Fractional Differencing in Discrete Time

John Elder

Department of Finance & Real Estate

College of Business

Colorado State University

Fort Collins, CO, USA 80523-1272

email: john.elder@colostate.edu

Robert J. Elliott

Haskayne School of Business

University of Calgary

Calgary, AB, Canada T2N1N4

and School of Mathematics

University of Adelaide

SA 5005, Adelaide, Australia

email: relliott@ucalgary.ca

Hong Miao

Department of Finance & Real Estate

College of Business

Colorado State University

Fort Collins, CO, USA 80523

email: hong.miao@colostate.edu

July 23, 2011

Abstract

This paper consists of two parts, a theoretical followed by an empirical contribution. We first give a new framework for fractional differencing in discrete time and show how the definition of fractional differencing that is commonly employed in empirical financial applications arises as a special case. We then use these methods to estimate the fractional differencing parameter in the return and volatility for three Comex metal futures contracts. The metal futures are sampled at very high frequencies - five minute intervals over a nearly eight year period.

1

1   Introduction

There is a large and growing theoretical and empirical literature on fractional differencing which provides a useful tool for parsimoniously modeling the persistence observed in many financial time series (e.g., Lo (1991); Campbell, Lo and Mackinlay (1997)). Fractional differencing was introduced in terms of an infinite filter that corresponds to the expansion of $(1-B)^d$ in Granger and Joyeux (1980) and by analogy with the continuous time fractional white noise in Hosking (1981). The definition in Granger and Joyeux (1980) and Granger (**1980**) is a modification of the original definition in Diaz and Osler (1974), which, as noted by Gray and Zhang (1988), has no general exponential law.

The first part of this paper consists of a theoretical contribution. A new framework for fractional differencing is defined by introducing some basic concepts in a discrete time framework. We define an algebra of operators and show it has no zero divisors. Consequently it can be embedded in a division algebra. All fractional powers of the summation operator are then naturally defined and the form of fractional differencing most often utilized in financial applications is a special case of our definition.

The second part of the paper consists of an empirical application. Using wavelet methods this investigates the estimation of the fractional differencing parameter in the high frequency return and volatility of three Comex metals futures contracts: gold, copper and silver. Evidence on fractional integrating dynamics in commodity futures is relevant for several reasons, including statistical tests of asset pricing models, forecasting, hedging, and related risk management techniques in In and Kim (2006), Dark (2007) and Elder and Jin (2009). The existing empirical evidence for fractional differencing in commodity futures, however, is mixed. Cheung and Lai (1993) find little evidence of fractional differencing in gold returns and Crato and Ray (2000) find little or no evidence of fractional differencing in seventeen commodity futures returns, while Panas (2001) finds evidence of fractional differencing in two of six metals futures contracts. More recently, Elder and Jin (2009) find little evidence of fractional differencing in metal futures, but there is evidence of long memory in about half of agricultural commodity futures, based on daily data over a thirty year period.

In our empirical application, we estimate the fractional differencing parameter in the

return and volatility to gold, copper and silver futures that are sampled at five minute intervals over a nearly eight year period from April 1, 2000 through December 31, 2008 **yielding more than 65,000 observations**. We use four methods to estimate the fractional differencing parameter. Two methods are based on frequency domain analysis, as developed in Geweke and Porter-Hudak (1983) (GPH) and Robinson (1995**a**) (GSP), and two methods are based on wavelet analysis, as developed by Wornell and Oppenheim (1992), Jensen (1999) and Jensen (2000). Our empirical results suggest some evidence of antipersistence in gold and copper returns, and strong evidence of fractional differencing in each of the volatility series, as measured by absolute returns.

2   Basic Concepts

$\mathbb{Z}$ will denote the group of integers with the discrete topology. $\mathbb{N}$ will denote the semi-group of non-negative integers, $\mathbb{N} = \{0, 1, 2, \cdots\}$. We shall consider the set of functions $\mathcal{F} = \{f\}$ whose support is contained in $\mathbb{N}$. That is, if $j < 0$ then $f(j) = 0$. These can be thought of as infinite sequences: $f(0) = a_0$, $f(1) = a_1, \cdots, f(n) = a_n, \cdots$, etc.

Definition 2.1 *If $f_1 \in \mathcal{F}$, and $f_2 \in \mathcal{F}$, their convolution product $f_1 * f_2 \in \mathcal{F}$, is the function defined by:*

$$(f_1 * f_2)(n) = \sum_{i=0}^{\infty} f_1(n - i) f_2(i) = \sum_{i=0}^{\infty} a_{n-i} b_i,$$

where $f_1(i) = a_i$, and $f_2(i) = b_i$.

Note that $f_1 * f_2 = f_2 * f_1$. If $f \in \mathcal{F}$ and $\sum_{i=0}^{\infty} |f(i)| < \infty$, its Fourier transform is $\hat{f}(y) = \sum_{n=0}^{\infty} f(n) e^{iny}$. Then $\widehat{f_1 * f_2}(y) = \hat{f}_1(y) \cdot \hat{f}_2(y)$ for $f_i \in \mathcal{F}$ and all $y \in \mathbb{R}$.

Lemma 2.2 *With convolution as multiplication, addition defined by $(f_1 + f_2)(i) = f_1(i) + f_2(i)$, and scalar multiplication defined by $(\lambda f_1)(i) = \lambda f_1(i)$, the functions $\mathcal{F}$ form an associative, commutative algebra with a unit.*

Proof. Immediate. The unit is the function: $f(0) = 1$, and $f(i) = 0$, if $i \neq 0$. The zero is the function $f(i) = 0$, for all $i \in N$. $\qquad\qquad\square$

Remark 2.3 *Care must be taken not to confuse the constant $\lambda \in \mathbb{C}$, and the constant function $f(i) = \lambda \in \mathcal{F}$, for all $i \in \mathbb{Z}^+$.*

Notation 2.4 $u \in \mathcal{F}$ *denotes the function* $u(i) = 1,$ *for all* $i \in \mathbb{Z}^+,$ *that is: the sequence* $(1, 1, 1, \cdots).$

Remark 2.5

$$(u * f)(n) = \sum_{i=0}^{n} f(i) = S_n.$$

We shall now prove that if $f_1, f_2 \in \mathcal{F}$, and $f_1 * f_2 = 0$, then either $f_1 = 0$, or $f_2 = 0$.

