Modeling international market correlations with high frequency data

Carmelo Genovese

Dipartimento di Fisica, Università di Pisa.

Roberto Renò

Dipartimento di Economia Politica, Università di Siena, Piazza S. Francesco 7, 53100, Siena, Italy.

Abstract. In this paper we describe a method for computing correlations with unevenly sampled data. Theory and Monte Carlo evidence suggest that the proposed method is well suited to the unevenly spaced time structure of high frequency data. We illustrate the method with an application to two years of high frequency financial data, the transactions on the Italian and French stock index futures market in 2000-2001. We show that correlations decrease when increasing sampling frequency (Epps effect) and we propose a dynamic model for observed correlations.

Keywords: Correlations, Epps effect, high-frequency data, futures market

1. Introduction

Correlations, as well as models of correlated data, are of tantamount importance in finance. As a simple example, the main results in asset allocation theory or portfolio management rely on the knowledge of the variance-covariance matrix of asset prices.

Recently, the advent of electronic systems and the increasing number of transactions in international financial markets provides researchers with a huge quantity of transaction data, with a typical frequency of few seconds. Typically these intraday transactions are known as high-frequency data, when compared to daily or lower frequencies. The new frequencies of data poses new econometric problems.

In this paper, we discuss the measurement of correlations with high-frequency data. We show that the main problem in this field is that financial transaction prices are unevenly sampled, since there is no reason for two different assets to trade at the same time. This non-sinchronicity leads to difficulties in implementing the classical Pearson estimator.

We discuss an estimator based on Fourier analysis, and we show, using simulated data, that it performs better than the classical estimator. Moreover, we show that when estimating correlations with unevenly sampled data we get a systematic bias in correlation measurements which is increasing when increasing the frequency of sampling (*Epps effect*).

We then illustrate the method by implementing it for the estimation of correlations among international futures prices, namely the Italian and French stock index futures (FIB30 and CAC40 Futures respectively). We show that correlations are inherently stochastic and that they show a high degree of persistence. We propose a diffusion model which takes into account such correlation, and we study the dependence of correlation on market volatility.

This paper is structured as follows: Section 2 describes the theory; Section 3 illustrates the performance of the estimator on Monte Carlo experiments; Section 4 shows the results on the data set; Section 5 concludes.

2. Measuring correlations

Since financial data are modeled in continuous time as semimartingales, in order to measure correlations we need the theory of quadratic variation, see e.g. Jacod and Shiryaev (1987). Let us denote by $S(t)$ the asset price at time t, and by $p(t) = \log S(t)$. The asset can be an equity price, a futures, an exchange rate, the price of any financial activity. We assume that $p(t) \in \mathbb{R}^d$ is the solution of the following stochastic differential equation (SDE):

$$\begin{cases} dp(t) = \mu(t)dt + \sigma(t)dW(t) \\ p(0) = x \end{cases} \tag{2.1}$$

where $\mu(t) : \mathbb{R} \to \mathbb{R}^d$ and $\sigma(t) : \mathbb{R} \to \mathbb{R}^d \times \mathbb{R}^d$ are time dependent stochastic functions such that a unique solution of the SDE exists in the interval $[0, 2\pi]$, and $W(t)$ is an $\mathbb{R}^d$-valued Brownian motion. In this case, with continuous trajectories, quadratic variation is simply:

$$[p,p](t) = \int_0^t \sigma^T(s)\sigma(s)ds \tag{2.2}$$

We will then set:

$$\Sigma_{ij}(t) = \sum_{k=1}^{d} \sigma_{ik}(t)\sigma_{kj}(t). \tag{2.3}$$

Quadratic variation can be estimated via the following theorem:

THEOREM 2.1. *Let X and Y be two semimartingales. Then for every Riemann sequence $\tau_{n,m}$ of adapted subdivisions, the process $S_{\tau_n}(X,Y)$ defined by:*

$$S_{\tau_n}(X,Y)_t = \sum_{m\geq 1} \left(X_{\tau_{n,m+1}\wedge t} - X_{\tau_{n,m}\wedge t}\right)\left(Y_{\tau_{n,m+1}\wedge t} - Y_{\tau_{n,m}\wedge t}\right) \tag{2.4}$$

converges, for $m \to \infty$, to the process $[X,Y]$, in probability and uniformly on every compact interval.

Proof. See Jacod and Shiryaev (1987). □

This theorem leads directly to the definition of realized co-volatility. Consider a subdivision $(t_i)_{1\leq i\leq n}$ of the interval $[0,T]$ with $t_i = in/T$. We then have:

$$\int_0^T \Sigma_{ij}(s)ds = \lim_{n\to\infty} \sum_{i=2}^{n} \left(p(t_i) - p(t_{i-1})\right)^2, \tag{2.5}$$

where the above limit is in probability. Realized volatility is now a very popular tool, see e.g. Andersen *et al.* (2002) and the references therein, while see Andersen *et al.* (2003) for a review of parametric and nonparametric methods for estimating volatility.

