Calibration of factor models with equity data: parade of correlations

Baranovski, Alexander L.

WestLB AG

30 January 2012

Calibration of factor models with equity data: parade of correlations

Alexander L. Baranovski
WestLB AG, Germany[1]

This paper describes the process of ML-estimating of the equity correlations which can be used as proxies for asset correlations. In a Gaussian framework the ML-estimators are given in closed form. On this basis the impact of the Lehman's collapse on the dynamics of correlations is investigated: after the Lehman failure in September 2008 the rise in correlations took place across all economic sectors.

Keywords: intra/inter asset correlations, maximum likelihood estimation, single risk factor model, normal mixture, VAR of equity portfolio.

[1] Email: Alexander.Baranovski@gmail.com
The opinions expressed in this paper are those of the author and do not necessarily reflect views of the WestLB AG.

1 Empirical analysis of equity time series

A natural way to estimate credit quality correlations using historical data is to examine price histories of stocks as the equity returns are one fundamental and very observable source of firm-specific correlation information. Our main data source is the Bloomberg data feed. As of 03.09.2010 it contains 7652 North-American issuers from 19 distinct ICB industry sectors listed in Table 1.

ID	ICB sector name	# of firms
1	Oil & Gas	495
2	Chemicals	144
3	Basic Resources	246
4	Construction & Materials	126
5	Industrial Goods & Services	1087
6	Automobiles & Parts	75
7	Food & Beverage	227
8	Personal & Household Goods	381
9	Health Care	900
10	Retail	363
11	Media	286
12	Travel & Leisure	249
13	Telecommunications	113
14	Utilities	125
15	Banks	805
16	Insurance	114
17	Real Estate	65
18	Financial Services	775
19	Technology	1070

Table 1: Industry sector classification

For each issuer we retrieve the 190 weekly log-returns V_t covering period: 12.01.2007 – 03.09.2010.

1.1 Kolmogorov-Smirnov test and mixture of distributions

We construct an empirical distribution function for T observations $V_t^{(i)}$ of i-th stock

$$F_T^{(i)}(x) = \frac{1}{T}\sum_{t=1}^{T} I_{V_t^{(i)} \le x}$$

and calculate the Kolmogorov-Smirnov statistic $D_T^{(i)} = \sup_x \left| F_T^{(i)}(x) - F(x) \right|$, where F is a theoretical cumulative distribution. Here we assume the normality of the data, i.e. $F(x) \equiv \Phi(x)$.

On the next step we compare KS-statistic for every obligor from sector s with the critical values of Kolmogorov distribution for a 5%-significance level and count obligor, if KS-test accepts the normality hypothesis. Table 2 shows the distributions of a number of firms having “normal” data across sectors for two groups of firms namely belonging to DJ STOXX Amer 600 or not.

	"nostoxx" firms			"stoxx" firms		
	#of firms *n*	# of "normal" firms *nks*	nks/n	#of firms *n*	# of "normal" firms *nks*	nks/n
Oil & Gas	460	122	0,27	35	30	0,86
Chemicals	129	44	0,34	15	12	0,80
Basic Resources	233	49	0,21	14	14	1,00
Construction & Materials	117	46	0,39	9	9	1,00
Industrial Goods & Services	1022	413	0,40	65	59	0,91
Automobiles & Parts	70	20	0,29	5	4	0,80
Food & Beverage	209	51	0,24	18	15	0,83
Personal & Household Goods	358	102	0,29	23	22	0,96
Health Care	848	255	0,30	52	34	0,65
Retail	322	128	0,40	41	39	0,95
Media	271	45	0,17	15	11	0,73
Travel & Leisure	235	61	0,26	14	10	0,71
Telecommunications	101	20	0,20	12	3	0,25
Utilities	91	50	0,55	34	29	0,85
Banks	786	203	0,26	19	7	0,37
Insurance	92	39	0,42	22	6	0,27
Real Estate	65	13	0,20	0	0	N/A
Financial Services	758	85	0,11	22	14	0,64
Technology	1013	317	0,31	57	51	0,89

Table 2: Distribution of number of "nostoxx" and "stoxx" firms vs KS-test

The calculations of KS-statistic for every sector specific empirical distribution function

$$F_s(x) = \frac{1}{n_s}\sum_{i=1}^{n_s} F_T^{(i)}(x)$$

lead to the following table

	nostoxx firms	stoxx firms
Oil & Gas	0	1
Chemicals	0	1
Basic Resources	0	1
Construction & Materials	0	1
Industrial Goods & Services	0	1
Automobiles & Parts	0	1
Food & Beverage	0	1
Personal & Household Goods	0	1
Health Care	0	1
Retail	0	1
Media	0	1
Travel & Leisure	0	1
Telecommunications	0	1
Utilities	1	1
Banks	0	0
Insurance	0	0
Real Estate	0	0
Financial Services	0	1
Technology	0	1

Table 3: KS-test results across sectors

Thus the Kolmogorov-Smirnov test with the 5%-significance level rejects ("0") both the normality and t-hypothesis for the data of "NO STOXX" firms (excluding "Utilities") and accepts ("1") the null hypothesis on normality of data of "STOXX" firms excluding three cases for sectors "Banks", "Insurance" and "Real Estate".

Quality of data

Here we introduce the following ratio

$$\theta := \frac{T_d}{T}$$

where T_d is a number of all the distinct elements that appear in a time series of log-returns. The parameter $0 \le \theta \le 1$ reflects a liquidity or tradability of share. Clear for the largest stocks $\theta \approx 1$. At the same time the stocks of low liquid names having repeated quotes are characterised by small $\theta << 1$.

Table 4 contains the histograms of θ for "no stoxx" firms before and after "KS"- adjustment with a 5% - significance level

theta	# of "no stoxx" firms	# of firms after KS-test
0<θ<10%	317	0
10<θ<20%	644	0
20<θ<30%	739	0
30<θ<40%	756	0
40<θ<50%	662	0
50<θ<60%	462	5
60<θ<70%	375	24
70<θ<80%	478	60
80<θ<90%	797	91
90<θ<=100%	1950	1883
Total	7180	2063

Table 4: distribution of number of "no stoxx" firms w.r.t. the parameter θ

Thus the more tradable a stock the more likely its price follows a geometric Brownian motion as well as the Kolmogorov-Smirnov test is a suitable tool to filter the data according to the assumption of a normality.

In sequel for purposes of the calibration of a single factor model we will use the dataset of the "normal" time series of log-returns of 2063 firms (see right column of Table 4). Here we plot their mean empirical distribution functions across sectors

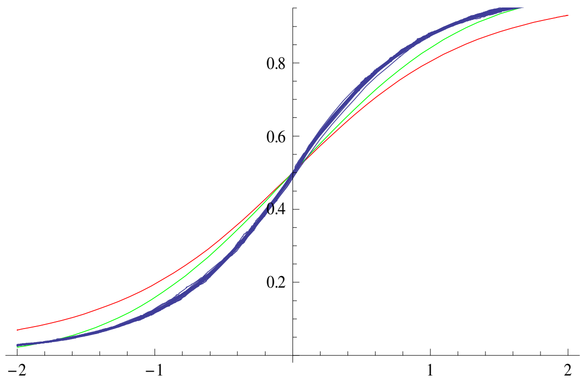


Fig. 1: the mean EDFs (blue curves) across sectors vs normal (green) and Student (red) t-distribution with 3 degrees of freedom

and calculate the average

$$F(x) = \frac{1}{19} \sum_{i=1}^{19} F_s(x)$$

for the set of 19 empiric CDFs from Fig. 1 as well as find its least-square fit in a family of normal distributions with zero mean and unknown variance. Fig. 2 depicts these two curves

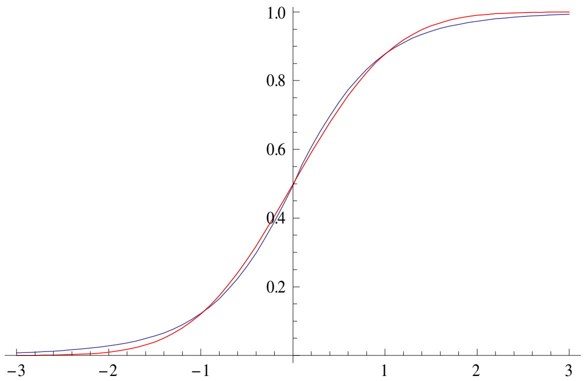


Fig. 2. Mean CDF (blue) and its fit (red)

Thus the standardized log-returns on a whole period T are surprisingly described by a normal distribution function with a nonunit variance:

$$F(x) \approx \Phi\left(\frac{x}{\sigma}\right), \sigma = 0.856974 \cdot$$

A such phenomenon can be explained by a mixture of normal distributions given on the subintervals T_i

$$F(x) \approx \sum_{i=1}^{l} \lambda_i \Phi\left(\frac{x - m_i}{\sigma_i}\right),$$

where $\lambda_i = \frac{T_i}{T}, \forall i$ and $\sum_{i=1}^{l} \lambda_i = 1$ as $T = \sum_{i=1}^{l} T_i$. The log-returns in an observation period T_i are assumed to be normally distributed with mean m_i and variance σ_i^2.

So, if we divide our data into two periods T_1 =90 weeks (12.01.07 – 3.10.08) and T_2 =100 weeks (10.10.08 - 03.09.10) and then estimate the mean CDFs for both periods we come to

$$F(x) \approx \lambda \cdot \Phi_{T1}\left(\frac{x}{0.698056}\right) + (1-\lambda) \cdot \Phi_{T2}(x) \approx \Phi\left(\frac{x}{\lambda \cdot 0.698056 + (1-\lambda)}\right) \equiv \Phi\left(\frac{x}{0.856974}\right),$$

where $\lambda = \frac{T_1}{T} = \frac{90}{190}$.

