Fakultät für Physik

UNIVERSITÄT DUISBURG ESSEN

Econophysics IV:

Market States and the Subtleties of Correlations

Thomas Guhr

XXVIII Heidelberg Physics Graduate Days, Heidelberg 2012

Outline

- mysterious vanishing of correlations: Epps effect
- Gaussian assumptions and correlations: copulae
- identification of market states
- time evolution of market states
- signatures of crisis

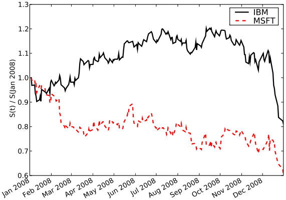
Epps Effect

Measurement of correlation coefficients

stock prices: $S^{(i)}(t), \; i = 1, \ldots, K$

returns $r_{\Delta t}^{(i)} = \dfrac{S^{(i)}(t+\Delta t) - S^{(i)}(t)}{S^{(i)}(t)}$

depend on the chosen return interval



$$C_{ij} = \mathrm{corr}(r_{\Delta t}^{(i)}, r_{\Delta t}^{(j)}) = \frac{\langle r_{\Delta t}^{(i)} r_{\Delta t}^{(j)} \rangle - \langle r_{\Delta t}^{(i)} \rangle \langle r_{\Delta t}^{(j)} \rangle}{\sigma^{(i)} \sigma^{(j)}} \,, \qquad \langle u \rangle = \frac{1}{T} \sum_{t=1}^{T} u(t)$$

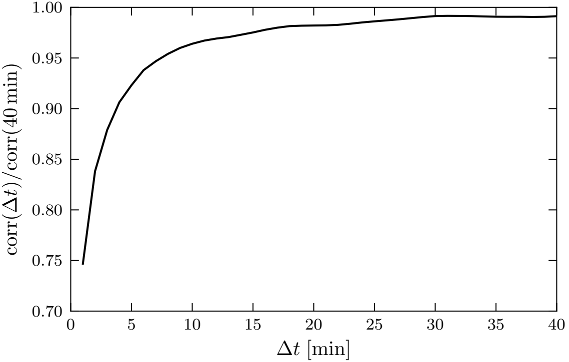
assume that T is long enough $\longrightarrow$ no noise dressing

... but what is the dependence on the return interval Δt ?

Empirical results

measured correlations suppressed towards small return intervals Δt

$\longrightarrow$ this is the Epps effect



ensemble of 50 stock pairs (normalized to saturation value)

Goal

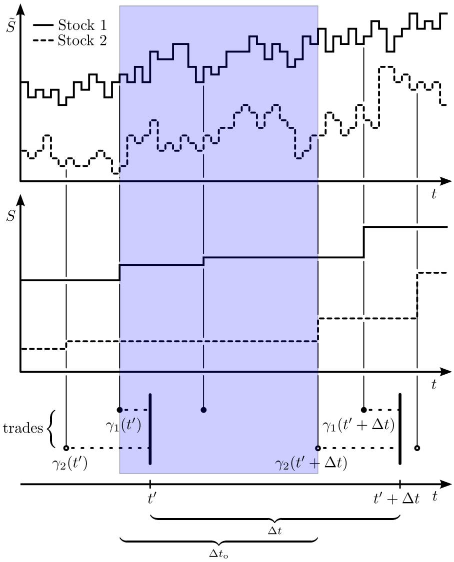
Variety of possible reasons discussed in finance, including highly speculative “emergence of correlations”.

Existing studies mostly aim at schematically compensating the Epps effect.

Being physicists, we ...

- look at the data — carefully,
- identify statistical causes,
- develop parameter free compensation,
- quantify what is left for other causes.

Asynchronity — formation of an overlap



underlying fictitious Markovian time series

actual time series

γ: last trading time

overlap $\Delta t_o(t)/\Delta t$ with synchronous information, outside random

Compensation of asynchronity

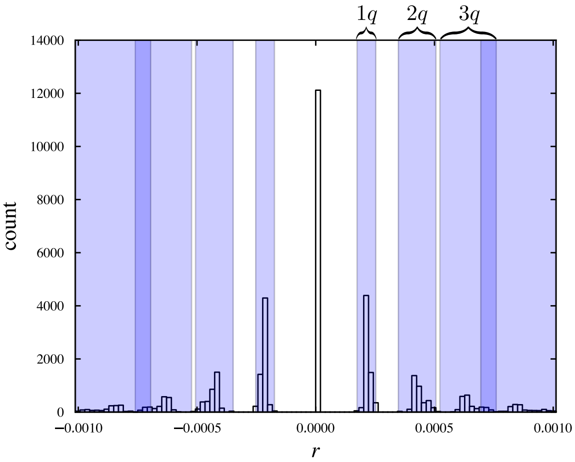
$$g_{\Delta t}^{(i)}(t) = \frac{r_{\Delta t}^{(i)}(t) - \langle r_{\Delta t}^{(i)}(t) \rangle}{\sigma_{\Delta t}^{(i)}}$$

