[14] Hurvich C. and K. Beltrao (1993) "Asymptotics for the Low-Frequency Ordinates of the Periodogram of a Long- Memory Time Series," *Journal of Time Series Analysis*, 14, 455-472.

[15] Kronland-Martinet, R., J. Morlet, and A. Grossman (1987) "Analysis of Sound Patterns Through Wavelet Transforms," *International Journal of Pattern Recognition and Artificial Intelligence*, 1, 273-301.

[16] Mallat, S. (1989) "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674-693.

[17] Mallat, S. and W. L. Hwang (1991) "Singularity Detection and Processing with Wavelets," *IEEE Transactions on Information Theory*, 38, 617-643.

[18] Mandelbrot, B. and J. Van Ness (1968) "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, 10, 422-437.

[19] McLeod, B. and K. Hipel (1978) "Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 14, 491-518.

[20] Meyer, Y. (1990) *Ondelettes ét Operérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs Multilinéaires*, Paris: Hermann.

[21] Morlet, J. (1983) "Sampling Theory and Wave Propagation," in *Acoustic Signal/Image Processing and Recognition*, ed. C. H. Chen (Springer-Verlag: Berlin), 233-261.

[22] Wornell, G. (1993) "Wavelet-Based Representation for the $1/f$ Family of Fractal Processes," *Proceedings of the IEEE*, 81, 1428-1450.

[23] Wornell, G. and A. Oppenheim (1992) "Estimation of Fractal Signals from Noisy Measurements Using Wavelets," *IEEE Transactions on Signal Processing*, 40, 611-623.

References

[1] Barnett, W. (1980) "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory," *Journal of Econometrics,* 48, 11-47.

[2] Beylkin, G., R. Coifman, and V. Rokhlin (1991) "Fast Wavelet Transforms and Numerical Algorithms," *Communications on Pure and Applied Mathematics,* 44, 141-183.

[3] Brockwell, P. J. and R. A. Davis (1987) *Time Series: Theory and Models*, (Springer-Verlag: New York).

[4] Burt, P. and E. Adelson (1983) "The Laplacian Pyramid as a Compact Image Code," *IEEE Transaction on Communications*, 31, 482-540.

[5] Cheung, Y. (1993) "Tests for Fractional Integration: A Monte Carlo Investigation," *Journal of Time Series Analysis,* 14, 331-345.

[6] Daubechies, I. (1988) "Orthonormal Bases of Compactly Supported Wavelets," *Communications on Pure and Applied Mathematics*, 41, 909-996.

[7] Daubechies, I. (1992) *Ten Lectures on Wavelets,* (SIAM: Philadelphia) 1992.

[8] Davies, R. and D. Harte (1987) "Tests for Hurst Effect," *Biometrika,* 74, 95-101.

[9] Diebold, F. and G. Rudebusch (1989) "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics,* 24, 189-209.

[10] Geweke, John and Susan Porter-Hudak (1983) "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis* 4, 221-238.

[11] Granger, C. and R. Joyeux (1980) "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-29.

[12] Hosking, J. R. (1981) "Fractional Differencing," *Biometrika*, 68, 165-176.

[13] Hosking, J. (1984) "Modeling Persistence in Hydrological Time Series Using Fractional Differencing," *Water Resources Research*, 20, 1898-1908.

the hypothesis of white noise. The choice of the wavelet and its order of regularity had little effect on the *z-stat* and did not seem to be vital in analyzing fractionally integrated processes.

Regardless of the level of contamination, the powers of the *z-stat* were found to be high for differencing parameters that were larger than 0.15 and fractionally difference series with at least 2^9 observations. The *z-stat*s for fractionally integrated series with differencing parameters near zero had very low power under both small and large time series, with and without contamination. The level of contamination had a relatively minor effect on the wavelet estimation of d as displayed by the slight regression towards zero of the calculated *z-stat*s.

We applied the wavelet estimation of d to the logarithmically differenced Divisia monetary aggregate at the level of aggregation found in M1, M2, M3, and L. In every case except M1 the null hypothesis of $d = 0$ was rejected at the 1% significance level. The asymptotic standard errors for the maximum likelihood estimates of d were found to be small. This helps in testing the impact of a unit shock on the process since a small standard error results in a tighter confidence interval for the cumulative impulse response of a fractionally integrated ARMA model.

Overall we found that wavelet analysis and estimation of the a fractionally integrated model were less subjective and provided stronger statistical results than the GPH method. Our method does not suffer from the subjective choice of the GPH of choosing the sample size for the periodogram, nor the questionable assumption about the distribution properties of the regression's residuals. In addition, our method incorporates noise into the fractionally integrated model, something the GPH method has not even considered.

	d	σ^2	σ_η^2	$\log L(\hat{\boldsymbol{\theta}})$
M1	1.25337*	3.5693E+6	2.6762E+1	-902.02
	(4.1146E-4)	(0.00000)	(1.108E+1)	
M2	1.22952*	2.2998E+6	1.8285E+1	-860.43
	(3.3559E-4)	(0.00000)	(5.53854)	
M3	1.24953*	3.3043E+6	2.4907E+1	-894.15
	(3.7226E-4)	(0.00000)	(9.29266)	
L	1.24329*	3.0957E+6	2.6670E+1	-898.94
	(4.0728E-4)	(0.00000)	(1.0396E+1)	

Asterisks denote significance at 1% level.

M1's OLS estimates for detrending: $a = 1.814$, $b = 1.428$

M2's OLS estimates for detrending: $a = 13.97$, $b = 1.571$

M3's OLS estimates for detrending: $a = 1.773$, $b = 1.761$

L's OLS estimates for detrending: $a = 3.507$, $b = 1.648$

Table 6: Maximum Likelihood Estimates of θ for Linear Detrended Divisia Monetary Data

of shock response more defined.

6 Conclusion

In this paper we have synthesized wavelet analysis with fractionally integrated processes to provide an alternative method of calculating the differencing parameter d from the GPH method [Geweke and Porter-Hudak (1983)]. We have shown that a fractionally integrated processes with a differencing parameter $d \in (-1/2, 1/2)$ is a member of the class of $1/f$ processes as defined by Wornell (1992). As a result of being a $1/f$ process, the wavelet coefficients from a $I(d)$ process will be distributed normal with mean zero and variance $\sigma^2 2^{2dm}$. Knowledge of the wavelet coefficient's distribution provides us with the likelihood function from which the maximum likelihood estimate of the differencing parameter d and its variance is found with the EM-algorithm of Wornell and Oppenheimer (1992).

We tested the robustness of the wavelet estimate of d to a variety of orthonormal wavelets with different orders of regularity, a wide range of $I(d)$ processes with and without contaminating noise, and to the length of the process by simulating fractionally integrated processes with the DH algorithm [Davies and Harte (1987)] and calculating the *z-stat* for

	d	σ^2	σ_η^2	$\log L(\hat{\boldsymbol{\theta}})$
M1	-1.9318E-5	1.2501E-2	1.2503E-2	235.88
	(3.2250E-9)	(5.332E-11)	(0.00000)	
M2	1.9545E-4*	1.2526E-2	1.2502E-2	235.90
	(2.6750E-9)	(5.008E-11)	(0.00000)	
M3	1.9351E-4*	1.2525E-2	1.2501E-2	235.92
	(2.4105E-9)	(4.5380E-11)	(0.00000)	
L	1.1426E-4*	1.2514E-2	1.2500E-2	235.95
	(1.1981E-9)	(2.2005E-11)	(0.00000)	

Asterisks denote significance at 1% level

Table 5: Maximum Likelihood Estimates of θ for Log Differenced Divisia Monetary Data

using nonlinear models that display long-term memory over linear models that do not. It also discourages differencing a series to obtain a finite variance since such action removes the process' important long-memory properties.

