COMPARISON OF DIFFERENCING PARAMETER ESTIMATION FROM ARFIMA MODEL BY SPECTRAL REGRESSION METHODS

(1)Gumgum Darmawan, (2)Nur Iriawan, (2)Suhartono

(1) Student of Statistics Magister Program of ITS;

gumstat_1973@yahoo.com,

(2) Lecturer of Statistics Magister Program of ITS;

Suhartono@statistika.its.ac.id

ABSTRACT

Spectral regression method is one of the popular methods for estimating the difference parameter of ARFIMA(p,d,q) model. Spectral density function of ARFIMA(p,d,q) model was formed to construct linear regression function for estimating the difference parameter d by Ordinary Least Square (OLS). This method has attracted many researchers because it could cope the difficulty in derivation of the autocovariance function of ARFIMA(p,d,q) model. The estimation of d by using regression method could be done directly without knowing p and q parameter. This method was proposed by Geweke and Porter-Hudak (1983) and modified by Reisen (1994) with smoothing periodogram by parzen window. Then, Robinson (1995) added l trimming on this periodogram. Hurvich and Ray (1995) and Velasco(1995a) used modified periodogram by cosine – bell tapered function, Velasco (1999) changed independent variable of spectral regression 2sin(ω_j/2) by j (index of periodogram frequency). In this paper we will compare the estimation accuracy among five methods by using simulation study in two conditions, i.e clean data and data with outlier.

From simulation results, GPH method shows a good performance in estimating the differencing parameter of ARFIMA model both clean data and data with outlier. Above all, estimation of spectral regression methods are better for ARFIMA(1,*d*,0) data than for ARFIMA(0,*d*,1) data.

Keywords : ARFIMA, Ordinary Least Square, Outliers, Periodogram

1. Introduction

Long range dependence or long memory means that observations far away each other are still strongly correlated. The correlation of a long-memory process decay slowly that is with a hyperbolic rate, not exponentially like for example ARMA-process (see figure.1).

The literature on ARFIMA processes has rapidly increased since early contribution by Granger and Joyeux (1980), Hosking (1981) and Geweke and Porter-Hudak (1983). This theory has been widely used in different fields such as meteorology, astronomy, hydrology and economics.

Geweke and Porter-Hudak (1983) presented a very important work on stationary long memory processes. Their paper gave rise to several other works, and presented a proof for the asymptotic distribution of the long memory parameter. These authors proposed an estimator of d as the ordinary least squares estimator of the slope parameter in a simple linear regression of the logarithm of the periodogram. Reisen (1994) proposed a modified form of the regression method, based on a smoothed version of the periodogram function. Robinson (1995a), making use of mild modifications on Geweke and Porter-Hudak's estimator, dealt simultaneously with $d \in (-0.5, 0.0)$ and $d \in (0.0, 0.5)$. Hurvich and Deo (1999), among others, addressed the problem of selecting the number of frequencies necessary for estimating the differencing parameter in the stationary case. Fox and Taqqu (1986) considered an approximated method, whereas Sowell (1992) presented the exact maximum likelihood procedure for estimating the fractional parameter. These two papers considered the estimation procedures for the stationary case. Simulation studies comparing estimates of *d* may be found, for instance, in Bisaglia and Guegan (1998), Reisen and Lopes (1999), Reisen et *al*. (2000,2001).

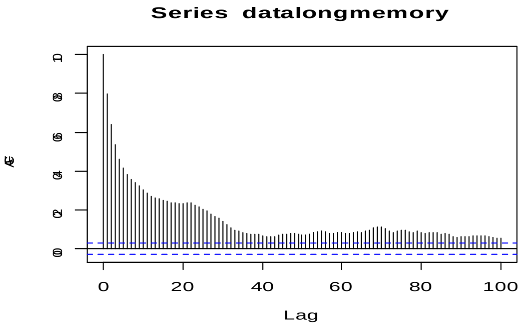


Figure 1. ACF plot of long memory processes

Here, by simulation study, we compare accuracy of estimation the differencing parameter of ARFIMA model by the spectral regression methods. The first estimator is proposed by Geweke and Porter-Hudak (1983), denoted in the following by GPH. The second estimator is the smoothed periodogram regression (SPR), suggested by Reisen (1994).