$$\widehat{\mathrm{corr}}(r_{\Delta t}^{(i)}, r_{\Delta t}^{(j)}) = \left\langle g_{\Delta t}^{(i)}(t) g_{\Delta t}^{(j)}(t) \right\rangle$$

$$\widehat{\mathrm{corr}}_{\mathrm{async}}(r_{\Delta t}^{(i)}, r_{\Delta t}^{(j)}) = \left\langle g_{\Delta t}^{(i)}(t) g_{\Delta t}^{(j)}(t) \frac{\Delta t}{\Delta t_{\mathrm{o}}(t)} \right\rangle$$

term–by–term compensation by multiplying with inverse overlap

Tick size and return distribution



tick-size q
discretizes prices

returns also affected

clustering

Correlation coefficient for discretized data

idea: discretization $r^{(i)}(t) \to \bar{r}^{(i)}(t)$ produces random errors $\vartheta^{(i)}(t)$

$$\begin{aligned} r^{(i)}(t) &= \bar{r}^{(i)}(t) + \vartheta^{(i)}(t) \\ \widehat{\mathrm{corr}}_{\mathrm{tick}}(r^{(i)}, r^{(j)}) &\approx \frac{\mathrm{cov}\left(\bar{r}^{(i)}, \bar{r}^{(j)}\right)}{\hat{\sigma}^{(i)}\hat{\sigma}^{(j)}} \end{aligned}$$

- compensation by correcting with normalization
- estimation using average discretization error

Combined compensation

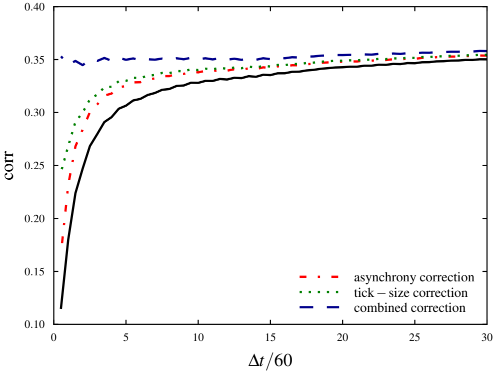
take both, asynchronity and discretization, into account

$$\widehat{\mathrm{corr}}(r^{(i)}, r^{(j)}) \approx \frac{\left\langle \bar{r}^{(i)} \bar{r}^{(j)} \frac{\Delta t}{\Delta t_{\mathrm{o}}} \right\rangle}{\hat{\sigma}^{(i)} \hat{\sigma}^{(j)}}$$

no interference, undisturbed superposition

Test with stochastic processes

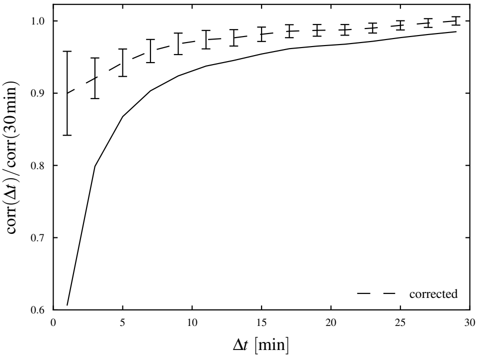
autoregressive GARCH(1,1) known to reproduce phenomenology of stock price time series and their distributions



full, parameter free compensation

Test with real data

stocks from Standard & Poor's 500, prices between \$10.01–\$20.00



parameter free combined compensation

$\longrightarrow$ rest has other causes

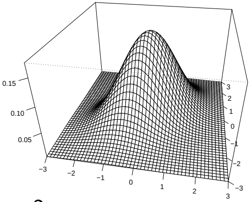
Results

- purely statistical causes have strong impact on Epps effect
- identified what is left for other causes (e.g. lags, etc)
- parameter free compensation
- significant better precision when estimating correlations
- can easily be applied