The point about differencing the series is strongly served by the results reported in table 6 for the linearly detrended monetary data. In table 6 the one-sided *z-stat* strongly rejects the hypothesis that the Divisia monetary aggregates are unit-roots. The small asymptotic variances found in table 6 strengthens the rejection of $d = 1$ by placing a tight confidence intervals over the estimates of d. The consequence of differencing is illustrated in M1 where the act of differencing removed the long-memory properties of M1, resulting in acceptance of the null hypothesis.

The tight variances calculated with the wavelet estimation is an additional improvement over the GPH method, particuliarly when deriving confidence intervals for the cumulative impulse response of a fractionally integration ARMA model. The confidence interval of a cumulative impulse response is found by calculating the cumulative impulse response for d's that are a multiple of its standard error. Diebold and Rudebusch (1989) showed that the cumulative impulse response of a fractionally integrated ARMA model is highly sensitive to the value of d and that a large standard error results in a wide confidence interval that can contain both persistence and dissipation responses to a unit shock. By reducing the standard error of d the confidence interval of the impulse response is tighter and the measure

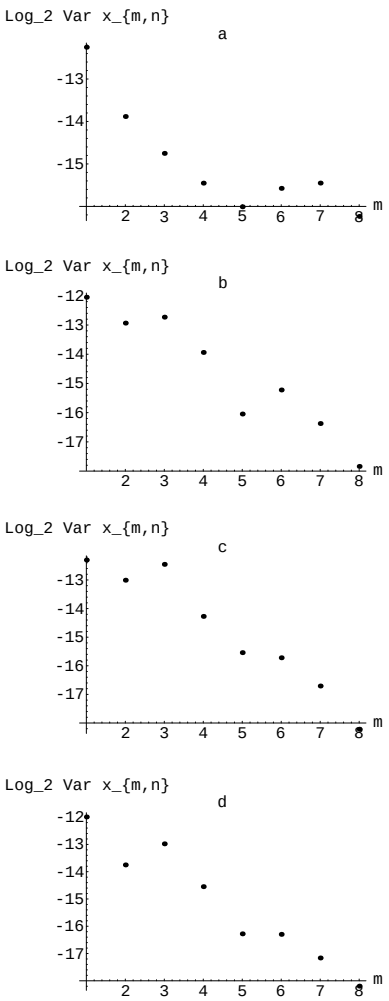

Figure 4: Scale-to-scale plots of wavelet coefficients sample variance for a) Divisia M1 b) Divisia M2 c) Divisia M3 and d) Divisia L

December 1990 that was supplied to us by the Federal Reserve Bank of St. Louis. Because the wavelet transform requires the length of the data to be a multiple of 2 we only use the first 256 (2^8) months to calculate d. The level of aggregation for the Divisia index is the same as the Federal Reserves' M1, M2, M3, and L.

Like most economic time series the Divisia monetary aggregates have a trend associated with them. To eliminate the effect of the trend we preform our calculations on the logarithmic difference of the Divisia aggregates. Many in spectral analysis express concern that this type of detrending is harmful to the low frequencies of the time series. This is especially important for long-memory processes like that of the fractionally integrated model. Hence, we also estimate and report d for the Divisia monetary aggregate detrended by the linear time regression $a + bt$.

5.1 Fractionally Integrated Monetary Data

In figure 4 the relationship of the wavelet coefficient's sample variance for the Divisia monetary data is plotted in the $m \times log_2 \text{var}\, x_{m,n}$ plane. Except for Divisia M1 the wavelet coefficient's sample variance for the monetary aggregate data exhibits a strong scale-to-scale geometric relationship. This relationship is more pronounced for those wavelet coefficients with larger scaling parameters. This is inpart due to the fact that at larger scales there are more coefficients to calculate the wavelet coefficient's sample variance. It may be possible to obtain a better estimate of d by limiting the wavelet coefficients to those wavelet coefficients with higher resolution, i.e. using only $\{x_{m,n} : m \in [min, max], n \in N(m)\}$ where $min \geq 1$. However, this idea has not been tested nor theoretically shown to be any better than the method we use here, but is an area for future research.

Table 5 reports the MLE for $\boldsymbol{\theta}$ and their asymptotic variance (in parenthesis) for the logarithmically differenced Divisia monetary aggregates. In every case except for M1 the one-sided *z-stat* for the null hypothesis $d = 0$ is rejected with a Type I error of 1% (denoted by asterisks). The significance of the *z-stat* lends support to the argument that economic time series possess fractional differencing parameters and strengthens the argument for

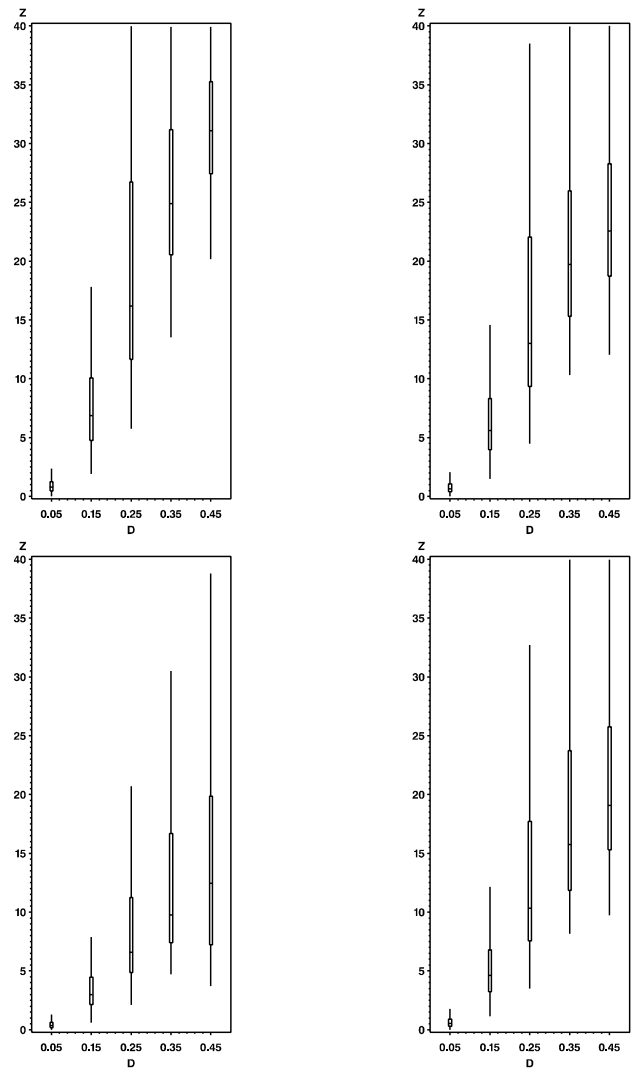

Figure 3: Box-Plots, $N = 2^{11}$ in clockwise order $\sigma_\epsilon^2 = 1.0, 0.95, 0.9, 0.8$

