Studies on the long range dependence in stock return volatility and trading volume

by

Chi-Liang Chen

Advisor

Mei-Hui Guo

Department of Applied Mathematics

National Sun Yat-sen University

Kaohsiung, Taiwan, 804, R.O.C

July 2004

Abstract

Many empirical studies show that both equity volatility and its trading volume have long range dependence and can be modeled as fractional integrated processes. The objective of this study is to investigate relationship between volatility and volume.

We adopt four estimators of volatility, which includes the squared log returns, historical volatility, iterative t estimators and $GARCH$ estimators. The results show that among the four estimators squared log returns usually have the largest integration orders and produce hightest ratios of fractional cointegration. The fractional integrated orders are estimated separately and jointly, and the cointegration parameters are estimated by ordinary least squares, a narrow band frequency domain least squares method and a semiparametric estimator of Whittle likelihood. Models are also established when volatility and volume are not fractional cointegrated.

Contents

1. Introduction

Many empirical studies show that both equity volatility and its trading volume have long range dependence and strong positive cross correlation during a long period of time. Unlike the exponential decay correlation of short run dependence models such as ARMA(p,q), the persistence of volatility and volume decay slowly which is usually modelled by the ARFIMA with hyperbolic decay autocorrelation.

Volatility is a measure of the standard deviation of the daily range of price movements. Why is volatility important? It is one of the important determinants of the riskiness of an asset and is also a crucial parameter in the pricing of contigement derivatives, such as options. The more one learn about volatility, the better one can value the option price. In the literature, there are different ways to estimate equity volatility, here we consider four estimators namely, the squared log returns, historical volatility, a non-parametric estimation and $GARCH$ estimators.

The availability of long series of asset price data at a daily or higher frequency have resulted in a large number of econometric volatility studies. A general finding across different asset markets is that volatility shocks are highly persistent. A shock which raises asset price volatility, as for example those experienced in equity and bond markets over the summer of 1998, will persist for several months after the initial impact of the shock. Volatility is typically stationary, in the technical sense that, over sufficiently long periods of time, it reverts back to a constant mean. However, volatility may depart from this mean for extended periods of time. This phenomenon is often referred to as ”mean reverting”.

There is a long established literature which links price volatility to trading volume (Bollerslev, T. and Jubinski, D. (1999)). What is less clear is whether changes in trading volume are the causes of volatility changes, or whether both are caused by a common unobserved(latent) variable generally interpreted as new information. Some research favors an intermediate view in which information arrival is the common cause of both volume and volatility changes, but the impacts of information on volatility come through trading volume (Bollerslev, T. and Jubinski, D. (1999)).

As mentioned above, both volatility and trading volume have a long range dependence. If the two processes have a linear regression relation such that the error term ϵ have less persistent correlation, then the two processes are said to be fractional cointegrated. The properties of the model in the standard $I(1) - I(0)$ cointegration case are well known, see for example Watson (1994). Yet the fractional cointegration framework has been examined only recently, see the short review in Robinson and Yajima (2002). Henry and Zaffaroni (2003) provide a survey of empirical applications of fractional integration and long memory in macroeconomics and finance. Compared with the standard time series regression model with weakly dependent regressors, the new complication is that, since the regressors and the errors both have long memory, they are potentially correlated even at very long horizons, thus rendering the ordinary least squares (OLS) estimator inconsistent, see Robinson (1994) and Robinson and Marinucci (2003). To deal with this issue, Robinson (1994) proposed a semiparametric narrow band frequency domain least squares(FDLS) estimator that assumes only a multivariate generation of (4.13), and essentially performs OLS on a degenerating band of frequencies around the origin. The consistency of the estimator in the stationary case is proved by Robinson (1994). Christensen and Nielsen (2001) show that its asymptotic distribution is normal when the collective memory of the regressors and the error term is less than 1/2, i.e. when $d + d_e < 1/2$. In contrast, Robinson and Marinucci (2003) consider several cases where the regressors are fractionally integrated and nonstationary, and show that the limiting distributions for the FDLS estimators are then functionals of fractional Brownain motion.

Throughtout this article, we are concerned with the stationary case $d \in (0, 1/2)$. This interval is relevant for many applications in finance, e.g. stock market trading volume (Lobato and Velasco (2000)), exchange rate volatility (Andersen, Bollerslev, Diebold and Labys(2001)) and stock return volatility (Andersen, Bollerslev, Diebold and Ebens (2001) and Christensen and Nielsen (2001)). In particular, it is also the relevant region for the volatility process in empirical applications.

Many estimators of the long memory parameters d and the regression parameters have been suggested in the literature. A semiparametric approach has been developed by Geweke and Porter-Hudak (1983), Robinson (1994, 1995*a*, 1995*b*), Lobato and Robinson (1996), and Lobato (1999), among others. The semiparametric estimators of the memory parameter assume only the form of the spectral density function, and use a degenerating part of the periodogram around the origin to estimate the model. This approach has the advantage of being invariant to any short and medium term dynamics (as well as mean terms since the zero frequency is usually left out). In particular, a local Whittle maximum likelihood estimator (QMLE) based on the maximization of a local Whittle approximation to the likelihood, see equation (4.17), to estimate the integeration orders of univariate and multivariate stationary fractionally integrated time series. Of course, a fully paremeteric approach is more efficient, using the entire sample, but is inconsistent if the parametric model is specified incorrectly, e.g. if the structure of the short term dynamics is misspecified.

Marinucci and Robinson (2001) and Christent and Nielsen (2001), further suggest conducting a fractional cointegration analysis in several steps. First, the integration orders of the data is estimated by local Whittle QMLE. Second, the narrow band FDLS estimator for the cointegrating parameters is calculated, and finally the integration order of the residuals is estimated. Hypothesis testing is then conducted on d_e and on β as if d_e (which enters in the limiting distribution of the FDLS estimator) were known. Although this is indeed a valid course of action, see Hassler, Marmol and Velasco (2000) and Velasco (2003) for the nonstationary case, and for the stationary case, a joint estimation method for the integration orders and the cointegration parameter would be preferable, see Nielsen (2004).

In this article, the main objective is to study whether volatility and trading volume are fractional cointegrated. In the next section we will give literature review. In section 3, we introduce the cross spectrum analysis. In section 4, we will introduce four estimators of volatility and estimators of fractional integrated order. Section 5, contains the data description and analysis results. Section 6 is the conclusion.

2. Literature review

In this section we briefly review the following three papers, Bollerslev and Jubinski (1999), Ray and Tsay (2000) and Nielsen (2004).

Over the past decade, a lot of literature has developed for modeling the temporal dependencies in financial market volatility. Clark (1973) provided a model called mixture of distributions hypothesis (MDH). It assumes that condictional on the latent intensity process, say $K_{j,t}$, the distribution for the daily returns may be expressed as

$$R_{j,t}|K_{j,t}\ N(0, \sigma_j^2 \cdot K_{j,t})$$

where $K_{j,t}$ has been standardized so that a value of unity will result in a daily variance for stock j equal to σ_j^2. In MDH model, the $K_{j,t}$ latent information-arrival process represents the intensity of "news" events. Andersen (1996) assumed that the daily trading volume, conditional on the information-arrival process, will be approximately Poisson distribution.

$$V_{j,t}|K_{j,t}\ P(\mu_{j,0} + \mu_{j,1} \cdot K_{j,t})$$

where $\mu_{j,0}$ and $\mu_{j,1}$ are normalizing constants related to the importance of liquidity or noise.

As long run proposition, the MDH implies that for large τ,

$$corr(|R_{j,t}|, |R_{j,t-\tau}|) \sim \tau^{2 \cdot d_j - 1}$$

$$corr(V_{j,t}, V_{j,t-\tau}) \sim \tau^{2 \cdot d_j - 1}$$

It shows that the correlations of τ step of the series of volatility and volume decay in the same rate like τ^d where d is order of the rate. By a direct extension of the arguments of Andersen (1996), it follows that this same long run dependence carries over to any positive power transform of the absolute returns and volume as defined by the MDH, here the absolute return is taken as a volatility estimator. We are interested in whether daily volatilities and volumes having the same fractional integrated order. That is to say, if their period of decay cycle is the same. Fractional cointegration implies the two series have a long run equilibrium relationship.

Breidt, Crato, and de Lima (1998) propose the following long memory stochastic volatility model (LMSV)

$$r_t = \sigma_t \psi_t, \sigma_t = \sigma exp(\nu_t/2)$$

where r_t is the stock return, σ_t is the volatility at time t, respectively, $\sigma > 0$, $\{\psi_t\}$ is a sequence of independent and identically distributed random variable with mean 0 and variance 1, $\{\nu_t\}$ is a fractionally integrated process with $0 < d < 0.5$, and $\{\nu_t\}$, $\{\psi_t\}$ are independent. By taking the logarithm of square of $\{r_t\}$, we have

$$\begin{aligned} y_t = log(r_t^2) &= [log(\sigma^2) + E(log\psi_t^2)] + \nu_t \\ &\quad +[log(\psi_t^2) - E(log\psi_t^2)] \\ &\equiv \mu + \nu_t + \epsilon_t \end{aligned}$$

that is, $\log r_t^2$ is a long memory process plus white noise innovation, which has the same fractional integrated order as $\log \sigma_t^2$. So we can take $\log r_t^2$ as volatility estimator.

In the literature, the implied volatilities and realized volatilities are shown to be fractionally cointegrated. The implied volatility $\sigma_{IV,t}$ is solved from the option pricing formula and the realized volatility $\sigma_{RV,t}$ is the standard deviation of the realized return from t to $t+1$.

Christensen and Prabhala (1998) considered the regression specification

$$y_t = \alpha + \beta x_t + e_t, \tag{2.1}$$

where $y_t = ln\sigma_R V, t$ and $x_t = ln\sigma_{IV,t}$ are the log volatilities and α, β are the intercept and slope coefficients. In practice, we work with the log volatilities, since they are close to Gaussian, see Andersen, Bollerslev, Diebold and Ebens (2001). Thus, if $x_t = E_t(y_t)$ with $E_t(\cdot)$ denoting conditional expectation as of time t, then β is unity and e_t is serially uncorrelated. For the detailed description of the implied-realized volatility relation and its implications, see Chirtensen and Prabhala (1998). Empirical literature suggests, volatility is fractionally integrated (Andersen, Bollerslev, Diebold and Ebens (2001) and Christensen and Nielsen (2001) find fractional integration order d around 0.35-0.45). For fractionally cointegrated x_t and y_t, the forecasting error e_t in (2.1) possesses only short memory.

In our study, in order to investigate fractional cointegration relationship between volume and volatility, first we need to define estimators of volatility. In particular, we will consider four estimators which will be introduct in Section 4.

3. Cross spectrum analysis

(The following materials are mainly from Wei, W. S. (1997).)

Suppose that x_t and y_t are jointly stationary with the cross-covariance function $\gamma_{xy}(k)$ for $k = 0, \pm 1, \pm 2, ...$ Define the cross-covariance generating function as

$$\gamma_{xy}(B) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) B^k$$

where the coefficient of B^k is the kth cross-covariance between x_t and y_t. This is the generalization of the autocovariance generating function $\gamma_x(B) = \sum_{k=-\infty}^{\infty} \gamma_x(k) B^k$ of single series. Specifically, $\gamma_{xx}(B) = \gamma_x(B)$ and $\gamma_{yy}(B) = \gamma_y(B)$ are autocovariance generating functions for x_t and y_t, respectively. If the cross-covariance sequence $\gamma_{xy}(k)$ is absolutely summable, i.e., $\Sigma_{k=-\infty}^{\infty} |\gamma_{xy}(k)| < \infty$, then its Fourier transform exists and is called the cross-spectrum between x_t and y_t and is given by

$$f_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_{xy}(k) e^{-i\omega k} = \frac{1}{2\pi} \gamma_{xy}(e^{-i\omega}).$$

This is the generalization of the spectrum for the univariate series. Specifically, $f_{xx}(\omega) = f_x(\omega)$ and $f_{yy}(\omega) = f_y(\omega)$ are spectrums for x_t and y_t, $f_{xy}(\omega)$ will in general be complex, since $\gamma_{xy}(k) \neq \gamma_{xy}(-k)$. Hence, we can write

$$f_{xy}(\omega) = c_{xy}(\omega) - i q_{xy}(\omega)$$

where $c_{xy}(\omega)$ and $-q_{xy}(\omega)$ are the real and imaginary parts of $f_{xy}(\omega)$. More precisely,

$$c_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_{xy}(k) \cos \omega k$$

and

$$q_{xy}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_{xy}(k) \sin \omega k$$

The function $c_{xy}(\omega)$ is called the cospectrum, and $q_{xy}(\omega)$ is called the quadrature spectrum, of x_t and y_t. However, these functions are difficult to interpret. Alternatively, we can express $f_{xy}(\omega)$ in the polar form

$$f_{xy}(\omega) = A_{xy}(\omega)e^{i\phi_{xy}(\omega)}$$

where

$$A_{xy}(\omega) = |f_{xy}(\omega)| = [c_{xy}^2(\omega) + q_{xy}^2(\omega)]^{1/2},$$

and

$$\phi_{xy}(\omega) = \tan^{-1}[-q_{xy}(\omega)/c_{xy}(\omega)].$$

The functions $A_{xy}(\omega)$ and $\phi_{xy}(\omega)$ are known as the cross-amplitude spectrum and the phase spectrum, respectively. Two other useful functions are the gain function and the coherence. The gain function is defined as

$$G_{xy}(\omega) = \frac{|f_{xy}(\omega)|}{f_x(\omega)} = \frac{A_{xy}(\omega)}{f_x(\omega)},$$

which is the ratio of the cross-amplitude spectrum to the input spectrum. The coherence is defined by

$$K_{xy}^2(\omega) = |f_{xy}(\omega)|^2/f_x(\omega)f_y(\omega)$$

which is essentially the standardized cross-amplitude spectrum.

