Working Paper 2012:15

Department of Economics
School of Economics and Management

Estimating Long Memory Causality Relationships by a Wavelet Method

Yushu Li

May 2012

LUND UNIVERSITY

Estimate long memory causality relationship by wavelet method

Yushu Li[1,2]

Department of Economics, Lund University

Abstract

The traditional causality relationship proposed by Granger (1969) assumes the relationships between variables are short range dependent with the same integrated order. Chen (2006) proposed a bi-variate model which can catch the long-range dependent among the two variables and the series do not need to be fractionally co-integrated. A long memory fractional transfer function is introduced to catch the long-range dependent in this model and a pseudo spectrum based method is proposed to estimate the long memory parameter in the bi-variate causality model. In recent years, a wavelet domain-based method has gained popularity in estimations of long memory parameter in unit series. No extension to bi-series or multi-series has been made and this paper aims to fill this gap. We will construct an estimator for the long memory parameter in the bi-variable causality model in the wavelet domain. The theoretical background is derived and Monte Carlo simulation is used to investigate the performance of the estimator.

JEL classification: C30, C51, C63

Keywords: Granger causality; Long memory; Monte Carlo simulation; Wavelet domain

[1] The author gratefully acknowledges funding from the Swedish Research Council (421-2009-2663)

[2] The author gratefully acknowledges comments from professor David Edgerton, Fredrik NG Andersson and Klas Fregert

1. Introduction

In time series econometric study, the causality relationship between subsets is an important subject and the Granger test proposed by Granger (1969) has been most frequently used in previous studies. The main idea of the Granger causality test is to measure whether the past information of a set of variables contains information on changes in another set of variables and helps to predict them. When excluding the self-lag influence, the framework in which the causality test is carried out can be represented in the following model: $Y_{\mathrm{t}} = \Psi(B)X_{\mathrm{t}} + \varepsilon_t$ where B is the lag parameter and $\Psi(B) = \sum_{k=0}^{n} \upsilon_k B^k$. The Granger causality test will check the significance of the linear coefficients of the lagged variables. In this framework, the lag specification plays an important role as it is the trade-off between bias and efficiency. The general principle is that the smaller lag length has smaller variance but confronts a risk of bias while larger lags will reduce the bias problem but may bring up inefficiency. A rule of thumb is that the lags on the independent variable should be relatively short to conserve degrees of freedom. Thornton and Batten (1985) pointed out that this rule of thumb 'fails to ensure that the results are not based on the capricious choice of the lag structure' and a safer approach is to perform an extensive search of the lag space. However, in practice there may exist a certain relationship where the impact effect decays quite slowly, thus the lag coefficients and cross-correlations will remain significant after the long term. In this situation, the lag searching procedure will be quite complicated and a model with too long lags will lose parsimony.

The discussion of the long memory dependence in the causality model continued with Hidalgo (2000), who mentioned that the long-range dependent process has already attracted immense attention in recent econometric study; Chen (2006) introduced a fractional transfer function which can model the long memory causality dependence between two series; Smirnov and Mokhov (2009) discussed the long-term causality effect from variation in atmospheric content, solar activity and volcanic activity which can affect global surface temperature up to 150 years. Among the studies, the model proposed by Chen (2006) is quite straight forward and parsimonious:

$$Y_{\mathrm{t}} = (1-B)^{-d}\Psi(B)X_{\mathrm{t}} + \varepsilon_t \tag{1}$$

where $0 < d < 0.5$ indicates a stationary long memory dependence and $\Psi(B)$ captures the short memory causality relationship. The out point of this model is that the slow decay pattern in the cross-correlation between two series of relationships is similar to the low speed auto-

correlation fading pattern in the long memory unit series. From equation (1), the specification of the long memory causality relationship will turn to be a problem in estimating the fractional difference parameter d and check how profoundly X_t will influence Y_t. To estimate d, Chen (2006) applied a cross-pseudo spectrum-based method and it is the bi-variate correspondence to the Geweke-Porter-Hudak (GPH) regression estimator (Geweke and Porter-Hudak, 1983) in unit long memory series.