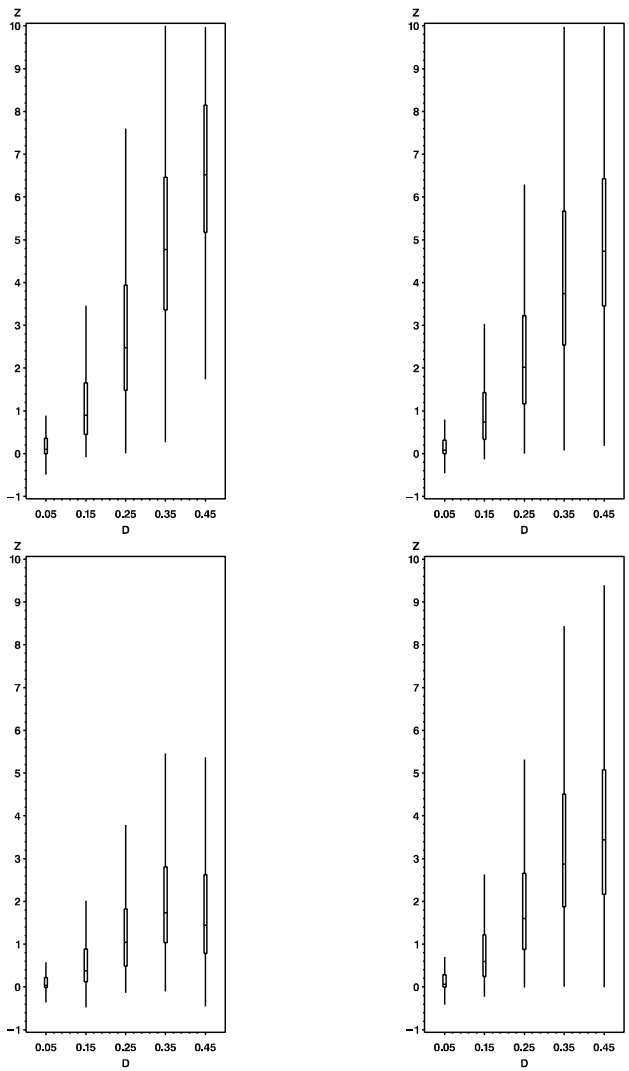

Figure 2: Box-Plots, for $N = 2^7$ in clockwise order $\sigma_\epsilon^2 = 1.0, 0.95, 0.9, 0.8$

regression by the *z-stat* to zero increases. In figure 3 the size of the drop in the *z-stat*'s 75% and 25% quantile range from one to fifteen. However, when comparing the size of the change with the *z-stat*'s overall distribution the relative magnitude of the decline in figure 2 and figure 3 are close to being the same regardless of the sample size.

Of particular interest in the box-plots of figure 2 and figure 3 is that after an initial drop in the *z-stat*'s distribution, increased levels of contamination only slightly lowers the *z-stat* for $I(d)$ with small values of d. One possible explanation is that contamination of a series has a limited effect on the wavelet estimation of d, i.e. past a certain contamination level the wavelet estimation of d is no longer effected by increases in the level of noise.

Comparing our results with those found with the GPH method [Cheung (1993)], both methods have a difficult time distinguishing I(0.05) from white noise. Other than this similarity the *z-stat* calculated with wavelets outperforms the GPH method for small and large sample size and for different values of d. The wavelet technique rejected white noise nearly three times more often than did the GPH method for $I(d)$ processes with $d \geq 0.25$ and approximately 100 observations. When the length of the series is approximately 500 observations the wavelet method continues to better the GPH results especially when $d = 0.25$ where the wavelet method rejects white noise at the 5% significance level in every simulation and the GPH method only 31%.

5 Application to Divisia Monetary Data

We apply the wavelet estimation of the fractional differencing parameter to the Divisia monetary aggregate to test an economic time series for a fractional differencing parameter. Our reason for choosing this particular economic time series rests on the strong statistical indice properties associated with the nonparameteric Divisia aggregate. Barnett (1980) showed that the Divisia monetary aggregate is a second-order approximation of the true monetary aggregate and hence, the error between the true aggregate and the Divisia aggregate is a third-order remainder term. This property of the Divisia monetary aggregate reduces the variance of the contamination η.

The Divisia monetary aggregate is calculated with monthly data from February 1969 to

d	N	Min	Max	Power 5%-Test	Power 1%-Test
0.05	2^7	-1.5401	4.2719	0.021	0.009
0.15	2^7	-0.4820	55.8e+7	0.094	0.048
0.25	2^7	-0.1378	12.4046	0.299	0.164
0.35	2^7	-0.1029	21.1e+6	0.528	0.349
0.45	2^7	-0.4560	14.0243	0.434	0.302
0.05	2^8	-3.1530	13.5023	0.016	0.006
0.15	2^8	-0.5685	6.5912	0.160	0.068
0.25	2^8	0.0611	12.2398	0.594	0.378
0.35	2^8	0.0783	21.3838	0.857	0.679
0.45	2^8	0.2957	37.6859	0.795	0.582
0.05	2^9	-0.9809	3.0555	0.017	0.006
0.15	2^9	-0.0545	9.5721	0.360	0.175
0.25	2^9	0.4877	31.9974	0.890	0.711
0.35	2^9	0.9886	78.8036	0.991	0.942
0.45	2^9	1.1505	33.9613	0.987	0.891
0.05	2^{10}	-0.2924	10.6362	0.013	0.005
0.15	2^{10}	0.0480	8.7721	0.641	0.404
0.25	2^{10}	1.3824	48.1856	0.994	0.948
0.35	2^{10}	2.4876	18.0e+7	1.000	1.000
0.45	2^{10}	2.3481	41.5209	1.000	1.000
0.05	2^{11}	-0.1087	2.5345	0.018	0.002
0.15	2^{11}	0.6052	12.4026	0.902	0.679
0.25	2^{11}	2.1014	86.7926	1.000	0.998
0.35	2^{11}	4.7185	41.2380	1.000	1.000
0.45	2^{11}	3.7038	11.4e+1	1.000	1.000

Table 4: Contaminated $I(d)$, $\sigma_\epsilon^2 = 0.8$

lengthed $I(d)$. For instance in table 1 the 5% power level for $I(0.25)$ with $N = 2^7$ equals 0.719 and drops in table 2 to 0.611. On the positive side, the $I(d)$ processes with $d = 0.25, 0.35, 0.45$ and at least 2^9 observations do not seem to be as ill effected by noise. In these cases the 5% power levels are all nearly greater than 90% regardless of the SNR.

