On measuring correlation of financial time series with high-frequency data

WCDM04, Torino, 9-10 January 2004

Roberto Renò

`reno@unisi.it`

Dipartimento di Economia Politica, Università di Siena

Introduction

Measuring linear correlation among assets is crucial in financial applications (e.g. Markovitz theory).

The last decade witnessed the advent of high-frequancy data.

High frequency data are very peculiar:

- Unevenly sampled
- Microstructure effects
- Huge quantity of data

New econometric techniques are needed!

Measuring linear correlation

We assume that $p(t) \in \mathbb{R}^d$ is the solution of the following stochastic differential equation (SDE):

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

Standard techniques need interpolation of the data to get an evenly spaced grid:

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

The Fourier method

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:

$$a_0^i(dp) = \frac{1}{2\pi}\int_0^{2\pi} dp_i(t)$$

$$a_k^i(dp) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)\, dp_i(t), \quad b_k^i(dp) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)\, dp_i(t),$$

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:

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

Computing the coefficients

Theorem

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

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

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

Implementation

The coefficients of dp are computed via integration by parts:

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

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; we use $p(t) = p(t_j)$ where t_j is the largest observation time before t.
For all computations, we set $n_0 = 1$.

Monte Carlo experiments

We simulate two correlated asset price diffusions with the bi-variate continuous GARCH(1,1) model:

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

and all other correlations between Brownian motions $x_1(t), x_2(t), x_3(t), x_4(t)$ set to zero.
We extract observation times drawing the durations from an exponential distribution with mean 60 seconds.