Recently, to estimate the fractional difference parameter in unit series, a wavelet domain-based method began to gain popularity and become a powerful tool due to its self-similarity (Masry, 1993; Wornell, 1993). Based on the property of de-correlating the long memory series in the frequency domain, McCoy and Walden (1996) proposed a wavelet Maximized Likelihood Estimation (MLE) where a reduced computational order of likelihood is approximated in wavelet domain. This method is useful to the large data set and quite robust to model specification under the assumption that the auto-covariance of the wavelet coefficients decays rapidly enough. Later, Jensen (1999) pointed out that the precision of the wavelet MLE depends on how rapidly the auto-covariance of the wavelet coefficients decays. Instead of the maximum likelihood based estimator, Jensen (1999) proposed a wavelet Ordinary Least Square (OLS) estimation of the long memory parameter and showed it over performance of the wavelet MLE in simplicity of implementing and non-numerical nature, with a slightly larger Mean Square Error (MSE) than the wavelet MLE method. The wavelet OLS estimation was further widely applied and developed by Whitcher and Jensen (2000), Hsu (2006), Gonzaga and Kawanaka (2007). However, little work has been done to expand the wavelet domain-based estimation for bi- or multi-series. This paper will then aim to combine the wavelet methodology to estimate the long memory dependence in the bi-variate causality model proposed by Chen (2006). We give out the theoretical background of the estimator and further use the Monte Carlo method to evaluate the performance of the estimation.

The paper will be divided as follows: the second section briefly introduces the long memory dependence model and the wavelet estimation for the unit long memory parameter; the third section is the theoretical illustration on how to combine the wavelet method to estimate the long memory parameter in the bi-variate causality model; the fourth section uses the Monte Carlo simulation to show the estimation procedure in the wavelet domain; and the conclusion can be found in the last section.

2. Model structure and wavelet method

2.1. Long memory causality model

The model which can capture the long memory Granger causality effect was recently proposed and investigated by Chen (2006). The main idea is that, for a bi-variate long range dependence model $Y_t = \upsilon(B)X_t + \varepsilon_t$ where $\upsilon(B) = \sum_{k=-\infty}^{+\infty} \upsilon_k B^k$, the regression coefficient for the non-negative power polynomial υ_k will not eliminate to zero even after large lag numbers. A fractional transfer function $(1-B)^{-d}$ is applied as a more parsimonious manner to catch this long-range dependence and the model will turn out to be $Y_t = (1-B)^{-d} X_t + \varepsilon_t$. In this way, the requirement to estimate the large number of explanatory lagged coefficient will then be simplified to estimate the fractional different parameter d. A straightforward example of this model can be illustrated by setting $d = 0.4$, with the independent variable X_t is generated separately from Auto-Regressive (AR) process $(1-0.3B)X_t = \varepsilon_t$ and Moving Average (MA) process $X_t = (1-0.3B)\varepsilon_t$. The cross-correlation between the two series will turn out to be:

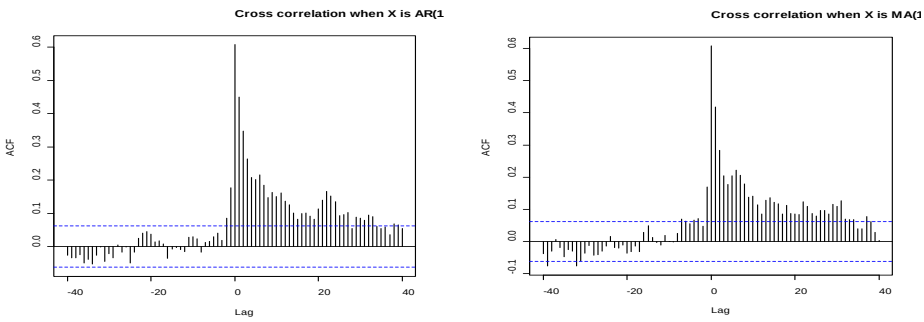

Figure 1. Cross-correlation of X and Y to lag 40

We see that there still exists a significant cross-correlation for data which are far apart even after 40 lags; this persistence behaviour mimics the slow decay auto-correlation in fractional difference data. This long-range dependence is predictable as the fractional difference operator $(1-B)^{-d}$ can be written as $\sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} B^k$ by the binomial expansion. The uni-direct effect of X_t on Y_t will last long and the traditional linear Granger causality test will give the following results:

Table 1. *P*-Values of Granger causality for an independent AR variable

Lag	1	2	3	4	5	6	7	8	9	10	…	31	32	33	34	35	36	37	38	39	40
x -> y	0	0	0	0	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0	0	
y -> x	0.74	0.32	0.45	0.50	0.60	0.70	0.80	0.67	0.37	0.32	…	0.78	0.64	0.67	0.65	0.64	0.71	0.72	0.77	0.76	0.73

