosh@iust.ac.ir

ABSTRACT

In this paper, we have applied the semi-parametric method for estimating the regression function to estimate first-order functional autoregressive models. To this end, the conditional nonlinear least squares estimation method is used to estimate the parameter and the nonparametric kernel approach is applied to estimate the adjustment factor. In this case, we have also proven some asymptotic behaviors of this estimator and shown its adequacy using simulations and a real series of financial data in Iran's Melli-Bank.

KEYWORDS

Semi-parametric estimation, the conditional nonlinear least squares estimation method, functional autoregressive model, nonparametric kernel.

1. INTRODUCTION

In this paper, we consider the following autoregressive model introduced by Yu et al. (2009):

$$Z_t = f\left(Z_{t-1}\right) + \varepsilon_t, \quad t = 1, \ldots, n \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d random variables with mean 0 and variance σ^2 and Z_t is independent of ε_t for all values of t.

If we have information about the structure under the study, we can assume that $f(.)$ has a parametric form as a previous choice as follows:

$$f(x) \in \{g(x,\theta); \theta \in \Theta\} \tag{2}$$

where, $\Theta \subseteq R^p$ is the parametric space. In this case, the regression function estimator is replaced by an unknown parameter vector estimator θ, which results in the regression function $f(.)$ being estimated as follows:

$$\hat{f}(x) = g(x, \hat{\theta}) \tag{3}$$

in which $\hat{\theta}$ is an estimator of θ. But if there exist little information about the nature of $f(.)$, using a nonparametric estimator would be deemed reasonable. See Comte (2004), Fan and Truong (1993) and Francesco (2005). For instance, an estimator of the auto-regression function $f(.)$, also known as the kernel estimator is as shown below:

$$\hat{f}\left(x\right) = \frac{\sum_{j=1}^{n} K\left(\frac{Z_{j-1} - x}{h_n}\right) Z_j}{\sum_{j=1}^{n} K\left(\frac{Z_{j-1} - x}{h_n}\right)} \tag{4}$$

where $K(.)$ is a kernel and h_n is the bandwidth dependent on n. This kernel estimator is a special case of the local polynomial estimator proposed by Hardel and Tibakov (1997).

In this paper, the parametric estimator of the regression function (3) is considered as an initial guess for $f(x)$. When this initial parametric approximation is corrected by a nonparametric modifier $\zeta(x)$, a semi-parametric estimator is obtained in the form of $g\left(x, \hat{\theta}\right)\zeta(x)$, in which a quadratic polynomial fitting criterion is used to estimate the adjustment factor. If the estimator of $\zeta(x)$ is represented by $\hat{\zeta}(x)$, the final estimator is obtained in the form of $\hat{f}(x) = g\left(x, \hat{\theta}\right)\hat{\zeta}(x)$. This is a special semi-parametric estimation method with a parametric estimate as the starting point and a nonparametric estimate as the adjustment factor.

Similar to this method, Farnoosh and Mortazavi (2011) used a semi-parametric method for estimating nonlinear function with dependent errors, and Nademi and Farnoosh (2014) used this method for estimation in a combination of autoregressive and autoregressive conditionally heteroscedastic models and also Farnoosh et al. (2014) use a semi-parametric estimation for regression functions in the partially linear autoregressive time series model. Farnoosh et al. (2017) applied a semi-parametric estimation for the nonlinear vector autoregressive time series model. Farnoosh et al. (2019) proposed a semi-parametric estimation for the first-order nonlinear autoregressive time series model with independent and dependent errors. Hajrajabi and Mortazavi (2019) used a semi-parametric approach for the first-order nonlinear autoregressive model with skew normal innovations. Samadi et al. (2019) applied a semi-parametric approach for modelling multivariate nonlinear time series.

We will continue to discuss the following. In Section 2, a least-squares estimate for estimating the parameter vector θ and the semi-parametric estimator by the local polynomial fitting criterion is introduced. The weak consistency of the semi-parametric estimator is proved in Section 3. In Section 4, we examine the adequacy of the proposed method, using simulation and an actual example.

2. SEMI-PARAMETRIC ESTIMATOR

Consider the following model:

$$Z_t = f\left(Z_{t-1}\right) + \varepsilon_t \qquad t = 1, 2, ..., n \tag{5}$$

where $\{\varepsilon_t\}$ is a sequence of random variables $(i.i.d)$ with mean zero and variance σ^2 and also ε_t and Z_t are independent for each value of t.

