

Wavelet Estimation of Time Series Regression with Long Memory Processes

Haibin Wu
University of Alberta

Abstract

This paper studies the estimation of time series regression when both regressors and disturbances have long memory. In contrast with the frequency domain estimation as in Robinson and Hidalgo (1997), we propose to estimate the same regression model with discrete wavelet transform (DWT) of the original series. Due to the approximate de-correlation property of DWT, the transformed series can be estimated using the traditional least squares techniques. We consider both the ordinary least squares and feasible generalized least squares estimator. Finite sample Monte Carlo simulation study is performed to examine the relative efficiency of the wavelet estimation.

* I am really grateful to Yanqin Fan for guidance and comments, and to Robert Taylor and two anonymous referees for valuable comments. All the errors are the author's.

Citation: Wu, Haibin, (2006) "Wavelet Estimation of Time Series Regression with Long Memory Processes." *Economics Bulletin*, Vol. 3, No. 33 pp. 1-10

Submitted: December 14, 2005. **Accepted:** December 6, 2006.

URL: <http://economicsbulletin.vanderbilt.edu/2006/volume3/EB-05C40004A.pdf>

1. Introduction

Long memory processes play an important role in time series analysis and can be found in many fields such as physics, engineering, environmental sciences and economics. Intuitively, a long memory process is a stochastic process whose autocovariance function decays very slowly as the distance between observations tends to infinity. Due to this property, traditional time series techniques are no longer applicable. In the current literature, estimation and regression analysis about long memory processes is usually done in the frequency domain¹.

In this paper, we look at the regression model where both regressors and disturbances have long memory property. Robinson and Hidalgo (1997) examine the asymptotic properties of the estimation of this model mainly in the frequency domain, and they show that feasible generalized least squares estimator using Fourier transform is asymptotically efficient. This paper proposes the estimation of the same model using discrete wavelet transform (DWT) of the independent and explanatory variables. DWT is a linear filter, when it is applied to a time series, the original time series is decomposed into coefficients at different time scales. Compared with Fourier transform in the frequency domain, DWT attains properties of the original series at different time scales, and at the same time gives information about the time locations at different scales. One important property of DWT which makes it ideal for analysis of long memory processes is the so-called de-correlation property. The de-correlation property ensures the coefficients obtained by applying DWT to the original series have very weak correlation (see Fan(2001)). So after the DWT, we can use the traditional econometric techniques to analyze long memory processes.

Since generalized least squares in the frequency domain is most efficient, we examine the relative efficiency achieved by the generalized least squares estimator using wavelet transformed series. After the DWT is applied to the dependent and explanatory variables, ordinary least squares is performed, then using the Cochrane-Orcutt method, we perform the generalized least squares regression. Though asymptotic theory cannot be established for the wavelet estimation, Monte Carlo simulation study shows that the relative efficiency of the estimator achieved by using the wavelet transform is reasonably good.

This remaining part of this paper is organized as follows: part II gives a brief introduction of DWT and the approximate de-correlation property; part III discusses the frequency domain estimation using Fourier transform in the current literature; part IV displays the estimation procedure using wavelet methods; part V gives the finite sample simulation results of the relative efficiency of the wavelet estimation, and part VI concludes.

2. DWT and the De-correlation Property

We consider the discrete wavelet transform (DWT) for a finite series². Let $T=2^J$ where J is a positive integer. Given a sequence $\{X_t\}_{t=1}^T$, let $W_{j,k}$ ($j=1, \dots, J; k=0, \dots, M_j-1$) denote the

¹ One exception is Robinson and Hidalgo (1997) which also considers the estimation in time domain, I would like to thank the editor and referees for pointing this out.

² Actually, y can be any finite data set (e.g., cross sectional sample). Since we are focusing on time series regression in this paper, we will use time series hereafter.

boundary independent DWT coefficients of a series where M_j is the number of boundary independent DWT coefficients at level j .

2.1 The Discrete Wavelet Transform³

A wavelet filter $\{h_l\}_{l=0}^{L-1}$ of length L where L is an even integer must have the following three properties:

Zero Sum: $\sum_{l=0}^{L-1} h_l = 0;$

Unit Energy; $\sum_{l=0}^{L-1} h_l^2 = 1;$

Orthogonality: $\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0$ for all nonzero integers n .