In general we found that contaminating a fractionally integrated series pulls down the overall distribution of the *z-stat*. Comparing the box-plots of the *z-stat* found in figure 2 the *z-stat*'s 75% and 25% quantiles for series with $N = 2^7$ drop between one to five points as a result of the increased level of contamination. With larger samples the size of this

d	N	Min	Max	Power 5%-Test	Power 1%-Test
0.05	2^7	-1.3793	4.7827	0.024	0.011
0.15	2^7	-0.2301	9.5931	0.160	0.079
0.25	2^7	-0.0167	22.3558	0.494	0.321
0.35	2^7	0.0039	28.6424	0.814	0.635
0.45	2^7	-0.0067	31.8e+1	0.858	0.731
0.05	2^8	-2.5597	14.8493	0.027	0.008
0.15	2^8	-0.2546	7.9156	0.302	0.153
0.25	2^8	0.3603	17.3638	0.835	0.653
0.35	2^8	0.7056	15.1e+9	0.991	0.946
0.45	2^8	1.2420	45.3249	0.991	0.967
0.05	2^9	-7.1470	3.4704	0.023	0.008
0.15	2^9	-0.0196	13.2220	0.598	0.385
0.25	2^9	1.0819	45.6429	0.988	0.939
0.35	2^9	2.3327	17.8e+1	1.000	1.000
0.45	2^9	5.2023	20.3e+1	1.000	1.000
0.05	2^{10}	-0.1346	13.6156	0.029	0.009
0.15	2^{10}	0.2663	11.5473	0.879	0.709
0.25	2^{10}	2.3212	73.9059	1.000	1.000
0.35	2^{10}	4.5912	15.1e+1	1.000	1.000
0.45	2^{10}	5.0981	63.2437	1.000	1.000
0.05	2^{11}	-0.0371	3.3985	0.058	0.015
0.15	2^{11}	1.1344	39.9e+7	0.997	0.944
0.25	2^{11}	3.5118	37.8774	1.000	1.000
0.35	2^{11}	8.1510	58.1637	1.000	1.000
0.45	2^{11}	9.7308	11.6e+1	1.000	1.000

Table 3: Contaminated $I(d)$, $\sigma_\epsilon^2 = 0.9$

power level of $I(0.45)$ increases from 0.434 to 1.0 as the sample size increases, whereas the power levels corresponding to $I(0.05)$ do not rise but shrink from 0.021 to 0.018.[8] As with the uncontaminated data most of the 5% power levels for contaminated series increase and equal 1 for series with $N = 2^{11}$.

Table 1 and table 2 shows that the level of contamination evidently decreases the power of the *z- stat*, but contamination more severely effects those *z-stat*s associated with shorter

[8]One of the reasons the *z-stat* might be preforming so poorly for $d = 0.05$ is that when d is close to zero $I(d)$ behaves very much like white noise itself. Hence, it is possible that the DH algorithm may actually be producing nearly whitenoise.

d	N	Min	Max	Power 5%-Test	Power 1%-Test
0.05	2^7	-1.2967	5.0678	0.027	0.013
0.15	2^7	-0.1377	11.4882	0.205	0.112
0.25	2^7	-0.0004	29.7618	0.611	0.422
0.35	2^7	0.0761	45.4537	0.923	0.800
0.45	2^7	0.1781	45.0117	0.975	0.926
0.05	2^8	-2.2845	15.6103	0.035	0.012
0.15	2^8	-0.1362	9.8759	0.389	0.223
0.25	2^8	0.5924	20.7127	0.916	0.786
0.35	2^8	1.4377	29.3e+1	0.999	0.990
0.45	2^8	2.0787	49.8645	1.000	0.999
0.05	2^9	-0.6703	3.7024	0.032	0.012
0.15	2^9	0.0140	15.4939	0.717	0.493
0.25	2^9	1.5246	54.7745	0.999	0.978
0.35	2^9	3.4898	27.6e+1	1.000	1.000
0.45	2^9	5.2023	20.3e+1	1.000	1.000
0.05	2^{10}	-0.0837	4.1314	0.040	0.013
0.15	2^{10}	0.4545	13.6403	0.953	0.817
0.25	2^{10}	2.9637	92.5852	1.000	1.000
0.35	2^{10}	6.0718	21.9e+1	1.000	1.000
0.45	2^{10}	8.1490	10.0e+1	1.000	1.000
0.05	2^{11}	-0.0143	4.1004	0.098	0.024
0.15	2^{11}	1.5000	20.7745	0.999	0.991
0.25	2^{11}	4.4884	44.8371	1.000	1.000
0.35	2^{11}	10.3081	70.8960	1.000	1.000
0.45	2^{11}	12.0541	15.2e+1	1.000	1.000

Table 2: Contaminated $I(d)$, $\sigma_\epsilon^2 = 0.95$

hypothesis of $d = 0$ for different values of σ_ϵ^2. The numbers found in the last two columns of each table suggests that the power function is an increasing function of the differencing parameter d and, except for the case $d = 0.05$, an increasing function of the time series' sample size. This is borne out in table 1 for the case $N = 2^7$ where the 5% power level is 0.032 for $d = 0.05$ and 1.00 for $d = 0.45$. As N increases to 2^{11} the 5% power levels for the fractional differenced processes with $0.15 \leq d \leq 0.45$ all reach 1.0. The only power level that does not is that of $I(0.05)$.

These same outcomes are also evident for the contaminated data. In table 4 the 5%

d	N	Min	Max	Power 5%-Test	Power 1%-Test
0.05	2^7	-71.7e+5	5.3735	0.032	0.015
0.15	2^7	-0.0849	11.4361	0.257	0.148
0.25	2^7	0.0088	39.6279	0.719	0.540
0.35	2^7	0.2680	74.0967	0.978	0.923
0.45	2^7	1.7383	87.7237	1.000	0.995
0.05	2^8	-2.0250	16.4351	0.042	0.014
0.15	2^8	-0.0509	12.2688	0.504	0.312
0.25	2^8	0.8773	28.7512	0.964	0.894
0.35	2^8	2.3721	89.1948	1.000	1.000
0.45	2^8	3.9651	6.5e+3	1.000	1.000
0.05	2^9	-0.5711	3.9515	0.045	0.018
0.15	2^9	0.0607	30.7e+9	0.822	0.630
0.25	2^9	2.1064	66.2333	1.000	0.994
0.35	2^9	5.0201	36.4e+7	1.000	1.000
0.45	2^9	8.4139	62.5e+1	1.000	1.000
0.05	2^{10}	-0.0432	4.5028	0.068	0.020
0.15	2^{10}	0.7074	16.9301	1.000	0.910
0.25	2^{10}	3.7761	11.7e+1	1.000	1.000
0.35	2^{10}	8.2442	30.8e+2	1.000	1.000
0.45	2^{10}	12.7710	15.3e+1	1.000	1.000
0.05	2^{11}	-0.0019	27.6216	0.132	0.047
0.15	2^{11}	1.9179	24.5112	1.000	0.998
0.25	2^{11}	5.7391	53.1920	1.000	1.000
0.35	2^{11}	13.5287	10.0e+1	1.000	1.000
0.45	2^{11}	20.1757	29.8e+1	1.000	1.000

Table 1: Uncontaminated $I(d)$, $\sigma_\epsilon^2 = 1.0$

Because of space restriction we only report the results from the Daubechies wavelet with 5 vanishing moments since they provide a good summary of the results found with the other wavelets. The results found with the other wavelets were almost indistinguishable from those found with the Daubechies wavelet with 5 vanishing moments. Neither the type of wavelet nor its regularity seemed to have any significant effect on the *z-stat*. Hence, the choice of the wavelet does not seem to be a critical issue when testing fractionally integrated series for the null hypothesis of $d = 0$.