In this paper, we adopt instead an estimator based on Fourier analysis. We define the Fourier coefficients of the i-th component dp_i in the usual way:

$$\begin{aligned} a_0^i(dp) &= \frac{1}{2\pi}\int_0^{2\pi} dp_i(t) \\ a_k^i(dp) &= \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dp_i(t) \\ b_k^i(dp) &= \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dp_i(t), \end{aligned} \tag{2.6}$$

and similar formulas hold for $a_k(\Sigma_{ij}), b_k(\Sigma_{ij})$; from the Fourier coefficients of Σ_{ij}, $\Sigma_{ij}(t)$ can be obtained pointwise by the Fourier-Fejer inversion formula:

$$\Sigma_{ij}(t) = \lim_{n\to\infty}\sum_{k=0}^{n}\left(1-\frac{k}{n}\right)\cdot\left[a_k(\Sigma_{ij})\cos(kt) + b_k(\Sigma_{ij})\sin(kt)\right]. \tag{2.7}$$

We then have the following:

THEOREM 2.2. *Consider a process $p(t)$ satisfying (2.1) , and define the Fourier coefficients of dp and Σ as in 2.6. Given an integer $n_0 > 0$, we have almost surely:*

$$a_0(\Sigma_{ij}) = \lim_{N\to\infty}\frac{\pi}{N+1-n_0}\sum_{k=n_0}^{N}\frac{1}{2}\left(a_k^i(dp)a_k^j(dp) + b_k^i(dp)b_k^j(dp)\right) \tag{2.8}$$

$$\begin{aligned} a_q(\Sigma_{ij}) = \lim_{N\to\infty}\frac{\pi}{N+1-n_0}\sum_{k=n_0}^{N}\frac{1}{2} &\Big(a_k^i(dp)a_{k+q}^j(dp) + a_k^j(dp)a_{k+q}^i(dp) + \\ &+ b_k^i(dp)b_{k+q}^j(dp) + b_k^j(dp)b_{k+q}^i(dp)\Big) \end{aligned} \tag{2.9}$$

$$\begin{aligned} b_q(\Sigma_{ij}) = \lim_{N\to\infty}\frac{\pi}{N+1-n_0}\sum_{k=n_0}^{N}\frac{1}{2} &\Big(a_k^i(dp)b_{k+q}^j(dp) + a_k^j(dp)b_{k+q}^i(dp) + \\ &+ b_k^i(dp)a_{k+q}^j(dp) + b_k^j(dp)a_{k+q}^i(dp)\Big) \end{aligned} \tag{2.10}$$

Proof. This is a slightly extended version of the main theorem in Malliavin and Mancino (2002), with almost the same proof. □

Thus, formulas (2.8-2.10) allow to compute the Fourier coefficients of Σ_{ij} from the Fourier coefficients of dp_i, then to reconstruct $\Sigma_{ij}(t)$ via (2.7). However, we are not interested in the pointwise covariance matrix Σ_{ij}, but in its integrated value over the time window, equation (2.2). In our framework, this is easily given by:

$$\hat{\sigma}_{ij}^2 = 2\pi a_0(\Sigma_{ij}), \tag{2.11}$$

where $a_0(\Sigma_{ij})$ is given by (2.8).
Implementation of this estimator is straightforward, see e.g Barucci and Renò (2002a),Barucci and Renò (2002b). The coefficients of dp are computed via integration by parts:

$$a_k(dp_i) = \frac{1}{\pi}\int_0^{2\pi}\cos(kt)dp_i(t) = \frac{p(2\pi)-p(0)}{\pi} - \frac{k}{\pi}\int_0^{2\pi}\sin(kt)p_i(t)dt. \tag{2.12}$$

In financial markets, $p_i(t)$ is not observed continuously, but it is unevenly sampled in the form of tick-by-tick observations, $p_i(t_k), k = 1, \ldots, T$. Thus, we need to make an assumption on the interpolation of prices when computing the integrals in (2.12); we use $p(t) = p(t_j)$ where t_j is the largest observation time before t. For all computations, we set $n_0 = 1$.
Formulas (2.8-2.10) cannot be computed at all orders and have to be stopped at some frequency N. We discuss the choice of N in the next Section.

3. Performance on simulated data

In order to illustrate the potentiality of the method outlined in the previous section, we will resort to Monte Carlo simulation of high frequency asset prices, extending the framework of Barucci and Renò (2002a) to a bi-variate analysis.
As in Renò (2003), we simulate two correlated asset price diffusions with the bi-variate continuous GARCH(1,1) model:

$$\begin{aligned} dp_1(t) &= \sigma_1(t)dx_1(t), \\ dp_2(t) &= \sigma_2(t)dx_2(t), \\ d\sigma_1^2(t) &= \lambda_1[\omega_1 - \sigma_1^2(t)]dt + \sqrt{2\lambda_1\theta_1}\sigma_1^2(t)dx_3(t), \\ d\sigma_2^2(t) &= \lambda_2[\omega_2 - \sigma_2^2(t)]dt + \sqrt{2\lambda_2\theta_2}\sigma_2^2(t)dx_4(t), \\ corr(dx_1, dx_2) &= \rho, \end{aligned} \tag{3.1}$$

and all other correlations between Brownian motions $x_1(t), x_2(t), x_3(t), x_4(t)$ set to zero. The choice of this particular model comes from the fact that it is the continuous time limit of the very popular GARCH(1,1) model, and it has been studied extensively in the literature, e.g. see Kroner and Ng (1998). We will use the parameter values estimated by Andersen and Bollerslev (1998) on foreign exchange rates data, i.e. $\theta_1 = 0.035, \omega_1 = 0.636, \lambda_1 = 0.296, \theta_2 = 0.054, \omega_2 = 0.476, \lambda_2 = 0.480$. We will instead analyze two mirror cases for the correlation coefficient: $\rho = 0.35$ and $\rho = -0.35$.
To get a representation of high-frequency tick-by-tick data, after discretizing (3.1) by a first-order Euler discretization scheme with a time step of one second, we extract observation times drawing the durations from an exponential distribution with mean 60 seconds. Observation times are the same for the two time series, thus we avoid the problems linked to the Epps effect, which appears when increasing frequency due to non-synchronous quotes, as will be discussed later.
After simulating the process (3.1), we compute daily (86400 seconds, corresponding to 24 hours of trading) variance-covariance matrix according to the Fourier method, and according to the realized volatility measure (2.5) The choice of n in (2.5) poses a problem similar to the choice of N in the Fourier method: it comes from a tradeoff between increasing precision and cutting out microstructure distortions. A typical value is $n = 288$, corresponding to five minute returns. Since the time series $p(t)$ is not observed in continuous time, one has to resort to interpolation techniques to obtain the values $p\left(\frac{i}{n}\right)$ in (2.5) at equally spaced times, when these values are not observed directly. Two alternatives have been followed in the literature: linear interpolation between adjacent observations, and previous-tick interpolation, i.e. the price at time t is set equal to the price of the last observation.
We implement both these interpolation schemes, for $n = 288, 96, 48$ (corresponding respectively to 5,15,30 minute returns), when measuring correlations on Monte Carlo experiments. Table 1 shows the results. First of all we notice that the Fourier estimator performs considerably better than realized volatility, which is biased toward zero. The bias in the correlation