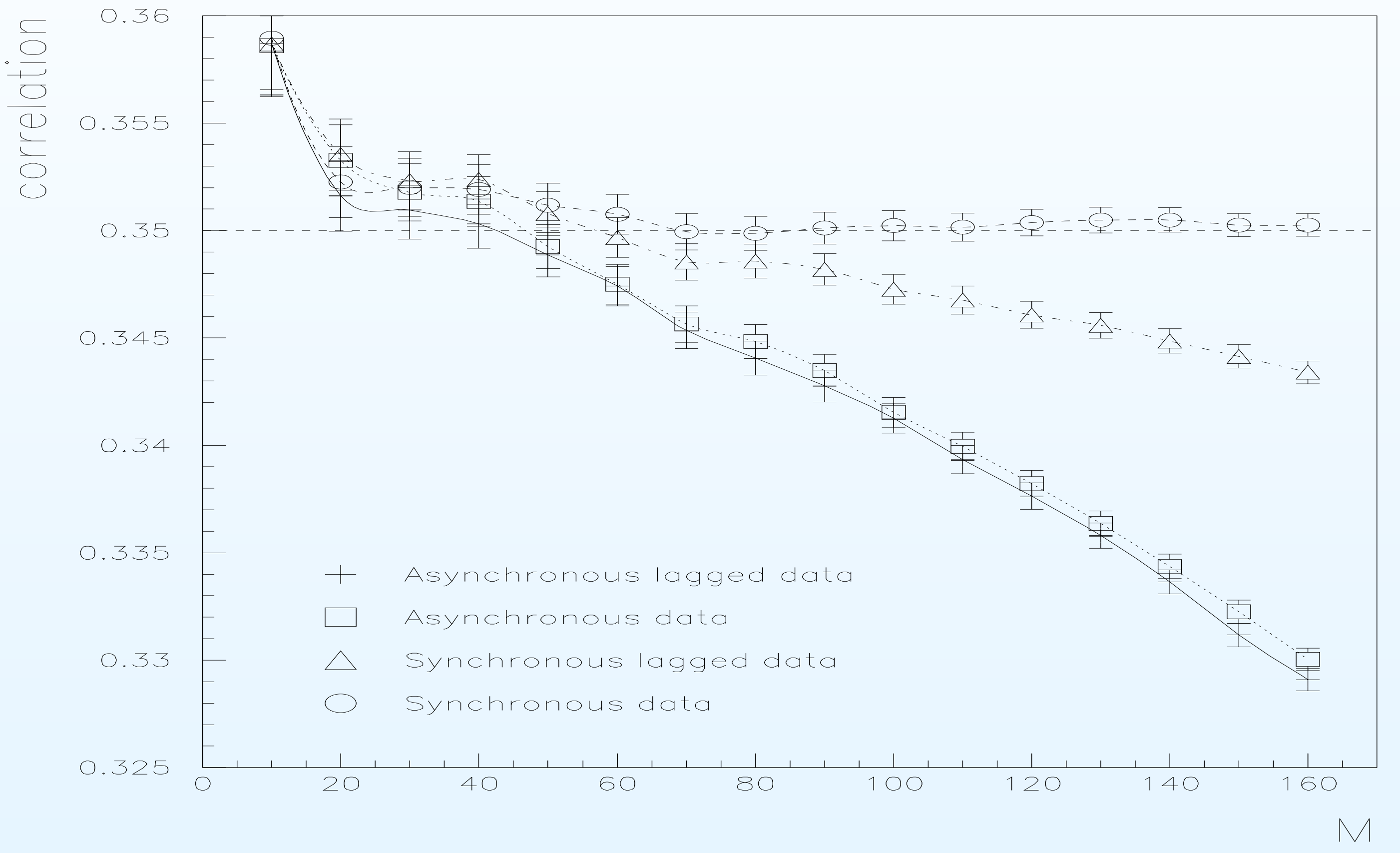
Results

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

Simulating high-frequency data

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, introducing asynchronous data and lagging.