This is aimed at estimating the regression function $f(x)$ that can be in the form of $g(x, \theta)$, which $g(x, \theta)$ is a known function of x and θ. The actual value of θ is shown by θ_0 as defined below:

$$\theta_\circ = \arg\min_{\theta\in\Theta} E\left(Z_t - E_\theta\left(Z_t \mid Z_{t-1}\right)\right)^2 \tag{6}$$

For model (5), the parameter θ is estimated using the conditional nonlinear least squares (CNLS) estimation method:

$$Q_n\left(\theta\right) = \sum_{j=1}^{n}\left\{\left(Z_j - g\left(Z_{j-1}, \theta\right)\right)^2\right\} \tag{7}$$

and as a result:

$$\hat{\theta}_n = \arg\min Q_n\left(\theta\right) \ , \ \theta \in \Theta \tag{8}$$

In fact, $\hat{\theta}_n$ is the CNLS estimator based on data $Z_0, Z_1, ..., Z_n$.

Now, we should estimate $\zeta(x)$ in $f\left(x\right) = g\left(x, \theta\right)\zeta\left(x\right)$ using the same idea of Hjort and Jones (2004) and Farnoosh and Mortazavi (2011). For this purpose, the local L_2-fitting criterion is defined using the kernel method as follows:

$$q\left(x, \zeta_n\right) = \frac{1}{h_n}\sum_{j=1}^{n} k\left(\frac{Z_{j-1} - x}{h_n}\right)\left\{f_n(Z_{j-1}) - g\left(Z_{j-1}, \hat{\theta}_n\right)\zeta_n\right\}^2 \tag{9}$$

where $f(.)$ is an autoregressive function with sample size n. Therefore, we obtain the estimator $\hat{\zeta}\left(x\right)$ by minimizing the local L_2-fitting criterion with respect to $\zeta(x)$, using the kernel method as follows:

$$\hat{\zeta}\left(x\right) = \frac{\sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1} - x}{h_n}\right) g\left(Z_{j-1}, \hat{\theta}_n\right) f\left(Z_{j-1}\right)\right]}{\sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1} - x}{h_n}\right) g^2\left(Z_{j-1}, \hat{\theta}_n\right)\right]} \tag{10}$$

Resulting estimator $f(x)$ to be as follows:

$$\hat{f}(x) = g\left(x, \hat{\theta}_n\right)\hat{\zeta}(x) \tag{11}$$

And given the fact that the errors in the model are small, we will have:

$$\sum_{j=1}^{n} K\left(\frac{Z_{j-1}-x}{h_n}\right) g(Z_{j-1}, \hat{\theta}_n) f(Z_{j-1}) \approx \sum_{j=1}^{n} K\left(\frac{Z_{j-1}-x}{h_n}\right) g\left(Z_{j-1}, \hat{\theta}_n\right) Z_{j-1} \tag{12}$$

therefore:

$$\tilde{\zeta}(x) = \frac{\sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1}-x}{h_n}\right) g\left(Z_{j-1}, \hat{\theta}_n\right) Z_j\right]}{\sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1}-x}{h_n}\right) g^2\left(Z_{j-1}, \hat{\theta}_n\right)\right]} \tag{13}$$

which could be defined as:

$$\tilde{\zeta}(x) = \frac{\tilde{S}(x)}{\tilde{T}(x)} \tag{14}$$

where,

$$\begin{aligned} \tilde{S}(x) &= \frac{1}{nh_n} \sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1}-x}{h_n}\right) g\left(Z_{j-1}, \hat{\theta}_n\right) Z_j\right] \\ \tilde{T}(x) &= \frac{1}{nh_n} \sum_{j=1}^{n}\left[k\left(\frac{Z_{j-1}-x}{h_n}\right) g^2\left(Z_{j-1}, \hat{\theta}_n\right)\right] \end{aligned} \tag{15}$$

Finally, the estimator of the autoregressive function is obtained as follows:

$$\tilde{f}(x) = g(x, \hat{\theta}_n)\tilde{\zeta}(x). \tag{16}$$

3. CONSISTENCY PROPERTIES OF THE SEMI-PARAMETRIC ESTIMATOR

In this section, we examine the consistency properties and asymptotic behavior of the estimator presented in the previous section. Consider the following assumptions:

a) The sequence $\{Z_t\}$ is a α-mixing and stationary ergodic sequence of integrable random variables and ε_t is a strictly stationary sequence of random variables.