Table 2. *P*-Values of Granger causality for an independent MA variable

Lag	1	2	3	4	5	6	7	8	9	10	…	31	32	33	34	35	36	37	38	39	40
x -> y	0	0	0	0	0	0	0	0	0	0	…	0	0	0	0	0	0	0	0	0	
y -> x	0.91	0.17	0.11	0.32	0.57	0.74	0.78	0.73	0.81	0.84	…	0.35	0.48	0.39	0.43	0.48	0.51	0.51	0.54	0.55	0.54

Table 1 and Table 2 show that if we just fit a linear model to the long memory dependence and use a traditional Granger causality test, there exists a uni-direct causality running from X to Y even to lag 40. Thus compared to the linear bi-variate model, the long memory dependence model shows an advantage of much more parsimony as the main parameter that needs to be specified is just the fractional difference parameter d. In Chen (2006), this parameter is estimated by a way which has its univariate correspondence in GPH Geweke-Porter-Hudak (GPH) regression estimator (Geweke and Porter-Hudak, 1983). Geweke and Porter-Hudak (1983) show that, in the unit case $(1-B)^{\delta} Z_t = \varepsilon_t$, the fractional difference parameter δ is estimated by the regression model from the linear relationship between the periodogram and low frequency ordinate as: $\ln(I(\lambda_k^{\mathrm{Z}})) \approx \beta_0 + \beta_1 \ln(4\sin^2(\lambda_k^{\mathrm{Z}}/2))$ where $I(\lambda_k^{\mathrm{Z}})$ is the pseudo spectrum at frequency λ_k and $p\lim \hat{\beta}_1 = -\delta$. In Chen (2006), the pseudo cross-spectrum instead of the pseudo auto-spectrum is then applied to estimate the long memory parameter in the transfer function. The main procedure is as follows: after using the same filter to pre-whiten X_{t} and Y_{t} into white noise α_{t} and β_{t}, the equation at the low band frequencies $\ln(\frac{\left| f_{\beta\alpha}^{A(T)}(\lambda_k) \right|}{f_{\alpha}^{(T)}(\lambda_k)}) \approx \alpha - d\ln(\lambda_k)$ is applied to estimate d, with $f_{\alpha}^{(T)}(\lambda_k)$ being the pseudo spectrum of white noise and the pseudo cross-spectrum $f_{\beta\alpha}^{A(T)}(\lambda_k)$. The cross-spectrum $f_{\beta\alpha}^{A(T)}(\lambda_k)$ is estimated from the cross-periodogram which is the Fourier transform of the cross-covariance $\hat{c}_{yx}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \overline{y})(x_{t-k} - \overline{x})$. However, it is well known that, in the unit series, using a periodogram as a estimation of the spectrum suffers from the problem of inconsistency and leads to a unsatisfactory asymptotic properties of the GPH estimator (see Jensen, 1999); the same problem will happen when using the cross-periodogram to estimate the cross-spectrum. Thus in this paper, instead of using the periogogram-based method, we will try to develop a wavelet domain-based estimation for the long memory parameter d in the bi-variate model.