A natural choice of T_1 as a point of regime change behaviour in a period September/October 2008 can be mathematically confirmed by a change point analysis of a kurtosis

$$\kappa(t) = 3 \frac{\frac{t}{T} \cdot \sigma_1^4(t) + \left(1 - \frac{t}{T}\right) \cdot \sigma_2^4(t)}{\left(\frac{t}{T} \cdot \sigma_1^2(t) + \left(1 - \frac{t}{T}\right) \cdot \sigma_2^2(t)\right)^2}$$

as well as KS-statistic

$$KS(t) = \sup_x \left| F(x) - \frac{t}{T} \cdot \Phi\left(\frac{x}{\sigma_1(t)}\right) - \left(1 - \frac{t}{T}\right) \cdot \Phi\left(\frac{x}{\sigma_2(t)}\right) \right|$$

where the fit parameters $\{\sigma_1(t), \sigma_2(t)\}$ on the observable periods (0, t) and (t +1, T), respectively, are shown in Fig. 3

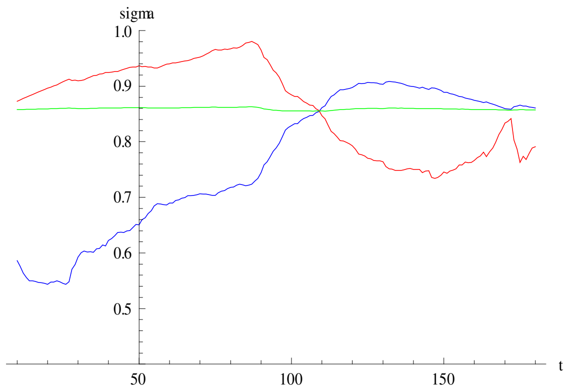


Fig. 3 Evolution of the std. deviations $\sigma_1(t)$ (blue), $\sigma_2(t)$ (red) and their weighted sum σ (green)

We depict both KS-statistic and kurtosis

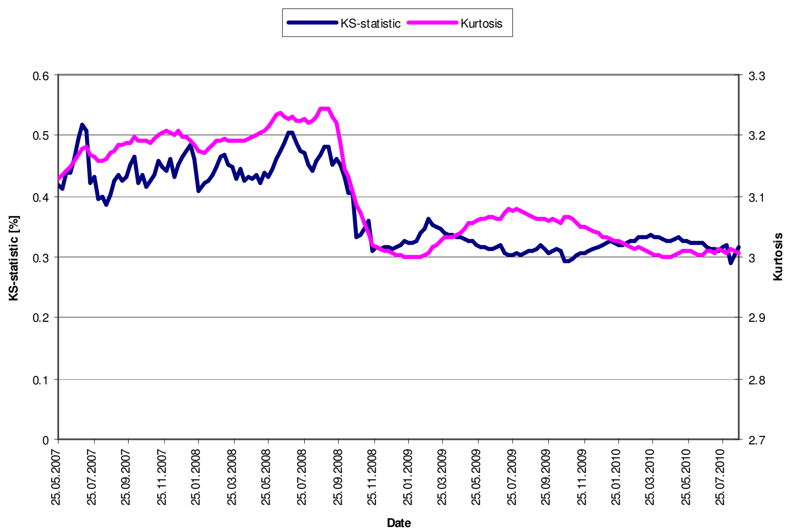


Fig.4 Kurtosis and KS-statistic of mixture of normal distributions over time

and note that the kurtosis is always greater than 3 since the mixture of two zero-mean normal densities always has a higher peak and heavier tails than the normal density of the same variance on the one hand. On the other hand the kurtosis as well as KS-statistic oscillates around two different values. Here we also define a period (T_1 = 87 (12.09.08), T_1 = 98 (28.11.08)) of a transition from one magnitude of oscillations of the statistics to another one. What was remarkable during this period was the Lehman Brother's collapse.

2 Correlation estimation

2.1.1 One Factor Model

Assume we have a set of n_s obligors (stocks) belonging to an industry sector ***s***. Associated with an obligor ***i*** is a latent variable $V_t^{(i)}$, which represents the normalized log-return on an obligor's assets at ***t***. $V_t^{(i)}$ is given by

$$V_t^{(i)} = \sqrt{\rho} f_{t,s} + \sqrt{1-\rho} \cdot \varepsilon_t^{(i)}, \tag{1}$$

where $f_{t,s}$ is a systematic risk factor (eg, industy ***s*** specific indice) at time ***t***. $\varepsilon_t^{(i)}$ represents ***i***-th obligor-specific risk. Based on above empirical evidence for the kurtosis for log-returns (Fig. 4) which can be approximated by a constant of 3 as well as according to the KS-test both $f_{t,s}$ and $\varepsilon_t^{(i)}$ are here assumed to have a standard normal distribution and are jointly independent and $\varepsilon_t^{(i)}$ is independent across obligors.

We also assume that obligors in a given industry have a single common risk factor and measure the sensitivity of each obligor to $f_{t,s}$ by a factor loading, ρ. For two industries ***i*** and ***j***, the corresponding factors f_i and f_j are assumed to be correlated and to possess a correlation coefficient $\rho_{i,j}$.

The correlation estimation procedure uses the two-step MLE method described in [Kalkbrenner, Onwunta 2009]. First the correlations of firms within each of the industry sectors are calculated (intra-sector correlations). Using these results, the correlations of firms within different industry sectors (inter-sector correlations) are calculated.

2.1.2 Estimation of the intra-sector correlations

Given a dataset of 2063 "normal" time series of log-returns $V_t^{(i)}$ we define the maximum likelihood estimator to the one-factor model (1) or more precisely to the model parameter ρ in the following three steps:

1. By construction (1) for an obligor ***i*** from sector ***s*** we get: $V_t^{(i)} - \sqrt{\rho} f_{t,s} \sim N\left(0, \sqrt{1-\rho}\right)$ or immediately in terms of a likelihood function at a time ***t***

$$L_{t,s}^{(i)}(f,\rho) = \frac{1}{\sqrt{2\pi}\sqrt{(1-\rho)}} e^{-\frac{\left(V_t^{(i)} - \sqrt{\rho} f\right)^2}{2(1-\rho)}}. \tag{2}$$

2. The marginal likelihood for V during an observation period T is thus:

$$\Lambda_s(\rho) = \prod_{t=1}^{T} \int_{-\infty}^{\infty} \prod_{i=1}^{n_s} L_{t,s}^{(i)}(f,\rho) d\Phi(f) \equiv (2\pi)^{-T/2} \prod_{t=1}^{T} \int_{-\infty}^{\infty} e^{-\frac{f^2}{2}} \prod_{i=1}^{n_s} L_{t,s}^{(i)}(f,\rho) df \tag{3}$$

3. Estimate of ρ can be obtained by maximizing the marginal likelihood for each sector

$$\rho_s = \arg\underbrace{\max}_{0\le\rho<1} \Lambda_s\left(\rho\right) \cdot \tag{4}$$

We note that both the integration schemes in calculation of likelihood (3) (e.g. Gauss-Hermite scheme) and numerical methods in searching of its extremes (4) can lead to significant errors. Fortunately the likelihood (3) can be both integrated and maximized analytically.

In Appendix we derive from (3)

$$\Lambda_s\left(\rho\right) \propto \frac{\left(1-\rho\right)^{T(1-n_s)/2}}{\left(1-\rho+n_s\rho\right)^{T/2}} \cdot Exp\left[\frac{\rho\cdot\mu\cdot n_s^2\left(T-1\right)+\left(1-\rho+n_s\rho\right)\cdot\left(T-1\right)n_s}{2\left(1-\rho\right)\cdot\left(1-\rho+n_s\rho\right)}\right], \tag{5}$$

where

$$\mu = \frac{1}{n_s^2}\Sigma\left(P_s\right) \equiv \frac{1}{n_s^2}\sum_i^{n_s}\sum_j^{n_s} r_{i,j} \tag{6}$$

is a mass of a correlation matrix P_s with elements

$$r_{i,j} = \frac{1}{T-1}\sum_{t=1}^{T} V_t^{(i)} V_t^{(j)} \,. \tag{7}$$

$r_{i,j}$ is a Pearson's correlation of weekly log-returns for a pair (i,j) of firms from sector ***s***.

Following Düllmann et al. (2008) asset correlations are estimated by μ, "the mean of the pair-wise correlation of all firms". It is referred as "direct" estimation method. For the parameter μ holds $n_s^{-1} \le \mu < 1$.

A first derivative of the likelihood (5) is factorized into product of cubic polynomial and exponential function.

Hence the maximizing of the MLE (5) leads to searching of the root of a cubic equation. Omitting technical details we present a Cardano's formula for optimum (4):

$$\rho_s = \arg\underbrace{\max}_{0\le\rho<1} \Lambda_s\left(\rho\right) = \sqrt[3]{-\frac{q}{2}+\sqrt{D}} + \sqrt[3]{-\frac{q}{2}-\sqrt{D}} - \frac{a}{3} \equiv \rho\left(n_s, T, \mu\right), \tag{8}$$

where

$$D = \left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2 > 0,\ \ p = b - \frac{a^2}{3};\ q = \frac{2a^3}{27} - \frac{ab}{3} + c;\ \ a = \frac{\left(n_s-1\right)\left(2+n_s-3T\right)+n_s^2\left(T-1\right)\cdot\mu}{\left(n_s-1\right)-T\left(n_s-1\right)^2};$$

$$b = \frac{3+n_s-2T}{\left(n_s-1\right)-T\left(n_s-1\right)^2};\ \ c = \frac{2n_s-T-1+n_s^2\left(T-1\right)\cdot\mu}{\left(n_s-1\right)^2-T\left(n_s-1\right)^3}$$

Tables 5 and 6 collect the results of (6)-(8) calculations for three groups of firms.