Table 1. Average correlation on 10,000 Monte Carlo replications of the model (3.1). Two generated values of the correlation are considered, $\rho = 0.35$ and $\rho = -0.35$. We compute the variance-covariance matrix via the Fourier estimator ($N = 160$ for the first series, $N = 500$ for the second and $N = 160$ for the covariance) and via the realized volatility estimator (2.5). L.I. means Linear Interpolation, while P.T. means Previous Tick interpolation. Standard deviations of in-sample measurements are reported in the columns named Std. The standard error on the mean is the standard deviation divided by 100.

Estimator	Generated correlation $\rho = 0.35$		Generated correlation $\rho = -0.35$	
	Measured	Std	Measured	Std
Fourier	0.350	0.039	-0.349	0.039
Realized $5'$, L.I.	0.204	0.058	-0.203	0.055
Realized $5'$, P.T.	0.181	0.060	-0.180	0.058
Realized $15'$, L.I.	0.338	0.090	-0.337	0.090
Realized $15'$, P.T.	0.329	0.091	-0.328	0.092
Realized $30'$, L.I.	0.345	0.127	-0.344	0.126
Realized $30'$, P.T.	0.342	0.127	-0.341	0.126

measurement of realized volatility is more and more severe as the sampling frequency increases. For the five-minute estimator with the previous-tick interpolation, we get a mean value of 0.181 (-0.180), which is quite far from the true value of 0.35 (-0.35); this bias is completely due to the use of interpolated prices on the evenly spaced grid. Realized volatility with linearly interpolated returns is closer to the right value, but this is because of the downward bias in the volatility measurement due to the linear interpolation documented in Barucci and Renò (2002a,b). In these papers, the authors show that the spurious positive serial correlation induced by the linear interpolation technique lowers the volatility estimates. Since variances are spuriously measured to be lower, correlations turn out to be spuriously higher, thus compensating in some way the bias due to non-synchronicity. This is also true, but to a much lower extent, for the 15-minute and 30-minute realized volatility estimator. The precision of the Fourier estimator, as measured by the standard deviation of measurements across Monte Carlo replications, is always better than that obtained with the realized volatility estimator. We implemented the Fourier estimator with $N = 500$ coefficients for the first time series, $N = 160$ coefficients for the second and $N = 160$ coefficients for the computation of covariance. It is worth noting that with this choice the estimator of the correlation is not guaranteed to be positive, while using the same N for the whole matrix guarantees a positive estimator, as it is straightforward to show.

Increasing the number of coefficients N would even increase the measurement precision. On the other hand, even the gain in precision of the realized covariance measurement obtained when increasing the sampling frequently is canceled out by the bias.

We want now to study the impact of two main features of high-frequency data: the fact that intraday asset prices are recorded in form of tick-by-tick transactions or quotes, which are unevenly spaced and whose frequency depends on the liquidity of the asset, and the fact that correlations may be lagged, due to different liquidity, economic significance or recording effects. Using the Monte Carlo simulation, we should be able to disentangle the impact of these two effects on correlation measurements. We will then build four different simulated time series using model (3.1), introducing asynchronous data and lagging, see

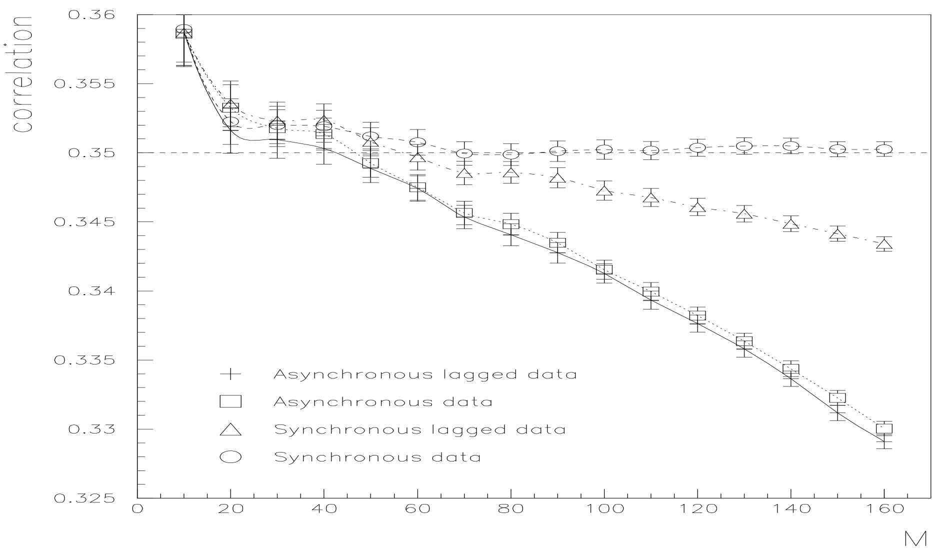

Figure 1. Average correlation between two Monte Carlo simulations of asset prices, according to (3.1), as a function of the sampling frequency N in (2.8). Boxes: simulated observation times are drawn independently. Crosses: simulated observation times are drawn independently and the correlation is lagged by 8 seconds. Triangles: simulated observation times are forced to be the same and the correlation is lagged by 8 seconds. Circles: simulated observation times are forced to be the same, no lag. The generated value of the correlation is $\rho = 0.35$, corresponding to the dashed line in the figure. Error bars are computed according to the Normal distribution. These results are obtained with 10,000 replications.