2.2. Wavelet transform for the unit long memory process

Wavelet methods have been widely applied in the field of signal and image processing after their theoretical development in the 1980s (Grossmann and Morlet, 1984; Mallat, 1989). The wavelet transformation can capture the characteristics of data series both in the frequency domain and the time domain using a two-dimensional resolution. Corresponding to sinusoidal waves in the Fourier transform, the wavelet bases $\{\psi_{k,a} : k, a \in R\}$ used in the wavelet transform are generated by translations and dilations of a basic mother wavelet $\psi \in L^2(R)$ and can be expressed as $\psi_{k,a}(z) = \frac{1}{\sqrt{a}}\psi(\frac{z-k}{a})$. For the signal $f(z)$, the wavelet transform is $\gamma(k,a) = \langle f, \psi_{k,a} \rangle = \int f(z)\psi^*_{k,a}(z)dz$. When the mother wavelet satisfies the condition $\int_0^\infty \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty$, with $|H(\omega)|$ as the Fourier transform of the $\psi(z)$, we can reconstruct $f(z)$ using the inverse $f(z) = \iint \gamma(k,a)\psi_{k,j}(z)dkda$. For the discrete series $\mathrm{Z} = \{Z_t, t = 1,...,N\}$, the level J maximal overlap discrete wavelet transform (MODWT) contains $J+1$ vectors $\tilde{\mathbf{W}}_1^{(\mathrm{Z})},...,\tilde{\mathbf{W}}_J^{(\mathrm{Z})},\tilde{\mathbf{V}}_J^{(\mathrm{Z})}$ with wavelet coefficients $\tilde{\mathbf{W}}_j^{(\mathrm{Z})} = \{\tilde{W}_{j,t}^{(\mathrm{Z})}, t = 1,...,N\}$ corresponding to changes of scale $\tau_j = 2^{j-1}$, while the wavelet scaling coefficients $\tilde{\mathbf{V}}_J^{(\mathrm{Z})} = \{\tilde{V}_{J,t}^{(\mathrm{Z})}, t = 1,...,N\}$ corresponds to averages on a scale of $\lambda_J = 2^j$. The N dimensional vectors $\tilde{\mathbf{W}}_j$ and $\tilde{\mathbf{V}}_J$ are computed by $\tilde{\mathbf{W}}_j = \tilde{\boldsymbol{W}}_j\mathrm{Z}$, $\tilde{\mathbf{V}}_J = \tilde{\boldsymbol{V}}_J\mathrm{Z}$ where $\tilde{\boldsymbol{W}}_j$ and $\tilde{\boldsymbol{V}}_J$ are $N \times N$ matrices defined separately as the wavelet high band pass filter at level j and low band pass filter at level J. Thus after the wavelet transform, the frequency resolution of Z can be decomposed as $\tilde{\mathbf{W}}_j$ contains the information in frequency band $(\frac{1}{2^{j+1}}, \frac{1}{2^j})$ and $\tilde{\mathbf{V}}_J$ contains information in the frequency band $(0, \frac{1}{2^J})$.

3. Wavelet-weighted least square estimation for the causality relationship

3.1. Wavelet estimation for the unit long memory process

The wavelet transform analysis and inverse synthesis are particularly suitable to measure the variability in fractional difference series with long memory dependence. For the long memory series $\mathrm{Z} = \{Z_t\}_{t\in\mathbb{Z}}$ with spectral density $f_{\mathrm{Z}}(\lambda)$, we have $f_{\mathrm{Z}}(\lambda) = \frac{\sigma_\varepsilon^2}{[4\sin^2(\pi\lambda)]^\delta} \approx \frac{\sigma_\varepsilon^2}{[2\pi\lambda]^{2\delta}}$ as

$\lambda \to 0$, where δ is the long memory parameter and the process is stationary for $0 < \delta < 0.5$. Furthermore, defining $\left|\tilde{H}_j(\lambda)\right|^2$ being the squared gain function for the wavelet filters at level j, Percival (1995) shows that the variance for the wavelet coefficients is:

$$\gamma_j^{\mathrm{Z}} = \int_{-1/2}^{1/2} \left|\tilde{H}_j(\lambda)\right|^2 f_{\mathrm{Z}}(\lambda) d\lambda \approx 2\int_{1/2^{j+1}}^{1/2^j} \frac{\sigma_\varepsilon^2}{(2\pi\lambda)^{2\delta}} d\lambda \approx \frac{2\sigma_\varepsilon^2}{(2\pi)^{2\delta}} \int_{1/2^{j+1}}^{1/2^j} \frac{1}{(\lambda)^{2\delta}} d\lambda = C\tau_j^{2\delta-1} \tag{2}$$

Thus a linear relationship can be built after the log transformation as:

$$\ln(\gamma_j^{\mathrm{Z}}) \approx \ln(C) + (2\delta - 1)\ln(\tau_j) \tag{3}$$

One important property of the wavelet coefficient $\tilde{\mathbf{W}}_{j,k}^{(\mathrm{Z})}$ is that they are zero mean and approximately de-correlated (see Percival, 1995). As it is well known, for the random variable Z, the variance $\hat{\sigma}_X^2 = \frac{1}{N}\sum_k (Z_t - \overline{\mathrm{Z}})^2$ will be seriously biased if $E(\mathrm{Z})$ is unknown. The zero mean character of $\tilde{\mathbf{W}}_{j,k}^{(\mathrm{Z})}$ will help us to construct an unbiased estimator of variance for the scale-based wavelet coefficient γ_j^{Z} as $\hat{\gamma}_j^{\mathrm{Z}} = \frac{1}{N - L_j + 1}\sum_{k=L_j-1}^{N-1} \left|\tilde{\mathbf{W}}_{j,k}^{(\mathrm{Z})}\right|^2$ with L_j being the boundary coefficients at j^{th} scale. Thus equation (3) can be further written as:

$$\ln(\hat{\gamma}_j^{\mathrm{Z}}) = C^* + (2\delta - 1)\ln(\tau_j) + \varepsilon_j \tag{4}$$

with ε_j being zero mean random errors, thus an unbiased and consistent estimator of δ is then available by OLS or Weighted Least Square (WLS) from equation (4). On the other hand, δ can also be estimated by MLE in wavelet domain after approximating the fractional difference time series Z by $\mathrm{Z}=\boldsymbol{W}^T\Lambda^{1/2}\mathrm{Z}^*$, where $\boldsymbol{W}$ is the Discrete Wavelet Transform (DWT) matrix, $\Lambda^{1/2}$ is a $K \times K$ diagonal matrix with diagonal elements being wavelet variance at different scale as $\left\{\underbrace{\gamma_1^{\mathrm{Z}},...,\gamma_1^{\mathrm{Z}}}_{K/2}, \underbrace{\gamma_2^{\mathrm{Z}},...,\gamma_2^{\mathrm{Z}}}_{K/4},...,\underbrace{\gamma_{J-1}^{\mathrm{Z}},\gamma_{J-1}^{\mathrm{Z}}}_{2}, \gamma_J^{\mathrm{Z}}, \gamma_{J+1}^{\mathrm{Z}}\right\}$ and Z^* is vector of deviations drawn from a standardized Gaussian distribution. One advantage of the wavelet-based MLE is that the computational order of the likelihood calculating is reduced and is suitable for a large data set (see McCoy and Walden, 1996). However, Jensen (1999) pointed out that the simplicity of the wavelet MLE is based on the assumption that the covariance of the wavelet coefficients decays rapidly enough. Jensen (1999) further proposed a wavelet OLS estimator which has a significantly lower MLE than the GPH estimator. The wavelet OLS estimator also over performs the wavelet MLE estimator in McCoy and Walden (1996)

with simplicity of implementing and non-numerical nature, although the wavelet MLE produces a slightly less MSE than the wavelet OLS estimation. Thus the wavelet OLS-based estimation is more frequently applied to estimate the long memory parameters further in Whitcher and Jensen (2000); Gonzaga and Kawanaka (2007); and Hsu (2006). As this paper deals with bi-variate data which will bring in the cross-correlation, the MLE method will be even more complicated in numeric computation; thus we concentrate on a least square-based method when we extend the wavelet method in unit series to bi-variate series when estimating the long memory parameter in the causality transfer function in the next section.

3.2. Wavelet weighted least square estimation for the causality relationship

The wavelet domain-based OLS estimator is shown to be asymptotically unbiased and consistent in the unit series. This paper aims to extend the wavelet-based methodology to the bi-variate model in Chen (2006) where there exists a long memory effect under Granger's causality sense. In this model, the cross-correlation of the two series X_t and Y_t is not possible to be summed and decays to zero slowly at a hyperbolic rate with $\gamma^{XY}(k) \sim \varsigma k^{2d-1}$ as $k \to \infty$, Chen (2006) proved this will lead to that $f_{XY}(\lambda) \sim \kappa\lambda^{-2d}$ as $\lambda \to 0$. In the wavelet domain, the covariance at scale j between the wavelet coefficient $\tilde{\mathbf{W}}_{j,k}^{(X)}$ and $\tilde{\mathbf{W}}_{j,k}^{(Y)}$ is: $\gamma_j^{XY} = \int_{-1/2}^{1/2} \tilde{H}_j(\lambda) f_{XY}(\lambda) d\lambda$, thus combined with $f_{XY}(\lambda) \sim \kappa\lambda^{-2d}$ we have the following equation:

$$\gamma_j^{XY} = \int_{-1/2}^{1/2} \left|\tilde{H}_j(\lambda)\right|^2 f_{XY}(\lambda) d\lambda \approx 2\int_{1/2^{j+1}}^{1/2^j} \kappa\lambda^{-2d}\, d\lambda \approx C'\tau_j^{2d-1} \tag{5}$$