In practice, several cross-spectral functions are needed to describe the relationship between two series in the frequency domain. In general, the cross-spectrum between two series is more easily analyzed through the gain function, the phase spectrum, and the coherence. Hence, they are possibly the three most commonly used cross-spectral functions in frequency domain analysis.

We assume that the amplitude spectrum and the phase spectrum are independent. Consequently, $A_{xy}(\omega)$ can also be thought as the average value of the product of amplitudes of the ω-frequency components of x_t and y_t, and the phase spectrum $\phi_{xy}(\omega)$ represents the average phase-shift, $[\phi_x(\omega) - \phi_y(\omega)]$, between the ω-frequency components of x_t and

y_t. In terms of a causal model $y_t = \alpha x_{t-\tau} + e_t$ or $x_t = \beta y_{t-\gamma} + a_t$, where there is no feedback relationship between x_t and y_t, the phase spectrum is measure of the extent to which each frequency component of one series leads the other. The ω-frequency component of x_t leads the ω-frequency component of y_t if the phase $\phi_{xy}(\omega)$ is negative. The ω-frequency component of x_t lags the ω-frequency component of y_t if the phase $\phi_{xy}(\omega)$ is positive.

4. Semiparametric estimation in long memory models

In this section, we first introduce three models to estimate the volatility in our analysis. In section 4.2, we introduce Whittle estimator of fractional integrated parameter.

4.1 Volatility models

Long memory stochastic volatility model

Recall the long memory stochastic volatility (LMSV) model defined as in Section 2,

$$r_t = \sigma_t \psi_t, \sigma_t = \sigma exp(\nu_t/2)$$

where ν_t is an $I(d)$ process. By taking the logarithm of r_t^2 we can obtain

$$y_t = log(r_t^2) = \mu + \nu_t + \epsilon_t \tag{4.1}$$

thus, y_t is a long memory process plus a white noise.

The log transformation will be a problem when r_t is 0 or a very small value. To deal with this problem, we use Fuller's (1996) transformation

$$y_t^* = log(r_t^2 + \lambda s_r^2) - \frac{\lambda s_r^2}{r_t^2 + \lambda s_r^2} \tag{4.2}$$

where λ is an pectoral small constant and s_r^2 is the sample variance of returns.

Non-parametric estimation of historical volatility

Randal, Thomson and Lally assume that X_t satisfies

$$X_t = \frac{\sigma Z_t}{\sqrt{S_t}} \quad (t = 1, ..., n)$$

where $Z_t's$ are independent $N(0, 1)$ random variables that are independent of S_t, $S_t's$ are independent random variables on $[0, \infty)$ each with distribution function $F_S(s)$, and $E(X_t^2) = \sigma^2$. In the case when S_t is proportional to a χ_v^2 random variable and X_t is a scaled t_v random variable. To ensure finite variance, the degree of freedom $\nu \geq 3$. The objective is to estimate the volatility σ^2 for arbitrary distribution function $F_S(s)$.

Dempster, Laird and Rubin (1997) use EM algorithm to construct an iterative formula for estimating σ^2. The log-likelihood of the complete data $X = (X_1, ..., X_N), S = (S_1, ..., S_n)$ is proportional to

$$log\tilde{L} = -log\sigma^2 - \frac{1}{\sigma^2}\frac{1}{n}\sum_{t=1}^{n} S_t X_t^2$$

Maximising $E_0(log\tilde{L}|X)$ with respect to σ^2 yields

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} E_0(S_t|X)X_t^2 \tag{4.3}$$

where the expectation is with respect to the conditional distribution evaluated at σ_0^2, a previous estimate of σ^2.

In the case when S_t is a $\chi_v^2/(n-2)$ random variable, X_t is a scaled t_v random variables with $E(X_t^2) = \sigma^2$. Evaluation of $E_0(S_t|X)$ gives the following recursive equation for the maximum likelihood estimator of σ^2

$$\hat{\sigma}_v^2 = \frac{v+1}{v-2}\frac{1}{n}\sum_{t=1}^{n}(1 + \frac{X_t^2}{(v-2)\sigma_0^2})^{-1}X_t^2. \tag{4.4}$$

We now consider the time series of daily returns $R_t(t = 1, ..., T)$ defined by $R_t = \sigma_t\epsilon_t$ where $R_t = lnP_t - lnP_{t-1}$, P_t is the underlying time series of prices and ϵ_t have a heavy-tailed distribution. We consider a general non-parametric historial volatility estimator $\hat{\sigma}_t^2$ given by the finite moving average of span $2r + 1$

$$\hat{\sigma}_t^2 = \hat{\tau}\tilde{\sigma}_t^2, \quad \tilde{\sigma}_t^2 = \sum_{j=-r}^{r} w_j q_{t+j} R_{t+j}^2 \tag{4.5}$$

where weights $w_j's$ satisfy $\sum_{j=-r}^{r} w_j = 1$ and control the precision and smoothness of $\hat{\sigma}_t^2$ as an estimator of the underlying time-varying volatility. The robustness weights $q_t \geq 0$ are also given. Here $\hat{\tau}$ is an estimate of the appropriate bias correction factor τ which depends on the choice of q_t, the distribution of the ϵ_t, and satisfies $E(\tilde{\sigma}_t^2) = \sigma_t^2/\tau$. We estimate τ globally using the sample variance of all the scale adjusted returns

$$\hat{\tau} = \frac{1}{T}\sum_{t=1}^{T}(\frac{R_t}{\tilde{\sigma}_t})^2. \tag{4.6}$$

The iterated t-estimator for $v \geq 3$ degrees of freedom, is a special case of (4.5) with $w_j = 1/(2r+1)$ for all j, and

$$q_t = \frac{v+1}{v-2}(1 + \frac{R_t^2}{(v-2)\sigma_{0,t}^{\hat{2}}})^{-1}$$

where $\hat{\sigma}_{0,t}^2$ is a prior estimator of volatility. The updated volatility estimate is

$$\hat{\sigma}_{1,t}^2 = \frac{1}{2r+1}\sum_{j=-r}^{r}\frac{v+1}{v-2}(1 + \frac{R_t^2}{(v-2)\hat{\sigma}_{0,t}^2})^{-1}R_t^2. \tag{4.7}$$

The GARCH model

(The following materials are mainly from Tsay, R. S. (2002).)

Bollerslev (1986) proposes a useful extension known as the generalized $ARCH(GARCH)$ model. For a log return series r_t, we assume that the mean equation of the process can be adequatedly described by an $ARMA$ model. Let $a_t = r_t - \mu_t$ be the mean-corrected log return. Then a_t follows a $GARCH(m,s)$ model if

$$a_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \alpha_0 + \Sigma_{i=1}^m \alpha_i a_{t-i}^2 + \Sigma_{j=1}^s \beta_j \sigma_{t-j}^2$$

where $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0 and variance 1, $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, and \;\; \Sigma_{i=1}^{max(m,s)}(\alpha_i + \beta_i) < 1$. Here it is understood that $\alpha_i = 0$ for $i > m$ and $\beta_j = 0$ for $j > s$. The latter constraint on $\alpha_i + \beta_i$ implies that the unconditional variance of a_t is finite, whereas its conditional variance σ_t^2 evolves over time. As before, ϵ is often assumed to be a standard normal or standardized Student-t distribution. Here we take the square root of conditional variance σ_t^2 as volatility.

4.2 Estimation of the fractional integrated parameter

(The following materials are mainly from Dzhaparidze K. (1986))

Let X_t be a Gaussian stationary process with mean 0 and finite variance. Denote by

$$B_f = [\beta(\tau - s)]_{\tau,s=1,...,n} = \begin{bmatrix} \beta(0) & \beta(1) & \cdots & \beta(n-1) \\ \beta(1) & \beta(0) & \cdots & \beta(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \beta(n-1) & \beta(n-2) & \cdots & \beta(0) \end{bmatrix}$$

the matrix associated with the spectral density function f, where

$$\beta(\tau) = \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda\tau} d\lambda$$

is the autocovariance function, that is

$$\beta(\tau) = E(X_t X_{t+\tau}).$$

The n-dimensional probability density $p_n(x_1, x_2, ..., x_n)$ of the random variables $X_1, X_2, ..., X_n$ is the form

$$p_n(x_1, x_2, ..., x_n) = \frac{1}{(2\pi)^{n/2}[det(B_f)]^{1/2}} e^{-\frac{1}{2}x' B_f^{-1} x} \tag{4.8}$$

where $x = (x_1, x_2, ..., x_n)' \in R_n$. Denote L_n the logaritm of the likelihood function

$$\begin{aligned} L_n &= \log p_n(X_1, X_2, ..., X_n) \\ &= -\frac{1}{2}\{n \log 2\pi + \log det(B_f) + X' B_f^{-1} X\} \end{aligned}$$

The Whittle log likelihood expression is an approximation to the exact Gaussian log likelihood, which can be written as

$$\tilde{L_n} = -\frac{n}{2}\{\log 2\pi + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[2\pi f(\lambda)] d\lambda\} + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\lambda, X)}{f(\lambda)} d\lambda\}. \tag{4.9}$$

Fox and Taqqu(1986) show that the Whittle estimator has the same asymptoptic properties as the exact MLE. Moreover, it still has some advantages over the exact MLE. (1)The asymptotic properties of the Whittle estimator hold even if the series is not Gaussian. (2)The computation of the Whittle likelihood is simple since the periodogram can be computed quickly using Fast Fourier Transform (FFT) at Fourier frequencies. So in general it is prefered to estimate parameter d instead of the exact MLE.

To find the maximum of the Whittle likelihood

$$\tilde{L_n} = -\frac{n}{2}\{\log 2\pi + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log[2\pi f(\lambda)]d\lambda\} + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\lambda, X)}{f(\lambda)}d\lambda\} \tag{4.10}$$

is equivalent to find the minimum of the following function

$$W_n = \int_{-\pi}^{\pi} (\log f(\lambda) + \frac{I_n(\lambda, X)}{f(\lambda)}d\lambda).$$

In practice we use the discrete version of the likelihood

$$\bar{W}_n = \frac{1}{m}\sum_{j=1}^{m}(\log f(\lambda_j) + \frac{I_n(\lambda_j)}{f(\lambda_j)})$$

where $\lambda_j = 2\pi j/n, j = 1, 2, ..., m$, to estimate the parameter d.

4.3 Fractional integrated model and generalized Whittle estimators

We call z_t an $I(d)$ process, denoted by $z_t \in I(d)$, if

$$(1-B)^d z_t = \epsilon_t, \tag{4.11}$$

where $\epsilon_t \in I(0)$ and $(1-B)^d$ is defined by

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} B^j, \Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \tag{4.12}$$

and B is the lag operator($Bz_t = z_{t-1}$). A process is labelled $I(0)$ if it is covariance stationary and has spectral density that is bounded and bounded away from zero at the origin.