ICB sectors/North America	all firms		DJ STOXX companies	
	# of firms	rho	# of firms	rho
Oil & Gas	152	37.44	30	71.58
Chemicals	56	38.99	12	64.01
Basic Resources	63	39.18	14	67.12
Construction & Materials	55	40.43	9	64.91
Industrial Goods & Services	472	31.95	59	58.61
Automobiles & Parts	24	38.01	4	92.67
Food & Beverage	66	24.12	15	45.99
Personal & Household Goods	124	34.77	22	48.60
Health Care	289	21.64	34	44.55
Retail	167	34.15	39	46.81
Media	56	30.61	11	62.23
Travel & Leisure	71	35.54	10	55.35
Telecommunications	23	31.95	3	-
Utilities	79	53.50	29	69.90
Banks	210	31.02	7	65.62
Insurance	45	40.75	6	64.53
Real Estate	13	38.41	0	-
Financial Services	99	34.84	14	61.50
Technology	368	27.81	51	49.65
Average	128.00	35.01	19.42	60.80

Table 5: Intra-sector correlations in %

	# of firms	mass mju	rhos
Oil & Gas	122	30.89	31.82
Chemicals	44	33.52	34.89
Basic Resources	49	32.94	34.28
Construction & Materials	46	36.60	37.67
Industrial Goods & Services	413	28.59	29.19
Automobiles & Parts	20	32.77	34.94
Food & Beverage	51	18.99	22.01
Personal & Household Goods	102	32.48	33.41
Health Care	255	18.93	20.16
Retail	128	31.18	32.07
Media	45	24.15	26.60
Travel & Leisure	61	32.69	33.90
Telecommunications	20	26.94	30.01
Utilities	50	47.10	47.48
Banks	203	29.68	30.45
Insurance	39	39.43	40.36
Real Estate	13	35.59	38.41
Financial Services	85	31.07	32.19
Technology	317	24.69	25.50
Average	108.58	30.96	32.39

Table 6: intra-sector correlations [%] of "no stoxx" firms

Assume for every sector

$$\rho_{all} = \nu_1 \rho_{stoxx} + \nu_2 \rho_{nostoxx} \tag{9}$$

where the weights are normalized to sum up to a parameter ν such that $\nu_1 = \frac{n_{stoxx}}{n_{all}} \cdot \nu,\ \nu_2 = \frac{n_{nostoxx}}{n_{all}} \cdot \nu$. From the results in Tables 6 – 7 we derive an almost uniform distribution of the weight coefficient ν across sectors as shown in Table 7.

ICB sectors/North America	weight coefficient
Oil & Gas	0,944
Chemicals	0,948
Basic Resources	0,942
Construction & Materials	0,960
Industrial Goods & Services	0,972
Automobiles & Parts	0,853
Food & Beverage	0,878
Personal & Household Goods	0,963
Health Care	0,940
Retail	0,962
Media	0,911
Travel & Leisure	0,963
Telecommunications	0,816
Utilities	0,960
Banks	0,981
Insurance	0,935
Real Estate	N/A
Financial Services	0,959
Technology	0,964

Table 7: distribution of the weight parameter across sectors

Denote $\gamma = \frac{n_{stoxx}}{n_{all}}$ and rewrite (9) in the following form

$$\rho_{all} = \gamma \cdot \nu \cdot (\rho_{stoxx} - \rho_{nostoxx}) + \nu \cdot \rho_{nostoxx}. \tag{10}$$

Hence a common intra-sector correlation linearly increases with γ (or a number of "stoxx" firms n_{stoxx}) on the interval $\nu\rho_{nostoxx} \le \rho_{all} \le \nu\rho_{stoxx}$. Thus e.g. to estimate a quintile-based credit/market risk measure for a portfolio containing both liquid and illiquid names two components of the intra-sector correlations are to be used according above weighted rule (9) - (10).

In Appendix: Tables A1 and A3 we also collect the MLE-results for different aggregations of "no stoxx" firms across sectors and note that the intra-sector correlations for companies with greater market capitalization / high credit quality / number of employees are bigger ones for companies with smaller capitalization / low credit quality / number of employees.

Fig. 6 gives a geometric interpretation of dependency of the MLE (8) on the mass μ (6) across ICB sectors as given in Table 7.

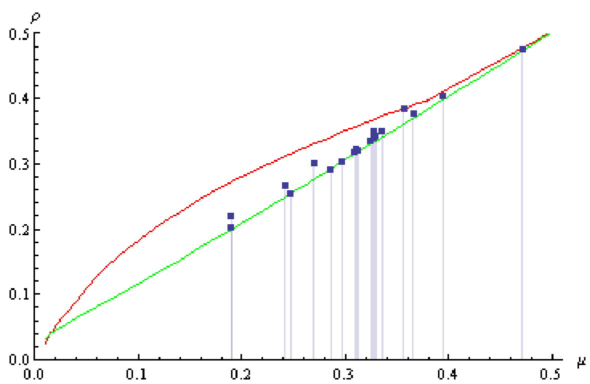


Fig. 6. MLE vs μ

A "cloud" of the intra-sector correlations is bounded by two curves $\rho_{nostoxx}(n_{17}=13, T=190, \mu)$ (red curve) and $\rho_{nostoxx}(n_5=413, T=190, \mu)$ (green curve) with minimal and maximal number of firms (stocks) in sector s=17 ("Real Estate") and s=5 ("Industrial Goods & Services"), respectively. Note also that for the same fixed number n_s^* of firms in every sector the all intra-correlations lie on a unique Cardano's curve $\rho\left(n_s^*, T=190, \mu\right)$.

The Cardano's curve approaches line $\rho(n_s, T, \mu) \underset{n_s\to\infty, T\to\infty}{\to} \mu$.

Proof. Due to L'Hôpital's rule we have the following chain of the limits:

$$b \underset{n_s\to\infty}{\to} 0;\ \ c \underset{n_s\to\infty}{\to} 0;\ \ a \underset{n_s\to\infty}{\to} \frac{2+2T\mu}{-2T} \underset{T\to\infty}{\to} -\mu \Rightarrow p \underset{n_s\to\infty, T\to\infty}{\to} -\frac{\mu^2}{3} \text{ and } q \underset{n_s\to\infty, T\to\infty}{\to} \frac{2\mu^3}{27}$$

leading first to zero discriminant D and then to a double zero root and a simple root

$$\frac{9c-4ab+a^3}{3b-a^2}, \text{ i.e. } \rho_s \underset{n_s\to\infty, T\to\infty}{\to} \frac{9\cdot 0-4\cdot(-\mu)\cdot 0-\mu^3}{3\cdot 0-(-\mu)^2} \equiv \mu \ \text{ as } a^2-3b \equiv \mu^2-3\cdot 0 \neq 0.$$

It means that asymptotically

$$\rho_s = \mu \tag{11}$$

or by other words an asymptotic MLE of the intra-sector correlation is given by a mass of a matrix with the Pearson's correlations (7).

If we admit the heavy tails for the risk factors distributions for each of the 19 industry sectors the estimates for intra-correlations can be also calculated in above three steps with the likelihoods (2)-(3) modified according to the distributions assumptions: e.g. the systematic factor follows a normal mixture distribution and the idiosyncratic factors are normally distributed and hence a latent variable $V_t^{(i)}$ by (1) has a normal mixture distribution.

2.1.3 Impact of Lehman Brother's collapse on the correlations

Taking into account the statistical analysis of log-returns time-series the intra-sector correlations can be decomposed into three components for (12.01.07, 12.09.08), (19.09.08, 28.11.08) and (4.12.08, 3.09.10) periods .

Applying the methodology (2)-(5) separately to every period one can obtain the following results:

		12.01.07 - 12.09.08		19.09.08 - 28.11.08		04.12.08 - 03.09.10	
	# of firms	mass mju	rhos	mass mju	rhos	mass mju	rhos
Oil & Gas	122	18.02	20.39	50.51	54.96	28.75	30.16
Chemicals	44	21.82	24.96	50.06	54.50	33.82	35.47
Basic Resources	49	21.72	24.73	49.32	53.84	30.92	32.77
Construction & Materia	46	25.18	27.80	56.85	60.64	35.22	36.71
Industrial Goods & Ser	413	17.35	18.85	46.76	51.59	27.65	28.67
Automobiles & Parts	20	19.01	24.75	51.37	55.67	29.87	32.74
Food & Beverage	51	13.92	18.03	34.78	40.87	17.32	20.89
Personal & Household	102	22.80	24.86	48.66	53.27	31.20	32.55
Health Care	255	9.63	12.15	39.09	44.63	17.26	19.01
Retail	128	22.78	24.65	53.51	57.68	27.32	28.79
Media	45	16.12	20.09	47.20	51.92	20.13	23.45
Travel & Leisure	61	24.99	27.32	51.69	55.99	28.47	30.39
Telecommunications	20	17.08	23.41	41.96	47.34	25.93	29.43
Utilities	50	37.36	38.66	59.10	62.69	49.81	50.35
Banks	203	24.86	26.29	38.49	44.09	30.32	31.43
Insurance	39	26.55	29.18	57.00	60.76	40.07	41.24
Real Estate	13	25.22	31.99	43.23	48.60	37.14	39.29
Financial Services	85	22.11	24.42	48.13	52.79	29.27	30.87
Technology	317	15.70	17.46	44.45	49.49	22.43	23.73

Table 8: Variation of intra-sector correlations over time

After the Lehman failure in September 2008 the rise in correlations took place across all economic sectors. In order to investigate an impact of the Lehman's perturbation on the dynamics of correlations we substitute into the marginal likelihood (3) the time series with a variable number of the log-returns $\{V_t, t=1,\ldots\tau\}$ covering period (12.01.07, 12.01.07 + τ weeks) and calculate the MLEs replacing in (8) T by τ. Fig. 7 shows a perturbed dynamics of the estimates of correlations across sectors for a period (25.05.07, 25.08.10), i.e. $\tau = 19,\ldots,T-1$.