The estimator of the covariance is $\hat{\gamma}_j^{XY} = \frac{1}{N - L_j + 1} \sum_{k=L_j-1}^{N-1} \tilde{\mathbf{W}}_{j,k}^{(X)} \tilde{\mathbf{W}}_{j,k}^{(Y)}$ where L_j is the boundary coefficient at j^{th} scale. Theorem 2 in Whitcher *et al.* (1999) shows that $\hat{\gamma}_j^{XY}$ is an asymptotic unbiased estimator for γ_j^{XY} with $\hat{\gamma}_j^{XY} \in N(\gamma_j^{XY}, N_j^{-1} S_j^{XY}(0))$. The *Delta method* will give $\ln \hat{\gamma}_j^{XY} \in N(\ln \gamma_j^{XY}, (\frac{\ln e}{\gamma_j^{XY}})^2 N_j^{-1} S_j^{XY}(0))$, then by setting $\varepsilon_j = \ln \hat{\gamma}_j^{XY} - \ln \gamma_j^{XY}$ we have $\varepsilon_j \in N(0, (\frac{\ln e}{\gamma_j^{XY}})^2 N_j^{-1} S_j^{XY}(0))$. Thus equation (5) will further lead to the following equation:

$$\ln \hat{\gamma}_j^{XY} = \ln C' + (2d-1)\ln \tau_j + \varepsilon_j \tag{6}$$

where ε_j will be approximately uncorrelated to the scale τ_j for large L_j (see Gonzaga and Kawanaka, 2007). By setting $\tilde{\mathrm{X}} = [1,\ \ln \tau_j]$, $\tilde{\mathrm{Y}} = \ln \hat{\gamma}_j^{\mathrm{XY}}$, $\beta = [\ln C',\ 2d-1]$, and Σ_e being the variance matrix with diagonal elements being $\sigma_j^2 = \mathrm{var}(\varepsilon_j)$, an asymptotically unbiased and consistent WLS estimator for β can be straightforwardly built up as $\beta^{WLS} = (\tilde{\mathrm{X}}^{\mathrm{T}}\Sigma_e^{-1}\tilde{\mathrm{X}})^{-1}\tilde{\mathrm{X}}^{\mathrm{T}}\Sigma_e^{-1}\tilde{\mathrm{Y}}$ with $\beta^{WLS} \in N(\beta,\ (\tilde{\mathrm{X}}^{\mathrm{T}}\Sigma_e^{-1}\tilde{\mathrm{X}})^{-1})$.

4. Monte Carlo experiment

This paper will apply the Monte Carlo experiment to show how the long memory parameter in the long-run relationship can be estimated. The model structure is $Y_{\mathrm{t}} = (1-B)^{-d} X_{\mathrm{t}} + N_t$, thus the data-generating process only needs to generate the independent variable X_{t}. Chen (2006) pre-whitens X_{t} and Y_{t} by filtering both series to a common pre-whitening function, and they estimate the parameter by using the white noise process. This will bring up two problems in application: the first is that we need to identify the structure of X_{t} to get the filter; the second is that, when we use the filter to Y_{t}, the long memory dependence between X_{t} and Y_{t} are destroyed which leads to serious under-bias problems in finite sample size. Thus in this paper, we will use the original data and investigate the model estimation across a range of parameter settings. Although Chen (2006) claims that the short memory part in X will converge to a constant in the frequency domain as the frequency approaches zero, the estimation will be tortured in finite samples when short memory auto-correlation appears in X. This can be shown in the following where we set $d = 0.3$, X is AR series $(1-\phi_1 B)X_{\mathrm{t}} = \varepsilon_t$ with ϕ_1 being separately 0.3, 0.5 and 0.7:

$\phi_1 = 0.3$ $\phi_1 = 0.5$ $\phi_1 = 0.7$

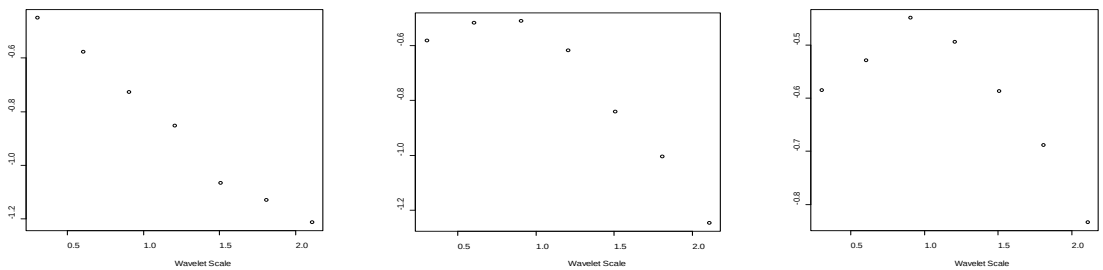

Figure 2. The linear relationship between $\ln \hat{\gamma}_j^{\mathrm{XY}}$ and $\ln \tau_j$ when ϕ_1 increase