A stochastic process satisfy (4.11) has the spectral density

$$f(\lambda) \sim g\lambda^{-2d}, as\lambda \to 0 \tag{4.13}$$

where g is a constant. and the symbol " $\sim$ " means that the ratio of the left and right hand sides tends to one in the limit. Because its autocorrelations decay very slowly at a hyperbolic rate so it is said to be long range dependent. On the contrast, that the autocorrelations in ARMA model decay quickly in exponential rate, thus is called weakly dependent. The parameter d determines the memory of the process. If $d > -1/2$, z_t is invertible and admits a linear representation, and if $d < 1/2$ it is covariance stationary. If $d = 0$, the spectral density (4.13) is bounded at the origin, and the process has only weak dependence. Sometimes, z_t is said to have intermediate memory, short memory, and long memory when $d < 0, d = 0$, and $d > 0$, respectively.

If two series are cointegrated, there is a long run equilibrium relationship between them. That is to say, they are in a dynamic equilibrium in the sense that they tend to move together in the long run.

Suppose x_t, y_t are two integrated processes satisfied the regression relation

$$y_t = \beta x_t + e_t, \tag{4.14}$$

where the error term is integrated of smaller order $d_e < d$, i.e $e_t \in I(d_e)$. A much studied special case is the standard $I(1) - I(0)$ cointegration model which arises when $d = 1$ and $d_e = 0$, see e.g. Watson (1994) for the view. When d and/or d_e are not integers the model is called a fractional cointegration model following the original ideal by Granger (1981). We call the model (4.14) with $0 \leq d_e < d < 1/2$ a stationary fractional cointegrated model, since it is concered with the long-run linear comovement between two stationary fractionally integrated processes.

Robinson (1995b), Lobato (1999) and Robinson and Yajima (2002) assume the spectral density of $z_t = (x_t, e_t)$ can be written as

$$f(\lambda) = \Lambda^{-1} G \Lambda^{-1}, as\lambda \to 0 \tag{4.15}$$

where $\Lambda = diag(\lambda^d, \lambda^{d_e})$ and G is a real symmetric matrix. Equation (4.15) is the multivariate extension of equation (4.13). It include multivariate ARFIMA models. The Whittle approximation to the likelihood is

$$W(\theta, G) = \int_{-\pi}^{\pi} (log|f(\lambda)| + tr[f^{-1}(\lambda) Re(I(\lambda))]) d\lambda,$$

where $I(\lambda) = (2\pi n)^{-1} |\Sigma_{t=1}^{n} w_t e^{it\lambda}|^2$ is the periodogram matrix of $w_t = (x_t, e_t)$ at the frequency λ.

The discrete version of the likelihood is

$$\bar{W}(\theta, G) = \frac{1}{m} \sum_{j=1}^{m} (log|f(\lambda_j)| + tr[f^{-1}(\lambda_j) Re(I(\lambda_j))]) \tag{4.16}$$

where $\lambda_j = 2\pi j/n, j = 1, ..., m$ are the Fourier frequencies.

The local Whittle estimator of (θ, G) is defined as

$$(\hat{\theta}, \hat{G}) = arg \min_{\theta, G} \bar{W}(\theta, G)$$

We can concentrate G out of the likelihood by setting $\hat{G}(\theta) = m^{-1}\Sigma_{j=1}^{m}\Lambda_j Re(I(\lambda_j))\Lambda_j$, and rewrite the concentrated likelihood as

$$L(\theta) = log|\hat{G}(\theta)| - \frac{2(d+d_e)}{m}\sum_{j=1}^{m} log\lambda_j. \tag{4.17}$$

The local Whittle estimator can be defined according to the concentrated likelihood as

$$\hat{\theta} = arg\min_{\theta} L(\theta). \tag{4.18}$$

Marinucci and Robinson (2001) and Christensen and Nielsen (2001) propose the following simple two step estimator (TSE) for the integration orders and the cointegrating vector,

$$\hat{\theta}^{(2)} = \hat{\theta}^{(1)} - (\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}|_{\hat{\theta}^{(1)}})^{-1}(\frac{\partial L(\theta)}{\partial\theta}|_{\hat{\theta}^{(1)}}) \tag{4.19}$$

where $\hat{\theta}^{(1)}$ is a consistent initial estimator, e.g. the local Whittle QMLE of Robinson (1995a) and the narrow band FDLS estimator of Robinson (1994) and Christensen and Nielsen (2001). We could iterate (4.19) until covergence for higher order gains, but that dose not change the first order asymptotics. It is well known that the TSE has the same asymptotic distribution as the QMLE, but TSE is prefered for its simplicity.

4.4 FDLS estimation of fractional cointegrating coefficient

When the nonstationary regressors are stochastic, OLS is often inconsistent even when the errors have nonzero mean or are not orthogonal to the regressors, so the regression is incompletely specified. Consider the equation

$$y_t = \beta z_t + x_t$$

linking the scalar stochastic variates x_t, y_t and z_t, where only y_t and z_t are observed at $t = 1, .., n$. As $n \to \infty$, one sufficient condition that $OLS\ \hat{\beta} = \Sigma_{t=1}^{n} y_t z_t / \Sigma_{t=1}^{n} z_t^2 \to_p \beta$ is $\Sigma_{t=1}^{n} x_t^2 / \Sigma_{t=1}^{n} z_t^2 \to_p 0$ (by Cauchy inequality), which indicates, along with the "cointegrating" relation, that there is a linear combination of y_t and z_t which is stationary or "less nonstationary" than y_t and z_t individually and holds in the special case, stressed in much recent econometric literature [e.g. Johansen (1988)].

The problem of consistently estimating β is more challenging, because $\hat{\beta}$ can be consistent only for $\beta + E(z_t x_t)/E(z_t^2)$, and OLS with an intercept can be consistent only for $\beta + Cov(z_t, x_t)/Var(z_t)$, neither of which equal β when x_t and z_t are not orthogonal. Robinson (1994) suggest the following more delicate narrow band frequency domain least squares approach, carrying out the regression in the frequency domain over only a degenerating band of low frequencies. Define the Fourier transforms $w_y(\lambda) = (2\pi n)^{-1/2} \Sigma_{t=1}^{n} y_t e^{it\lambda}$, $w_z(\lambda) = (2\pi n)^{-1/2} \Sigma_{t=1}^{n} z_t e^{it\lambda}$, and the spectrum functions $I_{yz}(\lambda) = w_y(\lambda) \bar{w}_z(\lambda)$, $I_z(\lambda) = |w_z(\lambda)|^2$. Choose m satisfies the condition $1/m + m/n \to 0\ \ as\ \ m, n \to \infty$; for example, we can set $m = \sqrt{n}$. Then define $\hat{F}_{yz}(\lambda_m) = (2\pi/n) \Sigma_{j=1}^{m} I_{yz}(\lambda_j)$, $\hat{F}_z(\lambda_m) = (2\pi/n) \Sigma_{j=1}^{m} I_z(\lambda_j)$. The FDLS estimator of β is defined as $\tilde{\beta} = Re\{\hat{F}_{yz}(\lambda_m)\}/\hat{F}_z(\lambda_m)$.

5 Results

Throughtout, we denote the volatility estimated by $\log r_t^2$, iterative t estimator, historical volatility and $GARCH$ models by $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$, respectively. In Section 5.1, we give the data description and study the cross-correlation structure of volatility and volume in Section 5.2. Fractional integrated orders and fractional cointegrated results are given in Section 5.3 and 5.4. For not fractional cointegrated series, we further construct models in Section 5.5.

5.1 Data description

We analyze daily returns and volumes for the companies listed on the S&P 100 index in 2004. Daily closing prices and tradind volumes were obtained from Yahoo Finance. Another data named "outstanding shares" were obtained from New York Stock Exchange(NYSE). The sample period includes April 15, 1996 to May 28, 2004. The prices are adjusted for dividends, stock splits and Monday and holiday effects. Since companies listed on S&P 100 changes over time, not all selected companies have the same number of observations. The sample size ranges from 1101 to 2046. Since trading volume is a relative large to the volatility thus we will scale the data before analysis. Following the arguments of Lo and Wang(1996), the daily trading volume was measured by the turnover ratios, $V_t \equiv S_t/N_t$,where S_t denotes the share volume for the stock and N_t refers to the total number of outstanding shares.

5.2 Cross correlation analysis of volatility and volume

In this section, we study the cross correlation structure of the two series. We use phase spectrum to analysis which is the leader, volatility or volume, which is shed light in our analysis. The phase spectrum, delay parameter τ and coherence of the Alcoa corporation are given in Figure 1-1 to 3-4. Most of the phase spectra have values around zero and the decay parameter τ is a curve near zero. Figure 1-2 and 1-3 show volatility leads volume at the frequencies near $0.5, 1$ and 1.7. That is to say that the phase spectra of volatility and volume are the same on the average, yet volatility leads volume at the three frequencies, which are corresponding to the periods 12, 6 and 3 days. Large volume will first drive volatility up, and large volatility will also increase trading volume, which can also be seen from the strong cross-correlation exhibited in Figure 3-1 to 3-4. The other companies in our study have the similar phenomenon. In fact, the reason is also easy to understand. American is a free trading country and its information liquid very quickly in the market.

Following RAY and Tsay (2000), we classify the 100 companies into 5 classes of industries. Class 1 is about technology and information, Class 2 is about finance and banks, Class 3 is about transportation, Class 4 is about biochemistry and the last class is about commodity. Table 1 show the numbers of classification in every frequency that volatility lead volume and the percentage of every class in each sector. We obtain that class 1, 3 and 5 have high proportion in the sector 7. That is to say volume leads volatility initially but afterwards volatility leads volume about the period 3, 6 and 12 days.

5.3 Estimations of individual volatility and volume

For estimation of d, we use quasi-likelihood method discussed by Breidt et al. (1998). Table 2 show the numbers of observations and the estimatiors of fractional integrated order for four volatility and volume by classification. We obtain that $\hat{\sigma}_1$ have the largest integrated order on average and $\hat{\sigma}_4$ usually have the smallest one, see from Figure 4 (boxplots of the 5 industries and the last is miscellanea) and table 2. Moreover, $\hat{\sigma}_2$ have the smallest integrated order in class 4 and $\hat{\sigma}_3$ have the similar integrated order except class 4.

5.4 Cointegration or not about volatility and volume

After seeing table 2, we want to understand whether volatility and volume are fractionally cointegrated. Then, the first method used is ordinary least square. By fitting a regression model for volatility and volume, we can obtain its residual. If the fractional integrated order of residual is less than that of volatility and volume. We say the two series are fractionally cointegrated. The second method is FDLS and it is dealt with the same step. The outcomes are on the Table 3, and Figure 5 is the plots with $min\{d_x, d_y\}$ versus d_e. It is convenient for us to judge whether they are fractionally cointegrated. If the spot is under the straight line $(x = y)$, it means the two series are fractionally cointegrated. On the contrary, they are not. In Figure 5 we receive the following result that the percentage of fractional cointegration is relatively large by taking $logr_t^2$ as volatility. The main reason may be that the fractionally integrated orders of $\hat{\sigma_2}$, $\hat{\sigma_3}$ and $\hat{\sigma_4}$ are on average too small. But we can't say it is good or bad because of this.

The result of the generalized Whittle estimatior of d, d_e and the slope of regression parameter β is shown on the table 4. We obtain very different values of the three parameters distinguished from OLS and $FDLS$. The most likely reason may be that we don't completely catch the correct structure about the two series. Because the generalized Whittle estimatior assume that the integrated order between volatility and volume are the same, but in fact they may be different. Another reason is that the main disadvantages of the Whittle estimator are (1) it need to know the parametric form of the spectral density and (2) a greater computational effect than some graphical methods. If the exact form of the spectral density is unknown, the Whittle estimator will become very biased. In opposition, if it is really a $I(d)$ or $ARFIMA$ structure, it will perform batter. In order to check this view point, we have to do some simulations. Table 5 is the simulation results with different fractional integrated orders and two kinds of errors. Each result is to run 100 times and take its mean. We obtain that if the true model is real a $I(d)$ or $ARFIMA$ and volatility and volume have the same integrated order, the estimation will perform not too bad.

5.5 The relation between two series

In many applications, the relationship between two series is the important. The Market Model in finance is an example that relates the return of a individual stock to the return of a market index. The least squares method is often used to estimate model

$$y_t = \alpha + \beta x_t + e_t.$$

Now we know that volatility and volume are not fractionally cointegrated in some companies. That is to say, they don't satisfy the regression relation. But we still want to establish a relation between them and regression model is the simplest one. So before we build up the regression relation, let them pre-whiting maybe a good choice. First, we pick up the series that are not fractionally cointegrated and let both of their volatility and volume pre-white. They should become two white noise series in theory then we fit a regression model for them. The result is on Table 6. The residuals fit well neither white noise nor $MA(1)$ model. It is saying we already construct a relation between volatility and volume among that are not fractionally cointegrated.