b) $f(x)$ and $g(x,\theta)$ are bounded and continuous with respect to x away from zero in a neighborhood of x, and also $\partial g/\partial\theta$ and $\partial^2 g/\partial\theta_i\,\partial\theta_j$ exist and are continuous for all values of $\theta \in \Theta$.

c) Z_0 has the distribution $\pi(.)$ in such a way that the density $\mu(.)$ of the existing $\pi(.)$, is continuous, bounded, and positive in a x-point neighborhood.

d) The kernel $K: R^1 \to R^+$ is a bounded, symmetric, and compact functions, such that $K(.) > 0$ and:

$$\sup_{x\in R} \left|K(x)\right| < \infty,\ \int \left|K(x)\right| dx < \infty,\ \int \left|K^2(x)\right| dx < \infty,\ \ |x|K(x) \to 0 \text{ as } |x| \to \infty.$$

We continue to prove the consistency theorems of the proposed estimator. Lemma (3.1) is proposed by Yamato (1970).

Lemma 3.1:

If a sequence of functions $\{\gamma_n(x)\}$ is convergent to a function $\gamma(x)$ at the point x when $n \to \infty$ it, then $\sum_{j=1}^{n} \gamma_j(x)/n \to \gamma(x)$ will be convergent when $n \to \infty$. If a sequence of functions $\{\gamma_n(x)\}$ is uniformly bounded in R and converges to a bounded function $\gamma(x)$ when $n \to \infty$ is uniformly in R, then $\sum_{j=1}^{n} \gamma_j(x)/n \to \gamma(x)$ converges uniformly in R, when $n \to \infty$.

Lemma 3.2:

If K applies to the above assumptions, then, when $n \to \infty$, at each continuous point x from S, we will have:

$$E\left(\tilde{S}_n(x)\right) \to S(x)$$

where $S(x) = g(x,\theta) f(x) \mu(x)$ or $S(x) = g^2(x,\theta) \xi(x) \mu(x)$.

Proof:

Lemma 3.1 and assumption of $E\left[Z_n \middle| Z_{n-1} = u\right] = f(u)$ suffice to show:

$$h_n^{-1} E\left[Z_n k\left(\frac{Z_{n-1} - x}{h_n}\right) g\left(Z_{n-1}, \hat{\theta}_n\right) \right] \to S(x)$$

at each continuous point x from $S(.)$. Now with regard to the above assumptions, we have:

$$h_n^{-1} E\left[Z_n k\left(\frac{Z_{n-1} - x}{h_n}\right) g\left(Z_{n-1}, \hat{\theta}_n\right) \right]$$
$$= h_n^{-1} \int_R E\left[Z_n k\left(\frac{Z_{n-1} - x}{h_n}\right) g\left(Z_{n-1}, \hat{\theta}_n\right) \middle| Z_{n-1} = u \right] \mu(u) du$$

$$= h_n^{-1} \int_R k\left(\frac{u-x}{h_n}\right) g(u,\hat{\theta}_n) E\left[Z_n \middle| Z_{n-1} = u\right] \mu(u) du$$

$$= h_n^{-1} \int_R k\left(\frac{u-x}{h_n}\right) g(u,\hat{\theta}_n) f(u) \mu(u) du$$

$$= \int_R k(t) g(x+th_n,\hat{\theta}_n) f(x+th_n) \mu(x+th_n) dt$$

$$\to g(x,\theta) f(x) \mu(x)$$

when $n \to \infty$. The fifth equality is obtained using the conversion $t = \frac{u-x}{h_n}$, and the final result is obtained in light of the fact that $S(.)$ is continuous in x,

$$h_n^{-1} E\left[Z_n k\left(\frac{Z_{n-1}-x}{h_n}\right) g\left(Z_{n-1},\hat{\theta}_n\right)\right] \to S(x)$$

So the proof is complete.