Figure 2 shows that the short memory in X_t destroys the linear relationship between $\ln \hat{\gamma}_j^{XY}$ and $\ln \tau_j$ at the low-scale level, which leads to an over-bias problem if we include all the scales in estimation. Another important issue is that equation (6) is an approximation relationship, and in the finite sample size the log–log relationship is not strictly linear in small samples. Especially when d increases to 0.5, the slope value $2d-1$ increases to zero and the linear relationship becomes more and more flat; this will lead to the large-scale end point derive from the linear relationship. A straightforward illustration is shown in the following figure when $X_t = (1-0.3B)\varepsilon_t$ and d is separately 0.2, 0.3 and 0.4:

$d = 0.2$ $\qquad$ $d = 0.3$ $\qquad$ $d = 0.4$

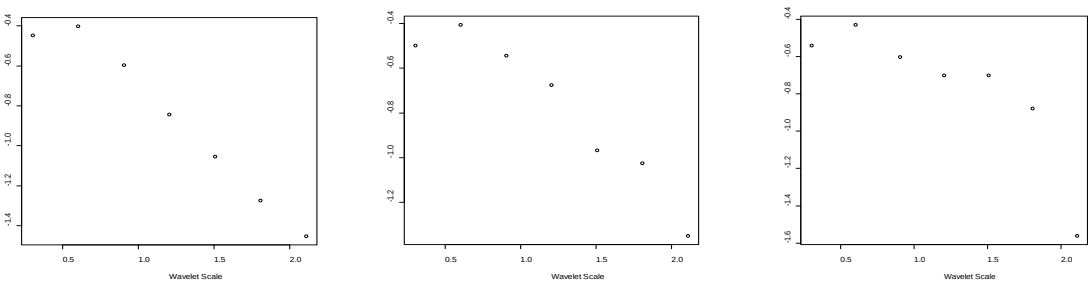

Figure 3. The linear relationship between $\ln \hat{\gamma}_j^{XY}$ and $\ln \tau_j$ when d increases

Figure 2 and Figure 3 show that, as the linear relation in equation (4) is an approximation under the assumption of infinite sample size, the estimation in the finite sample size will be disturbed by the deviation of end points in both sides: Figure 2 shows that the short memory part will influence the linear relationship in the lower scales and Figure 3 shows that, when d increases, the slope of the linear relationship will become flatter and the end points will deviate from the linear line. These two deviations bring in an important issue in the choice of the wavelet decomposition level J and the starting scale level K. The same issue of the scale level chosen in the unit series case is discussed in Hsu (2006), where it is suggested that J and K can be selected via certain model selection criterion such as Akaike information criterion (AIC). In the simulation, J and K can be decided based on the least mean square method. When setting the sample size as 1,000 the Monte Carlo experiment based on 1,000 replication is listed as follows:

Table 3. Mean and standard deviation of the estimation change AR parameter

d	$\phi_1 = 0.3$				$\phi_1 = 0.5$				$\phi_1 = 0.7$			
	Mean	*S.D.*	*MSE*	*(K J)*	*Mean*	*S.D.*	*MSE*	*(K J)*	*Mean*	*S.D.*	*MSE*	*(K J)*
0.2	0.192	0.037	0.0014	(2 7)	0.204	0.060	0.0036	(3 7)	0.230	0.098	0.0105	(4 7)
0.3	0.304	0.031	0.0009	(1 6)	0.317	0.078	0.0063	(3 6)	0.304	0.082	0.0067	(4 7)
0.4	0.350	0.027	0.0034	(1 5)	0.399	0.035	0.0012	(2 7)	0.430	0.055	0.0056	(3 7)