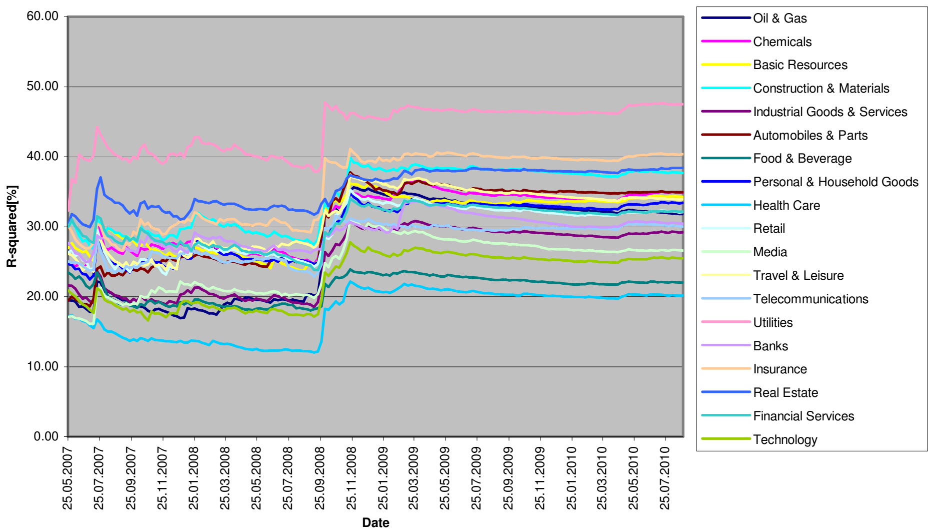


Fig. 7. Evolution of the intra-sector correlations across ICB sectors

as well as Fig.8 depicts their average curve (blue) in comparison to a dynamics of the correlations derived from the log-returns data in a post-Lehman episode.

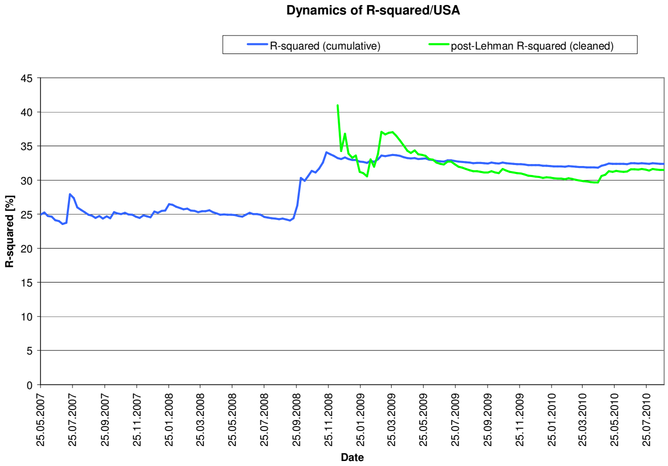


Fig. 8. Variation of average R-squared over time

Our research indicates that the estimates smoothly vary over time before as well as after Lehman failure. At the same time the increase in correlations during the Lehman episode goes exponentially. In a post-Lehman episode the correlations don't return to the values before the Lehman Brother's collapse and stabilize around a new magnitude in 32%.

2.1.4 Bias adjustment : parade of correlations

The ML-estimate (8) and the mass (6) have the following biases [see details in appendix]

$$bias_{\rho}(n)=\begin{cases}\begin{cases}\dfrac{-4.029\rho^2+0.588\rho+0.64}{\sqrt{n}}+2.159\rho^4-4.654\rho^3+2.65\rho^2-0.551\rho+0.00093, & 10\le n\le 100\\ \dfrac{-1.281\rho+0.546}{\sqrt{n}}-5.817\rho^3+2.672\rho^2-0.375\rho+0.0136, & 100\le n\le 1000\end{cases} & if\ \ \rho\le 0.3;\\ \begin{cases}\dfrac{5.129\rho^2-5.482\rho+1.66}{\sqrt{n}}+107.98\rho^6-85.624\rho^4+36.89\rho^2-14.12\rho+1.487, & 10\le n\le 100\\ \dfrac{1.31987\rho^2-1.74476\rho+0.608}{\sqrt{n}}+4.33\rho^7+0.492\rho^3-2.79346\rho^5-0.0068, & 100\le n\le 1000\end{cases} & if\ \ 0.3\le\rho\le 0.6\end{cases}$$

$$bias_{\mu}(n)=\begin{cases}\begin{cases}\dfrac{-0.4252\mu+0.411}{\sqrt{n}}-2.5\mu^3+1.265\mu^2-0.139\mu-0.0292, & 10\le n\le 100\\ \dfrac{42.168\mu^4-4.634\mu^2+0.492\mu+0.106}{\sqrt{n}}-4.075\mu^3+1.624\mu^2-0.18346\mu+0.00313, & 100\le n\le 1000\end{cases} & if\ \ \mu\le 0.3;\\ \begin{cases}\dfrac{0.903\mu^2-1.147\mu+0.553}{\sqrt{n}}+91.074\mu^6-70.812\mu^4+29.771\mu^2-11.264\mu+1.173, & 10\le n\le 100\\ \dfrac{0.249\mu^2-0.429\mu+0.219}{\sqrt{n}}+12.457\mu^6-12.605\mu^4+7.067\mu^2-3.16\mu+0.406, & 100\le n\le 1000\end{cases} & if\ \ 0.3\le\mu\le 0.6\end{cases}$$

which can be fitted from synthetic standard normally distributed time series f and ϵ generated by (1) with a given "true" parameter ρ. Note also that a standard deviation goes to zero as $\sim 1/\sqrt{n}$.

The bias adjustment of both parameters leads to a simple approximate $\hat{\rho} = \rho - bias_\rho(n) \approx \hat{\mu} = \mu - bias_\mu(n)$, as shown in Fig. 9 and Table 9.

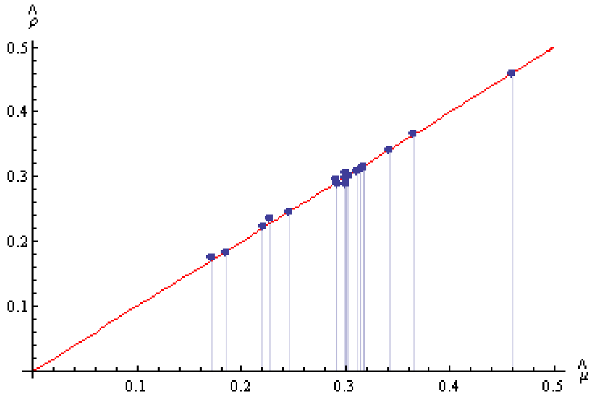


Fig. 9. Parade of intra-sector correlations

	# of firms	mass mju	rhos
Oil & Gas	122	29.87	29.99
Chemicals	44	31.69	31.57
Basic Resources	49	31.48	31.35
Construction & Materials	46	34.19	34.19
Industrial Goods & Services	413	29.03	29.63
Automobiles & Parts	20	29.14	28.91
Food & Beverage	51	17.07	17.62
Personal & Household Goods	102	31.49	31.51
Health Care	255	18.48	18.47
Retail	128	30.20	30.30
Media	45	21.97	22.46
Travel & Leisure	61	31.70	31.61
Telecommunications	20	22.68	23.72
Utilities	50	45.93	45.99
Banks	203	29.96	28.97
Insurance	39	36.49	36.62
Real Estate	13	30.00	30.88
Financial Services	85	31.07	31.05
Technology	317	24.57	24.75
Average	108.58	29.32	29.45

Table 9: Nostoxx firms: adjusted intra-sector correlations and masses of correlation matrices across sectors

2.1.5 Estimation of the inter-sector correlations

The above methodology can be extended to cross- correlations of obligors in different sectors, say ***i*** and ***j***. We assume that all obligors in sector ***i*** depend only on the systematic factor f_i and have the same $R^2 \equiv \rho_i$. The systematic factors f_i and f_j follow a bivariate normal distribution with correlation ρ. Thus we obtain the following likelihood function

$$\Lambda_{i,j}^{\text{inter}}(\rho) = \prod_{t=1}^{T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_{cross}(x, y, \rho_i, \rho_j)\, d\Phi_2(x, y, \rho) \tag{12}$$

where $L_{cross}(x, y, \rho_i, \rho_j) = \left(\prod_{k=1}^{n_i} L_{t,i}^{(k)}(x, \rho_i) \right) \cdot \left(\prod_{k=1}^{n_j} L_{t,j}^{(k)}(y, \rho_j) \right)$ with the pre-calculated sector-specific conditional likelihoods (2) at time t. In Appendix we derive from (12)

$$\Lambda_{i,j}^{\text{inter}}(\rho) \propto \frac{1}{\sqrt{\left(1-\rho_i\right)^{Tn_i}\left(1-\rho_j\right)^{Tn_j}\left(4p_i^2p_j^2\left(1-\rho^2\right)-\rho^2\right)^T}} Exp\left[\mathrm{X}+\frac{Q}{4\tilde{p}^2}\right] \quad . \tag{13}$$

Maximizing (13) leads to

$$\rho_{i,j} = \arg\underbrace{\max}_{0\le\rho<1}\Lambda_{i,j}^{\text{inter}}(\rho) \quad . \tag{14}$$

We apply (14) and calculate the inter-correlations for each pair of the 19 industries as shown in Table 10.