6. Conclusion

From cross-spectrum analysis we obtain the volatility and volume of companies listed in $S\&P100$ index on average have the same cross correlation structure. After classification we find out that the industries about technology and information, transportation and commodity seem notably that their volatility leads volume in some special frequencies. Then, the fractionally integrated order of $logr_t^2$ is larger than others and its percentage of fractionally cointegrated phenomenon is also the largest either by OLS or $FDLS$. On the contrary, the generalizd Whittle estimatior seem not so good. Maybe it not only possess the characteristics of $I(d)$ or $ARFIMA$ or don't have the same fractionally integrated order between volatility and vilume but we ignore it. We do some simulation that show if the two series are real $I(d)$ or $ARFIMA$, the generalized Whittle estimatior will perform not too bad. Furthermore, we establish a simple regression model between volatility and volume after pre-whiting if they are not fractionally cointegrated.

References

1. Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* **51**, 169-204.

2. Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics* **61**, 43-76.

3. Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). The distribution of exchange rate volatility, *Journal of the American Statistical Association* **96**, 42-55.

4. Bollerslev, T. and Jubinski, D. (1999). Equity trading volume and volatility: Latent information arrivals and common long-run dependencies, *Journal of Business and Economic Statistics* **17**, 9-21.

5. Breidt, F. J., Crato, N. and de Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* **83**, 325-348.

6. Christensen, B. J. and Nielsen, M. O. (2001). Semiparametric analysis of stationary fractional cointegration and the implied-realized volatility relation, Working Paper Department of Economics, University of Aarhus.

7. Christensen, B. J. and Prabhala, N. R. (1998). The relation between implied and realized volatility, *Journal of Financial Economics* **50**, 125-150.

8. Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**,135-155.

9. Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* **39**, 1-38.

10. Dzhaparidze, K. (1986). Parameter estimation and hypothesis testing in spectral analysis of stationary time series. Springer-Verlag, 35-66, New York.

11. Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics* **14**, 517-532

12. Fuller, W. A. (1996). Introduction to statistical time series (2nd ed.), Wiley, New York.

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

14. Granger, C. W. J. (1981). Some properties of time series dada and their use in econometric model specification, *Journal of Econometrics* **16**, 121-130.

15. Hassler, U., mARMOL, F. and Velasco, C. (2000). Residual log-periodogram inference for long-ren relationships, *Forthcoming in Journal of Econometrics*.

16. Henry, M. and Zaffaroni, P. (2003). The long range dependence paradigm for macroeconomics and finance, in Doukhan P., Oppenheim G. and Tauua M. S., eds, 'Theory and applications of long-range dependence', Birkhauser, Boston, pp.417-438.

17. Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics Control* **12**, 231-254.

18. Lo, A. W. and Wang, J. (1996). Trading volume, part 1: Definitions, data analysis and implacations of portfolio theory, unpublished manuscript, Massachusetts Institute of Technology, Sloan School of Management.

19. Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model, *Journal of Econometrics* **90**, 129-153.

20. Lobato, I. N. and Robinson, P. M. (1996). Averaged periodogram estimation of long memory, *Journal of Econometrics* **73**, 303-324.

21. Lobato, I. N. and Velasco, C. (2000). Long memory in stock-market trading volume, *Journal of Business and Economic Statistics* **18**, 410-427.

22. Marinucci, D. and Robinson, P. M. (2001). Semiparametric fractional cointegration analysis, *Journal of Econometrics* **105**, 225-247.

23. Nielsen, M. O. (2004). Local Whittle analysis of stationary fractional cointegration and the implied-realied volatility relation, Working Paper Cornell University.

24. Randal, J. A., Thomson, P. J. and Lally, M. T.. Non-parametric estimation of historical volatility, Working Paper.

25. Ray, B. K. and Tsay, R. S. (2000). Long-range dependence in daily stock volatilities, *Journal of Business and Ecnomic Statistics* **18**, 254-262.

26. Robinson, P.M. (1994). Semiparametric analysis of long-memory time series, *Annals of Statistics* **22**, 515-539.

27. Robinson, P.M. (1995a). Gaussian semiparametric estimation of long range dependence, *Annals of Statistics* **23**, 1630-1661.

28. Robinson, P.M. (1995b). Log-periodogram regression of time series with long range dependence, *Annals of Statistics* **25**, 1048-1072.

29. Robinson, P. M. and Marinucci, D. (2003). Semiparametric frequency domain analysis of fractional cointegration, in P. M. Robinson, ed, 'Time series with long memory', Oxford University Press, Oxford, pp.334-373.

30. Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems, *Journal of Economics* **106**, 217-241.

31. Tsay, R. S.(2002). Analysis of financial time series. Wiley Inter-Science, 93-95, United States of America.

32. Watson, M. W. (1994). Vector autoregressions and cointegration, in R. F. ENGLE and D. L. McFadden, eds, 'Handbook of Econometrics, Vol. IV', North-Holland, Amsterdam, chapter 47, pp. 2843-2915.