As a third method we consider the GPH, based on the trimming *l* and bandwith *m*, denoted hereafter by GPHT*r*, suggested by Robinson (1995a). The GPHT*r* method regresses $log\left\{ I\left(\omega_j\right)\right\}$ on $log\left\{ 2sin\left(\omega_j / 2\right)\right\}^2$ for *j* = 2,3,…,*m*, where *j* and *m* are index and bandwith of periodogram respectively. The fourth method is a modified form of the GPH method, denoted hereafter by MGPH, obtained by replacing in the regression equation the quantity $2sin\left(\omega_j / 2\right)$ by *j*.

The Cosine-bell tapered data method in the following by GPHTa, is the fifth approach considered here. In this method the modified periodogram function where the tapered data is obtained from the cosine-bell function, this estimator was also used in the works by Hurvich and Ray (1995) and Velasco (1999a).

2. ARFIMA Model

A well known class of long memory models is the autoregressive fractionally integrated moving average (ARFIMA) process introduced by Granger and Joyeux (1980) and Hosking (1981).

An ARFIMA model (*p*,*d*,*q*) can be defined as follows:

$$\phi(B)(1-B)^d\left(Z_t-\mu\right)=\theta(B)a_t \quad (1)$$

Where

t = index of observation

d = the degree of differential parameter (real number)

μ = mean of observation

$\phi(B)=1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p$, polynomial of AR(i)

$\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$, polynomial of MA(*q*)

$\left(1-B\right)^d$ = fractional differencing operator

a_t is IID$(0,\sigma^2)$.

3. Estimation of the Differencing Parameter

There have been proposed in the literature many estimators for the fractional differencing parameter. We shall concentrate in estimators based upon the estimation of the spectral density function. The semi-parametric estimators describe bellow are obtained taking the logarithm of the spectral density. Estimation of d from ARFIMA (p,d,q) model as follows

a) Construct spectral function of ARFIMA (p,d,q) model

For the ARFIMA model given in equation (1), let $W_t = (1-B)^d Z_t$, and let $f_W(\omega)$ and $f_Z(\omega)$ be the spectral density function of $\{W_t\}$ and $\{Z_t\}$, respectively. Then,

$$f_Z(\omega) = f_W(\omega)\{2\sin(\omega/2)\}^{-2d}, \quad 0 \le \omega \le \pi \tag{2}$$

Where

$$f_Z(\omega_j) = \frac{\sigma_a^2}{2\pi}\left|\frac{\theta_q(exp(-i\omega_j))}{\phi(exp(-i\omega_j))}\right|^2 ,$$

is the spectral density of a regular ARMA (p,q) model. Note that $f_Z(\omega) \to \infty$ as $\omega \to 0$.

b) Take logarithms on both sides of equation (2).

$$ln\{f_Z(\omega_j)\} = d\, ln|1 - exp(-i\omega_j)|^{-2} + ln\, f_W(\omega_j)$$

$$= ln\, f_W(0) + d\, ln|1 - exp(-i\omega_j)|^{-2} + ln\left(\frac{f_W(\omega_j)}{f_W(0)}\right) \tag{3}$$

c) Add $\ln I_Z(\omega_j)$, the natural logarithm of periodogram $\{Z_t\}$ to Both side of equation (3) above,

$$ln\{I_Z(\omega_j)\} = ln\{f_W(0)\} + d\, ln|1 - exp(-i\omega_j)|^2 + ln\left\{\frac{f_W(\omega_j)}{f_W(0)}\right\} + ln\frac{I_Z(\omega_j)}{f_Z(\omega_j)} \tag{4}$$

d) Determine the periodogram from equation (4)

Geweke and Porter-Hudak (1983) obtained periodogram by

$$I_Z(\omega_j) = \frac{1}{2\pi}\left\{\gamma_0 + 2\sum_{t=1}^{T-1}\gamma_t cos(t\omega_j)\right\}, \quad (\omega_j \in [-\pi,\pi]) \tag{5}$$

This periodogram was also used by GPHTr and MGPH methods.