Notation 2.6 *Write*

$$(x)^{(r)} = x(x-1)(x-2)\cdots(x-r+1),$$
$$(x)^{(0)} = 1, \quad (0)^{(r)} = 0 \text{ for } r \neq 0.$$

Note that if $n \in \mathbb{Z}^+$ *and* $r > n$, *then* $(n)^r = 0.$

We now prove the discrete form of Lerch's Theorem:

Theorem 2.7 *For* $f \in \mathcal{F}$, *if*

$$\sum_{i=0}^{M} (i)^{(n)} f(i) = 0, \tag{2.1}$$

for $n = 0, 1, \cdots M$, *then* $f(i) = 0,$ *for* $i = 0, 1, \cdots M.$

Proof. As equation (2.1) is true when $n = M$, and $(M)^{(M)} = M! \neq 0$, while $(i)^{(M)} = 0$ for $i < M$, we have:

$$M! f(M) = 0 \text{ so } f(M) = 0.$$

Similarly, when $n = M - 1$, as $f(M) = 0$,

$$(M-1)^{(M-1)} f(M-1) = 0. \text{ Therefore, } f(M-1) = 0,$$

and so on. Continuing M finite steps we have $f(i) = 0$ for $i = 1, 2, \cdots M$.
Finally as $(i)^{(0)} = 1, \forall i,$

$$\sum_{i=0}^{M} (i)^{(0)} f(i) = 0 \Rightarrow f(0) = 0.$$

$\square$

Theorem 2.8 *Suppose* $f * f = 0$, *that is* $\sum_{i=0}^{M} f(n-i) f(i) = 0$ *for all* n. *Then* $f(n) = 0$ *for all* n.

Proof.

$$f * f(0) = f(0) f(0) = 0 \Rightarrow f(0) = 0.$$

Suppose $f(i) = 0$, for $i = 1, 2, \cdots N - 1$. Consider

$$
\begin{aligned}
f * f(2N) &= \sum_{i=0}^{2N} f(2N-i) f(i) \\
&= f(N) f(N) = 0.
\end{aligned}
$$

Therefore implying $f(N) = 0$.

Consequently, $f(i) = 0$ for all i by induction. $\qquad\square$

Theorem 2.9 *if* $f * g = 0$, *then* $f = 0$ *or* $g = 0$.

Proof. $f * g = 0$ means for any $M \geq 0$

$$\sum_{i=0}^{M} f(M-i) g(i) = 0.$$

Therefore,

$$
\begin{aligned}
\sum_{i=0}^{M} (M-i) f(M-i) g(i) &+ \sum_{i=0}^{M} i f(M-i) g(i) \\
= M \sum_{i=0}^{M} f(M-i) g(i) &= 0.
\end{aligned}
\tag{2.2}
$$

Write $f'(i) = i f(i)$, and $g'(i) = i g(i)$. Then from (2.2)

$$f' * g + f * g' = 0.$$

so

$$
\begin{aligned}
f * g' * (f' * g + f * g') \\
= f * g' * f' * g + (f * g')^2 \\
= (f * g) * (f' * g') + (f * g')^2 \\
= 0.
\end{aligned}
$$

5

By hypothesis: $f * g = 0$. **Consequently,** we have shown $(f * g')^2 = 0$. By *Theorem 2.8*, $(f * g') = 0$. That is, for any $M \in \mathbb{Z}^+$,

$$\sum_{i=0}^{M} f(M-i) \times i \times g(i) = 0. \tag{2.3}$$

That is:

$$\sum_{i=0}^{M} (i)^{(1)} f(M-i) g(i) = 0.$$

Therefore, repeating the above argument

$$\sum_{i=0}^{M} f(M-i) \times i^2 \times g(i) = 0. \tag{2.4}$$

Subtracting (2.3) from (2.4), we have, for $M \in \mathbb{Z}^+$:

$$\sum_{i=0}^{M} (i)^{(2)} f(M-i) g(i) = 0.$$

Suppose we have shown for $M \in \mathbb{Z}^+$:

$$\sum_{i=0}^{M} (i)^{(n-1)} f(M-i) g(i) = 0. \tag{2.5}$$

Then by the above construction we have

$$\sum_{i=0}^{M} i \times (i)^{(n-1)} f(M-i) g(i) = 0, \tag{2.6}$$

Subtracting $(n-1) \times (2.5)$ from (2.6), we have

$$\sum_{i=0}^{M} (i)^{(n)} f(M-i) g(i) = 0. \tag{2.7}$$

By induction (2.7) is true for all n, and so in particular for $n = 0, 1, \cdots, M$.

By *Theorem 2.7,*

$$f(M-i) g(i) = 0, \forall i = 0, 1, \cdots, M. \tag{2.8}$$

This is true for any $M \in \mathbb{N}$. Suppose $g(k) \neq 0$ for some $k \in \mathbb{Z}^+$. Then from (2.8), we see $f(i) = 0$ for all i. Otherwise $g(i) = 0$, for all i. □

Definition 2.10 *From Theorem 2.8 we see that convolution multiplication in $\mathcal{F}$ has no zero divisors, so we can form its formal quotient field Q. The members of Q will be called operators. $\mathcal{F}$ can be embedded in its quotient field. We have seen the unit of the quotient field is $\{1\} = \{1, 0, 0, \cdots\}$. Q is formed by pairs (f, g), where $f, g \in \mathcal{F}$.*

3 Operators

3.1 The Summation Operator u

As above, consider $f \in \mathcal{F}$ as a sequence $(a_0, a_1, \cdots, a_i, \cdots)$. The operator u is defined as the sequence

$$u = (1, 1, \cdots, 1, \cdots).$$

Now for any $f \in \mathcal{F}$,

$$(u * f)(n) = a_0 + a_1 + \cdots + a_n = S_n,$$

so u is the integral or summation operator.