Non–Gaussian Dependencies

Correlation coefficient and joint probability density function

- correlation coefficient reduces complex statistical dependence to a single number



- only meaningful if dependence is multivariate Gaussian, e.g. bivariate

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-c^2}} \exp\left(-\frac{1}{2}\frac{x^2-2cxy+y^2}{1-c^2}\right)$$

- if not, have to retrieve better information from full joint probability density function $f_{X,Y}(x,y)$ which contains all information

Tools and definitions

- marginal distribution: $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy$
- comulative distribution: $F_X(x) = \int_{-\infty}^{x} f_X(x')dx'$
- u quantile: left of $x = F_X^{-1}(u)$ are u percent of events
- joint probability: $F_{X,Y}(x,y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' \, f_{X,Y}(x',y')$

Copulae

$$f_{X,Y}(x,y)$$

$$\swarrow \qquad \searrow$$

$$f_X(x),\ f_Y(y) \qquad\qquad \mathrm{cop}_{X,Y}(u,v)$$

separate statistical dependencies and marginal distributions

$$\mathrm{Cop}_{X,Y}(u,v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(v)\right)$$

$$\mathrm{cop}_{X,Y}(u,v) = \frac{\partial^2}{\partial u \partial v}\mathrm{Cop}_{X,Y}(u,v)\ .$$

(similar to “moving frame” or “unfolding” in quantum chaos)

Comparison true versus Gaussian copulae

- K return time series $r^{(i)}(t),\ i = 1, \ldots, K$
- calculate standard Pearson correlation coefficients C_{ij} for each pair (i, j)
- uniquely determines bivariate Gaussian distribution for pair (i, j)

$$f_{i,j}(x, y) = \frac{1}{2\pi\sqrt{1 - C_{ij}^2}} \exp\left(-\frac{1}{2}\frac{x^2 - 2C_{ij}xy + y^2}{1 - C_{ij}^2}\right)$$

- evaluate corresponding Gaussian copula $\mathrm{cop}_{i,j}^{(G)}(u, v)$

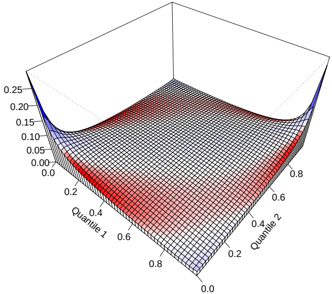
Comparison true versus Gaussian copulae — continued

- analyze true copula $\mathrm{cop}_{i,j}(u,v)$
- calculate distance

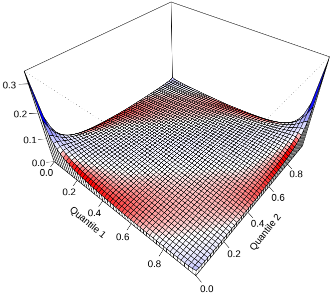
$$d(u,v) = \frac{1}{K(K-1)/2} \sum_{i<j} \Big(\mathrm{cop}_{i,j}(u,v) - \mathrm{cop}_{i,j}^{(G)}(u,v) \Big)$$

Empirical study

TAQ data 2007–2010, S&P 500, more than 12 billion transactions



$\Delta t = 2\text{h}$



$\Delta t = 30\text{min}$

- structure of copula stable when varying return interval
- bivariate Gaussian assumption drastically underestimates simultaneous extreme events

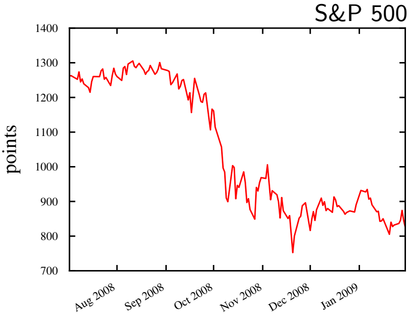
Results

- standard Pearson correlation coefficient problematic for data which are not bivariate Gaussian
- copulae provide alternative by extracting statistical dependencies independent of marginal distributions
- risk of simultaneous extreme events in real data much higher than usually assumed

Identifying Market States and their Dynamics

States of financial markets

- market is non–stationary
- different states before, during and after a crisis
- market can function in different modes
- qualitative/empirical — also quantitative ?