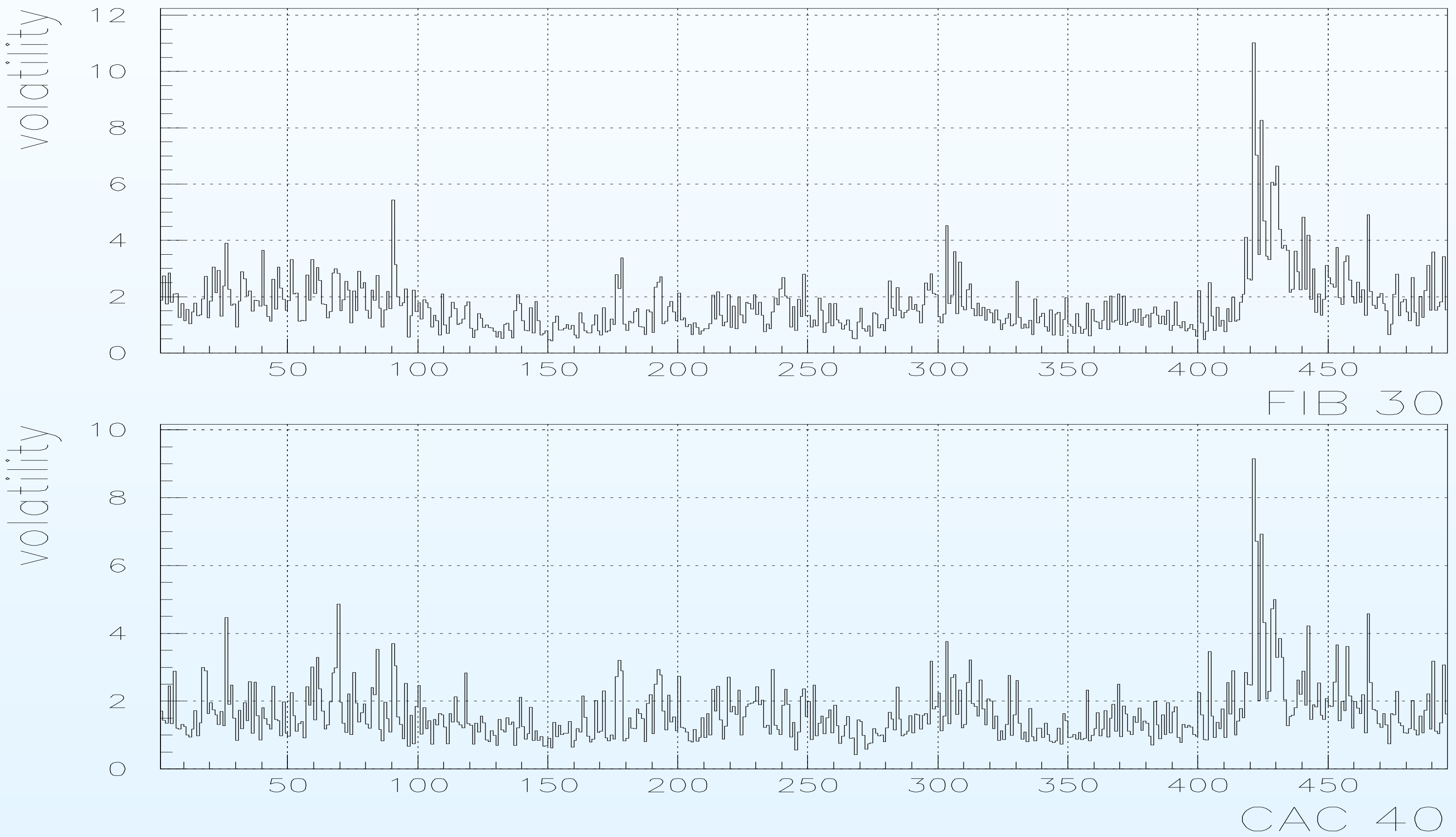
The Epps effect



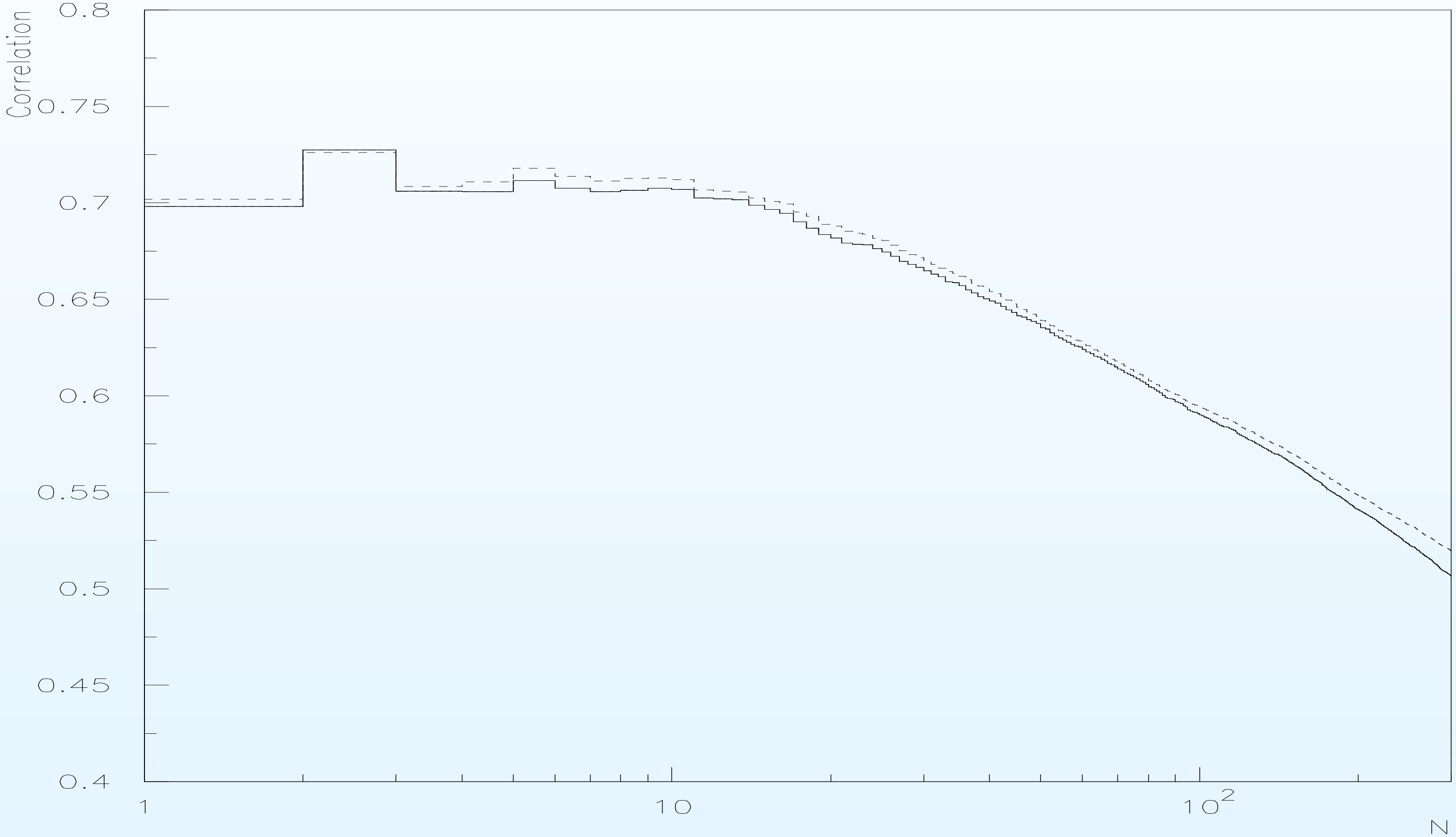
Data analysis

We analyze Italian and French stock index futures in the period 2001-2002 (11 millions trades).