Reisen (1994) used the smoothed periodogram estimate of spectral density function by

$$I_Z(\omega_j) = \frac{1}{2\pi}\left\{\gamma_0\tau_0 + 2\sum_{t=1}^{g^*(T)}\gamma_t\tau_t\, cos(t\omega_j)\right\}, \quad (\omega_j \in [-\pi,\pi]) \tag{6}$$

Where $\{\tau_t\}$ are a set of weights called the lag window. Various windows can be chosen to smooth the periodogram. The Parzen window is given by

$$\tau_t = \begin{cases} 1-6\left(t/g^*(T)\right)^2 + 6\left(t/g^*(T)\right)^3 , & 0 \le t \le \dfrac{g^*(T)}{2} \\ 2\left(1-\dfrac{t}{g^*(T)}\right)^3 , & \dfrac{g^*(T)}{2} \le t \le g^*(T) \end{cases}$$

and has the desirable property that it always produces positive estimates of the spectral density function.

Hurvich and Ray (1995) and Velasco(1999a) used modified periodogram function by

$$I_Z\left(\omega_j\right) = \frac{1}{2\pi \sum_{t=0}^{T-1}\left(tap\left(t\right)\right)^2}\left|\sum_{t=0}^{T-1} tap\left(t\right) Z_t \exp\left(-i t \omega_j\right)\right|^2 \qquad (7)$$

Where the tapered data is obtained from the cosine-bell function

$$tap(t) = \frac{1}{2}\left[1-\cos\left(\frac{2\pi\left(t+0,5\right)}{T}\right)\right]$$

e) Estimate differencing parameter (d)

From equation (4), for ω_j near zero, i.e., for $j = 1,\ldots,m \ll (T/2)$ such that $m/T \to 0$ as $n \to \infty$, we have $\ln\left(f_W\left(\omega_j\right)/f_W(0)\right) \approx 0$. Thus,

$$Y_j = \beta_0 + \beta_1 X_j + a_j , \qquad j = 1,2,\ldots,m$$

$$\hat{\beta}_1 = \hat{d} = \frac{\sum_{j=1}^{m}\left(x_j-\bar{x}\right)\left(y_j-\bar{y}\right)}{\sum_{j=1}^{m}\left(x_j-\bar{x}\right)^2} , \quad m = g(T^\alpha) = T^\alpha , \; 0<\alpha<1 \qquad (8)$$

Where,

$\hat{\beta}_0 = \ln f_W(0)$, $Y_j = \ln I_Z\left(\omega_j\right)$, $X_j = \ln\left|1-\exp(-i\omega_j)\right|^{-2}$ and $\bar{X} = \left(\frac{1}{m}\right)\sum_{j=1}^{m} X_j$

For computation, with Euler equation we have $X_j = \ln\left(\dfrac{1}{4\sin^2\left(\omega_j/2\right)}\right)$

4. Outlier in Time Series

Time series observations are sometimes inflenced by interruptive events, such as strikes, outbreaks of war, sudden political or economic crises, unexpected heat or cold waves, or event unnoticed error of typing and recording. The sequences of these interruptive events create spurious observations that are inconsistent with the rest of series. Such observations are usually referred to as outliers.

In this paper we focus on AO (Additive Outlier). Additive outlier is an event that effects a series for one time period only. An additive outlier model is defined as

$$Z_T = \begin{cases} X_t , & t \neq T \\ X_t + \omega , & t = T \end{cases}$$
$$= X_j + \omega I_j^{(T)},$$

$$= \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)} \qquad (9)$$

Where

$$I_t^{(T)} = \begin{cases} 1, & t = T, \\ 0, & t \neq T. \end{cases}$$

is indicator variable representing the presence or absence of an outlier at time T.

The likelihood ratio test for AO is, $\lambda_{1,T} = \hat{\omega}_{AT} / \sigma_a$, if $\hat{\lambda}_T = \left| \hat{\lambda}_{1,T} \right| > C$, where C is a predetermined positive constant usually taken to be some value between 3 and 4, then there is and AO at time T with its effect estimated by $\hat{\omega}_{AT}$. Where $\hat{\omega}_{AT}$ is the least square estimator of ω for the AO model.

5 Simulation Study

We have conducted simulation studies to obtain some information about the performance of the accuracy of spectral regression methods in estimating the degree of differencing parameter from ARFIMA model. In this simulation study we use five methods of spectral regression methods namely, $\hat{d}_{GPH}$, $\hat{d}_{SPR}$, $\hat{d}_{GPHTr}$, $\hat{d}_{GPHTa}$ and $\hat{d}_{MGPH}$. For this we consider T = 300 and 1000 time series with 1000 replications. In this study, time series data are generated according to the particular specification.

a) Data of ARFIMA (1,*d*,0) and ARFIMA(0,*d*,1) models
b) Data of ARFIMA (1,*d*,0) and ARFIMA(0,*d*,1) models with five outliers. The location of the outlier is set in the middle of the observational period, specifically T = {148,149,150,151,152} when T = 300 and T ={498,499,500, 501,502} for T = 1000.