Powers of u are:

$$\{u\} = (1, 1, \cdots, 1, \cdots),$$
$$\{u^2\} = u * u = (1, 2, 3, \cdots),$$
$$\{u^3\} = u * u^2 = (1, 3, 6, \cdots),$$
$$\{u^4\} = u * u^3 = (1, 4, 10, \cdots),$$
$$\cdots$$

If we write $C_r^n = \dfrac{n!}{r!(n-r)!}$, then

$$\{u\} = \left(C_0^1, C_0^1, \cdots, C_0^i, \cdots\right),$$
$$\{u^2\} = \left(C_1^1, C_1^2, \cdots, C_1^i, \cdots\right),$$
$$\{u^3\} = \left(C_2^2, C_2^3, \cdots, C_2^i, \cdots\right),$$
$$\{u^4\} = \left(C_3^3, C_3^4, \cdots, C_3^i, \cdots\right),$$
$$\cdots$$

So we see for general $n \in \mathbb{Z}^+ - \{0\}$,

$$u^n = \{f_n(i)\} = \left\{C_{n-1}^{n+i-1}\right\} = \left\{\frac{\Gamma(n+i)}{i!\,\Gamma(n)}\right\}.$$

That is, for positive integral n, u^n is the series:

$$\left\{1, \frac{n}{1!}, \frac{n(n+1)}{2!}, \cdots\right\}.$$

We can give the following definition:

Definition 3.1 *For general $r \in \mathbb{R}$, $\{u^r\}$ is the series:*

$$\left\{\frac{(r+1)^{(i)}}{i!}\right\} = \left(1, r, \frac{r(r+1)}{2!}, \frac{r(r+1)(r+2)}{3!}, \cdots\right).$$

and $\{u^0\} = (1, 0, 0, \cdots) = \{I\}$ is the identity operator, as $u^0 * f = f * u^0 = f$, for all $f \in \mathcal{F}$.

3.2   The Difference Operator Δ

Definition 3.2 *We define Δ to be the operator on $\mathcal{F}$, such that, for*

$$f = (f(0), f(1), f(2), \cdots, f(n), \cdots) \in \mathcal{F},$$
$$\Delta * f(n) = (f(n) - f(n-1)).$$

That is, $\Delta * f$ is the series

$$(f(0), f(1) - f(0), f(2) - f(1), \cdots) \in \mathcal{F}.$$

So, Δ is the difference operator.

Theorem 3.3 $\Delta = \{u^{-1}\}.$

Proof. By Definition 3.1, $\{u^{-1}\}$ is the series $(1, -1, 0, 0, \cdots)$. For any

$$f = (f(0), f(1), f(2), \cdots, f(n), \cdots) \in \mathcal{F},$$
$$u^{-1} * f = (f(0), f(1) - f(0), f(2) - f(1), \cdots) = \Delta * f.$$

That is

$$u^{-1} * f = \Delta * f, \ \forall f \in \mathcal{F}.$$

So

$$\left(u^{-1} - \Delta\right) * f = 0, \forall f \in \mathcal{F}.$$

Since $\mathcal{F}$ has no zero divisors, we have $\Delta = \{u^{-1}\}.$ $\qquad\square$

Remark 3.4

$$\Delta^2 * f = (f(0), f(1) - 2f(0), f(2) - 2f(1) + f(0), \cdots).$$

etc.

Definition 3.5 *For arbitrary $r \in \mathbb{R}$, we define*

$$\Delta^r = u^{-r} = \left(1, -r, \frac{-r(-r-1)}{2!}, \cdots\right).$$

3.3 The Operators L and L^{-1}

Definition 3.6 $L \in \mathcal{F}$, *is the function:* $L(0) = 0$, $L(1) = 1$, *and* $L(k) = 0$ *for all* $k > 1$. *That is,*

$$L = (0, 1, 0, 0, \cdots).$$

Remark 3.7 *For any $f = (a_0, a_1, \cdots, a_i, \cdots) \in \mathcal{F}$, we have*

$$L * f = (0, a_0, a_1, \cdots, a_i, \cdots).$$

That is $L = L^1$ is a translation one place to the right, so operator L^n translates n places to the right. L is sometimes called the lag operator.

Remark 3.8 *Note that $\Delta = I - L = u^{-1}$.*

In order to introduce L^{-1}, we must extend our operations from the class $\mathcal{F}$ of functions, (that is the class of f such that $f(i)$ is zero for negative i.), to the class $\mathcal{G}$ of functions with limited support on the left. That is, $f \in \mathcal{G}$ if there is an $n \in Z$ such that $f(i) = 0$, if $i < n$.

We note that f is now defined for all $i \in \mathbb{Z}$.

Definition 3.9 *Given $f, g \in \mathcal{G}$,*

$$f * g(n) = \sum_{i=-\infty}^{\infty} f(n-i)g(i).$$

For each n this is a finite sum, so $f * g$ is well defined and is in $\mathcal{G}$.