	Oil & Gas	Chemicals	Basic Reso	Constructio	Industrial G	Automobile	Food & Be	Personal &	Health Car	Retail	Media	Travel & Le	Telecommu	Utilities	Banks	Insurance	Real Estate	Financial S	Technology
Oil & Gas	100,00%	-	-	-	-	-	-	-	20,63%	69,06%	-	-	-	-	24,69%	-	-	-	13,28%
Chemicals	-	100,00%	81,25%	90,63%	-	-	80,47%	17,97%	-	11,09%	93,75%	52,81%	-	68,91%	4,06%	-	-	27,50%	-
Basic Resources	-	81,25%	100,00%	-	-	-	83,91%	20,63%	-	12,66%	97,34%	59,84%	-	73,28%	4,06%	-	-	32,50%	-
Construction & Materials	-	90,63%	-	100,00%	-	-	84,38%	20,63%	-	12,81%	-	59,84%	-	68,13%	4,53%	-	-	30,63%	-
Industrial Goods & Services	-	-	-	-	100,00%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Automobiles & Parts	-	-	-	-		100,00%	16,88%	3,91%	-	2,34%	21,72%	10,47%	90,31%	11,72%	0,78%	22,03%	-	5,63%	-
Food & Beverage	-	80,47%	83,91%	84,38%	-	16,88%	100,00%	23,44%	-	14,53%	-	68,91%	-	89,69%	5,31%	-	-	36,72%	-
Personal & Household Goods	-	17,97%	20,63%	20,63%	-	3,91%	23,44%	100,00%	-	69,69%	-	-	-	-	22,81%	-	-	-	-
Health Care	20,63%	-	-	-	-	-	-	-	100,00%	-	-	-	-	-	-	-	-	-	-
Retail	69,06%	11,09%	12,66%	12,81%	-	2,34%	14,53%	69,69%	-	100,00%	-	-	-	-	35,31%	-	-	-	-
Media	-	93,75%	97,34%	-	-	21,72%	-	-	-	-	100,00%	56,41%	-	67,03%	4,06%	-	-	28,75%	-
Travel & Leisure	-	52,81%	59,84%	59,84%	-	10,47%	68,91%	-	-	-	56,41%	100,00%	-	-	7,81%	-	-	55,16%	-
Telecommunications	-	-	-	-	-	90,31%	-	-	-	-	-	-	100,00%	13,28%	0,78%	23,91%	-	5,31%	-
Utilities	-	68,91%	73,28%	68,13%	-	11,72%	89,69%	-	-	-	67,03%	-	13,28%	100,00%	4,84%	-	-	31,25%	-
Banks	24,69%	4,06%	4,06%	4,53%	-	0,78%	5,31%	22,81%	-	35,31%	4,06%	7,81%	0,78%	4,84%	100,00%	-	-	-	-
Insurance	-	-	-	-	-	22,03%	-	-	-	-	-	-	23,91%	-	-	100,00%	-	21,25%	-
Real Estate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100,00%	2,66%	-
Financial Services	-	27,50%	32,50%	30,63%	-	5,63%	36,72%	-	-	-	28,75%	55,16%	5,31%	31,25%	-	21,25%	2,66%	100,00%	-
Technology	13,28%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			100,00%

Table 10: inter-sector correlations of "nostoxx" firms

The average inter-sector correlation is then equal to 17.82%.

3 Regression Analysis

An alternative way to estimate the intra/inter correlations ρ_{stoxx} **for "stoxx" firms** is to resolve the following standard linear regression equation

$$V_t^{(i)} = \beta_i f_{t,s} + \varepsilon_t^{(i)} \tag{15}$$

with respect to "beta" in analogy to the capital asset pricing model (CAPM). Thus the correlations will be calculated on the data of the largest North American stocks from STOXX Americas 600 Index.

The returns $V_t^{(i)}$ are exclusively correlated by means of their composite factors f which are modelled by industry specific indices. We denote the t-th week's return on the s-th index by $f_{t,s}$ and for each of the indices, we consider the last 190 weekly returns and compute the Pearson's correlations of weekly returns for all pairs of indices by

$$R_{i,j} = \frac{1}{T-1}\sum_{t=1}^{T} f_{t,i} f_{t,j}$$

Minimizing residuals $\varepsilon_t^{(i)}$ in (15) leads first to an optimal beta for i-th stock (obligor) in sector s:

$$\beta_s^{(i)} = \arg\underbrace{\min}_{\beta}\sum_t\left(V_t^{(i)} - \beta f_{t,s}\right)^2$$

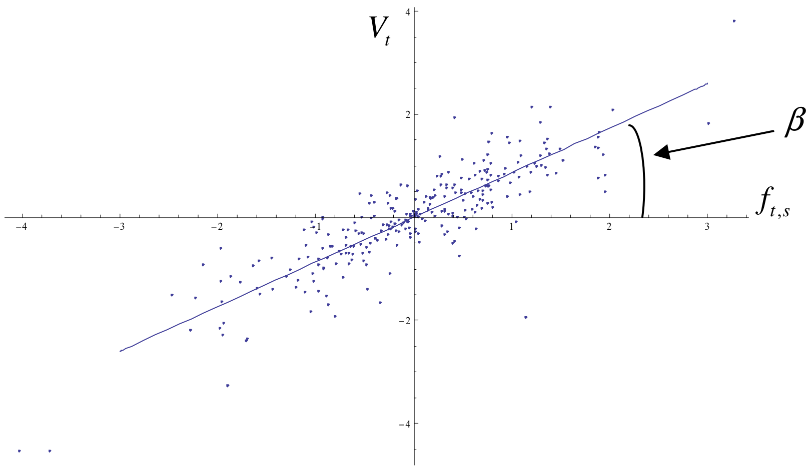


Fig. 10. The log-returns of Exxon Mobil Corp vs DJS Amer 600 Oil & Gas

and then to the natural estimate of the intra-sector correlation

$$\beta_s = \frac{1}{n_s}\sum_{i=1}^{n_s}\beta_s^{(i)} \tag{16}$$

and finally to the inter-sector covariate

$$\rho_{i,j} = \beta_i \cdot \beta_j \cdot R_{i,j} \quad . \tag{17}$$

Direct comparison of two models (1) and (15) gives an obvious linking equation:

$$\beta_s^2 = \rho_s$$

which can be used to test both models.
Tables 11 and 12 collect the intra- and inter-sector correlations computed by (16) and (17), respectively.

ID	ICB sector name	DJ STOXX companies: betas	
		11.11.03-10.12.09	**12.01.07-03.09.10**
1	Oil & Gas	71,68	74,45
2	Chemicals	75,23	75,61
3	Basic Resources	62,59	64,47
4	Construction & Materials	70,03	76,44
5	Industrial Goods & Servic	60,26	72,95
6	Automobiles & Parts	73,87	77,25
7	Food & Beverage	55,32	58,94
8	Personal & Household Go	51,34	63,03
9	Health Care	56,64	64,74
10	Retail	57,74	63,40
11	Media	67,85	72,07
12	Travel & Leisure	61,74	68,83
13	Telecommunications	59,17	62,04
14	Utilities	74,6	77,89
15	Banks	68,52	76,17
16	Insurance	63	67,97
17	Real Estate	87,04	86,07
18	Financial Services	68,45	72,94
19	Technology	63,06	66,85
	Average	65,69	70,64

Table 11: "beta"-version of the intra-sector correlations

	Oil & Gas	Chemicals	Basic Reso	Constructio	Industrial G	Automobile	Food & Be	Personal &	Health Car	Retail	Media	Travel & Le	Telecomm	Utilities	Banks	Insurance	Real Estate	Financial S	Technology
Oil & Gas	100,00	78,63	81,85	77,62	72,51	63,24	70,25	66,90	63,13	63,39	77,23	60,40	69,84	84,43	56,75	72,06	N/A	66,63	73,27
Chemicals	78,63	100,00	78,38	86,44	81,65	71,94	67,51	63,32	58,11	70,46	74,97	71,00	62,67	67,03	59,61	70,55	N/A	71,39	77,14
Basic Resources	81,85	78,38	100,00	74,26	64,65	57,64	52,42	47,74	45,52	50,75	61,30	52,77	49,90	63,59	39,89	54,11	N/A	52,85	63,49
Construction & Materials	77,62	86,44	74,26	100,00	88,21	80,19	67,06	69,74	63,03	78,27	82,92	80,92	69,02	66,94	72,64	80,05	N/A	83,10	80,18
Industrial Goods & Services	72,51	81,65	64,65	88,21	100,00	86,21	75,80	78,67	70,56	85,43	89,86	87,88	72,97	68,94	77,02	84,15	N/A	87,56	85,89
Automobiles & Parts	63,24	71,94	57,64	80,19	86,21	100,00	63,02	67,93	55,42	79,99	82,20	81,14	64,55	61,57	73,81	77,91	N/A	82,07	80,46
Food & Beverage	70,25	67,51	52,42	67,06	75,80	63,02	100,00	89,00	83,54	77,05	81,37	72,09	75,73	76,87	59,25	75,59	N/A	72,80	72,71
Personal & Household Goods	66,90	63,32	47,74	69,74	78,67	67,93	89,00	100,00	82,55	84,01	82,83	79,57	76,61	73,24	62,12	76,35	N/A	75,86	74,95
Health Care	63,13	58,11	45,52	63,03	70,56	55,42	83,54	82,55	100,00	74,43	76,88	69,67	73,42	73,44	59,14	77,27	N/A	71,83	69,46
Retail	63,39	70,46	50,75	78,27	85,43	79,99	77,05	84,01	74,43	100,00	86,80	88,38	76,15	66,98	68,97	78,61	N/A	86,15	84,06
Media	77,23	74,97	61,30	82,92	89,86	82,20	81,37	82,83	76,88	86,80	100,00	84,34	80,01	74,79	76,65	86,60	N/A	88,04	84,43
Travel & Leisure	60,40	71,00	52,77	80,92	87,88	81,14	72,09	79,57	69,67	88,38	84,34	100,00	69,04	60,40	73,22	78,29	N/A	84,70	80,19
Telecommunications	69,84	62,67	49,90	69,02	72,97	64,55	75,73	76,61	73,42	76,15	80,01	69,04	100,00	73,35	62,98	74,38	N/A	74,56	74,94
Utilities	84,43	67,03	63,59	66,94	68,94	61,57	76,87	73,24	73,44	66,98	74,79	60,40	73,35	100,00	52,51	73,51	N/A	64,72	71,23
Banks	56,75	59,61	39,89	72,64	77,02	73,81	59,25	62,12	59,14	68,97	76,65	73,22	62,98	52,51	100,00	85,99	N/A	88,96	62,24
Insurance	72,06	70,55	54,11	80,05	84,15	77,91	75,59	76,35	77,27	78,61	86,60	78,29	74,38	73,51	85,99	100,00	N/A	89,67	73,28
Real Estate	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	100,00	N/A	N/A
Financial Services	66,63	71,39	52,85	83,10	87,56	82,07	72,80	75,86	71,83	86,15	88,04	84,70	74,56	64,72	88,96	89,67	N/A	100,00	80,91
Technology	73,27	77,14	63,49	80,18	85,89	80,46	72,71	74,95	69,46	84,06	84,43	80,19	74,94	71,23	62,24	73,28	N/A	80,91	100,00

Table 12: "beta"-version of the inter-sector correlations on a period: 11.11.2003 – 10.12.2009

4 Value at Risk (VaR) of equity portfolio with normal mixtures

It is known that if all k equities of a portfolio are mapped to the same single risk factor (e.g. the market index, see details in section 3) the **normal** VaR at the (1-α)-confidence level is simply equal to

$$VaR_{\alpha} = -\Phi^{-1}(\alpha) \cdot P \cdot \frac{\sigma_N}{\sqrt{n}} \sum_{i}^{k} \omega_i \beta_i$$

where P is a total portfolio value, $\omega_i = \dfrac{y_i}{P}$, y_i is market value of i-th equity. At the same time the normal assumption could lead to the underestimation of VaR as it will be shown in sequel. We have already seen that the degree of excess kurtosis in the stocks return time series is considerably higher before than after crisis (Lehman failure). Here we qualitatively investigate an impact of "tail behaviour" of stocks returns distributions on VaR of a linear equity portfolio with a portfolio mean μ_N and a volatility of σ_N over n-day horizon.