33. Wei, W. S. (1997). Time series analysis. Addism Wesley, 316-322, New York.

Figure 1-1~1-4 : Phase spectra of $\hat{\sigma}_i$ and volume, for i=1,2,3,4

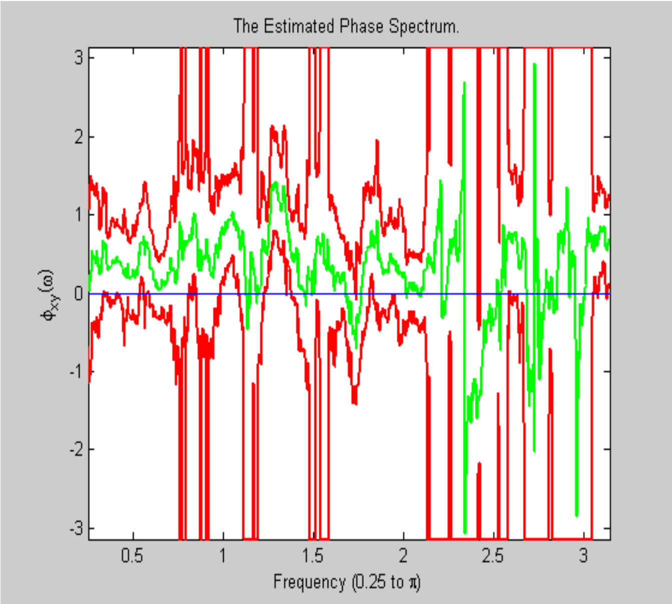


Figure 1-1

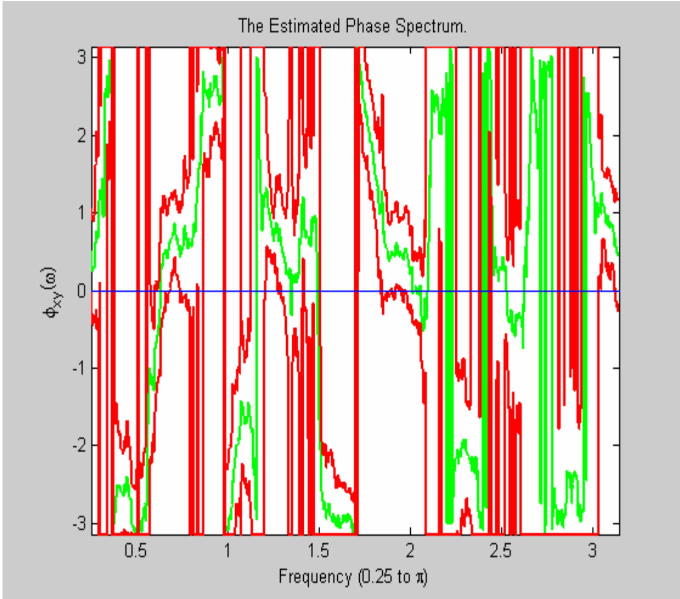


Figure 1-2

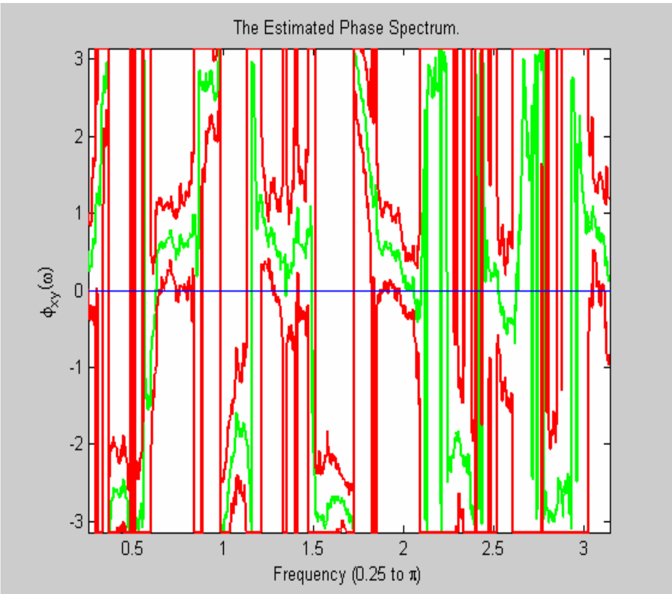


Figure 1-3

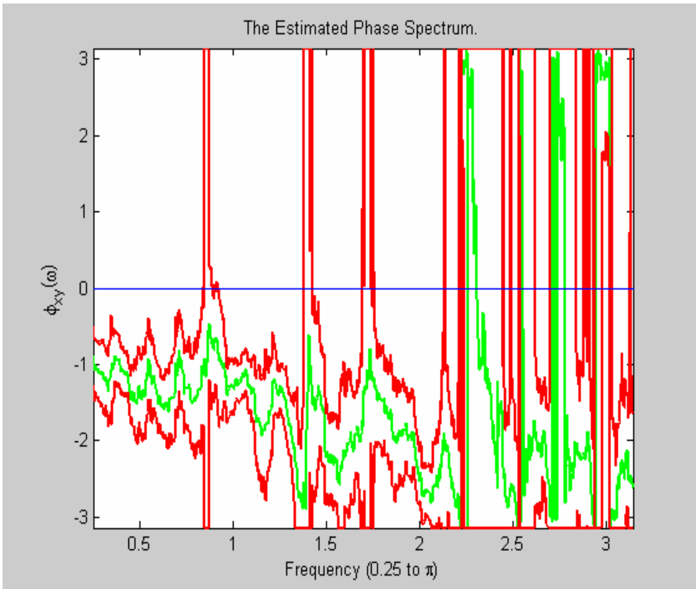


Figure 1-4

Figure 2-1~2-4 : Delay parameters of $\hat{\sigma}_i$ and volume, for i=1,2,3,4

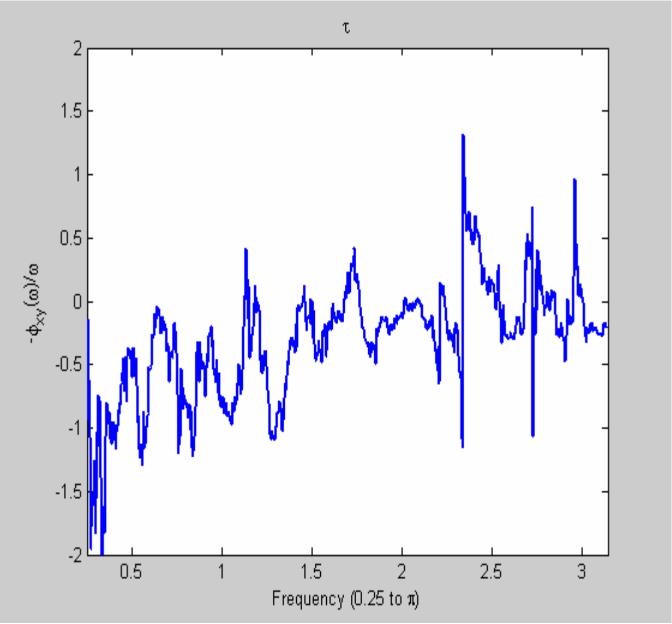


Figure 2-1

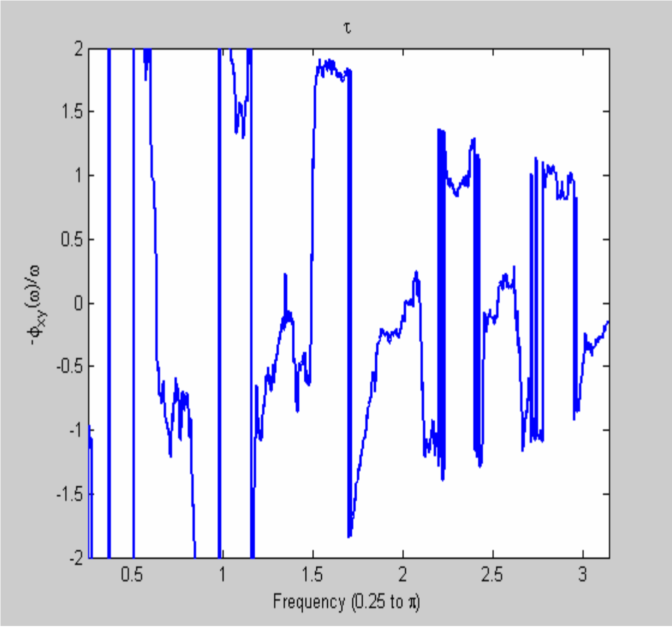


Figure 2-2

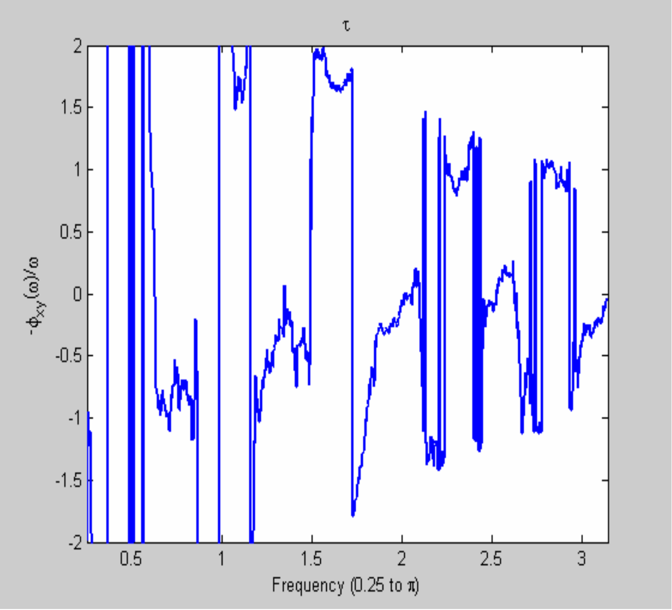


Figure 2-3

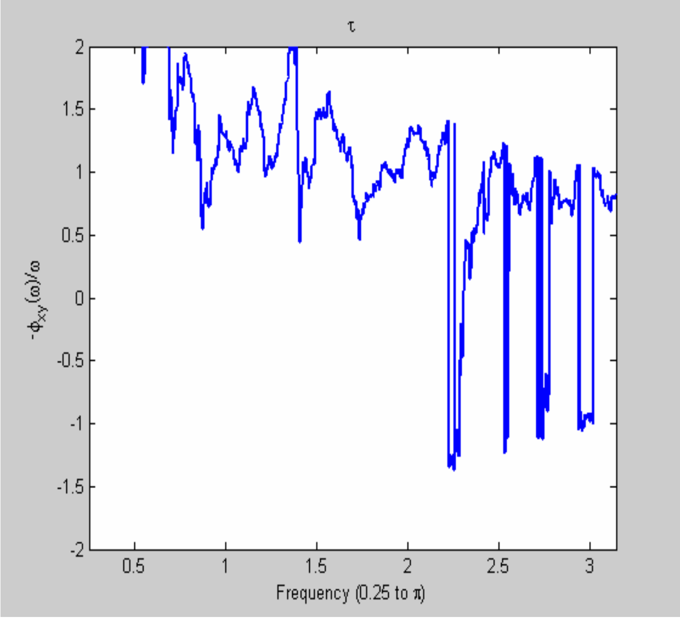


Figure 2-4

Figure 3-1~3-4 : Choerence spectra of $\hat{\sigma}_i$ and volume, for i=1,2,3,4

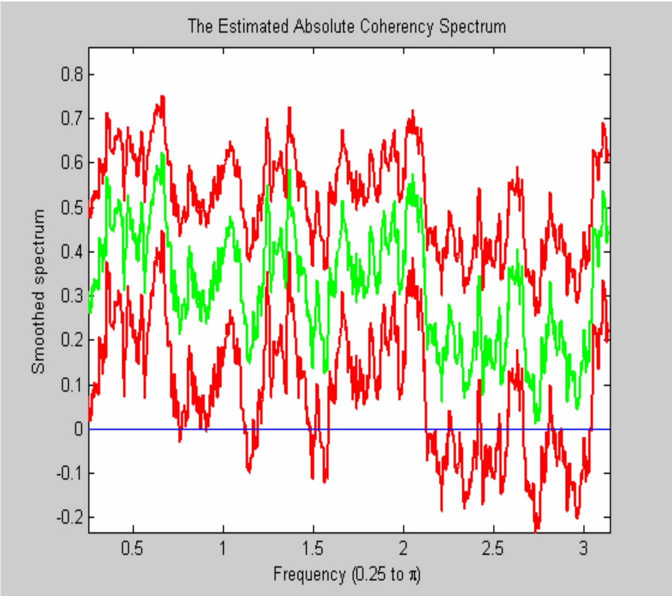


Figure 3-1

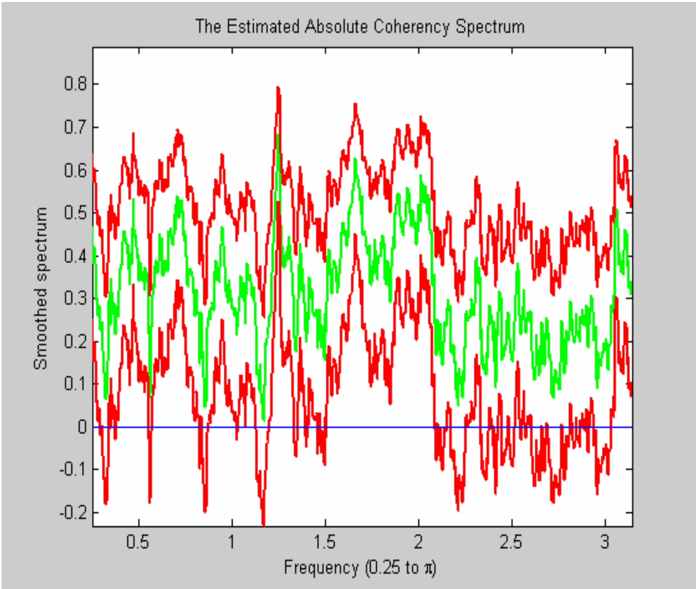


Figure 3-2

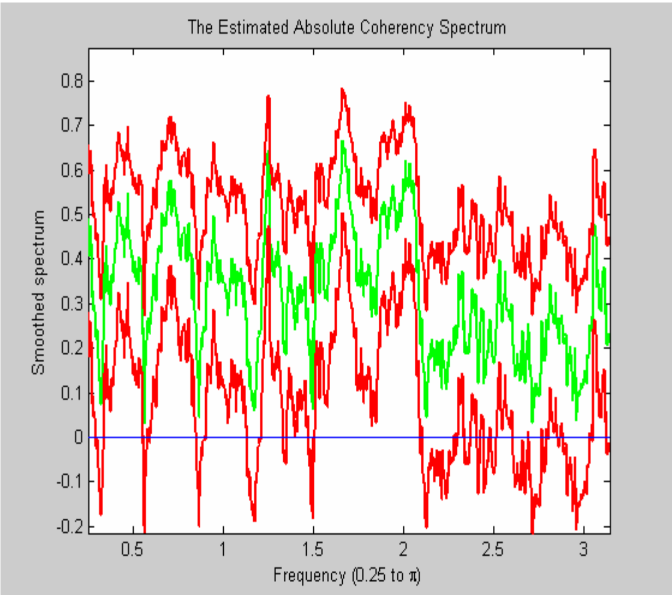


Figure 3-3

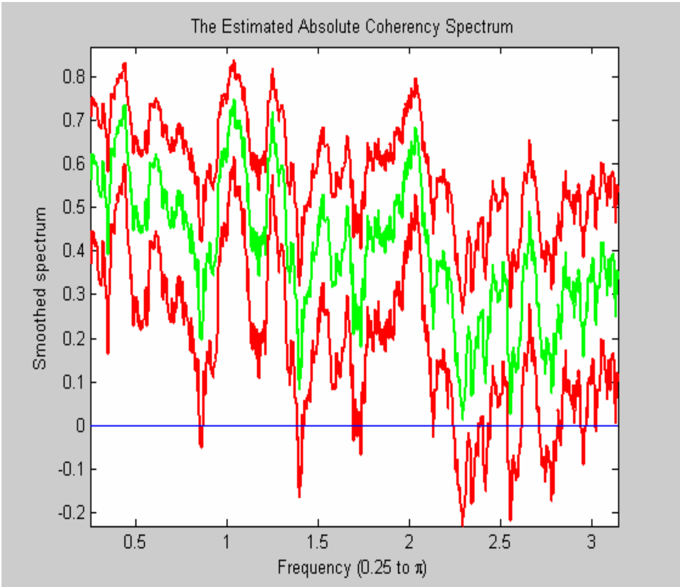
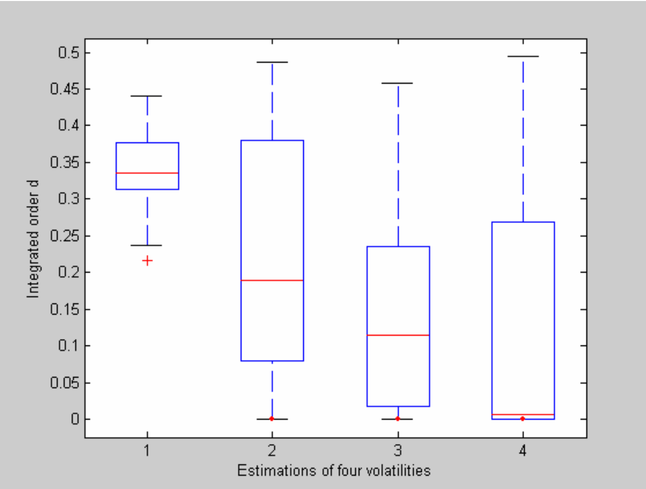


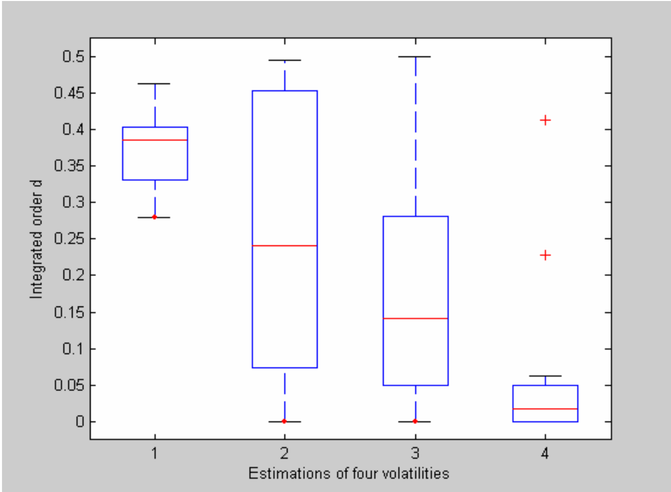
Figure 3-4

Figure 4 Boxplots of volatility $\hat{\sigma}_i$ fractional integrated orders, i = 1,2,3,4, in five classes