Lemma 3.10 *For $f, g \in \mathcal{G}$, if $f(i) = 0$ for $i < m$, and $g(i) = 0$ for $i < n$, then $f(i) * g(i) = 0$ for $i < m + n$.*

Definition 3.11 *L^{-1} is the function $f \in \mathcal{G}$ such that $L^{-1}(-1) = 1$, and $L^{-1}(i) = 0$, otherwise. The support of L^{-1} is $\{-1\}$. However,*

$$L^{-1}L = LL^{-1} = u^0 = \{I\}.$$

Theorem 3.12 *The algebra $\mathcal{G}$ has no zero divisors.*

Proof. Suppose $f, g \in \mathcal{G}$, and $f * g = 0$, then there is an n, such that $f(i) = 0$, for $i < n$, and there is an m, such that $g(i) = 0$ for $i < m$. By the above lemma, $f(i) * g(i) = 0$ for $i < m + n$. Therefore,

$$L^{m+n} * f * g = 0, \text{ and is in } \mathcal{F}.$$

That is

$$L^n * f * L^m * g = 0.$$

Then either $L^n * f = 0$ or $L^m * g = 0$. But trivially, if $L^r * h = 0$ then $h = 0$. Consequently $f = 0$, or $g = 0$. □

Thus, we can embed $\mathcal{G}$ in its quotient field $\mathcal{R}$ which contains Q.

Using Fourier transforms:

$$\hat{\Delta}(y) = \hat{u}^{-1}(y) = 1 - e^{iy},$$

and

$$\hat{\Delta}^r(y) = \hat{u}^{-r}(y) = \left(1 - e^{iy}\right)^r.$$

3.4 The Fractional Differencing Parameter

In finance the most popular application of fractional differencing is for an autoregressive moving-average process. The process $x(t)$ is said to be an autoregressive moving-average process integrated of order d and denoted $ARIMA(p, d, q)$, (here, p is the order of autoregression and q is the order of moving average), then

$$\Phi(L) x(t) = (1 - L)^{-d}\phi(L)z(t) = \Delta^{-d}\phi(L)z(t) = u^d\phi(L)z(t), \tag{3.1}$$

where $z(t)$ is sequence of i.i.d. random variables with zero mean and variance σ_z^2. Also, $\Phi(L)$ and $\phi(L)$ are finite polynomials in the lag operator with roots outside the unit circle. u^d is defined in Definition 3.1.

It is immediately seen that our definition agrees with that in Hosking (1981) and Granger and Joyeux (1980), or in terms of the gamma function as

$$(1 - L)^{-d} = 1 + \sum_{j=1}^{\infty} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} L^j.$$

The autocorrelation function of such a process decays hyperbolically for nonzero values of d when $|d| < 1.0$. For $|d| \leq 0.5$, the process $x(t)$ is both stationary and invertible, while for $0.5 \leq d \leq 1$, the process exhibits infinite variance.

It is also a fact that the autocovariances satisfy:

$$\gamma_u = c_1 u^{2d-1}, \text{ as } j \to \infty, \text{ for } |c_1| < \infty, \tag{3.2}$$

and, assuming that the process has absolute continuous spectral distribution, so that

11

it has a spectral function $f(\lambda)$, **defined as**

$$f(\lambda) = \frac{1}{2\pi}\left(\gamma_0 + 2\sum_{u=1}^{\infty}\gamma_u \cos(\lambda u)\right).$$

It can also be proved that

$$f(\lambda) \approx c_2\lambda^{-2d}, \text{ as } \lambda \to 0^+, \text{ for } 0 < c_2 < \infty, \tag{3.3}$$

where the symbol "$\approx$" indicates that the ratio of the left hand side and the right hand side tends to 1**, as** $j \to \infty$ **in equation (3.2) and as** $\lambda \to 0^+$ **in equation (3.3). These conditions are not always equivalent. However, Yong (1974), and Zygmund (1995) give conditions under which both expressions are equivalent.**

In equation (3.1), the general forms of $\Phi(L)$ and $\phi(L)$ are as following:

$$\Phi(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p, \text{ and}$$
$$\phi(L) = \beta_0 + \beta_1 L + \cdots + \beta_q L^q,$$

where $\alpha_i, 0 \le i \le p$, and $\beta_j, 0 \le j \le q$, are constants.

Consequently in our notation, equation (3.1) is

$$(\Phi(L) * x)(t) = \left(\Delta^{-d} * \phi(L) * z\right)(t). \tag{3.4}$$

Taking Fourier transforms:

$$\hat{\Phi}(y)\hat{x}(y) = (1 - e^{iy})^{-d}\hat{\phi}(y)\hat{z}(y).$$

as $\Delta = 1 - L$,

$$\hat{\Delta}(y) = 1 - e^{iy},$$

and

$$\hat{\Phi}(y) = \alpha_0 + \alpha_1 e^{iy} + \cdots + \alpha_p e^{ipy}.$$

Now $\widehat{(1-L)}^{-d}(y) = (1 - e^{iy})^{-d}$. Therefore, formally,

$$\log\hat{\Phi}(y) + \log\hat{x}(y) = -d\log(1 - e^{iy}) + \log\hat{\phi}(y) + \log\hat{z}(y),$$

so

$$d = \frac{1}{\log(1 - e^{iy})} \left[\log \hat{\phi}(y) + \log \hat{z}(y) - \log \hat{\Phi}(y) - \log \hat{x}(y) \right]$$

$$= \frac{1}{\log(1 - e^{iy})} \log \frac{\hat{\phi}(y)\hat{z}(y)}{\hat{\Phi}(y)\hat{x}(y)}. \tag{3.5}$$

To implement this expression, parameter estimates for the α_i and β_i must be obtained and the Fourier transforms of z and x approximated by finite sums. This will be investigated in future work. **It could be possible that the existing methods of estimating the fractional differencing parameter could be applied to the general framework after some extension or modification. In the remainder of this paper, we provide an illustration of fractional differencing in this context by estimating and interpreting the fractional differencing parameter for financial time series on the futures prices of gold, copper and silver.**

4 Empirical Applications

In this section, we apply two wavelet based and two frequency domain estimators to the three metals futures, gold, copper and silver. The series are sampled at five minute intervals from April 1, 2000 through December 31, 2008.