Using the (3.1) and (3.2) lemmas we will have:

$$E\left(\tilde{S}(x)\right) \to S(x) \ \ as \ \ n \to \infty$$

Lemma 3.3:

Suppose that the conditions of Lemma (3.1) exist, and also $nh_n \to \infty$ when $n \to \infty$. If $m(u) = E\left(Z_i^2 \middle| Z_{i-1} = u\right) < \infty$ then, at each continuous point x from S,

$$Var\left[\tilde{S}(x)\right] \to 0 \ \ as \ \ n \to \infty.$$

Proof:

$$Var[\tilde{S}(x)] = n^{-2} Var\left[\sum_{j=1}^{n} \frac{1}{h_n} Z_j k\left(\frac{Z_{j-1}-x}{h_n}\right) g\left(Z_{j-1},\hat{\theta}_n\right)\right]$$

$$= n^{-2} h_n^{-2} \sum_{j=1}^{n} Var\left[Z_j k\left(\frac{Z_{j-1}-x}{h_n}\right) g\left(Z_{j-1},\hat{\theta}_n\right)\right]$$

$$\le n^{-2} h_n^{-2} \sum_{j=1}^{n} \int_R E\left[Z_j^2 k^2\left(\frac{Z_{j-1}-x}{h_n}\right) g^2(Z_{j-1},\hat{\theta}_n) \middle| Z_{j-1} = u\right] \mu(u) du$$

$$= n^{-2} h_n^{-2} \sum_{j=1}^{n} \int_R k^2 \left(\frac{u-x}{h_n}\right) g^2(u,\hat{\theta}_n) m(u) \mu(u) du$$

where $m(u) = E(Z_j^2 \mid Z_{j-1} = u) < \infty$. Yet using Lemma (3.2) we will have:

$$h_n^{-1}\int_R k^2\left(\frac{u-x}{h_n}\right)g^2\left(u,\hat{\theta}_n\right)m(u)\mu(u)\,du \to m(x)\mu(x)g^2(x,\theta)\int_R k^2(t)\,dt < \infty .$$

So, using an application of Lemma (3.1), we will have:

$$(nh_n^2)^{-1}\int_R k^2\left(\frac{u-x}{h_n}\right)g^2\left(u,\hat{\theta}_n\right)m(u)\mu(u)\,du \to 0, \quad as\ n\to\infty .$$

The point-to-point (weak) consistency of the semi-parametric estimator is proved in the following theorems:

Theorem 3.1:

If K is valid under the assumptions (a) to (d), and $\{h_n\}$ holds true in the condition $nh_n \to \infty$ when $n\to\infty$, then at each continuous point x from $S(x)$ and $T(x)$, so that $T(x) > 0$,

$$\tilde{\zeta}(x) \to \zeta(x)$$

when $n\to\infty$, will be convergent in probability.

Proof:

In Lemma (3.2), we proved that $\tilde{S}(x)\to S(x)$ is convergent in probability at each continuous point x from $S(.)$ when $n\to\infty$. Also, $\tilde{T}(x)\to T(x) > 0$ will be convergent in probability at each continuous point x from $T(.)$ when $n\to\infty$, because

$$\begin{aligned}
E\left(\tilde{T}_n(x)\right) &= \frac{1}{h_n}E\left(\left[k\left(\frac{Z_{n-1}-x}{h_n}\right)g^2(Z_{n-1},\hat{\theta}_n)\right]\right)\\
&= \frac{1}{h_n}\int_R k\left(\frac{u-x}{h_n}\right)g^2\left(u,\hat{\theta}_n\right)\mu(u)\,du\\
&= \int_R k(t)\,g^2\left(x+th_n,\hat{\theta}_n\right)\mu\left(x+th_n\right)dt\\
&\to g^2(x,\theta)\mu(x) \ \ as\ n\to\infty
\end{aligned}$$

so,

$$E(\tilde{T}_n(x)) = T(x) \ \ as\ n\to\infty$$

where

$$T(x) = g^2(x,\theta)\mu(x)$$

And similar to Lemma (3.3):

$$Var(\tilde{T}(x)) \to 0 \;\; as \;\; n \to \infty$$

And consequently, using an application of Slutsky Lemma, we will have:

$$\tilde{\zeta}(x) \to \zeta(x)$$

will be convergent in probability, when $n \to \infty$.

Theorem 3.2:

Suppose $\tilde{f}(x)$ is the estimator introduced in (16). Under conditions of (a)-(d), $\tilde{f}(x) \to f(x)$ will be convergent in probability, when $n \to \infty$.

Proof:

Given the strong consistency of the conditional least squares estimator and theorem (3.1) and using Slutsky Lemma, the theorem is proved.