It is clear that a wavelet filter must be a difference filter. The scaling filter $\{g_l\}_{l=0}^{L-1}$ is the quadrature mirror filter corresponding to the wavelet filter $\{h_l\}_{l=0}^{L-1}$ by the relationship $g_l = (-1)^{l+1} h_{L-1-l}$.

Let $\{h_l\}_{l=0}^{L-1}$ be a Daubechies compactly supported wavelet filter, then the transfer function of $\{h_l\}$ is just the discrete Fourier transform:

$$H(f) = \sum_{l=0}^{L-1} h_l \exp(-i2\pi fl) \quad (1)$$

Similarly, let $G(f)$ be the transfer function of $\{g_l\}$. We can write the squared gain function of $\{h_l\}$ as:

$$H(f) = |H(f)|^2 = D^{L/2}(f)C(f) \quad (2)$$

Where $C(f) = \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f)$ is the squared gain function of a high-pass filter,

and $D(f) = 4\sin^2(\pi f)$ is the squared gain function of a first order backward difference filter. Thus a wavelet filter can be viewed as a two-stage filter. The first stage is a $L/2$ th order backward difference filter, while the second stage is a high-pass filter. Hence we know wavelet filter $\{h_l\}$ and scaling filter $\{g_l\}$ are approximately high-pass and low-pass filter, i.e., they preserve information at high or low frequencies of the original series.

The DWT can be performed using the so-called pyramid algorithm. The algorithm works as follows: At first stage, the time series $\{X_t\}_{t=1}^T$ with $T=2^J$ is filtered using the wavelet and scaling filters and downsampled by two. The resulting coefficients consist of a vector of wavelet coefficients \mathbf{W}_1 and scaling coefficients \mathbf{V}_1 each with length $T/2$. At the second stage, the vector of scaling coefficients are filtered with both wavelet and scaling filtered to obtain a vector of wavelet \mathbf{W}_2 and scaling coefficients \mathbf{V}_2 each with length $T/4$. The process continues until we reach level J to get a wavelet and scaling coefficients each with length 1.

To see more clearly the effects of DWT of the original series with pyramid algorithm. Consider the level 1 scaling coefficients \mathbf{V}_1 , which is obtained by filtering the original

³ This part follows basically the structure of Fan and Whitcher (2001) section 2.1.

series with scaling filter $\{g_l\}$ and downsampled by two. That is, the input is the original series, the operator is the scaling filter and the output is the level 1 scaling filter. At second stage, the level 1 scaling coefficients are filtered again by level 2 wavelet and scaling filters. In other words, the input is \mathbf{V}_1 , the operator is the level 2 wavelet and scaling filters, and the output is \mathbf{W}_2 and \mathbf{V}_2 . It is clear that we obtain the level 2 wavelet coefficients by a scaling filter first and then by a wavelet filter, and we can have the same wavelet coefficients \mathbf{W}_2 by a new filter which is the convolution of the level 1 scaling filter and level 2 wavelet filter. Define $H_2(f)$ as the transfer function of the new filter which is the convolution of the level 1 scaling filter and level 2 wavelet filter, we have

$$H_2(f) = H(2f)G(f) \quad |f| \leq 1/2$$

where the transfer function of level 2 wavelet filter $H(2f)$ captures the idea of downsampling by two since in the frequency domain which is equivalent to scaling the transfer function by two. Similarly, for the transfer function of the convolution of the level 1 scaling filter and level 2 scaling filter $G_2(f)$, we have

$$G_2(f) = G(2f)G(f) \quad |f| \leq 1/2$$

Continuing in this way, if $H_j(f)$ and $G_j(f)$ are the transfer functions of the equivalent filters which are the convolution of level 1 through $j-1$ scaling filters and level j wavelet and scaling filters respectively, we have

$$H_j(f) = H(2^{j-1}f) \prod_{i=0}^{j-2} G(2^i f) \quad |f| \leq 1/2$$

and

$$G_j(f) = \prod_{i=0}^{j-1} G(2^i f) \quad |f| \leq 1/2$$

2.2 Fractionally Integrated Processes and Long Memory

Before talking about the approximate de-correlation property of DWT, we briefly go over some important properties of long memory processes in this subsection for the purpose of illustrating the approximate de-correlation property of DWT of a long memory process.