4.1 Estimators of the Fractional Differencing Parameter

We utilize two wavelet-based estimators of the fractional differencing parameter d as described by Jensen (1999), Jensen (2000) and Wornell and Oppenheim (1992), and one frequency domain estimator. The two wavelet-based **methods and the frequency domain estimator applied here are** comparable to those employed by Elder and Jin (2007, 2009).

Let $\psi(t)$ be a real valued, square integrable function with finite oscillations that decrease to zero as t goes to $\pm\infty$. That is, let $\psi(t)$ satisfy $\int_{-\infty}^{\infty} \psi(t)dt = 0$ and $\int_{-\infty}^{\infty} \psi^2(t)dt = 1$. Then $\psi(t)$ is a wavelet if it also satisfies the admissibility condition described, for example, by Percival and Walden (2000). Let $\varphi(t)$ represent the associated orthogonal scaling function, which spans the spaces between the spaces spanned by the various scales of the wavelet function. The wavelet coefficients $[a_{m,n}]$ that link the wavelet function to the signal can be obtained by projecting the wavelet and scaling basis functions onto the signal $x(t)$

in the following manner,

$$x(t) = \sum_n c_n \varphi(t - n) + \sum_m \sum_n 2^{-\frac{m}{2}} a_{m,n} \psi(2^{-m} t - n),$$

where m and n are integer indices for the finite or infinite sum that scale the wavelet in size (m) and translate the wavelet in time (n).

The estimators of the fractional differencing parameter are derived from the property that the wavelet coefficients $a_{m,n}$ associated with a $ARFIMA(p, d, q)$ process $x(t)$ with $|d| \leq 0.5$ are distributed approximately $N(0, 2^{-2d(M-m)}\sigma^2)$ (Jensen (1999) and Jensen (2000)). Denoting the variance of the wavelet coefficients at scale m by $var(a_m, .) = 2^{-2d(M-m)}\sigma^2$, we can take logs to give

$$\ln var(a_m, .) = \ln \sigma^2 + d \ln \left[2^{-2d(M-m)} \right]. \tag{4.1}$$

A simple estimate of the fractional differencing parameter d can be obtained by applying OLS on equation (4.1), where $var(a_m, .)$ is estimated by the sample variance of the wavelet coefficients at scale index m.

An approximate MLE of the fractional differencing parameter, denoted $\hat{d}_{BWMLE}$, can be obtained by exploiting the multivariate normality of the wavelet coefficients of the data generating process and imposing the constraint that $cov\,(a_{m,n}, a_{j,k}) = 0$ (Jensen (2000) and Wornell and Oppenheim (1992)). In particular, let $\sigma^2_{m,n} = 2^{-2d(M-m)}\sigma^2$, then the approximate likelihood of the data is

$$L\left(d, \sigma^2, \sigma^2_\eta\right) = \prod_m \prod_n \frac{1}{\sqrt{2\pi\sigma^2_{m,n}}} \exp\left(-\frac{a^2_{m,n}}{2\sigma^2_{m,n}}\right).$$

Percival and Walden (2000) suggest trimming the largest and smallest scales prior to estimating equation (4.1) for the wavelet OLS estimator ($WOLS$). We trim the smallest scales in the approximation given by equation (4.1) to alleviate bias due to short-run dynamics. We also trim the largest scale, since the variance at the largest scale is estimated imprecisely. For the wavelet MLE ($BWMLE$), some authors suggest trimming only the small scales.

We also consider two frequency domain estimators, that proposed by Geweke and Porter-Hudak (1983) (denoted GPH) and Robinson's (1995b) Gaussian semi-parametric (GSP) estimator. **Intuitively, the frequency domain estimators capture the slope of the sample spectral density function, while the wavelet estimators exploit the structure of**

the wavelet variance at different scales for long memory processes.

The GPH estimator examines the slope of the sample spectral density near the origin. More formally, let

$$I(w_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} (y(t) - \bar{y}) e^{itw_j} \right|^2$$

denote the sample periodogram at the jth Fourier frequency of $y(t)$, where $w_j = 2\pi j/T, j = 1, 2, \cdots, J$. The GPH estimator of d is based on an OLS regression of the log periodogram on the (transformed) log frequency:

$$\ln [I(w_j)] = a + d \ln \left(2 \sin \left(\frac{w_j}{2} \right) \right)^{-2} + \varepsilon_j, \quad \text{for } j = 1, 2, \cdots, J.$$

The bandwidth parameter J is often set at $T^{0.5}$ to capture the behavior of the periodogram near the origin. The series $x(t)$ may be non-Gaussian, heteroskedastic, but with $d \in (-0.5, 0.5)$, plus certain other conditions, then $\hat{d}_{GPH}$ is consistent and asymptotically normal with $\sqrt{J} \left(\hat{d}_{GPH} - d \right) \rightarrow N \left(0, \frac{\pi^2}{24} \right)$ (Robinson (1995a), Hurvich, Deo and Brodsky (1998) and Kim and Philips (1999), Robinson and Henry (1999)). If $0.5 \leq d < 1$ that the GPH estimator is consistent, and it is asymptotic normality for $0.5 \leq d \leq 0.75$. The usual OLS standard errors are often used for inference (cf. Deo and Hurvich (2001)). We use the OLS standard errors in our application.