4. SIMULATION STUDIES AND EMPIRICAL EVIDENCE

4.1 Simulation Studies

In this section, we examine the adequacy of the semi-parametric method for estimating parameters and the regression function, using a simulated example and a real data series. Consider the following model:

$$Z_t = f(Z_{t-1}) + \varepsilon_t \;\; , \;\; t = 1, 2, ..., n$$

in which ε_t has an independent and identically distributed normal distribution with mean zero and variance $(0.125)^2$. Using the nonlinear function below, we generate data with different sample sizes:

$$f(x) = a \exp\left(-bx^2\right) + cx$$

For this model, suppose $g(x, \theta) = \theta_1 \exp\{-\theta_2 x\}$ and the values of a, b and c are the coefficients of the nonlinear regression model, respectively. We estimate the function $f(x)$ using the estimator introduced in (16) and the Gaussian kernel and the bandwidth proportional to the sample size. To evaluate the adequacy of the proposed estimation method, we calculate the average square error (ASE) as follows:

$$ASE = \frac{1}{P} \sum \left[\tilde{f}(x_i) - f(x_i) \right]^2$$

We show the square root of the ASE by MSE. The results are presented in the table below.

Table 1
MSE Values, $h_n = 0.006n$

n	MSE
150	0.0291
200	0.0359
250	0.0382
300	0.0455
350	0.0511
400	0.0543

Figure 1 shows the curves of function $f(x)$ and the semi-parametric estimator under the above model and the selected bandwidth. The blue and red lines indicate the regression function $f(x)$ and the semi-parametric estimation. The simulation results show that the semi-parametric estimator of the autoregressive function works well.

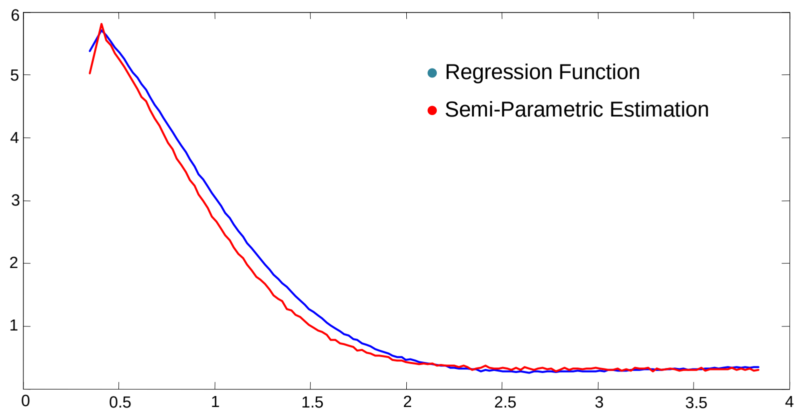

Figure 1: Functional Graph and Semi-Parametric Estimator for MSE = 0.0291, n = 150

4.2 Empirical Application

In order to demonstrate the adequacy of the proposed method in this study, the non-linear autoregressive model is used to predict the annual total deposit amount of Bank Melli Iran (National Bank of Iran) from 2000 to 2015.

In this regard, we set $Y_t^* = Y_t/100$ so that Y_t is the annual amount of the Bank Melli's deposit in billion Rials, at one of the branches, located in the north of Iran, at time t.

The initial model is considered as $g(x,\theta) = \alpha \exp\{\beta x\} + \gamma x$.

Using the proposed semi-parametric method, the regression function has been estimated in the autoregressive prediction model with an independent error for the bank deposit at time t using the Gaussian kernel and bandwidth 23n. The parameters of the function $g(x,\theta)$ are estimated to be 9 and 6.45. Table (2) contains bank real data along with their predicted values. The MSE value in this estimate is 0.0611. Figure (2) refers to the actual values (in blue) and the expected (in red) of Bank Melli's deposit using the nonlinear autoregressive model.

Table 2
Estimate of the Regression Function in the Deposit Amount of the National Bank of Iran

Year	Y_t^*	$\overset{*}{Y}_t$
2000	.68	1.3
2001	.96	1.52
2002	1.19	1.61
2003	1.54	2.01
2004	1.92	2.27
2005	2.10	2.61
2006	2.87	2.92
2007	3.63	3.13
2008	4.34	3.91
2009	4.38	3.67
2010	4.57	4.64
2011	4.93	4.76
2012	5.27	5.47
2013	5.79	6.26
2014	6.92	6.65
2015	7.48	7.59