Both data of ARFIMA models have specification as follow $a_t \sim N(0,1)$, $\theta, \phi = 0.5$, $d = 0.2$ and 0.4.

For each series we estimate the value of *d* through the different methods and later we take the arithmetic average of these values, that is

$$\bar{d}_i = \frac{1}{1000} \sum_{j=1}^{1000} \hat{d}_i(j)$$

Where $\hat{d}$ corresponds to $\hat{d}_{GPH}$, $\hat{d}_{SPR}$, $\hat{d}_{GPHTr}$, $\hat{d}_{GPHTa}$ and $\hat{d}_{MGPH}$, respectively, depending on the estimation method used. To compare the different estimator we considered the mean squared error value, denoted hereafter by MSE, i.e.,

$$MSE = \frac{1}{1000} \sum_{j=1}^{1000} \left(\hat{d}_i(j) - d \right)^2$$

Where *d* is the true parameter value.

6. Simulation Result

Table.1 Mean And MSE of estimate from *d* ARFIMA Model

MODEL	Statistic	*d*	GPH		SPR		GPHTr		GPHTa		MGPH	
			300	1000	300	1000	300	1000	300	1000	300	1000
[illegible] *d*,0)	Mean	0.2	0.23	0.22	0.19	0.19	0.40	0.35	0.27	0.22	0.39	0.31
		0.4	0.46	0.42	0.40	0.39	0.66	0.50	0.38	0.33	0.60	0.50
	MSE	0.2	0.19	0.13	0.15	0.11	0.13	0.07	0.20	0.17	0.09	0.06
		0.4	0.21	0.14	0.17	0.11	0.13	0.08	0.30	0.21	0.09	0.06
[illegible] *d*,1)	Mean	0.2	0.16	0.19	0.11	0.16	0.04	0.10	0.10	0.07	0.01	0.10
		0.4	0.37	0.40	0.31	0.66	0.27	0.33	0.13	0.10	0.21	0.30
	MSE	0.2	0.20	0.13	0.15	0.11	0.13	0.07	0.07	0.05	0.10	0.06
		0.4	0.20	0.13	0.16	0.11	0.13	0.08	0.09	0.07	0.10	0.06

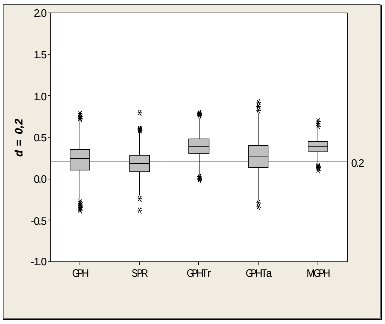
Table 2. Mean And MSE of estimate *d* from ARFIMA Model With Outlier

MODEL	Statistic	*d*	GPH		SPR		GPHTr		GPHTa		MGPH	
			300	1000	300	1000	300	1000	300	1000	300	1000
[illegible] *d*,0)	Mean	0.2	0.13	0.18	0.09	0.15	0.42	0.33	0.22	0.23	0.40	0.28
		0.4	0.33	0.40	0.28	0.36	0.50	0.55	0.34	0.33	0.54	0.46
	MSE	0.2	0.19	0.14	0.14	0.11	0.12	0.07	0.17	0.20	0.09	0.06
		0.4	0.21	0.14	0.17	0.11	0.13	0.08	0.26	0.26	0.09	0.06
[illegible] *d*,1)	Mean	0.2	0.05	0.02	0.02	0.02	0.36	0.16	0.07	0.07	0.40	0.16
		0.4	0.05	0.19	0.03	0.19	0.36	0.22	0.08	0.10	0.41	0.23
	MSE	0.2	0.08	0.11	0.06	0.08	0.05	0.06	0.05	0.05	0.05	0.04
		0.4	0.14	0.13	0.11	0.10	0.09	0.07	0.07	0.07	0.07	0.05

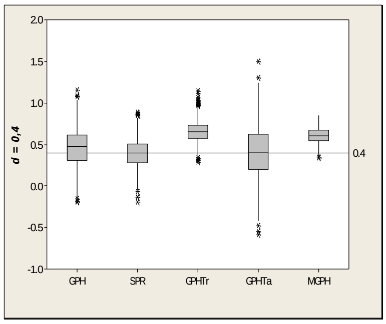
From table 1 we can report that accuracy of estimation with T = 1000 is better than T = 300 both ARFIMA(1,*d*,0) and ARFIMA(0,*d*,1) data. The best methods may be GPH and SPR methods, because estimate *d* from these methods are the nearest from true value *d*.