Table 1 through table 4 contain the simulation results of the *z-stat* under the null

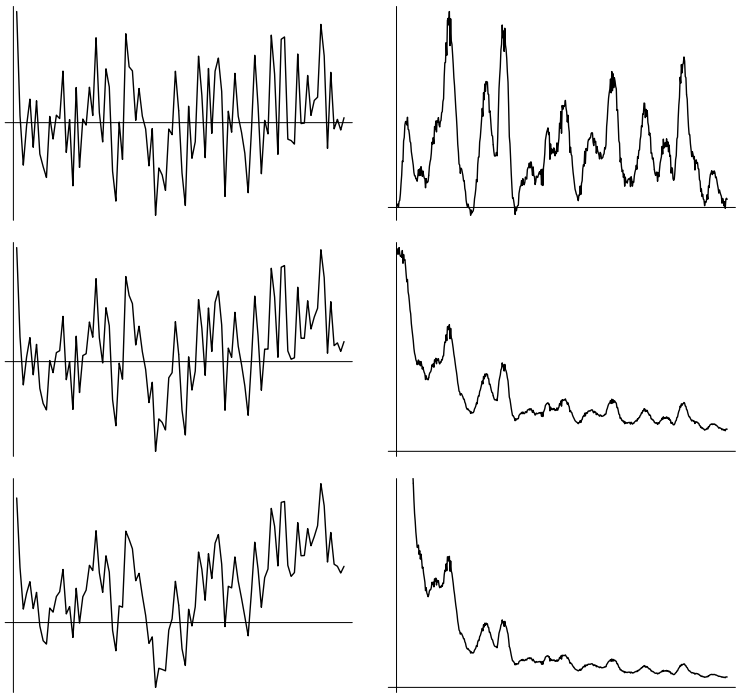

Figure 1: $I(d)$ Plots and Spectrum for $d = 0.05, 0.25, 0.45$

for each experiment as measured by the value of σ_ϵ^2 are, from no contamination to the highest level of contamination, $\sigma_\epsilon^2 = 1.0, 0.95, 0.9, 0.8$. In each simulation the EM algorithm's starting values are $\boldsymbol{\theta}' = (\log_2(0.1)/2, 0.1, 0.1)$ and its convergence criterion equal to 0.001.

For each experiment we used the following wavelets

- Daubechies with $1, 2, 3, \ldots, 10$ vanishing moments [Daubechies (1988)].
- Coiflet with $2, 4, 6, 8, 10$ vanishing moments [Daubechies (1992) p. 258].
- Biorthogonal with $5, 9, 13$ nonzero conjugate quadrature filters for ϕ [Daubechies (1992) p. 259].
- Beylkin with 18 nonzero conjugate quadrature filters for ϕ [Beylkin, Coifman, and Rokhlin (1991)].

to compute the wavelet coefficients, $\{< x, \psi_{m,n} >\}$.

the positive and negative time domain. Because $R(t,s)$ is an even function the Fourier transform of the autocovariance, $\{g(\omega_k)\}$ where $k = 0, 1, \ldots, 2N-1$, will be real. However, in order for the DH method to work $\{g(\omega_k)\}$ must also be nonnegative.[6]

Let $\{Z(k)\}$ for $k = 1, 2 \ldots, N-1$ be a sequence of independent complex normal random variables with unit variance who's real and imaginary parts are independent, i.e. $Z(k) \sim iid\mathcal{N}(0,1)$ for $k = 1, 2, \ldots, N-1$. And let $Z(k) = \bar{Z}(2N-k)$ for $k = N+1, N+2, \ldots, 2N-1$, where $\bar{Z}$ represents the complex conjugate of Z. Lastly, let Z_0 and Z_N be real random variables with mean zero and variance equal to two.

A Gaussian process with the autocovariances $\{c_k\}$ is then generated by taking

$$x(t) = \frac{1}{2N^{1/2}} \sum_{k=0}^{2N-1} Z(k) g(\omega_k)^{1/2} \exp\left[\frac{-2\pi i t k}{2N}\right] \tag{19}$$

for $t = 0, 1, \ldots, N$.

$x(t)$ will be a $I(d)$ process that is distributed Gaussian with variance $\sigma_\epsilon^2 + (c_0 - \sigma_\epsilon^2)$. If σ_ϵ^2 and c_0 are both set equal to 1 the DH algorithm produces a pure fractionally integrated process with parameter d. For σ_ϵ^2 less than 1 and $c_0 = 1$ the DH algorithm generates a contaminated $I(d)$ process. Hence, when σ_ϵ^2 and c_0 are both set equal to one $x(t)$ is a pure fractional differenced process with an infinite signal-to-noise ratio (SNR), whereas smaller values of σ_ϵ^2 decreases the SNR.

In figure 1 three DH generated $I(d)$ processes and their spectrums are graphed for the differencing parameters $d = 0.05, 0.25, 0.45$. The higher valued differencing parameter result in smoother fractionally integrated series that display their long-memory property with a distinguishable cycle and a steeper sloped spectrum.

4.3 Power of the Wavelet Estimate

We generate 1000 samples of $I(d)$ processes with $2^7, 2^8, 2^9, 2^{10}$ and 2^{11} observations for $d = 0.05, 0.15, 0.25, 0.35, 0.45$ with and without contamination.[7] The level of contamination

[6] The sufficiently slow autocovariance decay condition, $c_t - c_{t+1} \geq c_{t+2} - c_{t+3}$, insures that $\{g(\omega_k)\}$ are positive. Because of this condition $I(d)$ processes with a negative d cannot be simulated with the DH method.

[7] We chose to let the number of observation be a multiple of 2 so that the effect of zero-padding would not be a concern.

matrix, $\boldsymbol{I}(\boldsymbol{\theta}) =$

$$\sum_{m\in\mathcal{M}} \frac{2^{m-1}}{2\sigma_m^2} \begin{bmatrix} \left[\log(2^m)\sigma^2 2^{-\gamma m}\right]^2 & -\log(2^m)\sigma^2\left[2^{-\gamma m}\right]^2 & -\log(2^m)\sigma^2 2^{-\gamma m} \\ -\log(2^m)\sigma^2\left[2^{-\gamma m}\right]^2 & \left[2^{-\gamma m}\right]^2 & 2^{-\gamma m} \\ -\log(2^m)\sigma^2 2^{-\gamma m} & 2^{-\gamma m} & 1 \end{bmatrix}. \quad (18)$$

Hence, $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$ as $max \to \infty$, where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$. From the invariance property of the maximum likelihood estimators the differencing parameter's MLE is distributed $(\hat{d} - d) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_{11}^{-1}(\boldsymbol{\theta})/4)$, as $max \to \infty$, where $I_{11}^{-1}(\boldsymbol{\theta})$ is the first diagonal element of the inverse of $\boldsymbol{I}(\boldsymbol{\theta})$.

4.2 Simulation

To determine the robustness of the wavelet estimate's asymptotic properties we conduct a Monte Carlo experiment with fractionally integrated processes for different values of d using a broad variety of the currently known wavelets and test the null hypothesis of white noise, i.e. $d = 0$. Both pure and contaminated fractional differenced data is simulated to determine the effect noisy data has on the wavelet estimation of d.

Simulation of $I(d)$ processes has been a synthesis problem for a number of years. Many of the methods that generate processes with long-memory properties require large amounts of memory and are computationally slow.[5] Taking these issues into consideration we chose to use a method developed by Davies and Harte (1987) that is very efficient and requires small amounts of memory to generate $I(d)$ processes.

The DH algorithm can be generalized to produce any Gaussian process with a specific autocovariance structure as long as the process' autocovariance function has sufficiently slow decay. For our purpose we desire to generate a Gaussian time series $x(t)$ for $t = 0, 1, \ldots, N$ who's autocovariances

$$c_0, c_1, \ldots, c_{N-1}, c_N$$

are equal to those in Eq. (9) for some differencing parameter $d \in (0, 0.5)$. We first generate the autocovariances $\{c_t\}$ and take the Fourier transform of the entire autocovariance over

[5] The procedures most often used to generate fractionally integrated processes are those of McLeod and Hipel (1978) and Hosking (1984). Both of these methods require $\mathcal{O}(N^2)$ of memory and take $\mathcal{O}(N^3)$ computations to generate a $I(d)$ process of length N.