We assume the three distributions of the portfolio P&Ls: normal, normal mixture and normal with the weighted mean and deviation such that a n-day VaR at a significance level α has three possible values as roots of the following nonlinear algebraic equations:

$$\alpha = \Pr(P\&L < -VaR_\alpha) = \begin{cases} pl_1(VaR_\alpha) \equiv \Phi\left(\dfrac{-VaR_\alpha / P - \mu_N}{\sigma_N}\right), & V_t \sim N(0,1) \\ pl_2(VaR_\alpha) \equiv p \cdot \Phi\left(\dfrac{-VaR_\alpha / P - \mu_{NM}}{\sigma_{NM}}\right) + (1-p)\cdot\Phi\left(\dfrac{-VaR_\alpha / P - \mu_N}{\sigma_N}\right), & V_t \sim NM_\lambda \\ pl_3(VaR_\alpha) \equiv \Phi\left(\dfrac{-VaR_\alpha / P - [p\mu_{NM} + (1-p)\cdot\mu_N]}{p\sigma_{NM} + (1-p)\cdot\sigma_N}\right), & V_t \sim N(0,\sigma_\lambda) \end{cases} \quad (18)$$

where p is the probability of regime 1 (before crisis), μ_{NM} is a n-day portfolio return and σ_{NM} is a n-day standard deviation in regime 1. Regime 2 characterizes ordinary market circumstances with a pair μ_N and σ_N.

For fixed parameters: $P = 100Mio; \mu_{NM} = -0.3; \sigma_{NM} = \dfrac{1}{\sqrt{250/n}}; \mu_N = \dfrac{0.15}{250/n}; \sigma_N = \dfrac{0.2}{\sqrt{250/n}}; n = 10$ we resolve (18) w.r.t. VaR for two significance levels α = 5% (dashed) and 10% (solid) and arbitrary probability of crash p as shown on Fig. 11.

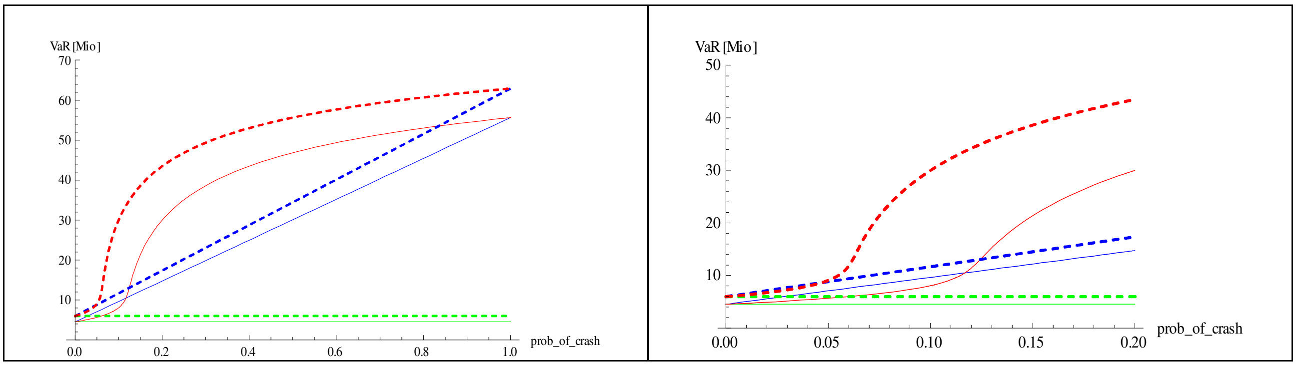
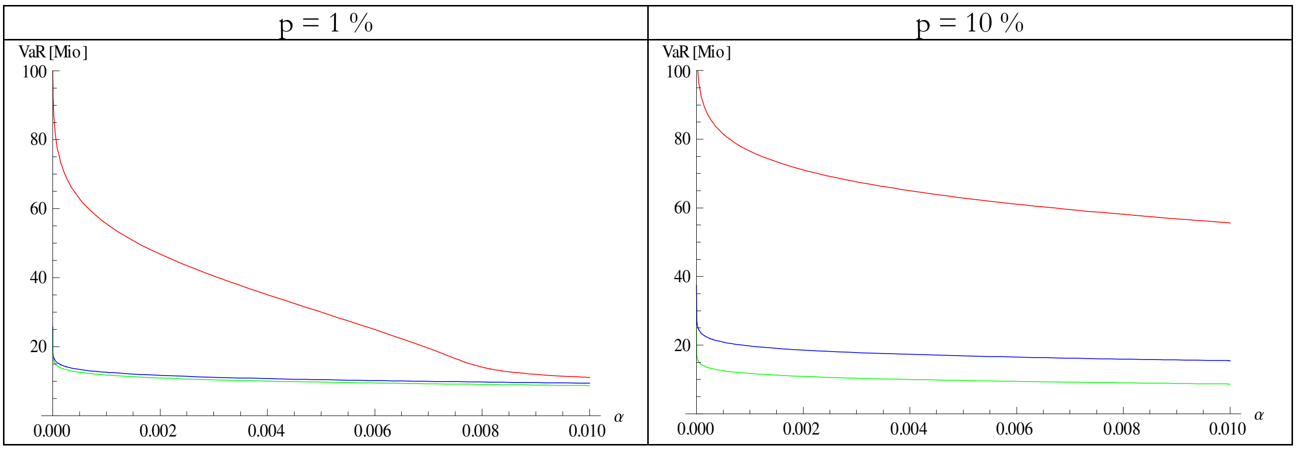


Fig. 11. VaR_α vs probability p for normal (green), NM (red) and equivalent normal (blue)

Thus ignoring the possibility of a crash can seriously underestimate the VaR. For low significance levels (e.g. 10% or 20%), the normal assumption ($pl_3(VaR_\alpha)$) can overestimate VaR if $p \le \alpha$.

The parameters of a normal mixture density function can be estimated from historical data by use of the expectation–maximization (EM) algorithm [7]. Thus we would be able to quantify the probability p.

Another case of study is to fix the probability p, e.g. 1% and 10%. Then we get



We see that for higher significance level NM VaR is considerably bigger both normal VaRs even for small probability of crash.

5 Conclusions

In this work we first carried out an empirical analysis of the equity time series covering a 4y period from 2007 to 2010. Then we consider a normal distribution assumption for the risk factors within a two-state version of the CreditMetrics framework and derive the maximum likelihood estimator in closed form. Concurrent to MLE asset correlations are estimated by mass μ or the mean of pair wise equity sample correlations. We show that the sample correlations are less biased than the ML-estimates and asymptotically both methods lead to the same correlations. Based on the ICB industry classification we computed the bias adjusted intra- and inter - correlations for the 19 industry sectors with an average value of 29.45% and 17.82%, respectively. We also investigated a dynamics of the correlations and correlation changes under stressed market conditions (Lehman failure in September 2008) and studied an impact of normal mixture assumption on the VaR of simple equity portfolio .

6 References

[1] M.Kalkbrenner and A. Onwunta, Validating structural credit portfolio models. In Model Risk - Challenges und Solution for Financial Risk Models, Risk Books, 2010

[2] Duellmann, M., J. Küll and M. Kunisch, Estimating Asset Correlations From Stock Prices or Default Rates – Which Method is Superior?, Journal of Economic Dynamics and Control, 34(11),2010

[3] Gradsteyn and Ryzhik, Tables of Integrals, series, and products, Seventh Edition, 2007

[4] A. L. Baranovski, Dynamical systems forced by shot noise as a new paradigm in the interest rate modelling, SFB 649, Humboldt-Universität zu Berlin, 2010 http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2010-037.pdf

[5] T. Altmann, T. Schmidt and W. Stute. "A Shot Noise Model for Financial Assets ", 2008. International Journal of Theoretical and Applied Finance, Vol 11, No. 1, p. 87 – 106

[6] A.P.Dempster, N.M.Laird, and D.B. Rubin. Maximum likelihood from incompletedata via the EM algorithm. Journal of the Royal Statistical Society B, 39(1):1-22,1977

Appendix.

Proof of Eq. (5):

The integrand in (3) is

$$e^{-\frac{f^2}{2}} \cdot \left(\frac{1}{\sqrt{2\pi(1-\rho)}} \right)^{n_s} e^{-\frac{1}{2(1-\rho_s)} \sum_{i=1}^{n_s} \left(V_t^{(i)} - \sqrt{\rho} f \right)^2} = e^{-\frac{f^2}{2}} \cdot \left(2\pi(1-\rho) \right)^{-n_s/2} \cdot e^{-\frac{1}{2(1-\rho)} \left[\sum_{i=1}^{n_s} \left[\left(V_t^{(i)} \right)^2 + \rho f^2 - 2\sqrt{\rho} f \cdot V_t^{(i)} \right] \right]} \equiv A(\rho) \cdot e^{-p^2 \cdot f^2 + q \cdot f},$$

where we denoted $A(\rho)=\left(2\pi(1-\rho)\right)^{-n_s/2}\cdot e^{-\frac{1}{2(1-\rho)}\sum_{i=1}^{n_s}\left(V_t^{(i)}\right)^2}$ and $p^2=\frac{1-\rho+n_s\rho}{2(1-\rho)};q=\frac{\sqrt{\rho}}{1-\rho}\sum_{i=1}^{n_s}V_t^{(i)}$.