We focus on the discrete time realization of a fractionally integrated auto-regressive and moving average (FIARMA(0, d , 0)) process $\{X_t\}_{t=1}^T$ with differencing parameter d and innovation variance σ^2 :

$$(1-B)^d X_t = \varepsilon_t \quad (3)$$

where B is backward shift operator with $BX_t = X_{t-1}$. If the differencing parameter satisfies $-1/2 \leq d < 1/2$, then the process is stationary and its spectral density can be represented as (see Taniguchi and Kakizawa (2000)):

$$S_X(f) = \frac{\sigma^2}{2\pi} |2 \sin(\pi f)|^{-2d} \quad (4)$$

If $0 < d < 1/2$, the process is a stationary long memory process, and if $d \geq 1/2$, the process is non-stationary, but we can difference the series to the $[d+1/2]$ th order to obtain stationarity.

When $0 < d < 1/2$, it is easy to show that

$$\text{cov}(X_t, X_{t+k}) = O(|k|^{2d-1}) \text{ as } |k| \rightarrow \infty \quad (5)$$

Compared with stationary short memory processes which have exponential rate of decay for the auto-covariance function. We can see the rate of decay for the long memory is much slower. That is where the name of long memory processes comes from.

2.3 The Approximate De-correlation Property of DWT

One important property of DWT which makes it ideal for long memory processes is the so-called approximate de-correlation property. It states that the DWT coefficients of a long memory process are approximately uncorrelated both with and across scales.

The covariance between DWT coefficients of a long memory process can be expressed as⁴:

$$\text{cov}(W_{j,k}, W_{j',k'}) = \int_{-1/2}^{1/2} \exp(i2\pi f(2^{j'}(k'+1) - 2^j(k+1))) H_j(f) H_{j'}^*(f) S_X(f) df \quad (6)$$

where $H_j(f)$ is the transfer function for the j th level wavelet filter, and $S_X(f)$ is the spectral density function for a long memory process.

Fan (2003) shows formally that for the within scale correlation at level j , we have the following property:

$$\text{cov}(W_{j,k}, W_{j,k'}) = O([2^j |k - k'|]^{-(L-2d)}), \text{ as } 2^j |k - k'| \rightarrow \infty \quad (7)$$

The above equation gives the relationship between the rate of decay of auto-covariance and the time periods separating the wavelet coefficients given any level j and length L of the wavelet filter.

While for the across scale correlation, we have

$$\text{cov}(W_{j,k}, W_{j',k'}) = O(L^{-3/4}) \text{ uniformly in } k \text{ and } k' \quad (8)$$

So when the length of wavelet filter tends to infinity, the across scale correlation is quite small even for fixed separating time periods between two wavelet coefficients.

Note the difference in rate of decay between the original long memory processes, and the DWT coefficients, compared with the original series, the DWT coefficients have much faster rate of decay.

3. DWT Estimation of Long Memory Processes

In this paper, we consider a regression model where both regressors and disturbances have long memory. Specifically, the model takes the following form:

$$Y_t = \alpha + \beta' X_t + u_t \quad (9)$$

where both X_t and u_t possess long memory property, and u_t is covariance stationary, having zero mean and absolute continuous spectral density function.

Robinson and Hidalgo (1997) derived the asymptotic property for the frequency domain estimator. In this section, we describe how to estimate the same model using DWT of the original series. As in Robinson and Hidalgo (1997), the regression model takes the form as in equation (9) where both X_t and u_t have long memory with differencing parameter c and d . Let $T=2^J$ be the number of observations, and $T_j=T/2^j$. Let $W_{j,k}^X$ denote the level j boundary independent DWT coefficients for $X=\{X_1, \dots, X_T\}$ obtained from the level j wavelet filter $\{h_{j,l}\}_{l=0}^{L_j-1}$, that is

⁴ For details of derivation here, see Fan (2003).

$$W_{j,k}^X = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(k+1)-l-1 \bmod N_{j-1}}, j=1, \dots, J, k=0, \dots, T_j-1$$

where $L_j=(2^j-1)(L-1)+1$ and M_j is the number of boundary independent coefficients at level j . Similarly define $W_{j,k}^Y$ and $W_{j,k}^U$ for $Y=\{Y_1, \dots, Y_{T_j}\}$ and $U=\{u_1, \dots, u_{T_j}\}$.

Consider the least squares regression with the DWT coefficients of Y and X :

$$W_{j,k}^Y = \alpha + \beta W_{j,k}^X + W_{j,k}^U \quad (11)$$

Due to the de-correlation property, we expect the OLS regression with DWT of the original series should perform reasonably well.