Another prominent frequency domain estimator is the Gaussian semiparametric (GSP) estimator proposed by Robinson (1995b). The GSP estimator maximizes the approximate frequency domain likelihood in which discrete averaging is carried out near frequency 0. In particular, the GSP estimator $\hat{d}_{GSP} = \arg \min_d R(d)$, where $R(d)$ is the concentrated local Whittle log-likelihood

$$R(d) = \ln \left[\frac{1}{J} \sum_{j=1}^{J} w_j^{2d} I(w_j) \right] - \frac{2d}{J} \sum_{j=1}^{J} \ln w_j \quad \text{for } \left(\frac{1}{J} + \frac{J}{T} \right) \rightarrow 0.$$

Robinson (1995b) shows that this estimator is consistent under relatively mild conditions and is more efficient than the GPH estimator for $d \in (-0.5, 0.5)$, with $\sqrt{J} \left(\hat{d}_{GSP} - d \right) \rightarrow N \left(0, \frac{1}{4} \right)$. **Further extensions of this procedure can be found in Velasco (1999), Velasco and Robinson (2000), Phillips and Shimotsu (2004), Shimotso and Phillips (2005), and Abadir et al. (2007). Some of those extensions (for instance, Phillips and Shimotsu, 2004) allow for estimation even in nonstationary regions.**

4.2 Data and Empirical Results

Our data consists of three metals futures, gold, copper and silver, which are traded on Comex, now a division of the New York Mercantile Exchange. The series are sampled at five minute intervals from April 1, 2000 through December 31, 2008, yielding 65536 observations. The futures prices are nearby futures prices, converted to continuously compounded returns at five minute intervals in basis points, $x(t)$, from the futures prices, $p(t)$, as $x(t) = 10,000 \times \ln \dfrac{p(t)}{p(t-1)}$. Volatility of the return series is calculated as the absolute value of the returns.

For the volatility series, it is well established that the autocovariances of $\ln(x^2(t))$, $x^2(t)$ and $|x(t)|$, display similar decay at long lags (cf. Harvey (1998)), justifying the use of semi-parametric estimators of the fractional differencing parameter on any of these matrixes. Wright (2002) suggests using absolute returns $|x(t)|$, based on Monte-Carlo evidence.

Summary statistics on the return and volatility series are presented in Table 1. The return series have small but significant arithmetic means, on the order of $0.10 - 0.20$ basis points per five minute interval. The returns also exhibit the negative skewness and excess kurtosis typical of financial data, and are correlated at long lags as would be consistent with fractional integration. The correlation at long lags in absolute returns is notably high, as evidenced by the $Q(20)$ statistic. The return and volatility series are plotted in Figure 1 at five minute intervals.

We first test the data for evidence of fractional integration against the null of a unit root. We use the procedure suggested by Lobato and Velasco (2007), which has been shown to be robust to unconditional heteroskedasticity in unrestricted forms by Kew and Harris (2009). The test suggested by Lobato and Velasco (2007) is a simple t-test that is asymptotically efficient against local alternatives. The test is implemented by a two-step procedure that relies on a preliminary estimate of the fractional differencing parameter, which we obtain by the local Whittle method. In the first step, the autoregressive parameters are estimated based on the fractionally differenced series. In the second step, the test equation is estimated, in which the differenced series is regressed on a filtered and rescaled regressors constructed from the estimated autoregressive parameters.

The results from this test are reported in Table 2 for the metal return and volatility series. According to Sephton (2009), the critical value at the 1% significance level for no trend test is about -2.37 and test statistics many times larger. For both re-

turns and volatilities, the null of unit root is strongly rejected at the 1% significance level. This result is particular interesting for the volatility series, which are typically highly persistent. The results suggest that the persistence is better captured by a long memory, rather than unit root, process.

The estimates of the fractional differencing parameters are reported in Table 3. For the return series, there is some evidence of antipersistence in gold and some evidence of persistence in copper. For gold, the GPH estimator for the fractional differencing parameter is negative and insignificant, while the GSP, $WOLS$ and $BWMLE$ estimators are negative and significant. The $WOLS$ estimator is negative and largest in magnitude, consistent with its large negative bias in Monte-Carlo studies (Jensen (1999), **Elder and Jubinski** (2010)).

For copper, three of the four estimates (GPH, GSP and $BWMLE$) are positive and significant, while the $WOLS$ estimator is again negative, but not significant, which is again in accordance with the large negative bias of the $WOLS$ estimator. The estimates of the fractional difference parameter for silver are each negative, although only the $WOLS$ estimate is significantly different from zero at conventional levels. Given the large negative bias of the $WOLS$ estimator, this result is more likely to be spurious.

In sum, there appears to be some modest evidence of antipersistence in gold returns and persistence in copper returns, according to both the GSP and $BWMLE$.

For the volatility series, each of the absolute return series displays evidence of very high persistence. Interestingly, each of the point estimates is greater than 0.5, indicating that the absolute return series may have infinite variance. The degree of persistence is greatest for gold return volatility, with the smallest point estimate being greater than 0.8 and the largest point estimate being greater than 0.9. Silver volatility displays the next highest degree of persistence, with the largest point estimate of 0.8. The largest point estimate for copper volatility is less than 0.7.

The very high level of persistence indicated for the volatility series in results in Table 3 is somewhat sensitive to the choice of bandwidth for the GPH and GSP estimators, and the number of trimmed scales for the wavelet estimators, as well as the choice of volatility transformation. For example, at a bandwidth of $T^{0.5}$, and trimming 8 small scales for wavelet estimators causes the point estimates to decline substantively, as reported in Table 3. Most of the point estimates in Table 4 are close to 0.5, indicating high persistence.

Different volatility transformations also tend to give somewhat different results. As an example, the Monte-Carlo results of Wright (2002) show that frequency domain estimators tend to be downward biased when volatilities are calculated based on squared returns,

rather than absolute returns. Accordingly, point estimates for the fractional differencing parameter based on squared returns are even lower, as reported in Table 5.

5 Conclusion

In this paper, we define fractional integration and difference in a natural way to facilitate a clear understanding of the concepts. Our definition is different from the popular used definition since by introducing convolution as multiplication, all our operators have well defined powers of any real order.

We then apply four approaches of estimating the fractional differencing parameter to three metal futures time series. We find that for the return series, there is some evidence of antipersistence in gold and some evidence of persistence in copper. However, for the volatility series, we find clear evidence of very high persistence.

References

[1] Abadir, K.M., Distaso, W., and Giraitis, L., 2007, Nonstationarity-extended local Whittle estimation. *Journal of Econometrics,* 141, 1353-1384.