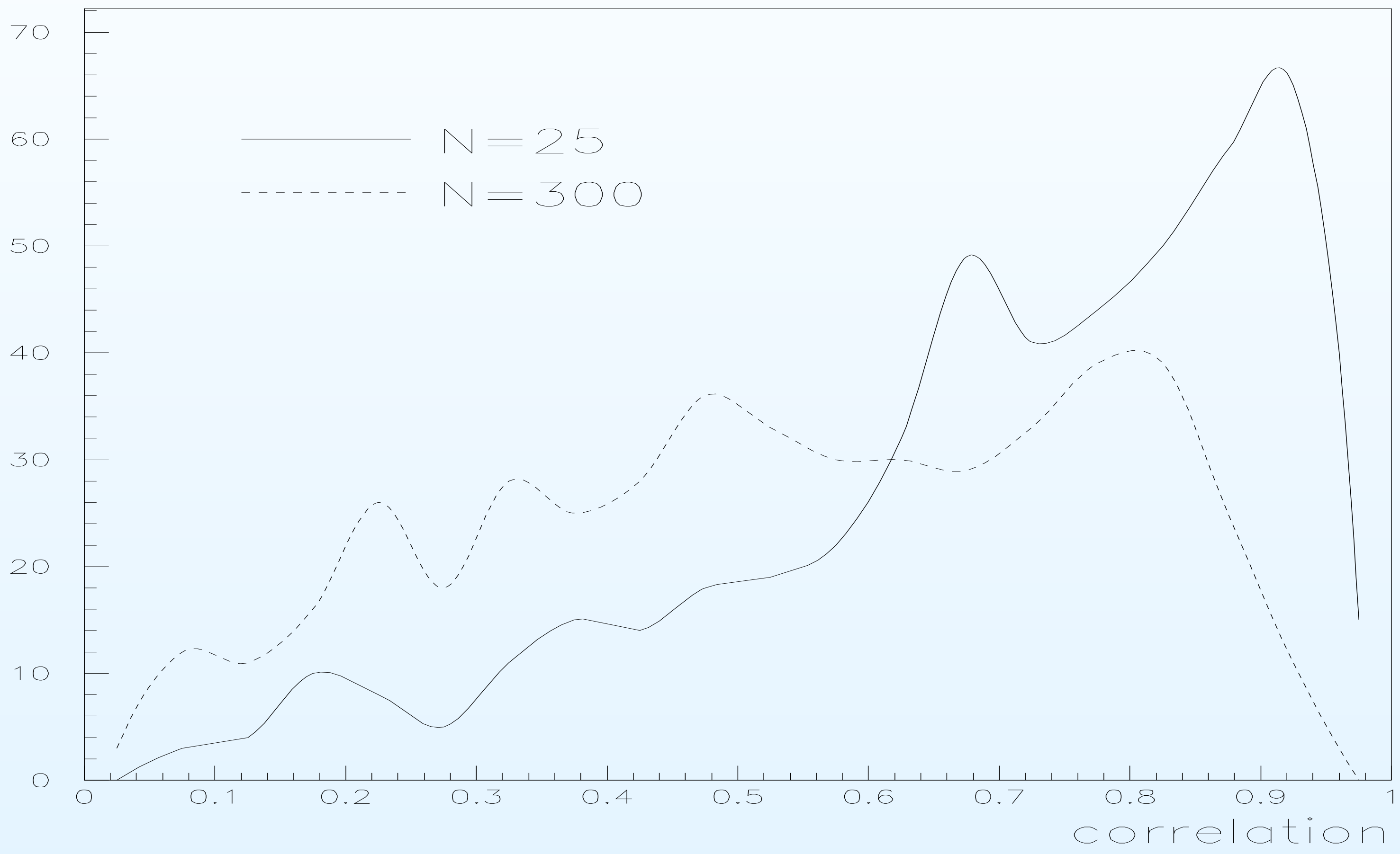
Volatilities



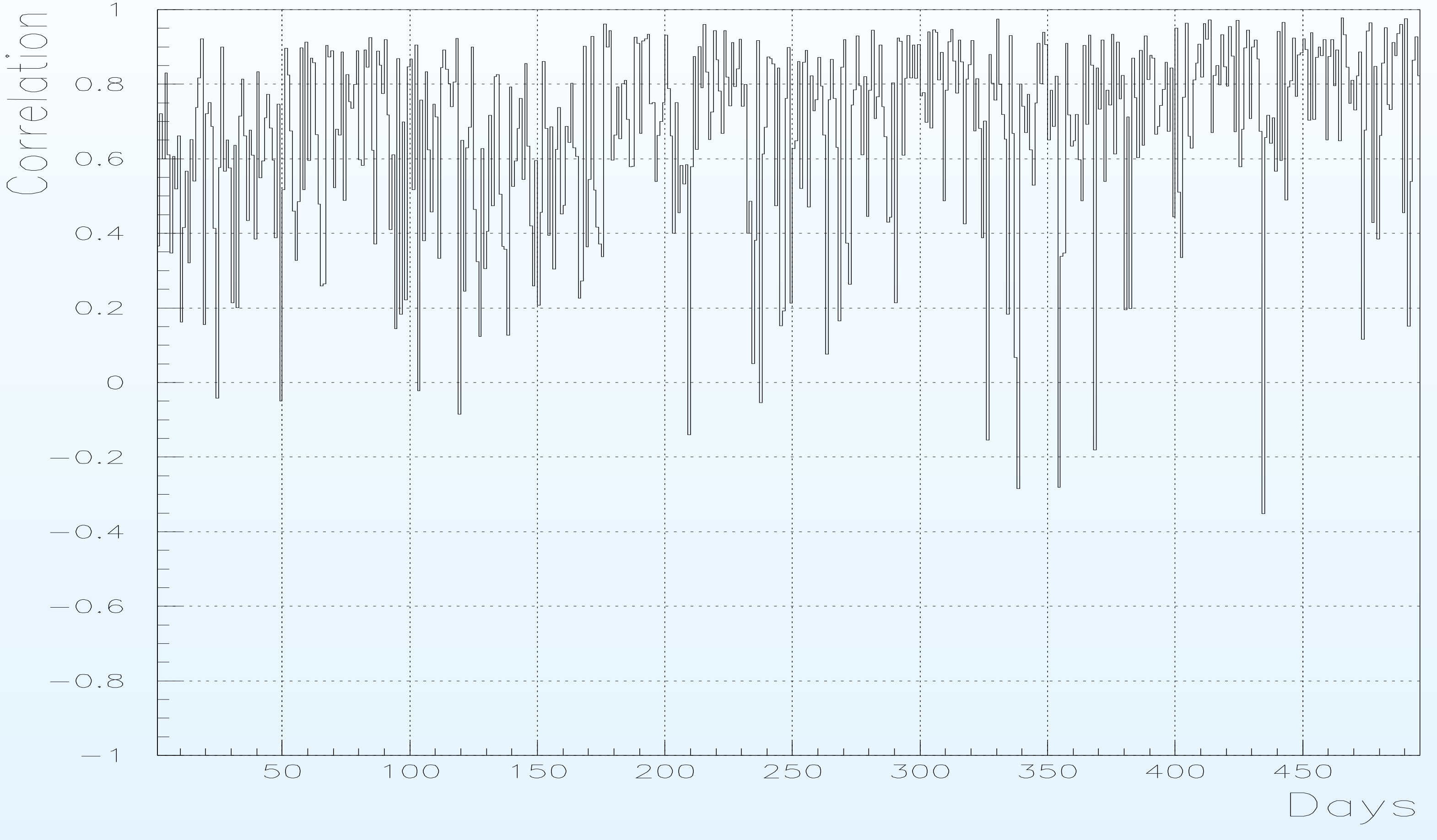
The Epps effect



Distribution of correlations



Time series



Correlation
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
50
100
150
200
250
300
350
400
450
Days

A model for correlations:

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 $f(\cdot)$:

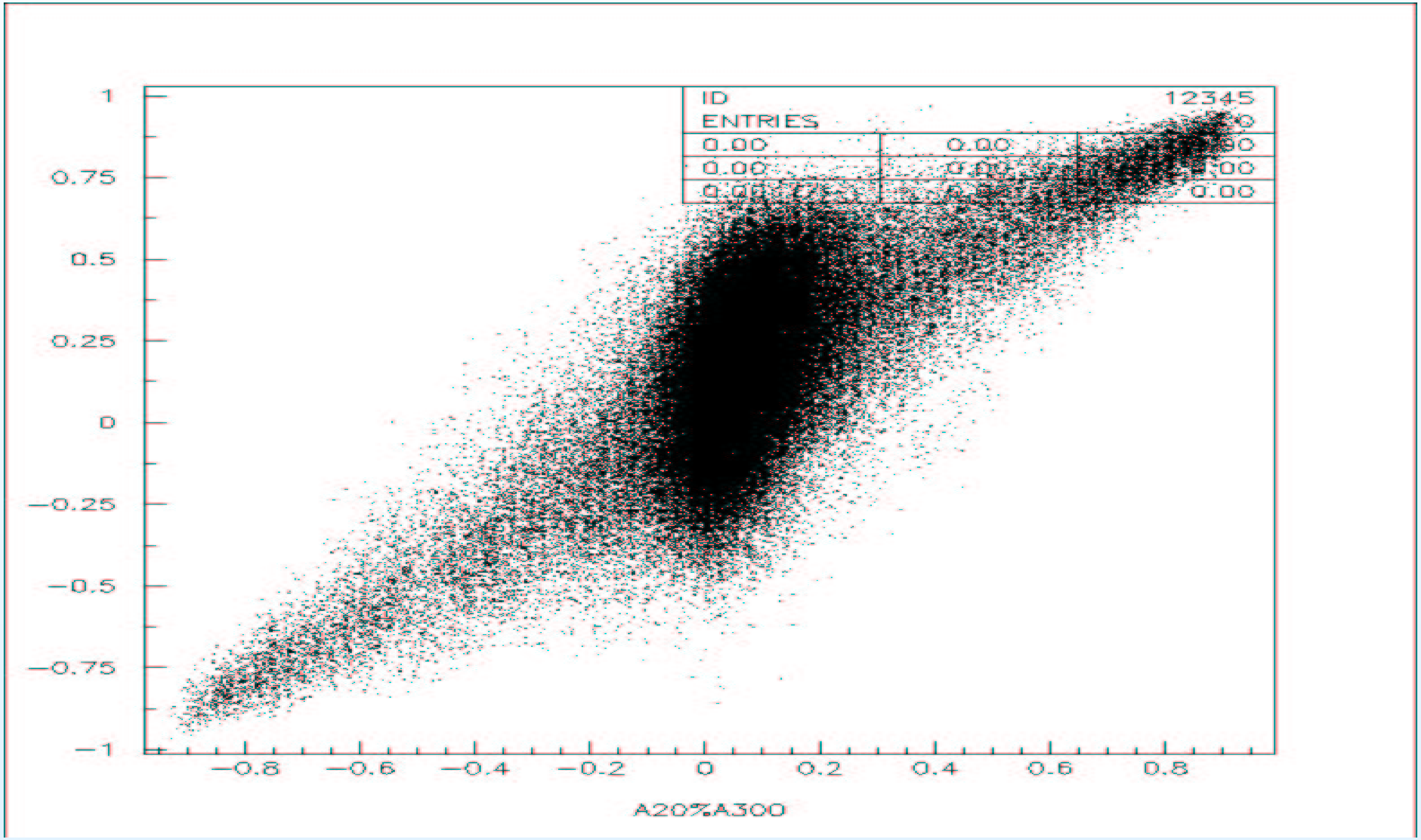
$$f(x) = \frac{1}{2} \log \left(\frac{1+x}{1-x} \right)$$

The model is then:

$$\begin{array}{cccccccccc}
\tilde{\rho}_t = & c & + & \beta & \cdot t & + & \gamma & \cdot \log \bar{\sigma}_t & + & \alpha & \cdot \tilde{\rho}_{t-1} + \varepsilon_t \\
& 1.24 & & 0.77 & & & 0.74 & & & 0.27 & \\
& (0.06) & & (0.15) & & & (0.05) & & & (0.05) &
\end{array}$$

A surprising galaxy

Correlation of 98 US stocks for 40 days:



Conclusions

- The Fourier estimator is a promising tool for variance-covariance matrix estimation
- We show that the Epps effect clearly displays
- The Epps effect looks “differential”
- Using realized covariances we can resort to standard econometrich techniques
- We provide evidence for persistence in correlations, as well as linkage with volatility