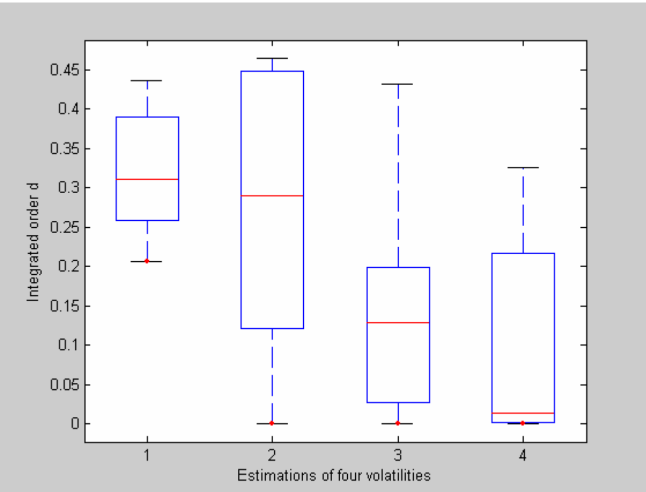
Class 1



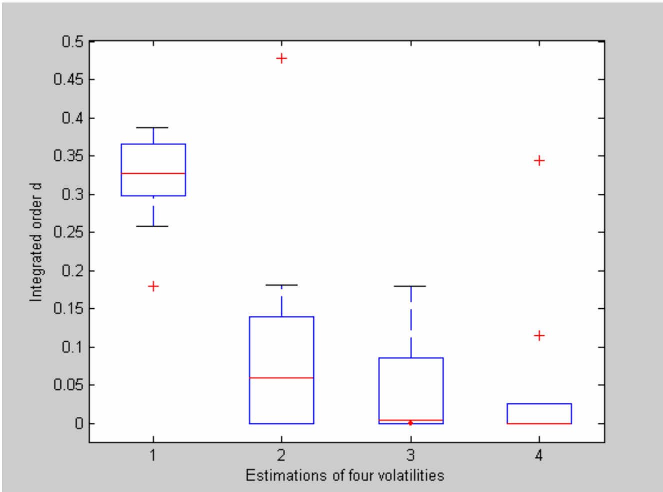
Class 2



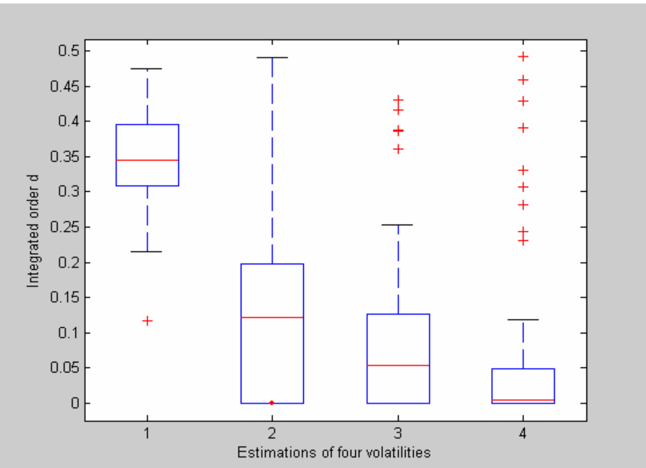
Class 3



Class 4



Class 5



Miscellanea

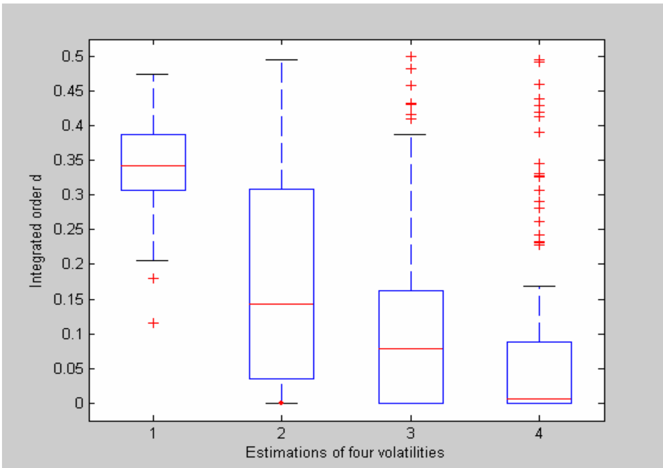
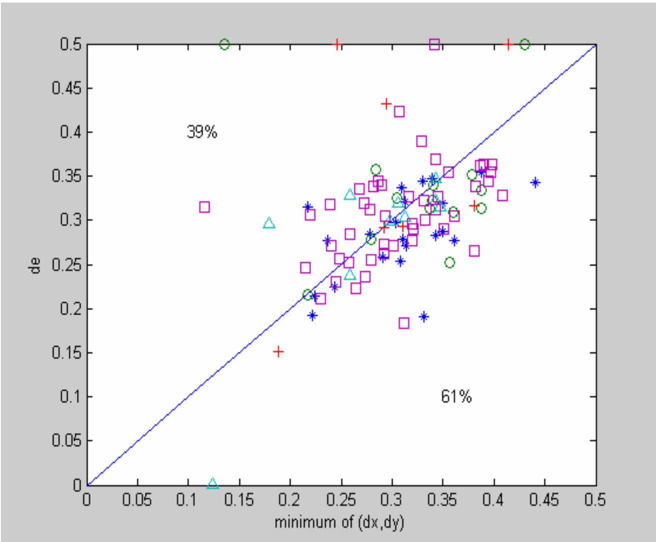


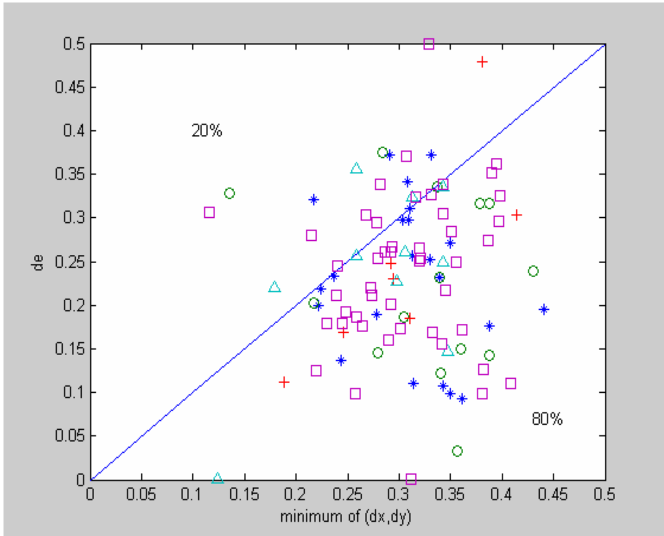
Figure 5 minimum of (d_x, d_y) versus d_e, x : volatility, y : volume, e : residual

Class 1 : * Class 2 : o Class 3 : + Class 4 : ^ Class 5 : square

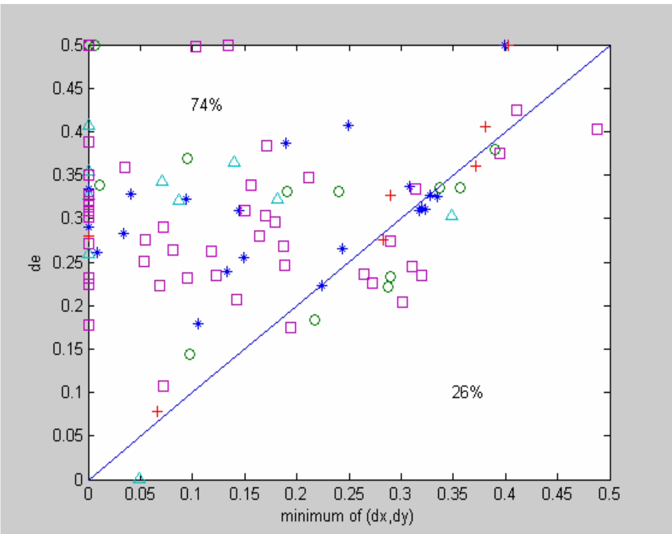
$\hat{\sigma}_1$ with OLS



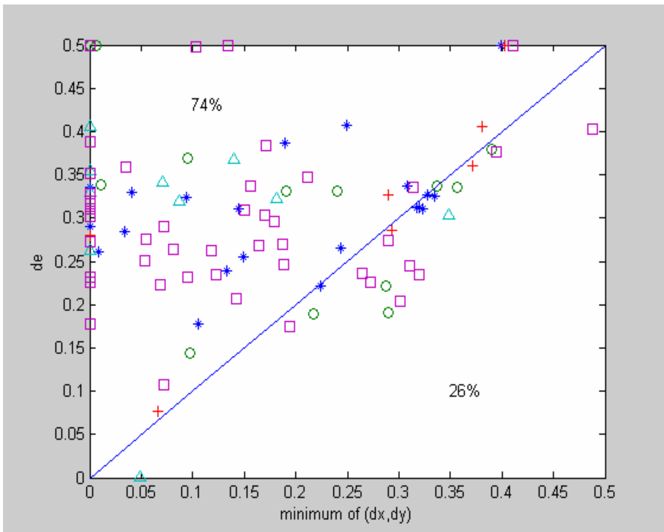
$\hat{\sigma}_1$ with FDLS



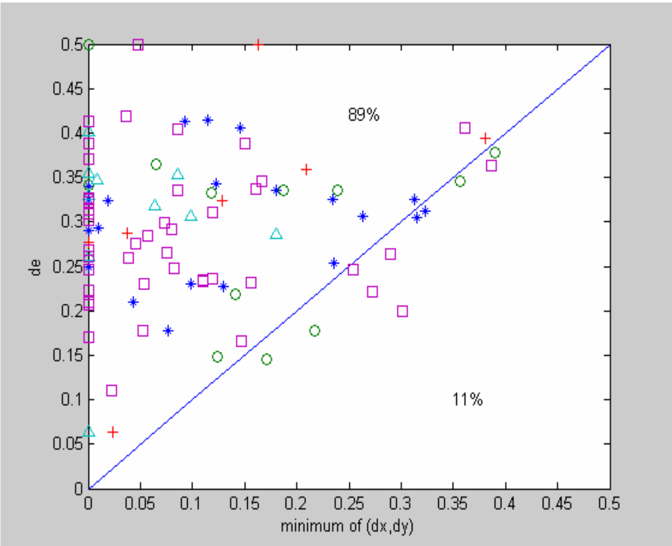
$\hat{\sigma}_2$ with OLS



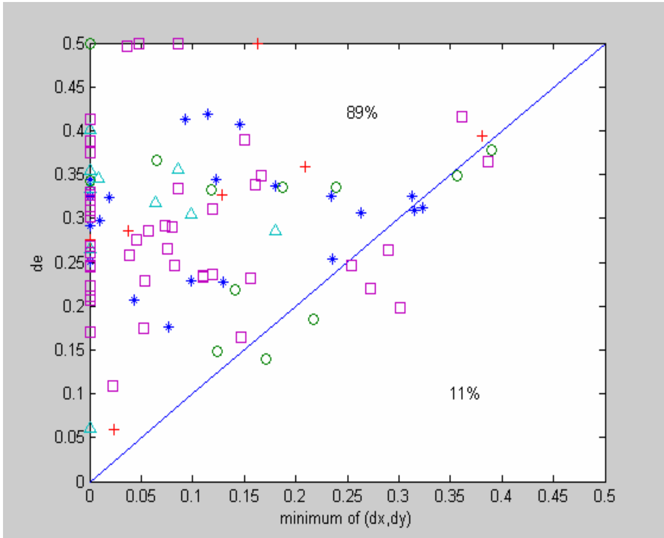
$\hat{\sigma}_2$ with FDLS



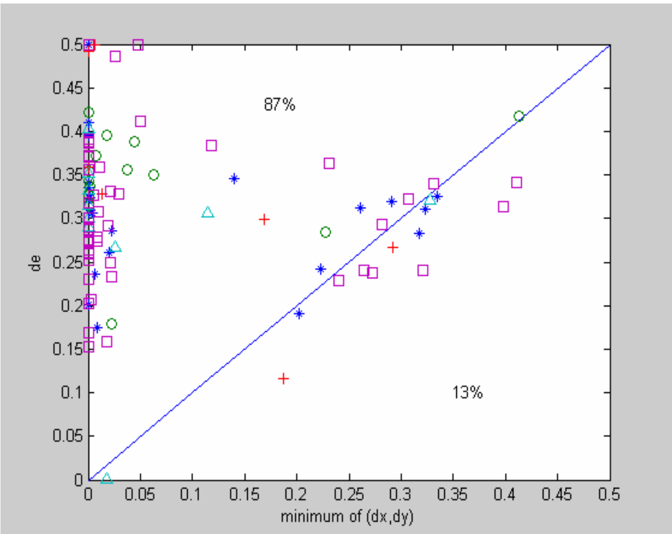
$\hat{\sigma}_3$ with OLS



$\hat{\sigma}_3$ with FDLS



$\hat{\sigma}_4$ with OLS



$\hat{\sigma}_4$ with FDLS

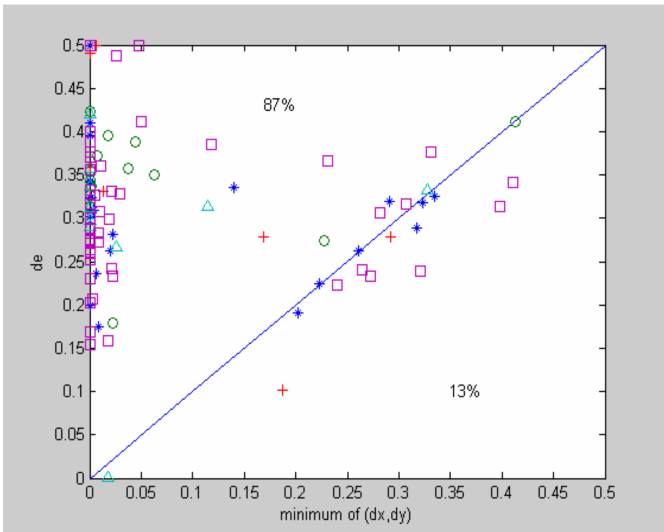


Table 1 Companies with $\hat{\sigma}_2$ and $\hat{\sigma}_3$ leads volume at frequencies $\omega =$ 0, 0.5, 1 and 1.7