Renò (2003) for details.
Figure 1 shows the resulting average daily realized correlation as a function of the sampling frequency N used in the computation, for the four different samples. Let us look at the asynchronous-lagged sample first, which is thought to be closer to actual data. Even if we do not go in the deep high frequency regime, the Epps effect (Epps, 1979) is clearly displayed. Correlation begins to drop above a certain frequency, going far from the "true" generated value. If we increase the sampling frequency, then correlation goes to zero (not displayed). If we look at asynchronous (not lagged) data, we see that the Epps effect is still present, but slightly less relevant. It is evident that, in this Monte Carlo setting, microstructure effects are not present, so we can see that asynchronous quoting can be a very important factor explaining the reduction of the correlation measurements by itself.
We now turn to the analysis of synchronous data. From Figure 1, we observe that the Epps effect is dramatically reduced. However, we find another relevant source of the Epps effect. When we consider lagged data, we still observe a substantial drop in correlation, even if it appears to be smaller than that caused by non-synchronicity. The effect is visible even if the introduced lag, i.e. 8 seconds, is very small, at least when compared with the frequency $N = 160$, which corresponds to 3.5 minutes in the time domain.
In contrast, if data are synchronous and not lagged, we clearly observe that correlations do

not drop at higher frequencies. Moreover, as expected, by increasing the sampling frequency we increase the measurement precision.
While Monte Carlo experiments can be useful in assessing the theoretical performance of the estimator, we need to implement it on real data to check its reliability; that is done in the next Section.

4. Data analysis

Our data set consists in all tick-by-tick transactions on the FIB30 contract and the CAC40 futures contract from January 2001 to December 2002. Both data samples consist in all the registered transactions. We have in total 5,236,000 (10,577 per day) transactions for the FIB30 and 5,971,000 (12,063 per day) transactions for the CAC40. FIB30 and CAC40 are the futures contract on MIB30, the weighted sum of 30 major Italian stocks according to their capitalization, and the 40 major French stocks respectively. We always consider nearest maturity contract; the FIB30 has quarterly expirations, while CAC40 has monthly expirations. Both contracts are by far the most liquid contract on stocks, being their volume more than the total volume traded on the individual stocks themselves. In order to assess correlation measurements, we consider only transactions in contemporaneous hours, from 9:15 to 17:30. We will not consider days in which one of the two markets is closed. In total we have 495 trading days.
Figure 2 shows the value of the futures contract for the Italian and the French contract. The correlation between the two time series is evident, and indeed the sample correlation coefficient between returns is 0.8232.
The correlation is much higher for volatilities, displayed in Figure 3. Daily volatilities have been computed with the Fourier method, using all intraday data, with $N = 25$. Correlation is visually very evident; the linear correlation coefficient is 0.9116. Figure 4 shows the scatter plot. The Italian and the French market look like integrated markets, affected by common factors, especially for volatility. In both figures 3 and 4 is evident the impact of 11 September 2001.
Finally, we compute covariances and correlations with the Fourier method. When computing correlations, we cannot use a value of N too large because of the Epps effect. The Epps effect is illustrated by Figure 5. When increasing the sample frequency N, average correlation decreases significantly. In the previous Section, a motivation for this effect is suggested: non-synchronicity of price records, as well as lead-lag relationship, even if of the order of few seconds. We then repeat correlation measurements by using only prices that are synchronous, that is recorded in the same second. This kind of quotes are nearly 20% of the whole sample. The result is the dashed line in Figure 5, and we found no differences with respect to the previous case. This suggests that for international assets, like those considered in this paper, the speed of transmission of information is slow when compared to the typical trading frequency. Then, even if the markets are nearly perfectly integrated, correlations need some time to fully establish.
Anyway, the Epps effect forces us to use a small value of N, which implies a quite low precision in measuring our variance-covariance matrix.
We then measure correlation at $N = 25$ and $N = 300$, corresponding to time scales of nearly 20 minutes and 1 minute, respectively. It is important to note that at the time scale corresponding at $N = 300$ we do not get rid of typical microstucture effects, such us autocorrelation of returns or bid-ask bounce. Figure 6 shows the distribution of the 495

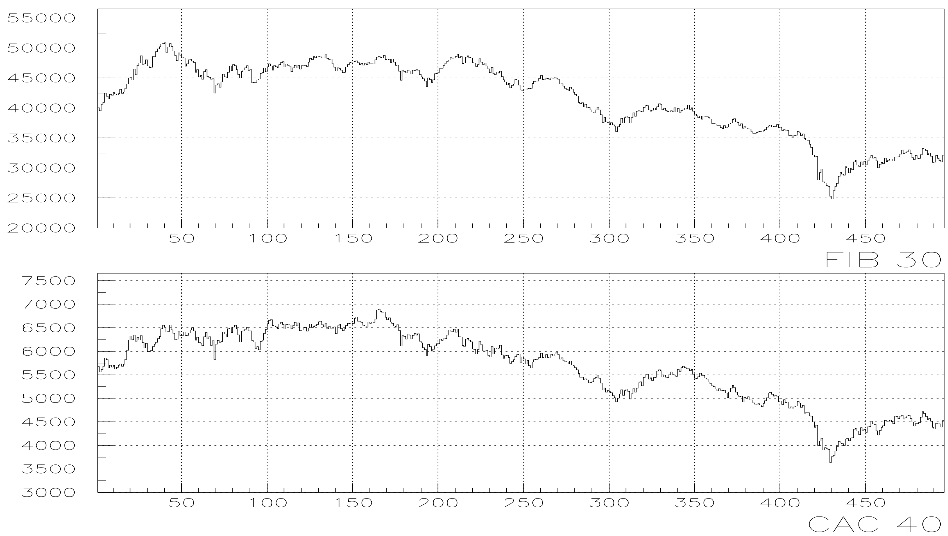

Figure 2. Daily time series of the futures price for the FIB30 (top) and the CAC40 (bottom), for the period 2000-2001.