Although DWT has good de-correlation property, it is possible the residuals from the above OLS regression with DWT coefficients have correlation among them. To see how the generalized least squares estimator improves the efficiency of the estimation, we consider a simple feasible generalized least squares estimator using Cochrane-Orcutt method. After the OLS regression with DWT coefficients, we regress the estimated residuals on its own first lag:

$$\hat{W}_{j,k}^U = \rho \hat{W}_{j,k-1}^U + \varepsilon_{j,k} \quad (12)$$

to get an estimated autocorrelation coefficient $\hat{\rho}$. Then using the estimated autocorrelation coefficient, we transform the DWT coefficients to estimate the following model:

$$\tilde{W}_{j,k}^Y = \alpha + \beta \tilde{W}_{j,k}^X + \tilde{W}_{j,k}^U$$

where $\tilde{W}_{j,k}^Y = W_{j,k}^Y - \hat{\rho} W_{j,k-1}^Y$ and similarly for X and U . If the specification in (12) is correct, i.e., there is just first order correlation among the residuals from OLD regression, then asymptotically this method yields efficient estimation.

4. Simulation Result

Simulation study similar to that in Robinson and Hidalgo (1997) is performed to examine the relative efficiency of wavelet estimation. Specifically, we compare the ratio of asymptotic variance $T^{-1}\Sigma_{g^{-1}}^{-1}$ of the generalized least squares estimator in Robinson and Hidalgo (1997):

$$\Sigma_{g^{-1}}^{-1} = \Gamma(1-c+d)^2 / \Gamma(1-2c+2d)$$

to the Monte Carlo mean squared errors of DWT estimator, where c and d are the order of fractional integration for independent variable and error term. Since the feasible GLS estimator in Robinson and Hidalgo is quite efficient, we want to see the relative efficiency of DWT estimator to the feasible estimator in Robinson and Hidalgo (1997).

In this paper, we use three wavelet filters from the family of the Daubechies compactly supported wavelet filters: Haar, D4 and LA8 wavelet filters. We generate 500 iterations. Table 1 shows the ratio of $T^{-1}\Sigma_{g^{-1}}^{-1}$ to the mean squared errors of ordinary least squares estimator using level 1 DWT coefficients. Several interesting features can be noticed from Table 1. First, as in Robinson and Hidalgo (1997), the efficiency of OLS estimator of DWT coefficients increases in d and decreases in c . Second, the efficiency of DWT OLS estimator is lower than the feasible GLS estimators in Robinson and Hidalgo

(1997). One possible reason is that we use only level DWT coefficients which is of length $T/2$. Third, the three wavelet filters perform quite similarly with Haar slightly better overall.

Table 2 shows the same ratios as in Table 1 using wavelet coefficients up to level 4. The significant improvement in efficiency with more levels of DWT coefficients is apparent compared with Table 1. The reason is simple, since with level 4 DWT coefficients, we use almost 94% of the complete DWT coefficients, so there is much more information used compared with estimation with only level 1 coefficients. The other general pattern is quite similar to that in Table 1.

Table 1: Ratios of $T^{-1}\Sigma_g^{-1}$ to MSE of DWT OLS (level 1)

T=128	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.4563	0.5327	0.5264	0.6129	0.6395
	0.15	0.3886	0.4602	0.5265	0.6015	0.6469
Haar	0.25	0.3923	0.4363	0.4353	0.5128	0.5597
	0.35	0.2947	0.3591	0.4230	0.4541	0.5349
	0.45	0.1709	0.2969	0.3909	0.4065	0.5110
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.4364	0.5097	0.5675	0.5736	0.8056
	0.15	0.3618	0.4555	0.4731	0.5839	0.6251
D4	0.25	0.3855	0.4080	0.4532	0.5354	0.5996
	0.35	0.2711	0.3095	0.4065	0.4778	0.5225
	0.45	0.1641	0.2900	0.3207	0.4102	0.4597
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.4006	0.5672	0.5174	0.5598	0.7085
	0.15	0.3843	0.4422	0.4800	0.5587	0.6173
LA8	0.25	0.3317	0.4036	0.4384	0.4995	0.5790
	0.35	0.2716	0.3326	0.3787	0.4506	0.5346
	0.45	0.1593	0.2483	0.3325	0.3532	0.4695

Table 3 gives the ratio of $T^{-1}\Sigma_{g^{-1}}$ to the mean square errors of feasible GLS estimator described above using level 4 wavelet coefficients. Compared with Table 2, an interesting observation is that the feasible GLS estimator performs relatively better compared with OLS estimator only when d is 0.35 or larger. The intuition is simple, since wavelet filter is two stage filter with de-correlation property, the usefulness of Cochrane-Orcutt method is evident only when the degree of long memory of disturbances is large enough.