Since

$\int_{-\infty}^{\infty}e^{-p^2x^2+q\cdot x}dx=e^{\frac{q^2}{4p^2}}\frac{\sqrt{\pi}}{p}$ [Gradsteyn&Ryzhik, p.337]

we get a likelihood in a form:

$$\Lambda_s(\rho)=(2\pi)^{-T/2}\prod_{t=1}^{T}\left(A(\rho)\cdot\frac{\sqrt{\pi}\sqrt{2(1-\rho)}}{\sqrt{1-\rho+n_s\rho}}\cdot e^{\frac{\rho\cdot\left(\sum_{i=1}^{n_s}V_t^{(i)}\right)^2}{2(1-\rho)(1-\rho+n_s\rho)}}\right).$$

$$=(2\pi)^{-T/2}\left(2\pi(1-\rho)\right)^{-Tn_s/2}\cdot\frac{\left[2\pi(1-\rho)\right]^{T/2}}{(1-\rho+n_s\rho)^{T/2}}Exp\left(-\frac{\sum_{t=1}^{T}\sum_{i=1}^{n_s}\left(V_t^{(i)}\right)^2}{2(1-\rho)}+\frac{\rho\cdot\sum_{t=1}^{T}\left(\sum_{i=1}^{n_s}V_t^{(i)}\right)^2}{2(1-\rho)(1-\rho+n_s\rho)}\right)$$

Thus

$$\Lambda_s(\rho)\propto\frac{(1-\rho)^{T(1-n_s)/2}}{(1-\rho+n_s\rho)^{T/2}}\cdot Exp\left[-\frac{(1-\rho+n_s\rho)\sum_{t=1}^{T}\sum_{i=1}^{n_s}\left(V_t^{(i)}\right)^2+\rho\cdot\sum_{t=1}^{T}\left(\sum_{i=1}^{n_s}V_t^{(i)}\right)^2}{2(1-\rho)(1-\rho+n_s\rho)}\right] \tag{A.1}$$

Taking into account that the log-returns time series are standardized, i.e. $\frac{1}{T-1}\sum_{t=1}^{T}\left(V_t^{(i)}\right)^2\approx 1$ we get:

1. $\sum_{t=1}^{T}\sum_{i=1}^{n_s}\left(V_t^{(i)}\right)^2\approx n_s\cdot(T-1)$

2.

$$\sum_{t=1}^{T}\left(\sum_{i=1}^{n_s}V_t^{(i)}\right)^2=\sum_{t=1}^{T}\left(\sum_{i=1}^{n_s}V_t^{(i)}\sum_{j=1}^{n_s}V_t^{(j)}\right)=\sum_{t=1}^{T}\sum_{i=1}^{n_s}\left(V_t^{(i)}\right)^2+2\sum_{t=1}^{T}\sum_{j=1}^{n_s-1}\sum_{i=1}^{n_s-j}V_t^{(i)}V_t^{(i+j)}\approx n_s\cdot(T-1)+2(T-1)\sum_{j=1}^{n_s-1}\sum_{i=1}^{n_s-j}r_{i,i+j}$$
$$=n_s\cdot(T-1)+(T-1)\left(\Sigma(P)-n_s\right)=(T-1)\cdot\Sigma(P)$$

then (A.1) transforms to (5).

MLE-results for different aggregations of "no stoxx" firms across sectors

Oil & Gas	29.52	59.76	-	29.06	59.43
Chemicals	27.39	54.29	-	29.95	63.44
Basic Resources	33.87	53.55	-	37.58	52.37
Construction & Materials	29.75	50.41	62.03	31.71	55.44
Industrial Goods & Services	17.15	41.53	52.24	24.28	51.74
Automobiles & Parts	40.25	52.66	84.27	36.70	84.27
Food & Beverage	21.09	30.40	83.30	22.01	37.36
Personal & Household Goods	26.20	40.72	64.85	29.66	52.85
Health Care	18.92	31.92	46.70	19.83	40.20
Retail	22.26	37.38	42.89	29.23	42.84
Media	26.08	36.50	88.95	27.44	45.68
Travel & Leisure	27.90	39.24	50.63	31.61	52.95
Telecommunications	30.54	59.61	-	33.00	65.74
Utilities	43.49	56.51	-	38.55	58.56
Banks	24.21	64.99	-	13.70	45.73
Insurance	37.58	54.81	-	34.84	49.13
Real Estate	38.56	100.00	-	45.41	85.57
Financial Services	27.60	46.76	-	33.13	48.74
Technology	21.77	37.20	59.88	24.95	45.19
Average	28.64	49.91	63.57	30.14	54.59

Table A1: Intra-sector correlations for different aggregations of "no stoxx" firms across sectors

	employees <1000	1000<employees <10000	10000<employees	1M<assets<1Mrd	assets >1Mrd
Oil & Gas	80	33	1	74	36
Chemicals	19	23	0	28	14
Basic Resources	19	22	2	26	16
Construction & Materials	14	24	7	25	20
Industrial Goods & Services	166	206	43	291	112
Automobiles & Parts	7	9	4	15	4
Food & Beverage	21	27	3	36	13
Personal & Household Goods	39	57	6	76	26
Health Care	184	54	10	216	31
Retail	26	63	40	81	45
Media	19	19	4	28	13
Travel & Leisure	19	23	21	44	16
Telecommunications	11	7	1	12	6
Utilities	19	31	0	11	38
Banks	152	50	0	55	147
Insurance	27	12	1	11	28
Real Estate	11	0	2	8	4
Financial Services	36	26	0	36	23
Technology	188	114	7	255	47
Total	1057	800	152	1328	639

Table A2: number of firms in buckets

	# of firms rated < BBB	sub-investment grade	investment grade	# of firms rated >=BBB
Oil & Gas	30	60.86	74.31	5
Chemicals	7	58.45	73.17	6
Basic Resources	9	53.08	78.77	5
Construction & Materials	9	56.12	74.16	5
Industrial Goods & Services	55	46.63	56.61	29
Automobiles & Parts	6	65.21	-	0
Food & Beverage	5	61.04	55.00	5
Personal & Household Goods	23	51.21	57.95	9
Health Care	19	34.51	77.32	4
Retail	24	43.38	-	1
Media	12	42.32	-	3
Travel & Leisure	18	48.22	-	1
Telecommunications	6	60.87	-	2
Utilities	6	59.60	63.24	26
Banks	4	92.89	72.33	14
Insurance	0	-	-	1
Real Estate	0	-	-	1
Financial Services	3	-	54.17	8
Technology	19	45.29	71.48	5
Total	255			130
Average		54.98	67.37	

Table A3: Intra-sector correlations vs number of firms in buckets

Comments to bias for MLE (8) and for a mass μ (6):

1. Set a triple $\{\rho, n, T\}$ and simulate n **mutually independent** standard normally distributed time series for idiosyncratic factors $\varepsilon_t^{(i)}, \{t=1,...,T; i=1,..,n\}$ and a time series for a common factor $f_t, \{t=1,...,T\}$. Generate the "log-returns" $\{V_t^{(i)}, t=1,...,T; i=1,...,n\}$ by (1).
2. For the given set of time series find both an optimal ρ from (4) and a mass μ from (6)-(7), keep them and then repeat (1) N times.
3. Calculate the mean values $\bar{\rho}(n), \bar{\mu}(n)$ and standard deviations $\sigma_\rho(n), \sigma_\mu(n)$ from the samples $\{\rho_n^{(i)}, i=1,...,N\}$ of ρ and $\{\mu_n^{(i)}, i=1,...,N\}$ of μ, respectively.
4. Set a new value for n and repeat (1)-(3) K times.
5. Set a new value for ρ and repeat (1)-(4) L times.

As a result we get the following tables

number of firms\rhos	5,00%	10,00%	15,00%	20,00%	25,00%
100	9,82%	14,09%	18,37%	22,69%	27,03%
200	8,48%	12,75%	17,17%	21,77%	26,24%
300	7,84%	12,19%	16,80%	21,40%	25,94%
400	7,46%	11,82%	16,49%	21,17%	25,73%
500	7,23%	11,59%	16,30%	21,11%	25,58%
600	7,03%	11,46%	16,20%	21,00%	25,53%
700	6,87%	11,34%	16,12%	20,93%	25,48%
800	6,75%	11,25%	16,07%	20,87%	25,47%
900	6,67%	11,18%	16,00%	20,83%	25,46%
1000	6,57%	11,12%	15,96%	20,82%	25,42%

Table A4. **Empirical distribution of a MLE (8) vs number of firms and given rhos**

number of firms\mju	5,00%	10,00%	15,00%	20,00%	25,00%
100	6,00%	10,89%	15,87%	20,78%	25,58%
200	5,56%	10,45%	15,44%	20,47%	25,24%
300	5,38%	10,31%	15,41%	20,35%	25,11%
400	5,30%	10,21%	15,30%	20,26%	25,00%
500	5,29%	10,15%	15,23%	20,29%	24,92%
600	5,24%	10,16%	15,22%	20,24%	24,91%
700	5,20%	10,13%	15,21%	20,22%	24,89%
800	5,19%	10,12%	15,22%	20,19%	24,90%
900	5,19%	10,11%	15,19%	20,17%	24,91%
1000	5,16%	10,11%	15,18%	20,19%	24,89%

Table A5. **Empirical distribution of mass μ (6) vs number of firms and given rhos**

We first find the best fits to these data in a form $b(\rho)n^{-1/2}+\rho+a(\rho)$. Then from the fits we get the series $B=\{b(5\%),b(10\%),b(15\%),b(20\%),b(25\%)\}$ as well as $A=\{a(5\%),a(10\%),a(15\%),a(20\%),a(25\%)\}$ which can be easily approximated by suitable polynomials as shown on Fig. A1 and Fig. A2 for MLE (8) and Fig. A3 and Fig. A4 for MLE and mass μ, respectively.