daily correlations for both frequencies. The Epps effect is clearly visible. At $N = 25$ we observe a strong correlation between the Italian and the French futures prices. At $N = 300$ correlation is strongly reduced. Figure 7 shows the scatter plot of the two: most of the observations lie under the line $y = x$, an other signature of the Epps effect.
Figure 6 leads us to an important consideration. We can notice that the variance of the two distributions is nearly the same, and indeed standard deviation of the correlations at $N = 25$ is 0.2433, while at $N = 300$ it is 0.2733 (the increase is due to the boundary at $\rho = 1$). The variance of the distribution is given by the inherent variance of the correlations plus the measurement error. Since the measurement error at $N = 300$ is negligible with respect to that at $N = 25$, apart from the bias caused by the Epps effect, we conclude that stochasticity is inherent in the series and has to be modeled.
Figure 8 shows the time series of correlations at $N = 25$. Correlation is significant and very high throughout all the period. Moreover it shows a slight linear positive trend: correlations increased during the period $2000 - 2001$ and its volatility looks reduced. The highest correlation is reached on September, 12^{th}, 2001, the day after tragic events in New York, when it is $\rho = 0.9774$.
When modeling correlations, a limitation is in the fact that $|\rho| \leq 1$. We can overcome this difficulty, by using the transformed variable $\tilde{\rho} = f(\rho)$ according to the sigmoid function $f(\cdot)$:

$$f(x) = \frac{1}{2} \log \left(\frac{1+x}{1-x} \right) \tag{4.1}$$

which transforms the interval $[-1, 1]$ in $[-\infty, \infty]$ being anyway nearly linear around 0, see

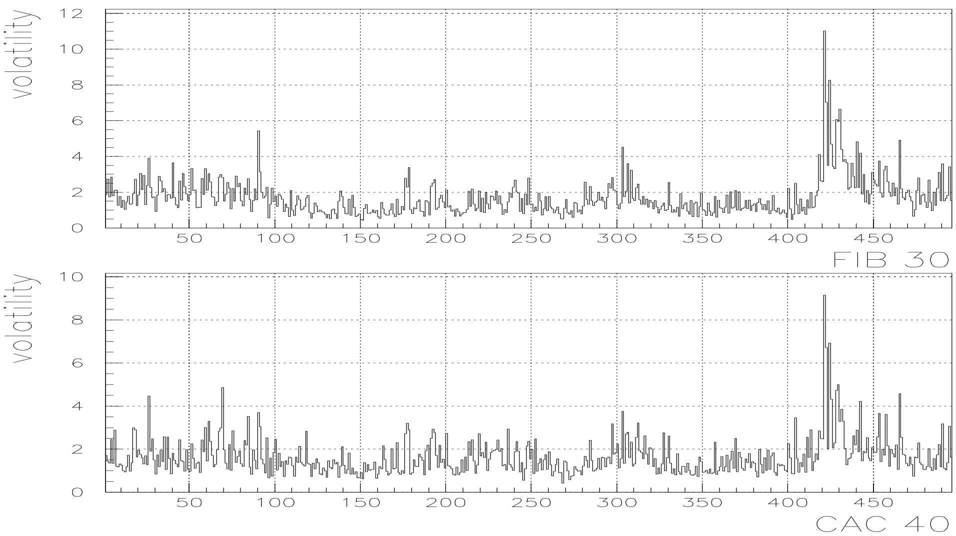

Figure 3. Daily time series of integrated volatilities, percentage, daily basis, measured with the Fourier method. Top: FIB30 volatilities, Bottom: CAC40 volatilities.

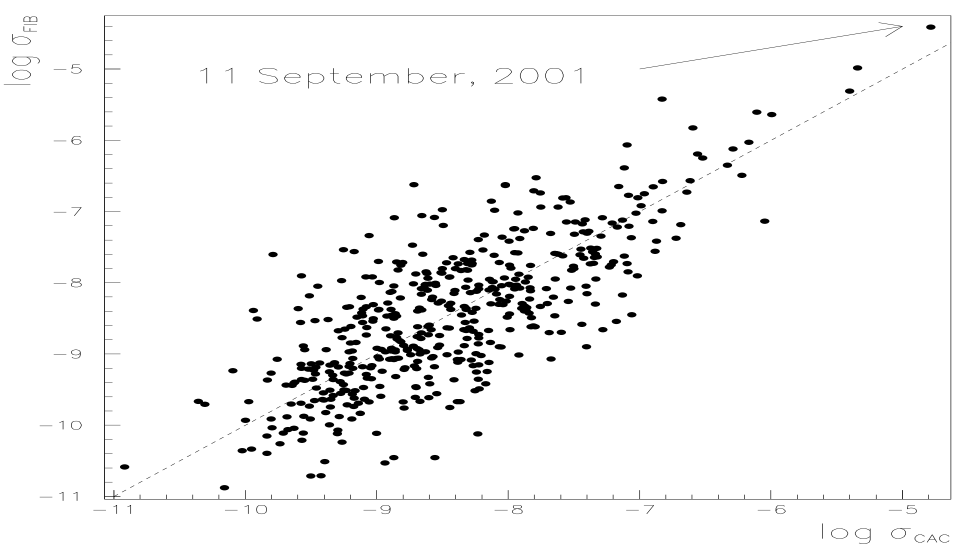

Figure 4. Scatter plot of FIB30 log-volatilities and CAC40 log-volatilities.