[2] Campbell, J.Y., Lo, A.W., and MacKinlay, C., 1997. The econometrics of financial Markets, Princeton University Press, Princeton, New Jersey, USA.

[3] Cheung, Y., and Lai, K.S., 1993. Do gold market returns have long memory? *Financial Review*, 28, 181-202.

[4] Crato, N., and Ray, B.K., 2000. Memory in returns and volatilities of futures contracts, *Journal of Futures markets*, 20, 525-543.

[5] Deo, R.S., and Hurvick, C.M., 2001. On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, 17, 686-710.

[6] Diaz, J.B., and Osler, T.J., 1974. Differences of fractional order. *Mathematics of Computation*, 28, 182-201.

[7] Dark, J., 2007. Basis convergence and long memory in volatility when dynamic hedging with futures. *Journal of Financial and Quantitative Analysis*, 42, 1021-1240.

[8] Elder, J., and Jin, H.J., 2007. Long memory in commodity futures volatility: A wavelet perspective. *Journal of Futures Markets*, 27, 411-437.

[9] Elder, J., and Jin, H.J., 2009. Fractional integration in commodity futures returns. *Financial Review*, 44, 583-602.

[10] Elder, J., and Jubinski, D., 2010. Wavelet estimators of the fractional differencing parameter: some Monte-Carlo results and an application to individual stock returns and volatility, working paper, Colorado State University.

[11] Geweke, J., and Poerter-Hudak, S., 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4, 221-238.

[12] Granger, C.W.J., and Joyeux, R., 1980. An introduction to long memory time series and fractional differencing. *Journal of Time Series Analysis*, 1, 1-15.

[13] Granger, C.W.J., 1980. Long memory relationship and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227-238.

[14] Gray, H.L., and Zhang, N.F., 1988. On a new definition of the fractional difference. *Mathematics of Computation*, 50, 513-529.

[15] Harvey, A.C., 1998. Long memory in stochastic volatility. In: Forecasting Volatility in Financial Markets, Knight, J. and Satchell, S., (Eds) Butterworth-Heinemann, Oxford.

[16] Hosking, J.R., 1981. Fractional differencing. *Binometrika*, 68, 165-176.

[17] Hurvich, C.M., Deo, R.S., and Brodsky, J., 1998. The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis*, 19, 19–46.

[18] In, F., and Kim, S., 2006. The hedge ration and the empirical relationship between the stock and futures markets: a new approach using wavelet analysis. *Journal of Business*, 79, 799-820.

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

[20] Jensen, M.J., 2000. An alternative maximum likelihood estimator of long-memory precesses using compactly supported wavelets. *Journal of Economic Dynamics and Control*, 24, 361-387.

[21] Kew, H., and Harris, D., 2010. Impulse responses of fractionally integrated processes with long memory. *Econometric Theory*, 26, 1855-1861.

[22] Lo, A.W., 1991. Long-term memory in stock market prices. *Econometrica*, 59, 1279-1331.

[23] Lobato, I.N., and Velasco, C., 2007. Efficient Wald tests for fractional unit roots. *Econometrica,* 75, 575-589.

[24] Panas, E., 2001. Long memory and chaotic of prices on the London metal exchange. *Resources Policy*, 27, 235-246.

[25] Percival, P.B., and Walden A., 2000. Wavelet methods for time series analysis. Cambridge University Press, Cambridge.

[26] Phillips, P.C.B., and Shimotsu, K., 2004. Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics,* 32, 656-692.

[27] Robinson, P.M., 1995a. Log-periodogram regression for time series with long range dependence. *Annals of Statistics*, 23, 1048-1072.

[28] Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23, 1630-1661.

[29] Robinson, P.M., and Henry, M., 1999. Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory*, 15, 299–336.

[30] Shimotsu, K., and Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. *Annals of Statistics,* 33, 1890-1933.

[31] Sephton, P.S., 2009. Critical values for the augmented efficient Wald test for fractional unit roots. *Empirical Economics*, 37, 615-626.

[32] Velasco, C., 1999. Gaussian semiparametric estimation of nonstationary time series. *Journal of Time Series Analysis,* 20, 87-127.

[33] Velasco, C., and Robinson, P.M., 2000. Whitle pseudo maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* 95, 1229-1243.

[34] Wornell, G., and Oppenheim, A., 1992. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Precessing*, 40, 611-623.

[35] Wright, J.H., 2002. Log-periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. *Econometric Reviews*, 4, 397–417.

[36] Yong, C.H.,1974. Asymptotic behaviour of trigonometric series. Chinese University of Hong Kong, Hong Kong.

[37] Zygmund, A., 1995. Trigonometic series. Cambridge University Press, Cambridge.

Table 1: Summary Statistics of the Futures Returns and Volatilities. This table presents summary statistics on metal futures returns and volatilities. Units are basis points per five minute interval continuously compounded. The sample is April 2000 through December 31, 2007. Returns are calculated from futures prices on nearby contracts. Q(20) represents the Ljung-Box portmanteau test statistics for up to the 20th order autocorrelation. **, * indicate statistical significance at the 0.05 and 0.10 levels, respectively.

Series	Mean	Std Dev	Min	Max	Skew	Kurtosis	Q(20) on returns
Returns							
Gold	0.18**	20.71	-404.07	494.80	-0.10**	42.27**	47.91**
Copper	0.10**	34.90	-1001.18	847.01	-1.59**	83.54**	64.41**
Silver	0.12**	35.24	-823.89	967.54	-0.47**	67.49**	108.86**
Volatilities							
Goldv	12.07**	16.83	0.00	494.80	6.27**	79.55**	13,909.68**
Copperv	19.00**	29.27	0.00	1001.18	8.49**	147.39**	14,566.05**
Silverv	19.62**	29.28	0.00	967.54	7.59**	121.72**	18,925.43**

Table 2: T-statistics for the Wald Test of Unitroot vs. Fractional Integration. This table reports the t-statistics of the Wald test of the null hypothesis of a unit root against fractional alternatives introduced by Lobato and Velasco (2007). The Wald test results for both the return sereis and the volatility series (absolute return) are reported. Using up to 10 lags, the null hypothesis is rejected at all signifiant levels.