The same as table 1, from table 2 we can report that accuracy of estimation with T = 1000 is better than T = 300 both ARFIMA(1,*d*,0) and ARFIMA(0,*d*,1) data. The best methods may be GPH and GPHTa methods, because estimate *d* from these methods are the nearest from true value *d*. If we compare the result of estimation from ARFIMA(1,*d*,0) and ARFIMA(0,*d*,1) models then estimation from ARFIMA(1,*d*,0) data is better than ARFIMA(0,*d*,1) data.

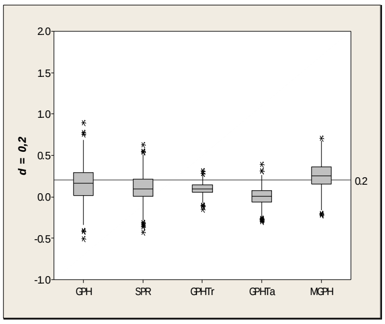
From figure 2 , GPH and SPR methods show a good performance in estimating the differencing parameter especially from ARFIMA(1,*d*,0) model. The centre of these Boxplots next to reference line as true value. GPHTa has a good accuracy but still wide of range.



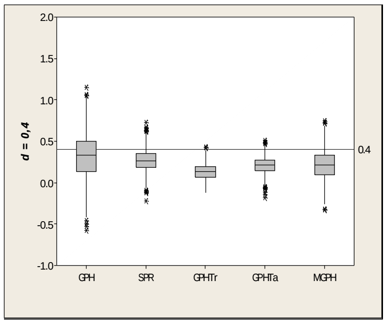
(i) ARFIMA(1,*d*,0), *d* = 0.2



(ii) ARFIMA(1,*d*,0), *d* = 0.4

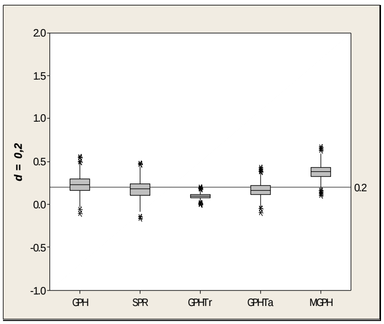


(iii) ARFIMA(0,*d*,1), *d*=0,2

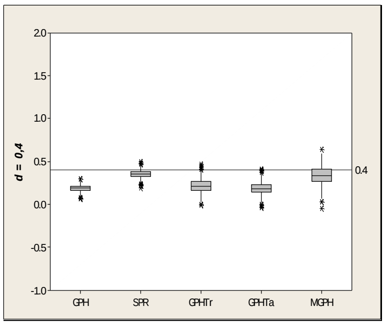


(iv) ARFIMA(0,*d*,1), *d* = 0.4

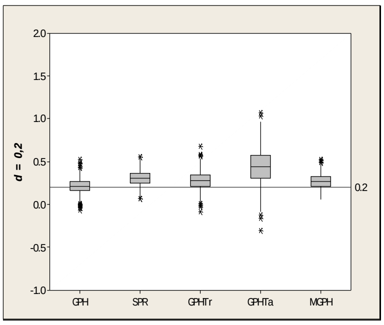
Figure 2. Box plot of estimate *d* from ARFIMA Model Data With T = 300.