Table 4. Mean and standard deviation of the estimation change MA parameter

d	$\phi_1 = 0.3$				$\phi_1 = 0.5$				$\phi_1 = 0.7$			
	Mean	*S.D.*	*MSE*	*(K J)*	*Mean*	*S.D.*	*MSE*	*(K J)*	*Mean*	*S.D.*	*MSE*	*(K J)*
0.2	0.212	0.028	0.0009	(1 7)	0.180	0.039	0.0023	(2 6)	0.194	0.042	0.0018	(2 6)
0.3	0.284	0.030	0.0011	(1 5)	0.330	0.026	0.0009	(1 6)	0.333	0.028	0.0023	(1 7)
0.4	0.308	0.029	0.0096	(1 5)	0.396	0.026	0.0007	(1 5)	0.409	0.028	0.0009	(1 6)

Table 3 and 4 show that the estimation performs well with low bias, small variance and low MSE. However, there exists underestimate problems when $d = 0.4$ for θ_1=0.3, The reason lies behind that, when d increases to 0.5, the slope of the linear relationship $2d - 1$ will tend to be zero, which results in that any deviation from the linear relationship will tend to make the estimation for $2d - 1$ deviate from zero and then results in the underbias estimation. Table 3 also shows that, when ϕ_1 grows the lower-scale information should be abandoned to avoid a serious over-bias problem. Our suggestion in the empirical example is to draw a linear graph before estimation and gain a first impression of the linear slope. Then the initial choice of J and K can be selected based on the performance of the end points in both sides. The final selection of J and K can be decided after an iterative procedure based on certain model selection criteria.

5. Conclusion

This paper has talked about the long memory causality relationship between two series – the model framework proposed by Chen (2006). In Chen (2006), the long memory parameter is estimated by a spectrum-based estimator, while this paper proposes a wavelet WLS estimator. The main gain in this paper is to extend the widely applied wavelet estimation in unit series to a bi-variate series. The derivation of the estimator is presented and we further use Monte Carlo simulation to investigate separately how the AR and MA parameter will influence the estimation. We show that attention should be paid to the choice of decomposition level J and starting scale K in the finite sample application. In general, the wavelet based estimation is straightforward and quite easy to apply as well as a decent result.

6. References

Chen, W. D. (2006) Estimating the long memory granger causality effect with a spectrum estimator. *Journal of Forecasting*, 25 (3), 193–200.

Geweke, J. & Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4, 221–238.

Gonzaga, A. & Kawanaka, A. (2007) Wavelet-based estimation of generalized fractional process. *Methods of Information in Medicine*, 46 (2), 117–120.

Granger, C. W. J. (1969) Investigating causal relation by econometric and cross-sectional method. *Econometrica*, 37 (3), 424–438.

Grossman, A. & Morlet, J. (1984) Decomposition of hardy functions into square integrable wavelets of constant shape. *Society for Industrial and Applied Mathematics Journal on Mathematical Analysis*, 15, 732–736.

Hidalgo, J. (2000) Nonparametric test for causality with long-range dependence. *Econometrica*, 68 (6), 1465–1490.

Hsu, N. J. (2006) Long-memory wavelet models. *Statistica Sinica*, 16, 1255–1271.

Jensen, M. J. (1999) Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting*, 18 (1), 17–32.

Mallat, S. G. (1989) A theory for multiresolution signal decomposition: The wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Trans.*, 11 (7), 674–693.

Masry, E. (1993) The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Trans. Inform. Theory*, 39, 260–264.

McCoy, E. J. & Walden, A. T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics*, 5, 1–31.

Percival, D. B. (1995) On estimation of the wavelet variance. *Biometrika*, 82 (3), 619–631.

Smirnov, D. A. & Mokhov, I. I. (2009) From Granger causality to long-term causality: Application to climatic data. *Physical Review*, E 80: 1.

Thornton, D. L. & Batten, D. S. (1985) Lag-length selection and tests of Granger causality between money and income. *Journal of Money, Credit, and Banking*, 17, 164–178.

Whitcher, B. Guttorp, P. & Percival, D. (1999) Mathematical background for wavelet estimators of cross-covariance and cross-correlation. *Technical Report*, 38. National Research Center for Statistics and the Environment, United States.

Whitcher, B. & Jensen, M. J. (2000) Wavelet estimation of a local long memory parameter. *Exploration Geophysics*, 31, 94–103.

Wornell, G. W. (1993) Wavelet-based representations for the $1/f$ family of fractal processes. *Proc. IEEE*, 81 (10), 1428–1450.