4.1 Estimation of d

Let $x(t)$ be a mean zero fractionally integrated series with differencing parameter $d \in (-0.5, 0.5)$ and suppose that it is contaminated by measurement and aggregation errors that take the form of additive white noise disturbance.[4] The observed series is

$$\tilde{x}(t) = x(t) + \eta(t) \tag{15}$$

where $\eta \sim WN(0, \sigma_\eta^2)$. Because the additive noise is independent of $x(t)$ and has a flat spectrum the wavelet coefficients of $\tilde{x}(t)$ are independently distributed Gaussian processes with variance $\sigma^2 2^{-2dm} + \sigma_\eta^2$.

The knowledge of the wavelet coefficient's distribution enables us to write the likelihood function as a function of the unknown parameters $\boldsymbol{\theta}' = (d, \sigma^2, \sigma_\eta^2)$

$$L(\boldsymbol{\theta}) = \prod_{m \in \mathcal{M}} \prod_{n \in \mathcal{N}(m)} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{< x(t), \psi_{m,n} >^2}{2\sigma_m^2} \right] \tag{16}$$

where $\sigma_m^2 = \sigma^2 2^{-2dm} + \sigma_\eta^2$. If

$$2^{2d} \geq 0, \quad \sigma^2 \geq 0, \quad \sigma_\eta^2 \geq 0 \tag{17}$$

Eq. (16) will be bounded from above and will have a maximum. Occasionally Eq. (16) will have more than one maximum in which case those maximums in addition to the desired maximum likelihood estimate correspond to the trivial cases where the parameter estimates are on the boundary of the parameter space, i.e. either $\sigma^2 = 0$ or $\sigma_\eta^2 = 0$. This insures that a numerical algorithm searching for the likelihood function's maximum will converge to the true maximum likelihood estimates if initialized with values for which Eq. (17) barely hold.

Wornell and Oppenheimer (1992) provide an EM-algorithm, where $\{< x, \psi_{m,n} >\}$ is assumed to be the complete data and $\{< \tilde{x}, \psi_{m,n} >\}$ the censored data, that maximizes Eq. (16) over the parameter space defined by Eq. (17) (except with d replaced by γ). They also compute the Cramér-Rao lower bound to be equal to the inverse of the Fischer Information

[4] Estimation of the parameters d and σ^2 without contaminated data is a simpler problem, but since most economic time series suffer from either aggregation problems or measurement error the assumption of additive disturbance is a much more realistic assumption. In addition, the assumption of contaminated data adds credence to using our approach of estimating d over that of the GPH method since the GPH method has never been tested with contaminated data.

The autocovariance of $y(t)$ is

$$\begin{aligned} R_y(t,s) &= E[y(t)y(s)] \\ &= \frac{\sigma_\epsilon^2 \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \int_{-\infty}^{\infty} b(\tau_1)\, d\tau_1 \\ &\quad \int_{-\infty}^{\infty} \frac{\Gamma(|\tau_2 - \tau_1 + s - t| + d)}{\Gamma(|\tau_2 - \tau_1 + s - t| + 1 - d)} b(\tau_2)\, d\tau_2 \end{aligned} \tag{13}$$

Since Eq. (13) is a function of only $|s-t|$ the filtered process $y(t)$ is stationary.

By Sheppard's formula the ratio of the Gamma functions will equal $|\tau_2 - \tau_1 + s - t|^{2d-1}$ for large values of $|\tau_2 - \tau_1 + s - t|$. The spectrum of $y(t)$ is then

$$S_y(\omega) = |B(\omega)|^2 \frac{1}{|\omega|^{2H+1}} \qquad \text{as } \omega \to 0 \tag{14}$$

where $H = d - 1/2$. If $B(\omega)$ is the frequency response for a ideal bandpass filter then $y(t)$ will have a finite variance. These two results satisfy Definition 1. Q.E.D.

Thus, fractionally integrated series are $1/f$ processes, as defined in Definition 1, who's filtered processes satisfy the results of Theorem 1 with the differencing parameter d related to γ by $d = \gamma/2$.

4 Wavelet Analysis of $I(d)$

Now that we have shown that fractionally integrated series are $1/f$ processes we use wavelet analysis of $I(d)$ processes to calculate the differencing parameter d. Wornell (1993) showed that if a $1/f$ process, as defined in Definition 1, is projected onto an orthonormal wavelet basis with $R \geq 1$ vanishing moments the wavelet coefficients at each scale will have a variance of $\sigma^2 2^{-\gamma m}$ if $0 < \gamma < 2R$. Applying Wornell's result to $I(d)$ processes with d restricted to $(-0.5, 0.5)$, we find that the wavelet coefficients of a fractionally integrated process have a variance of $\sigma^2 2^{-2dm}$.[3] Knowledge of the wavelet coefficient's distribution for $1/f$ processes provides the foundation from which parameter estimates of γ, d, and σ can be calculated with regular maximum likelihood techniques.

[3] Wornell's result can be extended to include $\gamma < 0$ provided enough vanishing moments exist in the wavelet. This dispels any problem with a fractional differencing parameter that is less than zero.

those statistically self-similar random processes that are stationary when filtered by an ideal bandpass filter.

We use the following definition and theorem of $1/f$ processes by Wornell (1993) to prove that fractionally integrated series are $1/f$ processes.

Definition 1 *A wide-sense statistically self-similar zero-mean random process $x(t)$ shall be said to be a* **1/f *process*** *if there exists an ω_0 and ω_1 satisfying $0 < \omega_0 < \omega_1 < \infty$ such that when $x(t)$ is filtered by an ideal bandpass filter with frequency response*

$$B_1(\omega) = \begin{cases} 1, & \omega_0 < |\omega| \leq \omega_1 \\ 0, & otherwise \end{cases} \tag{10}$$

the resulting process $y_1(t)$ is wide-sense stationary and has finite variance.

Theorem 1 *A $1/f$ process $x(t)$, when filtered by an ideal bandpass filter with frequency response*

$$B_1(\omega) = \begin{cases} 1, & \omega_L < |\omega| \leq \omega_H \\ 0, & otherwise \end{cases} \tag{11}$$

for any $0 < \omega_L < \omega_H < \infty$, yields a wide-sense stationary random process $y(t)$ with finite variance and having power spectrum, for some σ_x^2

$$S_y(\omega) = \begin{cases} \sigma_x^2/|\omega|^\gamma, & \omega_L < \omega < \omega_H \\ 0, & otherwise \end{cases} \tag{12}$$

where the spectral exponent γ is related to the self-similarity parameter H according to $\gamma = 2H + 1$.

Proof: See Wornell (1993) p. 1446.