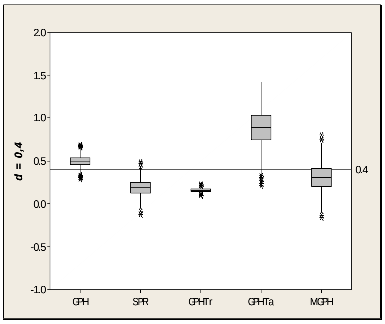
(i) ARFIMA(1,*d*,0), *d* = 0.2



(ii) ARFIMA(1,*d*,0), *d* = 0.4

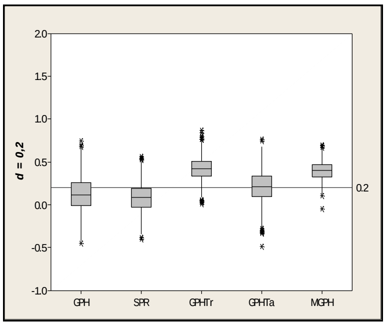


(iii) ARFIMA(0,*d*,1), *d*=0,2

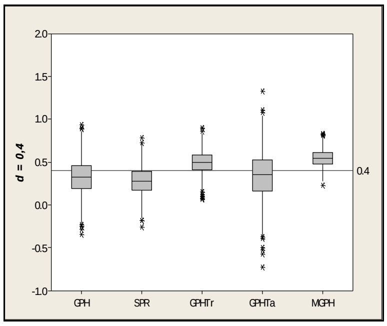


(iv) ARFIMA(0,*d*,1), *d* = 0.4

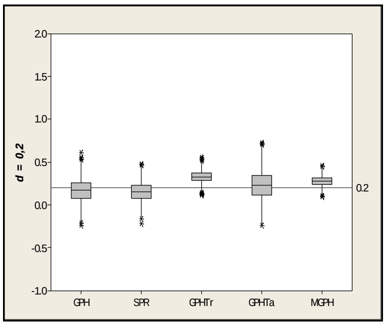
Figure 3. Box plot of estimate *d* from ARFIMA Model Data With T = 1000.



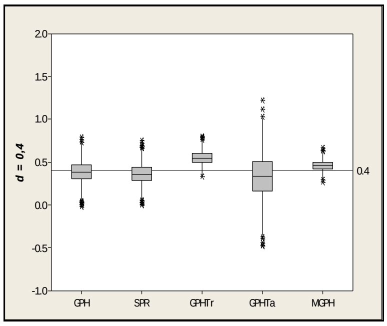
(i) ARFIMA(1,*d*,0), *d* = 0.2



(ii) ARFIMA(1,*d*,0), *d* = 0.4

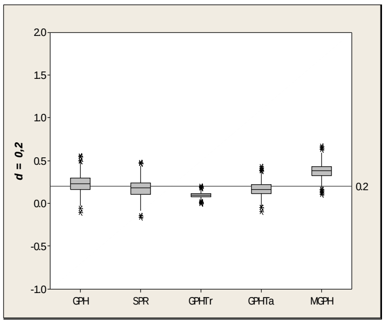


(iii) ARFIMA(0,*d*,1), *d* = 0.2

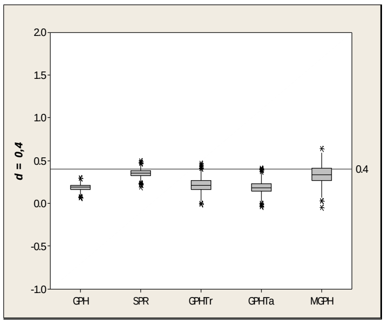


(iv) ARFIMA(0,*d*,1), *d* = 0.4

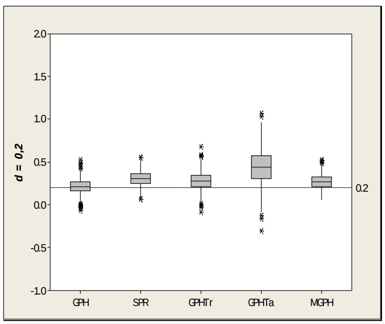
Figure 4. Box plot of estimate *d* from ARFIMA Model Data With Outlier (T = 305)



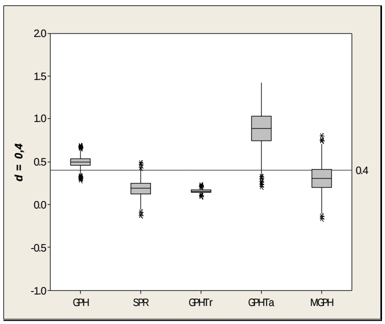
(i) ARFIMA(1,*d*,0), *d* = 0.2



(ii) ARFIMA(1,*d*,0), *d* = 0.4



(iii) ARFIMA(0,*d*,1), *d* = 0.2



(iv) ARFIMA(0,*d*,1), *d* = 0.4

Figure 5. Box plot of estimate *d* from ARFIMA Model Data With Outlier (T = 1005)