	Return			Volatility		
Lag	Gold	Copper	Silver	Goldv	Copperv	Silverv
0	-241.47	-259.77	-245.75	-220.23	-229.99	-214.43
1	-166.57	-183.00	-174.92	-155.13	-161.26	-155.27
2	-136.25	-148.67	-140.87	-127.47	-132.93	-128.26
3	-118.78	-129.09	-122.11	-111.67	-119.02	-112.10
4	-106.41	-115.37	-109.39	-101.39	-107.35	-102.55
5	-97.38	-105.21	-99.23	-92.71	-98.61	-94.27
6	-89.67	-97.32	-90.99	-84.70	-89.55	-86.63
7	-81.18	-91.04	-82.71	-77.09	-81.82	-79.39
8	-71.19	-85.26	-73.41	-69.39	-74.18	-72.49
9	-59.57	-80.91	-62.95	-61.77	-65.20	-66.08
10	-47.55	-77.94	-51.29	-54.52	-56.09	-59.44

Table 3: Estimates of Fractional Differencing Prameters. The values in parentheses are asymptotic t-statistics for the null hypothesis $d = 0$. $\hat{d}_{GPH}$ represents point estimates based on the GPH estimator, $\hat{d}_{WOLS}$ represents point estimates based on the wavelet OLS estimator of Jensen (1999), and $\hat{d}_{BWMLE}$ represents point estimates based on the banded wavelet MLE estimator of Jensen (2000). **, * indicate statistical significance at the 0.05 and 0.10 levels, respectively.

Return Series	$\hat{d}_{GPH}$ $T^{0.4}$	$\hat{d}_{GSP}$ $T^{0.4}$	$\hat{d}_{WOLS}$ Daub-8	$\hat{d}_{BWMLE}$ Daub-8
Gold	-0.081	-0.111**	-0.216 **	-0.121**
	(-1.22)	(-2.02)	(-3.51)	(-2.44)
Copper	0.211**	0.122 **	-0.173	0.123**
	(2.47)	(2.23)	(-1.135)	(2.58)
Silver	-0.008	-0.012	-0.147**	-0.014
	(-0.12)	(-0.22)	(-2.86)	(-0.29)
Volatility Series				
Goldv	0.847**	0.827**	0.695**	0.921**
	(12.58)	(15.16)	(5.21)	(11.31)
Copperv	0.568**	0.55**	0.563**	0.662**
	(8.01)	(10.08)	(5.53)	(7.38)
Silverv	0.700**	0.716**	0.542**	0.853**
	(10.21)	(13.12)	(3.73)	(10.01)

Table 4: Estimates of Fractional Differencing Parameters. The values in parentheses are asymptotic t-statistics for the null hypothesis $d = 0$. $\hat{d}_{GPH}$ represents point estimates based on the GPH estimator, $\hat{d}_{WOLS}$ represents point estimates based on the wavelet OLS estimator of Jensen (1999), and $\hat{d}_{BWMLE}$ represents point estimates based on the banded wavelet MLE estimator of Jensen (2000). **, * indicate statistical significance at the 0.05 and 0.10 levels, respectively. The bandwidth for the frequency domain estimators is $T^{0.5}$ and the 8 smallest scales are trimmed for the wavelet estimators. The volatility transformation is absolute returns.

Volatility Series	$\hat{d}_{GPH}$ $T^{0.5}$	$\hat{d}_{GSP}$ $T^{0.5}$	$\hat{d}_{WOLS}$ Daub-8	$\hat{d}_{BWMLE}$ Daub-8
Goldv	0.531**	0.540**	0.678 **	0.713**
	(13.33)	(17.28)	(7.13)	(15.65)
Copperv	0.623**	0.604 **	0.557**	0.582**
	(15.12)	(19.31)	(7.76)	(11.63)
Silverv	0.550**	0.523**	0.562**	0.687**
	(14.15)	(16.93)	(5.43)	(14.45)

Table 5: Estimates of Fractional Differencing Parameters. The values in parentheses are asymptotic t-statistics for the null hypothesis $d = 0$. $\hat{d}_{GPH}$ represents point estimates based on the GPH estimator, $\hat{d}_{WOLS}$ represents point estimates based on the wavelet OLS estimator of Jensen (1999), and $\hat{d}_{BWMLE}$ represents point estimates based on the banded wavelet MLE estimator of Jensen (2000). **, * indicate statistical significance at the 0.05 and 0.10 levels, respectively. The bandwidth for the frequency domain estimators is $T^{0.5}$ and the 8 smallest scales are trimmed for the wavelet estimators. The volatility transformation is squared returns.

Volatility Series	$\hat{d}_{GPH}$ $T^{0.5}$	$\hat{d}_{GSP}$ $T^{0.5}$	$\hat{d}_{WOLS}$ Daub-8	$\hat{d}_{BWMLE}$ Daub-8
Goldv	0.468**	0.474**	0.647 **	0.664**
	(9.82)	(15.17)	(4.05)	(3.36)
Copperv	0.468**	0.449 **	0.423**	0.399*
	(13.56)	(14.37)	(4.05)	(1.89)
Silverv	0.419**	0.417**	0.562**	0.560**
	(10.41)	(13.34)	(3.41)	(2.79)

Figure 1. The Return and Volatility Series for the Three Metal Futures.

Figure 1 presents the return and volatility series for the three metal futures, gold, copper and silver. The returns are in percentage and volatility of the return series is measured as the absolute value of the returns. The time frequency of the data is five minutes.