ω	none	0.5	1	1.7	0.5 & 1	0.5 & 1.7	0.5, 1 & 1.7
class1	MSFT ORCL VIAB	INTC LU	T VZ		TXN	ROK	DIS EMC CSC CSCO HPQ NSM NXTL RSH SBC UIS CCU CL
Total	3	2	2	0	1	1	12
Percentage	14%	9%	9%	0%	4%	4%	57%

ω	none	0.5	1	1.7	0.5 & 1	0.5 & 1.7	0.5, 1 & 1.7
class2	GS LEH	AIG HIG	WFC	ONE	BNI CI	MER MWD	BAC C JPM
Total	2	2	1	1	2	2	3
Percentage	15%	15%	7%	7%	15%	15%	23%

ω	none	0.5	1	1.7	0.5 & 1	0.5 & 1.7	0.5, 1 & 1.7
class3	GM	NSC				DAL	AXP BA F FDX
Total	1	1	0	0	0	1	4
Percentage	14%	14%	0%	0%	0%	14%	57%

ω	none	0.5	1	1.7	0.5 & 1	0.5 & 1.7	0.5, 1 & 1.7
class4	BAX	DOW		AMGN			JNJ
	MEDI	MDT		HCA			MRK
				PFE			BMY
Total	2	2	0	3	0	0	3
Percentage	20%	20%	0%	30%	0%	30%	30%

ω	none	0.5	1	1.7	0.5 & 1	0.5 & 1.7	0.5, 1 & 1.7
class5	HNZ	BCC	EP	HET	ETR		AA
	LTD	GD	WMB		EXC		AEP
	MAY	MCD			TYC		AES
	PG						BDK
	SLE						EK
	SO						G
	TWX						GE
	WY						HAL
							HD
							HON
							IP
							KO
							MMM
							MO
							PEP
							RTN
							S
							SLB
							TOY
							UTX
							WMT
							XRX
							ATI
							AVP
							BHI
							ALL
							CPB
							DD

Total	8	3	2	1	3	0	28
Percentage	17%	6%	4%	2%	6%	0%	62%

Check percentages of companies with volatility leads volume at $\omega = 0.5$, 1 and 1.7

	sector1	sector2	sector3	Sector4	sector5	sector6	sector7	total
class1	14%	9%	9%	0%	4%	4%	57%	21
class2	15%	15%	7%	7%	15%	15%	23%	13
class3	14%	14%	0%	0%	0%	14%	57%	7
class4	20%	20%	0%	3%	0%	30%	30%	10
class5	17%	6%	4%	2%	6%	0%	62%	45

sector 1: non, sector 2: 0.5, sector 3: 1, sector 4: 1.7
sector 5: 0.5 & 1, sector 6: 0.5 & 1.7, sector 7: 0.5, 1 & 1.7

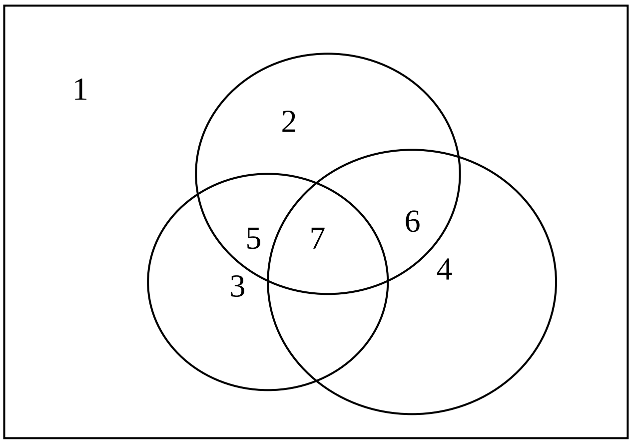


Table 2 Integrated order of the 4 volatility estimators and volume

Class	Company	n	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_3$	$\hat{d}_4$	$\hat{d}_e$
Class 1	CCU	2046	0.393	0.093	0.019	0.001	0.34
	CL	2039	0.329	0.259	0.130	0.494	0.224
	CSC	2039	0.237	0.149	0.099	0.006	0.279
	CSCO	2046	0.388	0.034	0.009	0.000	0.331
	DIS	2046	0.34	0.105	0.077	0.009	0.222
	EMC	2044	0.331	0.000	0.000	0.023	0.308
	HPQ	2046	0.303	0.418	0.315	0.328	0.318
	IBM	2046	0.35			0.314	0.361
	INTC	2046	0.291	0.000	0.000	0.004	0.35
	LU	2039	0.397	0.249	0.146	0.000	0.388
	MSFT	2046	0.335	0.447	0.263	0.291	0.31
	NSM	2046	0.33	0.367	0.313	0.439	0.335
	NXTL	2046	0.441	0.190	0.093	0.000	0.474
	ORCL	2046	0.373	0.040	0.000	0.000	0.343
	ROK	2038	0.217	0.484	0.235	0.140	0.328
	RSH	2046	0.351	0.009	0.000	0.020	0.309
	SBC	2039	0.35	0.430	0.114	0.000	0.399
	T	2046	0.313	0.144	0.123	0.261	0.392
	TXN	2039	0.314	0.487	0.458	0.419	0.323
	UIS	2039	0.342	0.133	0.043	0.000	0.278
	VIAB	1990	0.314	0.293	0.235	0.000	0.243
	VZ	2044	0.387	0.308	0.180	0.000	0.362
Class 2	AIG	2039	0.36	0.191	0.118	0.063	0.36
	BAC	2039	0.463	0.097	0.124	0.000	0.136
	BNI	2039	0.34	0.095	0.066	0.000	0.469
	C	2046	0.43	0.007	0.000	0.413	0.5
	CI	2044	0.28	0.290	0.172	0.228	0.313
	GS	1269	0.386	0.494	0.499	0.022	0.217
	HIG	2039	0.305	0.240	0.187	0.000	0.349
	JPM	2044	0.397	0.000	0.000	0.018	0.388
	LEH	2040	0.397	0.011	0.000	0.000	0.388
	MER	2039	0.283	0.474	0.410	0.045	0.39
	MWD	2046	0.393	0.494	0.481	0.037	0.357
	ONE	2039	0.34	0.288	0.141	0.000	0.353

	USB	2045	0.387			0.008	0.378
	WFC	2038	0.418	0.446	0.239	0.002	0.337
Class 3	AXP	2039	0.414	0.402	0.163	0.006	0.5
	BA	2046	0.206	0.067	0.023	0.232	0.188
	DAL	2039	0.436	0.464	0.432	0.000	0.381
	F	2046	0.246	0.283	0.038	0.168	0.315
	FDX	2039	0.318	0.000	0.000	0.326	0.292
	GM	2046	0.294	0.464	0.210	0.000	0.371
	NSC	2039	0.311	0.290	0.128	0.013	0.321
Class 4	AMGN	2046	0.375	0.000	0.000	0.000	0.347
	BAX	2039	0.179	0.087	0.064	0.115	0.322
	BMY	2039	0.298	0.181	0.098	0.000	0.329
	DOW	2044	0.387	0.049	0.000	0.017	0.124
	HCA	2044	0.306	0.478	0.180	0.000	0.348
	JNJ	2046	0.343	0.140	0.086	0.000	0.343
	MDT	2039	0.312	0.000	0.000	0.345	0.328
	MEDI	2039	0.348	0.000	0.000	0.026	0.258
	MRK	2039	0.258	0.000	0.000	0.000	0.389
	PFE	2044	0.365	0.071	0.009	0.000	0.343
Class 5	AA	2046	0.367	0.035	0.000	0.000	0.351
	AEP	2044	0.425	0.460	0.430	0.000	0.301
	AES	2046	0.474	0.188	0.054	0.492	0.32
	ALL	2046	0.367	0.490	0.110	0.391	0.265
	ATI	1101	0.311	0.314	0.166	0.021	0.314
	AVP	2039	0.44	0.452	0.156	0.000	0.319
	BCC	2039	0.267	0.000	0.000	0.307	0.326
	BDK	2039	0.271	0.000	0.000	0.243	0.24
	BHI	2039	0.412	0.069	0.000	0.000	0.245
	BUD	2039	0.383			0.381	0.331
	CPB	2039	0.345	0.187	0.076	0.000	0.28
	DD	2046	0.307	0.134	0.047	0.048	0.5
	EK	2046	0.116	0.055	0.082	0.118	0.3
	EP	2039	0.455	0.081	0.000	0.000	0.382
	ETR	2039	0.321	0.072	0.022	0.000	0.32
	EXC	2044	0.336	0.164	0.073	0.009	0.293
	G	2044	0.362	0.000	0.000	0.000	0.38
	GD	2038	0.239	0.000	0.000	0.282	0.341
	GE	2046	0.342	0.103	0.036	0.026	0.5

HAL	2039	0.346	0.179	0.046	0.000	0.316
HD	2044	0.423	0.418	0.387	0.000	0.29
HET	2039	0.278	0.072	0.057	0.000	0.291
HNZ	2039	0.362	0.118	0.038	0.000	0.273
HON	2046	0.258	0.477	0.416	0.330	0.273
IP	2039	0.319	0.096	0.110	0.029	0.23
KO	2046	0.343	0.000	0.000	0.000	0.343
LTD	2039	0.343	0.150	0.120	0.010	0.343
MAY	2039	0.215	0.000	0.000	0.019	0.269
MCD	2039	0.258	0.000	0.000	0.000	0.305
MMM	2046	0.328	0.411	0.361	0.000	0.5
MO	2046	0.341	0.194	0.147	0.018	0.219
PEP	2044	0.44	0.400	0.386	0.000	0.394
PG	2044	0.427	0.171	0.151	0.000	0.391
RTN	2044	0.355	0.310	0.254	0.231	0.401
S	2046	0.348	0.169	0.081	0.000	0.346
SLB	2046	0.396	0.054	0.000	0.005	0.332
SLE	2039	0.287	0.212	0.161	0.000	0.365
SO	2039	0.344	0.000	0.000	0.009	0.293
TOY	2039	0.309	0.000	0.000	0.003	0.248
TWX	2046	0.396	0.488	0.086	0.050	0.5
TYC	2045	0.386	0.000	0.000	0.459	0.398
UTX	2038	0.409	0.123	0.120	0.429	0.411
WMB	2038	0.381	0.143	0.052	0.000	0.396
WMT	2046	0.401	0.000	0.000	0.000	0.398
WY	2038	0.273	0.156	0.086	0.021	0.35
XOM	2046	0.282			0.011	0.358
XRX	2046	0.356	0.000	0.000	0.022	0.293