From Figure 3, all of methods have under estimate in estimatıng the differencing parameter for ARFIMA(1,*d*,0) models with *d* = 0,4. For ARFIMA(1,*d*,0) model with d = 0,2 GPH and SPR show a good accuracy in estimating the differencing parameter, their centre of boxplot near reference line d = 0,2. For ARFIMA(0,d,1) model with d=0,2, only GPH method has a good accuracy with the centre of boxplot exactly at reference line and another methods have higher estimate. For ARFIMA(0,d,1) model with d = 0,4, all methods fail to identify value of the differencing parameter, no boxplot has centre at reference line.

Figure 4, we can report that GPHTa show a good performance in estimating the differencing parameter for all models, especially for ARFIMA(1,d,0) model with d = 0,2. The second method is GPH, this method has enough consistent in estimating the differencing parameter. GPHTr and MGPH methods have a bad estimate for all models.

Figure 5, GPH method has a good performance in estimating the differencing parameter for ARFIMA model with d=0,2, the centre of boxplot has near the reference line. All methods have under estimste in estimating the differencing parameter for ARFIMA(1,d,0) model with d=0,4. For ARFIMA(0,d,1) model with d = 0,4 and have five outliers all of methods fail to identify value of the dfferencing parameter for ARFIMA(0,d,1) model with d = 0,4 and five outliers.

7. Conclusion

From simulation results, GPH method shows a good performance in estimating the differencing parameter of ARFIMA model both clean data and data with outlier. Above all, estimation of spectral regression methods are better from ARFIMA(1,d,0) data than from ARFIMA(0,d,1) data.

REFERENCES

Barnett, V dan Lewis, T. (1994), *Ouliers in Statistical Data*, J. Wiley, New York.

Beran, J. (1994), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models", *Journal of the Royal Statistical Society*, Vol. 57, hal. 659-672.

Geweke J dan Porter-Hudak,S. (1983), "The Estimation and Application of Long Memory Time Series Models", *Journal of Time series Analysis,*Vol. 4, hal. 221-238.

Granger, C. W. J. dan Joyeux,R (1980), "An Introduction to Long-Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis*, Vol. 1, hal. 15-29.

Hosking, J.R.M. (1981), "Fractional Differencing", *Biometika*, Vol. 68, hal. 165-176.

Hurst, H.E. (1951), "Long-Term Storage of Reservoirs: An Experimental Study", *Transactions of the American Society of Civil Engineers*, Vol. 116, hal. 770-799.

Hurvich, C.M. dan Ray, B.K. (1995), "Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes", *Journal of Time series Analysis*, Vol. 16, hal.17-42.

Lopes, S.R.C dan Nunes,M.A. (2006), "Long Memory Analysis in DNA Sequences", *Physica A*, Vol. 361, hal. 569-588.

Lopes, S.R.C.,Olberman,B.P dan Reisen,V.A. (2004), "A Comparison of Estimation Methods in Non-Stationary ARFIMA Processes", *Journal of Statistical Computation & Simulation*, Vol. 74, No. 5, hal. 339-347.

Reisen, V.A. (1994), " Estimation of the Fractional Parameter for ARIMA(*p*,*d*,*q*) Model Using the Smoothed Periodogram", *Journal of Time Series Analysis*, Vol.15, hal. 335-350.

Robinson, P.M. (1995), "Log-Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics*, Vol. 23, hal. 1048-1072.

Sowell, F. (1992), "Maximum Likelihood Estimation of Stationary Univariate Fractonally Integrated Time Series Models", *Journal of econometrics*, Vol.53, hal.165 – 188.

Velasco, C. (1999a), "Non-Stationary Log-Periodogram Regression", *Journal of Econometric*, Vol. 91, hal. 325-371.

Velasco, C. (1999b), "Gaussian Semiparametric Estimation of Non-Stationary Time Series", *Journal of Time Series Analysis*, Vol. 20, No.1, hal. 87-127.

Wei, S.W.W. (1994), *Time series Analysis Univariate and Multivariate Methods,* Addison-Wesley Publishing Company, Canada.