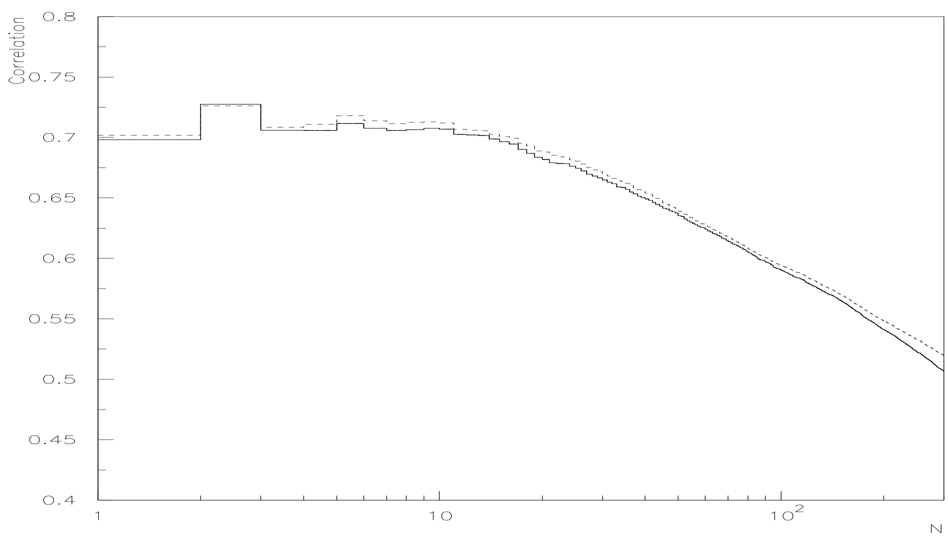

Figure 5. Full sample average correlation between FIB30 and CAC40 as a function of the frequency N used in the computation. Solid line: all transactions are used in the computation. The Epps effect is evident. Dashed line: only sinchronous transactions are used. There is no difference with the previous case.

Figure 9. We get correlations from the inverse function $\rho = f^{-1}(\tilde{\rho})$ given by:

$$f^{-1}(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{4.2}$$

We then consider the time series of transformed variables $\tilde{\rho}_t,\ t = 1, \ldots, 495$. We first estimate, via maximum likelihood, a simple autoregressive model (standard errors are in brackets):

$$\begin{array}{lccl} \tilde{\rho}_t = & c & + \quad \alpha & \cdot\tilde{\rho}_{t-1} + \varepsilon_t \\ & 0.9620 & \quad 0.1306 & \\ & (0.0245) & \quad (0.0446) & \end{array} \tag{4.3}$$

Our estimates indicate significant persistence in correlations. This persistence disappears if we add a linear trend in our model:

$$\begin{array}{lcccl} \tilde{\rho}_t = & c & + \quad \beta & \cdot t \cdot 10^{-3} \quad + \quad \alpha & \cdot\tilde{\rho}_{t-1} + \varepsilon_t \\ & 0.7182 & \quad 0.9832 & \quad 0.0497 & \\ & (0.0433) & \quad (0.1512) & \quad (0.0449) & \end{array} \tag{4.4}$$

Estimates of model (4.4) show that there's a significant linear trend which obscure persistence, since α is no longer significant.

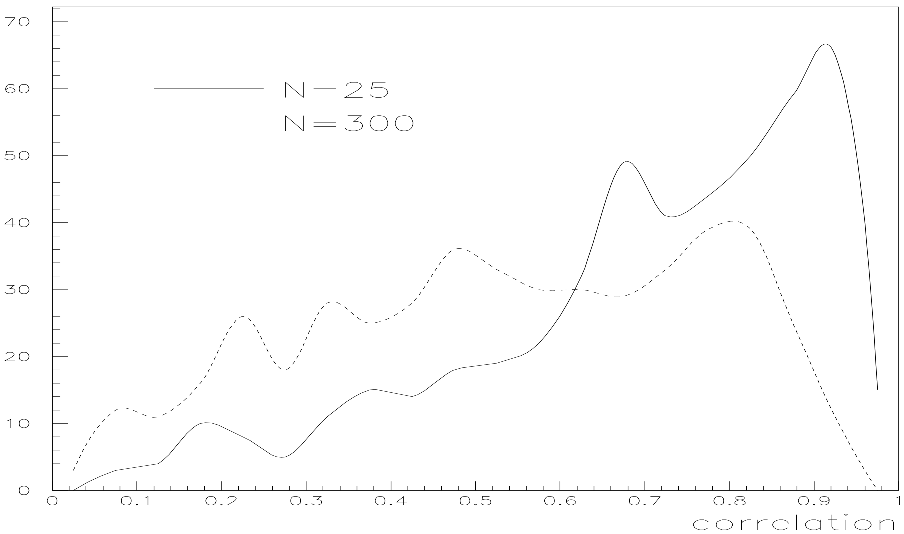

Figure 6. Distribution of correlations at $N = 25$ (solid line) and $N = 300$ (dashed line).