Table 3 Fractional cointegrated companies

1.Integrated order of $\hat{\sigma}_1$, volume and residual by OLS and FDLS

Class	Company	$\hat{d}_1$	$\hat{d}_v$	$\hat{d}_e$ *OLS*	$\hat{d}_e$ *FDLS*
Class1	CCU	0.393	0.340		0.232
	CL	0.329	0.224	0.215	0.219
	CSC	0.237	0.279		0.233
	CSCO	0.388	0.331	0.191	
	DIS	0.340	0.222	0.192	0.200
	EMC	0.331	0.308	0.254	
	HPQ	0.303	0.318	0.297	0.298
	IBM	0.350	0.361	0.287	0.270
	INTC	0.291	0.350	0.259	
	LU	0.397	0.388	0.355	0.176
	MSFT	0.335	0.310	0.278	
	NSM	0.330	0.335		0.252
	NXTL	0.441	0.474	0.343	0.195
	ORCL	0.373	0.343	0.283	0.108
	RSH	0.351	0.309		0.298
	SBC	0.350	0.399	0.320	0.099
	T	0.313	0.392		0.257
	TXN	0.314	0.323	0.271	0.111
	UIS	0.342	0.278		0.189
	VIAB	0.314	0.243	0.225	0.137
	VZ	0.387	0.362	0.277	0.093
Class 2	AIG	0.360	0.360	0.309	0.150
	BNI	0.340	0.469		0.122
	C	0.430	0.500		0.239
	CI	0.280	0.313	0.278	0.145
	GS	0.386	0.217	0.216	0.203
	HIG	0.305	0.349		0.187
	JPM	0.397	0.388	0.335	0.143
	LEH	0.397	0.388	0.314	0.317
	MWD	0.393	0.357	0.252	0.033
	ONE	0.340	0.353	0.322	0.232
	USB	0.387	0.378	0.351	0.317

	WFC	0.418	0.337	0.313	0.335
Class 3	AXP	0.414	0.500		0.304
	BA	0.206	0.188	0.152	0.111
	DAL	0.436	0.381	0.317	
	F	0.246	0.315		0.169
	FDX	0.318	0.292	0.291	0.247
	GM	0.294	0.371		0.230
	NSC	0.311	0.321	0.294	0.185
Class 4	AMGN	0.375	0.347	0.314	0.147
	BMY	0.298	0.329		0.228
	DOW	0.387	0.124	0.000	0.000
	HCA	0.306	0.348		0.261
	JNJ	0.343	0.343	0.323	0.336
	MDT	0.312	0.328	0.303	
	MEDI	0.348	0.258	0.238	0.257
	PFE	0.365	0.343		0.250
Class 5	AA	0.367	0.351	0.291	0.285
	AEP	0.425	0.301	0.271	0.174
	AES	0.474	0.320	0.290	0.253
	ALL	0.367	0.265	0.223	0.175
	ATI	0.311	0.314	0.183	0.000
	AVP	0.440	0.319	0.277	0.251
	BHI	0.412	0.245	0.231	0.179
	BUD	0.383	0.331	0.323	0.327
	CPB	0.345	0.280	0.255	0.253
	EP	0.455	0.382	0.339	0.126
	ETR	0.321	0.320	0.296	0.266
	EXC	0.336	0.293	0.273	0.261
	G	0.362	0.380	0.304	0.172
	GD	0.239	0.341		0.211
	GE	0.342	0.500		0.155
	HD	0.423	0.290		0.160
	HNZ	0.362	0.273	0.236	0.211
	HON	0.258	0.273	0.252	0.099
	IP	0.319	0.230	0.212	0.179
	KO	0.343	0.343		0.338
	LTD	0.343	0.343	0.312	0.306
	MCD	0.258	0.305		0.186

	MO	0.341	0.219		0.126
	PEP	0.440	0.394	0.345	0.362
	PG	0.427	0.391	0.364	0.351
	RTN	0.355	0.401	0.354	0.250
	S	0.348	0.346	0.326	0.217
	SLB	0.396	0.332	0.300	0.168
	SLE	0.287	0.365		0.261
	SO	0.344	0.293	0.262	0.201
	TOY	0.309	0.248		0.192
	TWX	0.396	0.500	0.354	0.295
	TYC	0.386	0.398	0.362	0.275
	UTX	0.409	0.411	0.328	0.110
	WMB	0.381	0.396	0.265	0.099
	WMT	0.401	0.398	0.363	0.325
	WY	0.273	0.350		0.220
	XRX	0.356	0.293		0.267

2.Integrated orders of $\hat{\sigma}_2$, volume and residual by OLS and FDLS

Class	Company	$\hat{d}_2$	$\hat{d}_v$	$\hat{d}_e$ *OLS*	$\hat{d}_e$ *FDLS*
Class 1	CL	0.259	0.224	0.222	0.222
	HPQ	0.418	0.318	0.310	0.312
	MSFT	0.447	0.320	0.312	0.312
	NSM	0.367	0.335	0.325	0.325
	ROK	0.484	0.328	0.326	0.326
	TXN	0.487	0.323	0.310	0.310
Class 2	CI	0.290	0.313	0.233	0.191
	GS	0.494	0.217	0.183	0.189
	MER	0.474	0.390	0.380	0.380
	MWD	0.494	0.357	0.335	0.335
	ONE	0.288	0.353	0.222	0.222
	WFC	0.446	0.337	0.336	0.336
Class 3	F	0.283	0.315	0.276	0.286
	GM	0.464	0.371	0.361	0.361
Class 4	DOW	0.049	0.124	0.000	0.000
	HCA	0.478	0.348	0.303	0.304

Class 5	AEP	0.460	0.301	0.204	0.203
	ALL	0.490	0.265	0.237	0.236
	AVP	0.452	0.319	0.235	0.234
	HD	0.418	0.290	0.275	0.275
	HON	0.477	0.273	0.226	0.226
	MO	0.194	0.219	0.175	0.174
	PEP	0.400	0.394	0.376	0.376
	RTN	0.310	0.401	0.244	0.244
	TWX	0.488	0.500	0.402	0.403

3.Integrated order of $\hat{\sigma}_3$, volume and residual by OLS and FDLS

Class	Company	$\hat{d}_3$	$\hat{d}_v$	$\hat{d}_e$ *OLS*	$\hat{d}_e$ *FDLS*
Class 1	HPQ	0.315	0.318	0.306	0.309
	TXN	0.458	0.323	0.312	0.313
Class 2	CI	0.172	0.313	0.145	0.140
	GS	0.499	0.217	0.178	0.184
	MER	0.410	0.390	0.379	0.378
	MWD	0.481	0.357	0.346	0.348
Class 5	AEP	0.430	0.301	0.199	0.197
	HD	0.387	0.290	0.264	0.264
	HON	0.416	0.273	0.221	0.220
	PEP	0.386	0.394	0.364	0.364
	RTN	0.254	0.401	0.246	0.246

4.Integrated order of $\hat{\sigma}_4$, volume and residual by OLS and FDLS

Class	Company	$\hat{d}_4$	$\hat{d}_v$	$\hat{d}_e$ *OLS*	$\hat{d}_e$ *FDLS*
Class 1	HPQ	0.328	0.317	0.283	0.288
	IBM	0.314	0.202	0.191	0.191
	NSM	0.439	0.334	0.325	0.325
	TXN	0.419	0.323	0.311	0.318
Class 2	C	0.413	0.500	0.417	0.411
Class 3	BA	0.232	0.187	0.117	0.102
	FDX	0.326	0.292	0.266	0.279

Class 5	AES	0.492	0.320	0.241	0.239
	ALL	0.391	0.265	0.241	0.241
	BDK	0.243	0.240	0.229	0.223
	HON	0.330	0.273	0.237	0.234
	TYC	0.459	0.398	0.313	0.313
	UTX	0.429	0.410	0.341	0.342

Summary table for fractional cointegration percentage

	OLS	FDLS
$\hat{\sigma}_1$	61%	80%
$\hat{\sigma}_2$	26%	26%
$\hat{\sigma}_3$	11%	11%
$\hat{\sigma}_4$	13%	13%

Table 4 Gnneralized Whittle estimators for fractional cointegration

Company	$\hat{d}_1$	$\hat{d}_e$	β	class
CPB	0.438	0.332	0.0005	5
ETR	0.443	0.354	-0.0012	5
G	0.426	0.392	-0.0013	5
XRX	0.475	0.400	-0.0011	5

Table 5 Simulation results of generalized Whittle estimators

We let $\beta = 1$ and $d_e = 0$

d	$\sigma_e = 0.5$		$\sigma_e = 1$	
0.2	0.1948	(0.0674)	0.1910	(0.0809)
	0.0206	(0.0875)	0.0009	(0.0863)
	1.0144	(0.1647)	1.0370	(0.3089)
0.3	0.2961	(0.0884)	0.2904	(0.0864)
	0.0222	(0.0758)	0.0253	(0.0944)
	0.9894	(0.1392)	1.0530	(0.2989)
0.4	0.3894	(0.0891)	0.3970	(0.0899)
	-0.0019	(0.0906)	-0.0020	(0.0803)
	0.9924	(0.0644)	1.0185	(0.2074)

Table 6 Regression model with pre-whiting

AXP	MA(1)	BCC	WN	BAC	MA(1)	AA	WN
BAX	MA(1)	BNI	WN	BAX	MA(1)	AMGN	WN
BDK	MA(1)	DD	WN	BDK	MA(1)	ATI	WN
BMY	MA(1)	GD	WN	BHI	MA(1)	AXP	WN
C	MA(1)	GE	WN	BMY	MA(1)	BCC	WN
CCU	MA(1)	HCA	WN	C	MA(1)	BNI	WN
CSC	MA(1)	HET	WN	CCU	MA(1)	CSCO	WN
EK	MA(1)	MAY	WN	CPB	MA(1)	DD	WN
F	MA(1)	MCD	WN	EK	MA(1)	ETR	WN
GM	MA(1)	MER	WN	EMC	MA(1)	GD	WN
HAL	MA(1)	MRK	WN	FDX	MA(1)	GE	WN
HD	MA(1)	NSM	WN	G	MA(1)	HIG	WN
HIG	MA(1)	ROK	WN	HAL	MA(1)	HNZ	WN
KO	MA(1)	RSH	WN	HET	MA(1)	INTC	WN
MMM	MA(1)	SLE	WN	IP	MA(1)	JNJ	WN
MO	MA(1)	T	WN	KO	MA(1)	JPM	WN
PFE	MA(1)	WY	WN	LU	MA(1)	MAY	WN
TOY	MA(1)	XOM	WN	MMM	MA(1)	MCD	WN
UIS	MA(1)	XRX	WN	MRK	MA(1)	MDT	WN
		BAC	x	PFE	MA(1)	NSM	WN
				TOY	MA(1)	NXTL	WN
				TYC	MA(1)	ORCL	WN
				UIS	MA(1)	PEP	WN
				LEH	x	PG	WN
				LTD	x	RSH	WN
				MEDI	x	SBC	WN
				ROK	x	SLB	WN
						SLE	WN
						SO	WN
						T	WN
						UTX	WN
						VIAB	WN
						WMT	WN
						WY	WN

AEP	MA(1)	AA	WN	MCD	WN
AES	MA(1)	AIG	WN	MDT	WN
BAC	MA(1)	ALL	WN	MER	WN
BMY	MA(1)	AMGN	WN	MMM	WN
C	MA(1)	ATI	WN	MRK	WN
CL	MA(1)	AVP	WN	NSC	WN
CPB	MA(1)	AXP	WN	NSM	WN
CSC	MA(1)	BDK	WN	NXTL	WN
DIS	MA(1)	BHI	WN	ONE	WN
EK	MA(1)	BNI	WN	ORCL	WN
F	MA(1)	CCU	WN	PEP	WN
FDX	MA(1)	CI	WN	PFE	WN
GM	MA(1)	CSCO	WN	PG	WN
GS	MA(1)	DAL	WN	ROK	WN
HAL	MA(1)	DD	WN	RTN	WN
HD	MA(1)	EMC	WN	SBC	WN
IP	MA(1)	ETR	WN	SLB	WN
KO	MA(1)	EXC	WN	SLE	WN
LTD	MA(1)	G	WN	SO	WN
MEDI	MA(1)	GD	WN	TWX	WN
MSFT	MA(1)	GE	WN	VIAB	WN
S	MA(1)	HCA	WN	VZ	WN
T	MA(1)	HET	WN	WFC	WN
TOY	MA(1)	HIG	WN	WMT	WN
TYC	MA(1)	HNZ	WN	WY	WN
UIS	MA(1)	INTC	WN	BAX	WN
XRX	MA(1)	JNJ	WN	BCC	WN
HPQ	MA(1)	JPM	WN	LU	WN
		LEH	WN	MAY	WN
				RSH	WN

Summary table of percentage for WN and MA(1)

	WN	MA(1)
$\hat{\sigma}_1$	50%	50%
$\hat{\sigma}_2$	60%	40%
$\hat{\sigma}_3$	66%	34%