Theorem 2 *Fractionally integrated processes, I(d), with $-0.5 < d < 0.5$, are $1/f$ processes in the sense of Definition 1.*

Proof: Let $x(t)$ be a fractionally integrated process with $-0.5 < d < 0.5$, zero mean and autocovariances defined by Eq. (9), and define $b(t)$ to be the regular finite-energy impulse response function for any LTI filter who's frequency response $B(\omega) = 0$. Define the output of the filter applied to $x(t)$ to be

$$y(t) \quad = \quad \int_{-\infty}^{\infty} b(t-\tau)x(\tau)d\tau.$$

many times series have spectrums that increase without bound as the level of frequency approaches the origin, but when differenced this unbounded spectrum equals zero. Such behavior occurs in models with power spectrums of the form

$$S(\omega) \sim 1/|\omega|^{2d} \qquad \text{as } \omega \to 0. \tag{8}$$

In addition, by allowing d to be a real number a wide variety of processes with different shaped spectrums are made available.

From Eq. (8) Mandelbrot and Van Ness' (1968) statistical self- similarity property is evident. For every a, $S_x(\omega) = |a|^{2d} S_x(a\omega)$, i.e. the statistical properties of the series x_t remain the same regardless of expanding or contracting the time intervals. This property of $I(d)$ will be needed in the next subsection.

The autocovariance function for Eq. (7) is

$$\begin{aligned} R_x(t,s) &= E[x(t)\,x(s)] \\ &= \frac{\sigma_\epsilon^2 \Gamma(1-2d)\Gamma(|t-s|+d)}{\Gamma(d)\Gamma(1-d)\Gamma(|t-s|+1-d)} \end{aligned} \tag{9}$$

which is proportional to $a|t-s|^{2d-1}$ for some constant a as $|t-s| \to \infty$. The statistical self-similarity parameter derived from Eq. (9) is $H = d - 1/2$ since $R(t,s) = a^{-2H} R(at, as)$. Eq. (9) also shows the interesting property of long-term memory found in fractional integrated models. When $d > 0$ the fractional differenced model's autocovariance decays logrythmically as opposed to the exponential decay associated with short-term processes. This same slow decay is evident in the autocovariance when $d < 0$, however, instead of having positive long-term dependance the process has negative dependance.

3.1 1/f Processes

The properties and characteristic of a fractionally integrated series are the same as those found in many $1/f$ processes. Some of the more commonly known $1/f$ processes are fractional Brownian motion and fractional Gaussian noise [Mandelbrot and Van Ness (1968)]. In this subsection we desire to show that the fractionally integrated series is also a member of the $1/f$ family. We use a new notion of the $1/f$ process introduced by Wornell (1993) that draws on wavelet analysis and ideal bandpass filters to show that $1/f$ processes are

If we normalize the time interval of the time series to the unit interval the translation parameters n will be a function of the scaling parameter m. This is illustrated by the support of the wavelet $[(n-1)2^{-(m-1)}, n2^{-(m-1)}]$. If $m = 1$ then $n = 1$ will cover the entire time domain, and at the other extreme $m = max$, $n = 1, 2, \ldots, 2^{max-1}$ is needed to cover the unit interval. Hence, the translation parameter need only take on the values $n = 1, 2, \ldots, 2^{m-1}$ for a given scaling parameter, m, in order for the wavelet to cover the time series. In general, the wavelet coefficients for a time series of length $N2^{max}$ will consist of

$$\left\{ < x, \phi_{m,n(m)} > : m \in \{1, \ldots, max\},\ n(m) \in \{1, 2, \ldots, 2^{m-1}\} \right\}.$$

For future reference let $\mathcal{M} = \{1, 2, \ldots, max\}$ and $\mathcal{N}(m) = \{1, 2, \ldots, 2^{m-1}\}$.

2.2 Example

One of the simplest and well know wavelet is the Haar function

$$\psi(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The subspaces V_m associated with this wavelet contain $L^2(\Re)$ signals that are piecewise constant on intervals of length 2^{-m}. This means that the scaling function, $\phi(t)$, is the indicator function $\chi_{[0,1)}$ since it is constant over the unit interval.

3 Fractionally Integrated Series

Let $x(t)$ represent a time series with the form

$$(1-L)^d x(t) = \epsilon(t) \tag{7}$$

where $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and L is the lag operator. The time series $x(t)$ is an integrated series with differencing parameter d, denoted as $I(d)$, where d is normally considered to be an integer. If d is an integer differencing the series d times results in a stationary process with a nearly flat spectrum.

Granger and Joyeux (1980) and Hosking (1981) assume that the differencing parameter d can also take on non-integer values, in particular $d \in (-0.5, 0.5)$. Their reasoning is that

Let W_m be the orthogonal complement of V_m in V_{m+1}, i.e. $V_{m+1} = V_m \otimes W_m$. W_m can be viewed as what is added to a function in V_m so that twice as much of the function's detail can seen. Translations of the mother wavelet $\psi(t) = 2\sum_n (-1)^n a_{n+1}\phi(2t+n)$ will form an orthogonal basis for W_0 and dilations of the mother wavelet $\{\psi_{m,n} : n \in \mathbf{Z}\}$ an orthogonal basis for W_m. It is easily shown that $L^2(\Re) = \overline{\bigcup_{m\in\mathbf{Z}} V_m}$, and $\{0\} = \bigcap_{m\in\mathbf{Z}} V_m$. Combining the definition of W_m with these two properties results in $L^2(\Re) = \otimes_{m\in\mathbf{Z}} W_m$ and since the W_m's are mutually orthogonal $\{\psi_{m,n} : m,n \in \mathbf{Z}\}$ will form an orthogonal basis for $L^2(\Re)$. Thus, if $x(t) \in L^2(\Re)$ then

$$x(t) = \sum_m \sum_n < x, \psi_{m,n} > \psi_{m,n}(t) \tag{4}$$

and is completely characterized by the wavelet coefficients $\{< x, \psi_{m,n} >\}_{m,n\in\mathbf{Z}}$.

Since time series are not continuous and have a limited number of observations the above wavelet decomposition can be reduced by dropping the coarser and finer resolution spaces. For example, suppose that the time series $x(t)$ consists of $N = 2^{max}$ observations. If the coarses resolution desired is 2^{min}, where $0 \leq min < max$, $x(t)$ can be projected onto V_{min} and $L^2(\Re) = V_{min} \otimes \sum_{m>min} W_m$. At this level of resolution the series $x(t)$ can be represented by the linear combination of the translates of the scaling function at scale $m = min$ and translates of the wavelet function for scales greater than min as

$$x(t) = \sum_n < x, \phi_{min,n} > \phi_{min,n}(t) + \sum_{m\geq min} \sum_n < x, \psi_{m,n} > \psi_{m,n}(t) \tag{5}$$

Furthermore, for resolutions finer than 2^{max} the wavelet coefficients are just trivial since no additional information on $x(t)$ is available in between the signal's observations. This allows us to discard translates of the wavelet with scales larger than max.

This representation of $L^2(\Re)$ can be interpreted by the Laplacian pyramid data structure [Burt and Adelson (1983)] in the frequency domain. For each scale the wavelet coefficients of $x(t)$ decomposes the signal into a set of independent frequency channels. The first level of the pyramid is the projection of $x(t)$ onto V_{min} i.e. $\{< x, \phi_{min,n} >\}_{n\in\mathbf{Z}}$. Each subsequent level is the projection of the signal onto W_m, i.e. $\{< x, \psi_{m,n} >\}_{n\in\mathbf{Z}}$ for $m = min+1, \ldots, max$. Because of the orthogonality of the wavelet each coefficient in the pyramid will be independent.

$\pm 2^m\omega_0$ in frequency. Hence, by increasing and decreasing m and n the wavelet covers different levels of frequency and is shifted over the time horizon. At higher frequencies (larger m) the translation steps, $2^{-m}n$, are smaller since the tighter wavelet requires shorter translation steps to cover the entire time scale, as opposed to large translation steps for low frequencies.