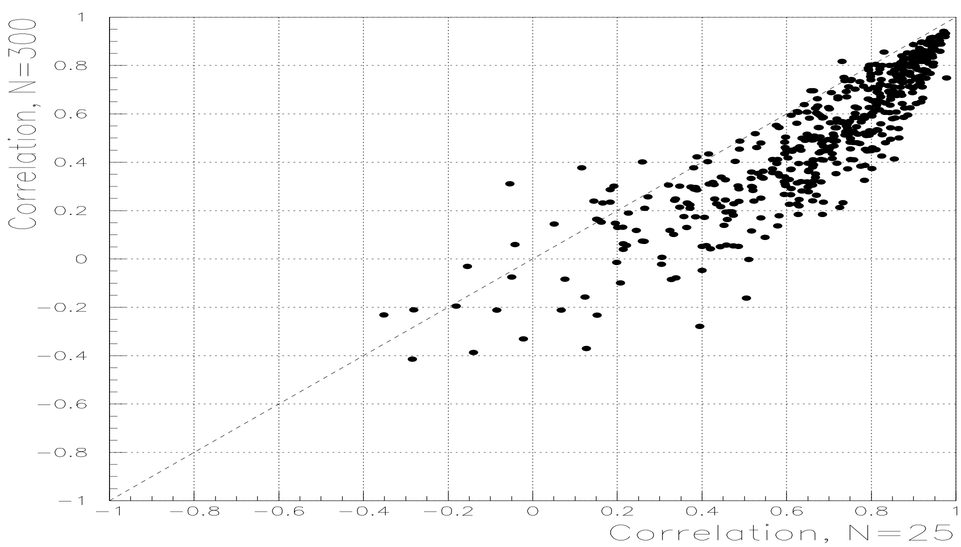

Figure 7. Scatter plot of correlation at $N = 25$ vs. correlation at $N = 300$.

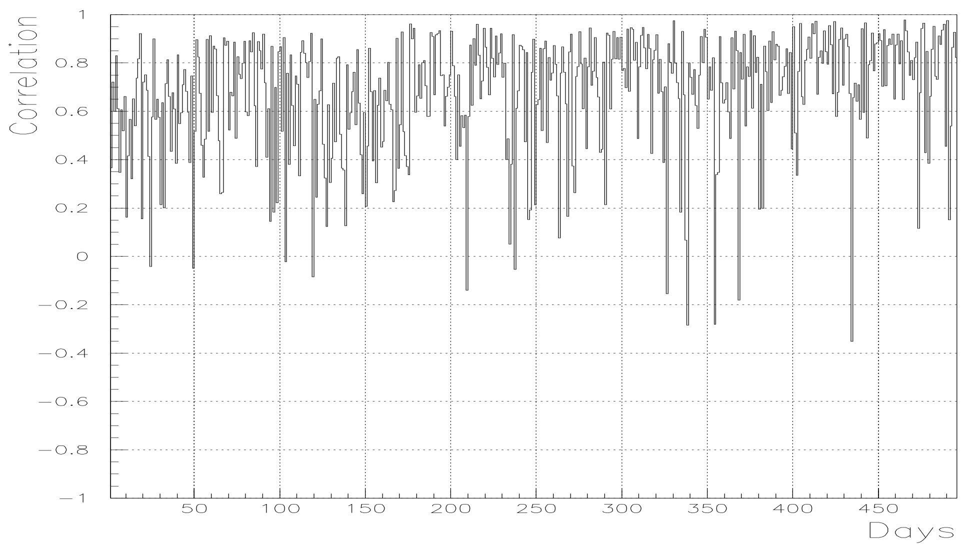

Figure 8. Time series of correlations at $N = 25$.

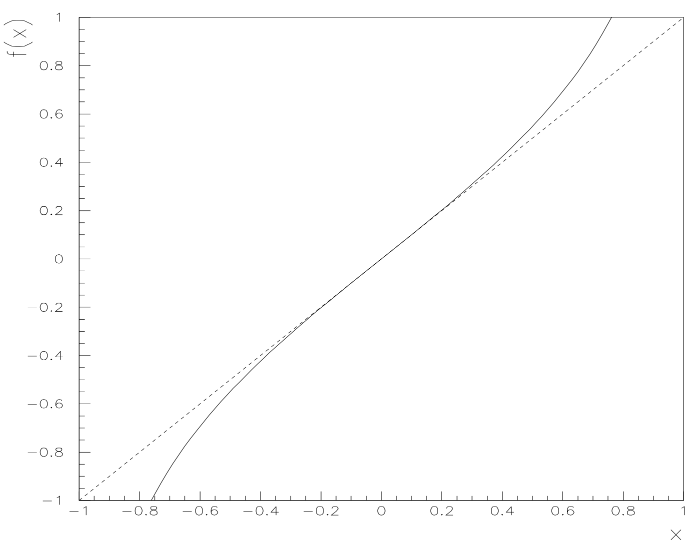

Figure 9. Plot of the function $f(x)$ in (4.1).