5. Conclusion

This paper examines the relative efficiency of wavelet estimator of a regression model with both regressors and disturbances having long memory. Though not so efficient as the feasible GLS estimator proposed by Robinson and Hidalgo (1997), the wavelet estimator is quite useful due to its computational convenience. Much can be still done in this direction. Further work includes (1) comparing the efficiency performance for different sample size; (2) studying the performance when regressors or disturbances are nonstationary, i.e., c or d is greater than 0.5.

Table 2: Ratios of $T^{-1}\Sigma_g^{-1}$ to MSE of DWT OLS (level 4)

T=128	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.9428	0.9942	0.9328	0.9001	0.8420
	0.15	0.9110	0.9195	0.7585	0.8006	0.7350
Haar	0.25	0.8545	0.8331	0.8006	0.7125	0.6800
	0.35	0.6958	0.7477	0.7223	0.7463	0.6486
	0.45	0.4618	0.7706	0.7114	0.7527	0.5677
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.8441	0.9582	0.8673	0.8837	0.8725
	0.15	0.7857	0.8872	0.7457	0.7815	0.6863
D4	0.25	0.7428	0.7328	0.7471	0.6764	0.6509
	0.35	0.5845	0.6922	0.7015	0.7637	0.7070
	0.45	0.4299	0.6972	0.6904	0.7063	0.5679
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.6942	0.8602	0.7942	0.8902	0.8502
	0.15	0.7036	0.7720	0.7091	0.7333	0.7026
LA8	0.25	0.6234	0.6841	0.6632	0.6456	0.5671
	0.35	0.4557	0.6016	0.6209	0.6332	0.6786
	0.45	0.3547	0.5485	0.5603	0.6146	0.5460

Table 3: Ratios of $T^{-1}\Sigma_g^{-1}$ to MSE of DWT FGLS (level 4)

T=128	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.9184	0.9978	0.8844	0.8897	0.8620
	0.15	0.8806	0.8997	0.7568	0.8233	0.7299
Haar	0.25	0.8316	0.8030	0.7909	0.7218	0.7062
	0.35	0.6765	0.7245	0.7029	0.7401	0.6579
	0.45	0.4524	0.7366	0.6944	0.7725	0.5897
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.8204	0.9550	0.8744	0.8659	0.8613
	0.15	0.7678	0.8607	0.7483	0.7608	0.7281
D4	0.25	0.7317	0.7549	0.7447	0.7047	0.7064
	0.35	0.5555	0.6765	0.6916	0.7722	0.7198
	0.45	0.4284	0.6737	0.6701	0.7402	0.6042
	c/d	0.05	0.15	0.25	0.35	0.45
	0.05	0.6664	0.8433	0.7869	0.9178	0.9112
	0.15	0.6681	0.7642	0.7209	0.7384	0.7503
LA8	0.25	0.5994	0.6592	0.6632	0.6681	0.6237
	0.35	0.4344	0.5849	0.6321	0.6241	0.6839
	0.45	0.3351	0.5214	0.5571	0.6202	0.5738

References

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

Fan, Yanqin (2003). "On the Approximate De-correlation Property of the Discrete Wavelet Transform for Fractionally Differenced Processes", *IEEE Transactions on Information Theory* 49, 516-521

_____ and Brandon Whitcher (2001) "A Wavelet Solution to the Spurious Regression of Fractionally Integrated Processes", working paper, Department of Economics, Vanderbilt University

Percival, D.B. and A.T. Walden (2000). *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic mathematics, Cambridge University Press

Robinson, P.M. and F.J. Hidalgo (1997) "Time Series Regression with Long Range Dependence", *The Annals of Statistics*, 25(1), 77-104

Taniguchi, Masanobu and Yoshihide Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series*, Springer Series in Statistics

Tasy, Wen-Jen and Ching-Fan Chung (2000) "The Spurious Regression of Fractionally Integrated Processes", *Journal of Econometrics*, (96), 155-182