Similarity measure

correlations provide detailed information about the market

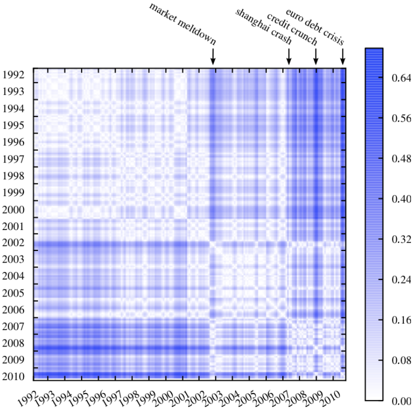
introduce distance of two correlation matrices

$$\zeta^{(T)}(t_1, t_2) = \left\langle \left| C_{ij}^{(T)}(t_1) - C_{ij}^{(T)}(t_2) \right| \right\rangle_{ij}$$

i, j running index of risk element or company
t_1, t_2 times at which the two correlation matrices calculated
T sampling time backwards

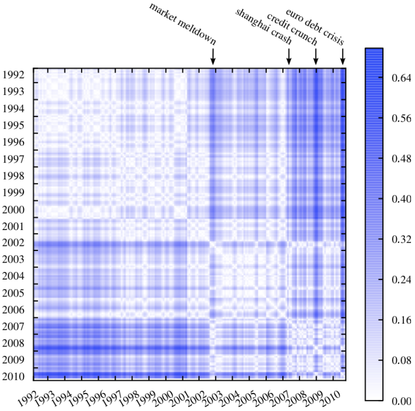
$\longrightarrow$ distances $\zeta^{(T)}(t_1, t_2)$ array or matrix in points (t_1, t_2)

Empirical results

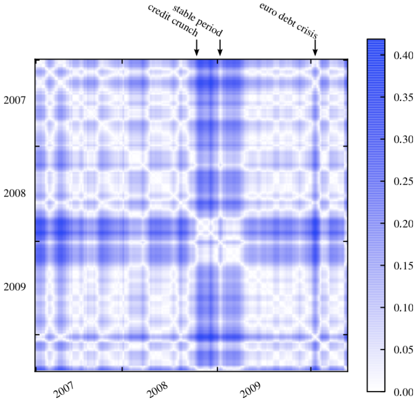


$T = 2$ months

Empirical results



$T = 2$ months



$T = 1$ week

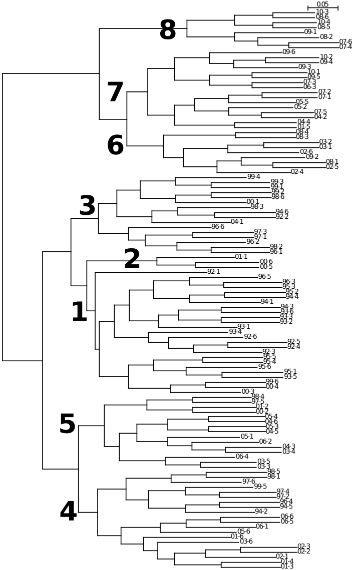
Identification of market states

array $\zeta^{(T)}(t_1, t_2)$ reveals changes in market structures over long time horizons

identify states by cluster analysis

- ensemble of correlation matrices $C_{ij}^{(T)}(t),\ t = t^{(a)}, \ldots, t^{(b)}$
- find two disjunct clusters where distance $\zeta^{(T)}$ from average within each cluster is smallest
- repeat that for these two clusters, and so on
- stop when distances within groups comparable to distances between grop

States of US financial market 1992–2010



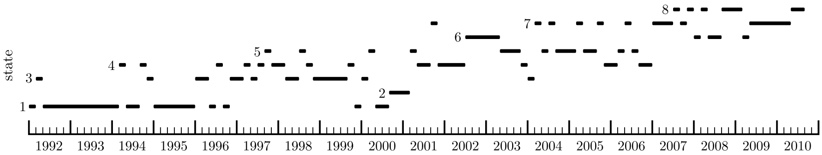
bold face numbers label market states

if no threshold $\longrightarrow$

division process continued until all correlation matrices are identified

small numbers label year and two–months period

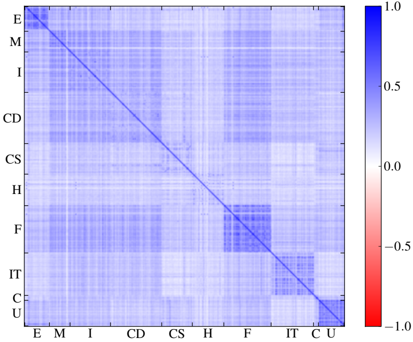
Time evolution of US market states



- subsequent formation of US market states 1992–2010
- market jumps between different states
- old states die out $\longrightarrow$ states have a lifetime
- how does this lifetime relate to other time scales (e.g. times between crashes) ?