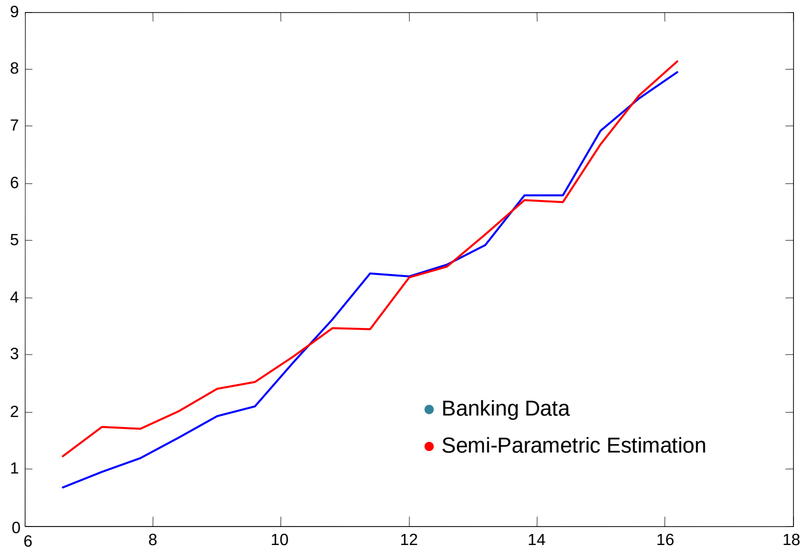

Figure 2: Semi-Parametric Estimation for Banking Data

5. CONCLUSION

In this paper, we examined the semi-parametric method for estimating the first-order functional autoregressive model that has recently been used in economic and financial studies. We proved the weak consistency (convergence in probability) of this estimator and examine the adequacy of the proposed method using the MSE criterion resulted from simulation studies. Finally, we used real data of annual total deposit amount of Bank Melli Iran to confirm the accuracy of the semi-parametric estimation method. but there is a lot of possible extensions and open questions to be addressed in future work. Investigating more general models and nonlinear estimation methods are possible projects to be done in future. Also, the asymptotic behavior of the parameter estimators, and hypothesis tests for the parameters are topics have to be investigated.

REFERENCES

1. Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21, 1900-1925.
2. Farnoosh, R. and Mortazavi, S.J. (2011). A semiparametric method for estimating nonlinear autoregressive model with dependent errors. *Nonlinear Analysis*, 74, 6358-6370.
3. Farnoosh, R., Hajebi, M. and Mortazavi, S.J. (2014). A semiparametric estimation for regression functions in the partially linear autoregressive time series model. *Applications and Applied Mathematics: An International Journal*, 9.2, 573-591.
4. Farnoosh, R., Hajebi, M. and Mortazavi, S.J. (2017). A Semiparametric Estimation for the Nonlinear Vector Autoregressive Time Series Model. *Applications and Applied Mathematics: An International Journal*, 12.1, 78-97.

5. Farnoosh, R., Hajebi, M. and Samadi, Y. (2019). A Semiparametric Estimation for the First-Order Nonlinear Autoregressive Time Series Model with Independent and Dependent Errors. *Iranian Journal of Science and Technology, Transactions A: Science*, 43.3, 905-917.
6. Francesco, A. (2005). Local likelihood for non-parametric ARCH(1) models. *J. Time Series Analyses*, 26, 251-278.
7. Hajrajabi, A. and Mortazavi, S.J. (2019). The first-order nonlinear autoregressive model with skew normal innovations: A semiparametric approach. *Iranian Journal of Science and Technology, Transactions A: Science*, 43.2, 579-587.
8. Hardle, W. and Tsybakov, A. (1997). Local polynomial estimators of volatility function in nonparametric autoregression. *J. Econometrics*, 81, 223-243.
9. Hjort, N.L. and Jones, M.C. (1996). Locally parametric nonparametric density estimation. *Annals of Statistics*, 24, 1619-1647.
10. Nademi, A. and Farnoosh, R. (2014). Mixtures of autoregressive-autoregressive conditionally heteroscedastic models: semi-parametric approach. *Journal of Applied Statistics*, 41.2, 275-293.
11. Samadi, Y., Hajebi, M. and Farnoosh, R. (2019). A semiparametric approach for modelling multivariate nonlinear time series. *Canadian Journal of Statistics*. https://doi.org/10.1002/cjs.11518.
12. Yamato, H. (1971). Sequential Estimation of a Continuous Probability Density Function and Mode. *Bull. Math. Statist.*, 14, 1-12.
13. Yu, Z., Wang, D. and Shi, N. (2009). Semiparametric estimation of regression functions in autoregressive models. *Statistics & Probability Letters*, 79(2), 165-172.