$b(\rho)\approx -1.28119\rho+0.546$	$a(\rho)\approx -5.817\rho^3+2.672\rho^2-0.375\rho+0.0136$

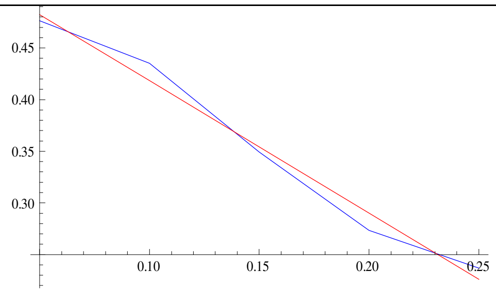


Fig.A1 Drift B of MLE and its polynomial fit

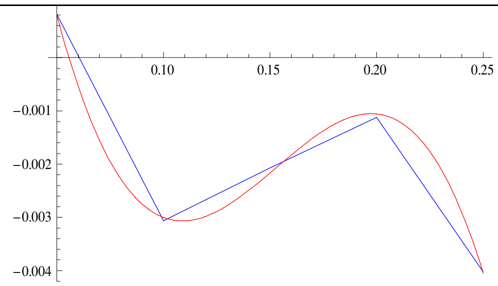


Fig.A2 Shift A of MLE and its polynomial fit

Table A6: Plots of the MLE-fit's components and their approximates

$b_\mu(\rho) \approx 42.168\rho^4 - 4.634\rho^2 + 0.492\rho + 0.106$	$a_\mu(\rho) \approx -4.075\rho^3 + 1.624\rho^2 - 0.18346\rho + 0.00313$

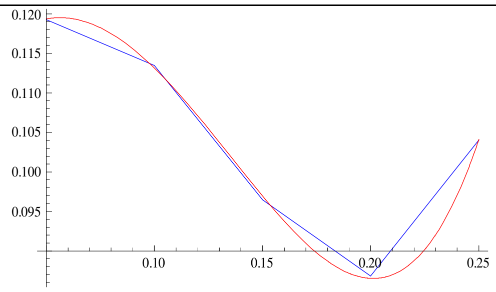


Fig.A3 Drift B of mass μ and its polynomial fit

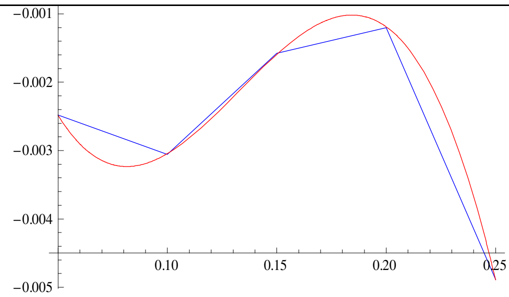


Fig.A4 Shift A of mass μ and its polynomial fit

Table A7: Plots of the mass-fit's components and their approximates

Proof of Eq. (13):

First we derive

$$L_{cross}\left(x, y, \rho_{s_i}, \rho_{s_j}\right) = \prod_{k=1}^{n_{s_j}} \frac{e^{-\frac{\left(V_{t,j}^{(k)}-\sqrt{\rho_j}x\right)^2}{2(1-\rho_j)}}}{\sqrt{2\pi}\sqrt{(1-\rho_j)}} \prod_{m=1}^{n_{s_i}} \frac{e^{-\frac{\left(V_{t,i}^{(m)}-\sqrt{\rho_i}y\right)^2}{2(1-\rho_i)}}}{\sqrt{2\pi}\sqrt{(1-\rho_i)}} \propto (1-\rho_i)^{-n_i/2}(1-\rho_j)^{-n_j/2}\tilde{L}_{cross}\left(x, y, \rho_{s_i}, \rho_{s_j}\right),$$

where

$$\tilde{L}_{cross}\left(x, y, \rho_i, \rho_j\right) = Exp\left(-\frac{1}{2(1-\rho_j)}\left[\sum_{k=1}^{n_j}\left[\left(V_{t,j}^{(k)}\right)^2 + \rho_j x^2 - 2\sqrt{\rho_j}\,x\cdot V_{t,j}^{(k)}\right]\right] - \frac{1}{2(1-\rho_i)}\left[\sum_{m=1}^{n_i}\left[\left(V_{t,i}^{(m)}\right)^2 + \rho_i y^2 - 2\sqrt{\rho_i}\,y\cdot V_{t,i}^{(m)}\right]\right]\right).$$

Then an internal integral in (12) calculates as

$$I = \int_{-\infty}^{\infty} \tilde{L}_{cross}\left(x, y, \rho_i, \rho_j\right) e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dx = Exp\left(-\frac{\alpha_j}{2(1-\rho_j)} - \frac{\alpha_i + \rho_i y^2 n_i - 2\sqrt{\rho_i}\,y\cdot\beta_i}{2(1-\rho_i)} - \frac{y^2}{2(1-\rho^2)}\right)\int_{-\infty}^{\infty} e^{-p_j^2x^2+q_jx}dx$$

$$= \frac{\sqrt{\pi}}{p_j} Exp\left(\frac{q_j^2}{4p_j^2} - \frac{\alpha_j}{2(1-\rho_j)} - \frac{\alpha_i + \rho_i y^2 n_i - 2\sqrt{\rho_i}\,y\cdot\beta_i}{2(1-\rho_i)} - \frac{y^2}{2(1-\rho^2)}\right)$$

where

$\alpha_j(t)=\sum_{k=1}^{n_j}\left(V_{t,j}^{(k)}\right)^2$; $\beta_j(t)=\sum_{k=1}^{n_j}V_{t,j}^{(k)}$ and $p_j^2=\dfrac{\rho_j n_j}{2(1-\rho_j)}+\dfrac{1}{2(1-\rho^2)}$; $q_j=\dfrac{\sqrt{\rho_j}\beta_j}{1-\rho_j}+\dfrac{\rho y}{1-\rho^2}$.

An external integral is then given by

$$II=\int_{-\infty}^{\infty} I\cdot dy=e^{\xi}e^{\frac{\tilde{q}^2}{4\tilde{p}^2}}\frac{\sqrt{\pi}}{\tilde{p}}\frac{\sqrt{\pi}}{p_j},$$

where

$$\tilde{p}^2=p_i^2-\frac{\rho^2}{4p_j^2\left(1-\rho^2\right)^2};\tilde{q}=\frac{\sqrt{\rho_j}\beta_j\rho}{2p_j^2(1-\rho_j)(1-\rho^2)}+\frac{\sqrt{\rho_i}\beta_i}{(1-\rho_i)}\text{ and }\ \xi=\frac{\rho_j\beta_j^2}{4p_j^2\left(1-\rho_j\right)^2}-\frac{\alpha_i}{2(1-\rho_i)}-\frac{\alpha_j}{2(1-\rho_j)}.$$

Since

$\Lambda_{i,j}^{\text{inter}}(\rho)=\prod_{t=1}^{T}\dfrac{(1-\rho_i)^{-n_i/2}\left(1-\rho_j\right)^{-n_j/2}}{2\pi\sqrt{1-\rho^2}}II$ we come to $\Lambda_{i,j}^{\text{inter}}(\rho)\propto\left[(1-\rho_i)^{n_i}\left(1-\rho_j\right)^{n_j}\tilde{p}^2\cdot p_j^2\left(1-\rho^2\right)\right]^{-T/2}\prod_{t=1}^{T}e^{\xi+\frac{\tilde{q}^2}{4\tilde{p}^2}}$.

Denoting

$$\begin{aligned}
&\overline{\alpha}_j=\sum_{t=1}^{T}\alpha_j(t);\quad \overline{\beta}_j^2=\sum_{t=1}^{T}\beta_j^2(t);\ \overline{\beta}_{i,j}=\sum_{t=1}^{T}\beta_i(t)\beta_j(t);\\
&\mathrm{X}=\frac{\rho_j\overline{\beta}_j^{\,2}}{4p_j^2\left(1-\rho_j\right)^2}-\frac{\overline{\alpha}_i}{2(1-\rho_i)}-\frac{\overline{\alpha}_j}{2(1-\rho_j)}\\
&Q=\sum_{t=1}^{T}\tilde{q}^2=\frac{\rho_j\overline{\beta}_j^2\rho^2}{4p_j^4\left(1-\rho_j\right)^2\left(1-\rho^2\right)^2}+\frac{\overline{\beta}_{i,j}\sqrt{\rho_j\rho_i}\cdot\rho}{p_j^2(1-\rho_j)(1-\rho^2)(1-\rho_i)}+\frac{\rho_i\overline{\beta}_i^2}{(1-\rho_i)^2}
\end{aligned}$$

we get

$$\cdot\ \Lambda_{i,j}^{\text{inter}}(\rho)\propto\frac{1}{\sqrt{(1-\rho_i)^{Tn_i}\left(1-\rho_j\right)^{Tn_j}\left(4p_i^2p_j^2\left(1-\rho^2\right)-\rho^2\right)^T}}Exp\left[\mathrm{X}+\frac{Q}{4\tilde{p}^2}\right].$$

As the log-returns time series are standardized we have

$$\begin{aligned}
&\overline{\alpha}_j=\sum_{t=1}^{T}\alpha_j(t)\approx n_j\cdot(T-1);\\
&\overline{\beta}_j^2=\sum_{t=1}^{T}\beta_j^2(t)\approx n_j^2\cdot(T-1)\cdot\mu_j;\\
&\overline{\beta}_{i,j}=\sum_{t=1}^{T}\beta_i(t)\beta_j(t)\approx n_i\cdot n_j\cdot(T-1)\cdot\mu_{i,j}
\end{aligned}$$

where $\mu_{i,j}=\dfrac{1}{n_j\cdot n_i}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}r_{k,l}$ is a mass of a matrix of the Pearson's cross-correlations $r_{k,l}=\dfrac{1}{T-1}\sum_{t=1}^{T}V_{t,i}^{(k)}V_{t,j}^{(l)}$ for a pair

(***k,l***) of firms from sectors ***i*** and ***j***, respectively.

Collecting the above formulas we finally come to (13).