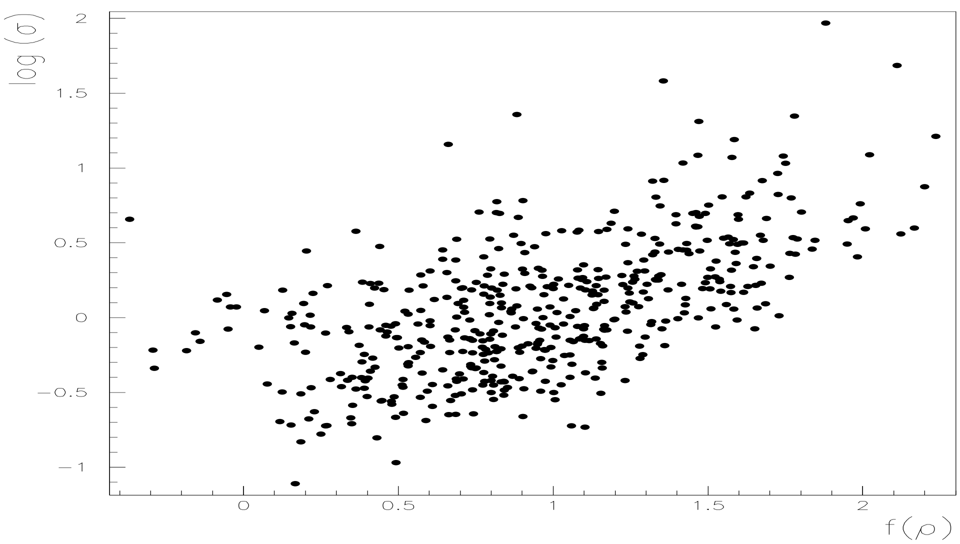

Figure 10. Scatter plot of $\tilde{\rho} = f(\rho)$ vs. $\log \bar{\sigma}$ defined in equation (4.5).

It is sometimes maintained that correlations are higher in periods of higher volatility (Ball and Torous, 2000; Longin and Solnik, 2001). We check this by adding the following regressor:

$$\bar{\sigma}_t = \frac{1}{4}\sqrt{\sigma_{1,t}^2 + \sigma_{2,t}^2} \tag{4.5}$$

where σ_1^2 and σ_2^2 are integrated volatilities of the Italian and French stock index futures respectively, measured with the Fourier method. Using $\bar{\sigma}$ instead of σ_1 or σ_2 makes no difference since the two volatilities are strongly correlated. Resulting volatility is measured in percentage, on daily basis. Our last model is then:

$$\begin{array}{lccccccccc} \tilde{\rho}_t = & c & + & \beta & \cdot t \cdot 10^{-3} & + & \gamma & \cdot \log \bar{\sigma}_t & + & \alpha & \cdot \tilde{\rho}_{t-1} + \varepsilon_t \\ & 1.2365 & & 0.7725 & & & 0.7450 & & & 0.2675 & \\ & (0.0566) & & (0.1553) & & & (0.0509) & & & (0.0541) & \end{array} \tag{4.6}$$

In model (4.6) all estimates are largely significant. The coefficients γ is positive, meaning that for an extra larger volatilities imply larger correlations. It is worth noting that this result is obtained with a very simple econometric technique, exploiting daily measurements of ρ with high frequency data. Figure 10 shows the scatter plot of $\tilde{\rho}$ and $\log \bar{\sigma}$, displaying a strong linear relation. Persistence is again significant; the trend is robust to the new specification.

5. Conclusions and directions for future research

In this paper, we describe a method to compute correlations with high frequency data, based on Fourier analysis. Monte Carlo evidence shows that this method performs better than classical realized volatility. We then use the method to analyze two years of tick-by tick transactions on the Italian and French market. We show that the Epps effect clearly displays, so we select a low frequency ($N = 25$) to compute volatility. For our data, the Epps effect cannot be explained by non-synchronicity. We show that correlations are high throughout all the period, and they are dynamically changing over time. We then model correlations, and we provide strong evidence for a linear trend in the period, as well as for a positive link with volatility. Moreover, we find a weak evidence of correlation persistence. Our paper shows that the use of high frequency data, and in particular the Fourier method, can be fundamental tools in modeling volatilities and correlations. We think this advancement will thoroughly take place in future research.
Moreover, it is clear that there's a typical time scale which is needed by international correlations to fully establish. Assessing the measurement of this time scale can be fundamental for high-frequency (daily or intradaily) portfolio rebalancing.

6. Acknowledgments

We wish to acknowledge the organizing committee of the second *Workshop on Correlated Data Modeling*, held in Turin, Italy, in January 2004, as well as Workshop participants. We would also like to thank the participants at the *Workshop on High Frequency Data* held in Siena, Italy, in November 2003.
We are grateful to Borsa Italiana S.P.A for providing us the data set, and in particular to Dr. Concetta Ricciardi and Dr. Ada De Roma.
We thank Maria Pasquale and Simone Bianco for outstanding research assistance.

References

Andersen T. and Bollerslev T. (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, 39, 885–905.

Andersen T., Bollerslev T., Diebold F. and Labys P. (2002) Modeling and forecasting realized volatility, *Econometrica*, 71, 529–626.

Andersen T., Bollerslev T. and Diebold F.X. (2003) Parametric and nonparametric volatility measurement, in: *Handbook of Financial Econometrics*, Hansen L.P. and Ait-Sahalia Y., eds., Amsterdam: North-Holland.

Ball C. and Torous W. (2000) Stochastic correlation across international stock markets, *Journal of Empirical Finance*, 7, 373–388.

Barucci E. and Renò R. (2002a) On measuring volatility and the GARCH forecasting performance, *Journal of International Financial Markets, Institutions and Money*, 12, 183–200.

Barucci E. and Renò R. (2002b) On measuring volatility of diffusion processes with high frequency data, *Economics Letters*, 74, 371–378.

Epps T. (1979) Comovements in stock prices in the very short run, *Journal of the American Statistical Association*, 74, 291–298.

Jacod J. and Shiryaev A.N. (1987) *Limit Theorems for Stochastic Processes*, Springer.

Kroner K. and Ng V. (1998) Modeling asymmetric comovements of asset returns, *Review of Financial Studies*, 11, 4, 817–844.

Longin F. and Solnik B. (2001) Extreme correlation of international equity markets, *Journal of Finance*, 56, 2, 649–676.

Malliavin P. and Mancino M. (2002) Fourier series method for measurement of multivariate volatilities, *Finance & Stochastics*, 6, 1, 49–61.

Renò R. (2003) A closer look at the Epps effect, *International Journal of Theoretical and Applied Finance*, 6, 1, 87–102.