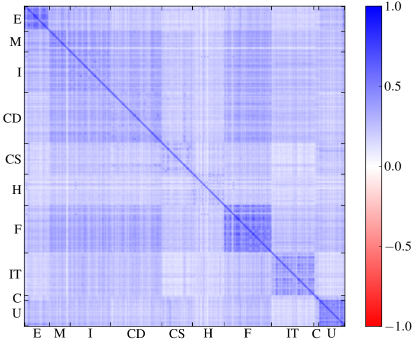
Market states in basis of industrial sectors

overall average correlation matrix

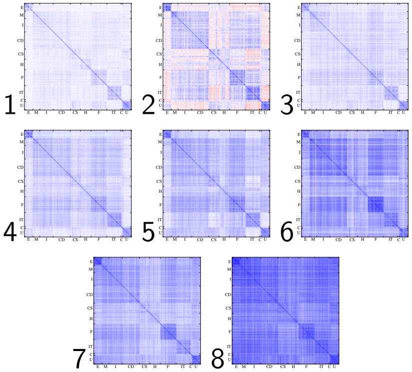


Market states in basis of industrial sectors

overall average correlation matrix

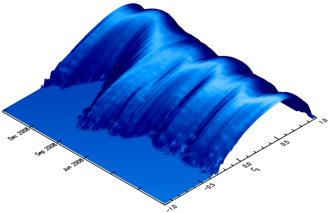


the eight states:

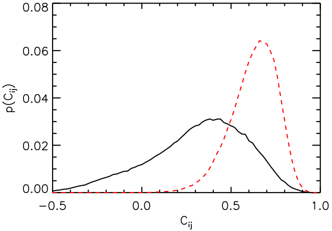


Time evolution of correlation coefficients distribution

time resolved analysis of distribution $p(C_{ij})$ during the 2008 crisis



surface plot



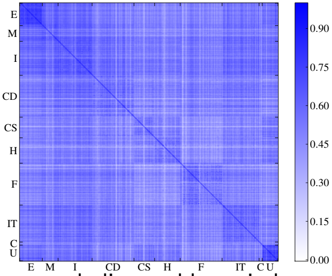
black: September 2008, red: December 2008

- distribution became broader before the crisis started in October 2008, partly due to decoupling of energy sector
- very narrow during the crisis with large mean value $\longleftrightarrow$ panic

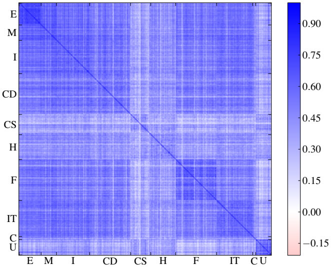
Stable period within 2008–2009 crisis

correlations during crisis October 15th, 2008, to April 1st, 2009

three–week stable period January 1st, 2009, to January 21st, 2009



excluding stable period



stable period

stable period very similar to market state number 7

Results

- usually stock market data analyzed "as if they were thrown on the floor"
- similarity measure reveals structural changes
- clear identification of market states
- time evolution of states followed $\longrightarrow$ dynamical information
- changes during 2008–2009 crisis: correlation matrix and distributions
- early warning system ?

Conclusions

- Epps effect: it helps to look at the data
- copulae: the world of finance is not Gaussian
- market states: they exist, have time evolution and lifetime

M.C. Münnix, R. Schäfer and T. Guhr,
Compensating Asynchrony Effects in the Calculation of Financial Correlations, Physica **A389** (2010) 767
Impact of the Tick–size on Financial Returns and Correlations, Physica **A389** (2010) 4828
Statistical Causes for the Epps Effect in Microstructure Noise, Int. J. Theoretical and Applied Finance **14** (2011) 1231

M.C. Münnix and R. Schäfer,
A Copula Approach on the Dynamics of Statistical Dependencies in the US Stock Market, Physica **A390** (2011) 4251

M.C. Münnix, T. Shimada, R. Schäfer, F. Leyvraz, T.H. Seligman, T. Guhr and H.E. Stanley, *Identifying States of a Financial Market,* `arXiv:1202.1623`, submitted (2011)