2.1 Multiresolution Analysis

With multiresolution analysis Mallat (1989) has shown that a linear combination of dilations and translation of the wavelet forms an orthogonal basis of the set of squared integrable functions, $L^2(\Re)$.[2] Multiresolution analysis decomposes $L^2(\Re)$ into a chain of closed subspaces $\cdots \subset V_{m-1} \subset V_m \subset V_{m+1} \subset \cdots$ where V_m consists of $L^2(\Re)$ functions with resolution 2^m. V_m can be thought of as a set containing functions who's details smaller than 2^{-m} can not be seen. This characteristic of the subspaces V_m allows us to go back and forth among the subspaces by rescaling, i.e. $x(t) \in V_m$ if and only if $x(2^{k-m}t) \in V_k$.

If we define $\phi(t)$ to be the scaling function with the property

$$\phi(t) = 2\sum_n a_n \phi(2t - n) \tag{2}$$

where $\int \phi(t)dt = \sum_n a_n = 1$, then ϕ is an orthonormal basis of V_0, and by the scaling property of multiresolution analysis $\{\phi_{m,n}(t) = 2^{m/2}\phi(2^m t - n) : n \in \mathbf{Z}\}$ forms an orthonormal basis of V_m. Hence, projections of $x(t) \in L^2(\Re)$ onto V_m, i.e. 2^m resolution approximation of $x(t)$, can be represented as

$$Proj_{V_m}\, x(t) = \sum_n < x, \phi_{m,n} > \phi_{m,n}(t) \tag{3}$$

where $< x, \phi_{m,n} >= \int_{-\infty}^{\infty} x(t)\phi_{m,n}(t)dt$. The projection is totally characterized by the set of inner products $\{< x, \phi_{m,n} >\}_{n\in\mathbf{Z}}$.

Since the orthonormal basis $\{\phi_{m,n}\}$ has distinct elements that are not orthogonal to those contained in the basis for $m+1$, a linear combination of dilations and translations of the scaling function will not form a orthonormal basis for $L^2(\Re)$. However, the additional properties of multiresolution analysis enables us to utilize the wavelet to obtain such a basis.

[2] In this section we limit ourselves to signals in $L^2(\Re)$. However, wavelets have be applied to other function spaces like Sobolev and Lipschitz spaces (See Meyer (1990) and Mallat and Hwang (1991)).

of computer design that addresses the problems of graphing continuous functions in the discrete world of computers, and for our purpose provides a nice pedagogical tool for introducing wavelets. Section 3 defines both fractionally integrated and $1/f$ processes and proves that the class of fractionally integrated models is a member of the large family of $1/f$ processes. Section 4 contains the asymptotic properties of the wavelet's maximum likelihood estimate of the fractional differencing parameter, along with an extensive array of simulations showing the robustness of these estimates to sample size, differencing parameter, and wavelet type. Lastly, in Section 5 we estimate the fractional differencing parameter of the Divisia monetary aggregates with wavelets to determine if monetary data is a fractionally integrated process.

2 Wavelet Theory

A wavelet is defined as

$$\begin{aligned} \psi_{m,n}(t) &= a_0^{m/2}\psi\left(a_0^m(t - nb_0a_0^{-m})\right) \\ &= a_0^{m/2}\psi\left(a_0^m t - nb_0\right) \end{aligned} \tag{1}$$

where $a_0 > 0$, $b_0 > 0$, and m and n are respectively the dilation and translation parameters that are elements of $\boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$. The function $\psi(t)$ is often referred to as the mother wavelet and must satisfy the admissability condition $\int \psi(t)dt = 0$. The admissability condition is a necessary condition that insures that the wavelet possesses the properties of smoothness and localization in the frequency and time domain. To strengthen these properties additional regularity conditions can be placed on the wavelet such as increasing the number of vanishing moments of the wavelet, i.e. $\int x^{(r)}\psi(t)dt = 0$ for $r = 0, 1, 2, \ldots, R$, and/or constructing $\psi(t)$ to be a C^k-function.[1] As commonly found in the wavelet literature we set $a_0 = 2$ and $b_0 = 1$, i.e. $\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$.

The strength of the wavelet as discussed earlier are now much more apparent. Suppose that $\psi(t)$ is well localized around zero in time and $\pm\omega_0$ in frequency. The dilated and translated wavelets $\psi_{m,n}(t)$ will also be well localized around $2^{-m}n$ in time and around

[1] Daubechies(1988) shows that one can construct a wavelet in C^k, $k = 0, 1, \ldots$ which will also have $K \geq k$ vanishing moments. Such wavelets are referred to as Daubechies wavelets with order k.

1 Introduction

Wavelet analysis is a recent development in the area of applied mathematics. They were first applied in the area of seismology [Morlet (1983)], but because of the wavelet's simplistic nature they have been used in a number of areas ranging from signal analysis [Kronland- Martinet, Morlet and Grossman (1987)] to numerical analysis [Beylkin, Coifman, and Rokhlin (1991)]. As to our knowledge wavelet analysis has not been applied in a true economic context.

The strength and usefulness of the wavelet rests on its ability to localize a process in frequency-time space. The properties of the wavelet enables it to zoom in on a processes' behavior at a certain frequency level with a time interval that is a function of the frequency. As a result of this time-frequency relationship at high frequency levels the wavelet is tight in shape and is able to focus on short lived phenomenas, while at lower frequencies the wavelet has a broader time window. This behavior is a desirable one when analyzing nonstationary time series since the series' wavelet transform provides frequency information that is localized in time and will also vary over time.

In this paper we apply wavelet analysis to the class of fractionally integrated processes to produce an alternative method of estimating and testing the fractional differencing parameter. In a manner similar to Geweke and Porter-Hudak (1983) we develop a nonparametric method of estimating the fractional differencing parameter. However, unlike the GPH approach which uses the periodogram of the series to approximate the series' spectrum, we draw on the distribution properties of the wavelet coefficients for a $1/f$ processes to calculate a maximum likelihood estimate (MLE) of the fractional differencing parameter. As a result our method does not require the strong assumption of the GPH approach that the normalized periodogram ordinates are asymptotically distributed exponential [Hurvich and Beltrao (1993)]. Furthermore, estimating the fractional differencing parameter with the wavelet does not suffer from the subjective choice as to how many of the frequency points are to be used in the GPH regression.

The paper is arranged as follows. In Section 2 we provide an introduction to wavelet analysis through the area of multiresolution analysis. Multiresolution analysis is an area

Wavelet Analysis of Fractionally Integrated Processes

Mark J. Jensen
Department of Economics, Washington University
St. Louis, MO 63130

May 24, 1994

Abstract

In this paper we apply wavelet analysis to the class of fractionally integrated processes to show that this class is a member of the $1/f$ family of processes as defined by Wornell (1993) and to produce an alternative method of estimating the fractional differencing parameter. Currently the method by Geweke and Porter-Hudak (1983) is used most often to estimate and test the fractional differencing parameter. The GPH approach, however, has been shown to have poor statistical properties and suffers from subjective decisions that the users must make. The wavelet analysis estimate of the fractional differencing parameter is shown to be more straightforward and to provide results that are more robust than the GPH method.

Keywords: Long-Memory, Wavelets, Spectral Analysis, $1/f$ Processes.