MIXED-STABLE MODELING OF HIGH-FREQUENCY FINANCIAL DATA: PARALLEL COMPUTING APPROACH

Igoris Belovas[1,2], Vadimas Starikovičius[2]
[1] *Vilnius University Institute of Mathematics and Informatics*
[2] *Vilnius Gediminas Technical University*
Vilnius, Lithuania
e-mail: Igoris.Belovas@mii.vu.lt

Abstract

In this paper we apply the mixed-stable model for the analysis of high-frequency German DAX stock return data. We demonstrate the inadequacy of the classical Gaussian model as well as standard α-stable models. We develop efficient parallel numerical algorithms for the maximum likelihood estimation of mixed-stable parameters. The research has showed that the application of modern parallel technologies allows a fast estimation of mixed-stable parameters even for large amounts of data. We have studied the influence of the accuracy of probability density function calculation and maximum likelihood optimization on the results of the modelling and processing time and constructed mixed-stable models for all 29 DAX companies. The adequacy of the modelling was verified with Koutrouvelis goodness-of-fit test.

1 Introduction

Modern electronic trading provided market analysts with huge quantity of trading data. High-frequency data contain all transactions of the financial market and can reveal events and laws impossible to uncover with monthly, weekly or even daily data [3]. However, this comes with its price. The processing time increases with the amount of the data. Meanwhile in business decision-making the time is crucial.

Handling huge amounts of high-frequency data is virtually impossible without the application of modern parallel technologies. Common perception is that the most accurate method of processing (maximum likelihood) is very time-consuming and therefore it is often rejected. However, as we show, the application of parallel technologies makes this method both precise and practical. In this research, we develop and expand our parallel computing approach [2] to the mixed-stable modelling [4] of high-frequency data. A summary of the literature covering high-frequency and intra-daily data research is presented in Sun et al. [8] and the references therein.

2 The data

In practice we often observe a large number of zero returns in the high-frequency return data due to the fact that the underlying asset price does not change at given short-time intervals. The mixed-stable model is designed to deal with these unique features, often observed in the high-frequency return data. In this research we apply the mixed-stable model for the analysis of high-frequency German DAX stock return data (from January 1 to December 27 of 2007). Inhomogeneous raw intra-daily data were aggregated into the equally-spaced homogeneous intra-daily time series with previous-tick interpolation [4, 9]. Having processed yearly high-frequency returns data for 29 German DAX companies with different time steps, we have obtained that almost all data series are asymmetric, and the empirical kurtosis shows that density functions of the series are more peaked than Gaussian.

However, it should be pointed out, that rather often high-frequency empirical data exhibit the stagnation effect, i.e. series contain numerous zero returns [4]. We have examined the obtained returns series for the effect and discovered that a strong stagnation effect (43 to 82 percent zeros at 10 sec time step) manifested itself in most high-frequency series. To take the zero effect into account, we apply a generalized mixed-stable model [4].

3 Mixed-stable model

The probability density function of a mixed-stable random variable is

$$f(x,\Theta) = (1-r)p(x,\Theta) + r\delta(x), \tag{1}$$

where $p(x,\Theta)$ is a probability density function of an α-stable distribution [6, 7] and $\delta(x)$ is the Dirac delta function. $\Theta = (\alpha, \beta, \mu, \sigma)$ is a vector of stable parameters and the coefficient $r \in (0,1)$ is the index of stagnation. The empirical cumulative distribution functions of data series with the stagnation effect exhibits jumps at $x = 0$. Model (1) enables us to accommodate to these jumps. For each return series we estimate the vector of stable parameters Θ by maximizing corresponding log-likelihood function.

Precise and fast calculation of stable densities $p(x,\Theta)$ is a nontrivial task [1]. It is crucial part of the computational algorithm. To deal with the integral representation of the probability density function, we substitute the improper integral with a definite integral using specially designed upper integration bound $\Delta = \Delta(\alpha, \varepsilon)$. To optimize the log-likelihood function we use the Nelder-Mead method.

Having estimated the parameters of the mixed-stable law we must test the adequacy of the modelling. Since we have a discontinuous distribution function, classic methods for continuous distributions (ex. Kolmogorov-Smirnov, Anderson-Darling) do not work. Therefore we apply a special goodness-of-fit test based on characteristic functions (characteristic functions are uniformly continuous on the entire space), proposed by Koutrouvelis [5].

4 Parallelization and results

Several hierarchical levels of parallelism can be distinguished and defined: multiple independent returns sets, optimization method, maximum likelihood target function, and probability distribution function integral calculation. In this research we have chosen a fine-grained parallelization at the maximum likelihood target function calculation level and implemented it with OpenMP and MPI. The performed test showed that this parallel algorithm is very efficient and scalable (see Table 1).

Table 1: Performance and scalability of maximum likelihood parallelization with the MPI on IBM SP6; time step $\Delta t = 10$ and set size $n = 436143$.

proc	1	32	64	128	192	256	320	384	512
T_p	6598.01	210.81	107.05	55.42	36.50	27.86	22.29	18.80	14.03
S_p	-	31.30	61.64	119.06	180.79	236.83	295.95	350.96	470.13
E_p	-	98%	96%	93%	94%	93%	93%	91%	92%

We have studied the influence of the accuracy of probability density function calculation (ε_{pdf}) and maximum likelihood optimization (ε_{ML}) on the results of the modelling and processing time. We have found that insufficient accuracy results in faulty outcomes. We need at most $\varepsilon_{pdf} = 10^{-9}$ and $\varepsilon_{ML} = 10^{-6}$ to achieve plausible results.

We have constructed the mixed-stable models for 29 DAX companies. Table 2 contains estimates of model parameters: stagnation r and stable parameters $\Theta = (\alpha, \beta, \mu, \sigma)$. We must stress that we could obtain these results in a reasonable time only using parallel computations. This supports the employment of mixed-stable modelling in high-frequency finance analysis with the use of modern parallel technologies.

Acknowledgement

The work has been performed under the Project HPC-EUROPA2 (Project number 1020), with the support of the European Community - under the FP7 "Research Infrastructures" Programme.

References

[1] Belov I. (2005). On the computation of the probability density function of α-stable distributions. *Mathematical modelling and analysis. Proceedings of the 10th International Conference MMA 2005*, pp. 333-341.

[2] Belov I., Starikovičius V. (2007). Parallelization of α-stable modelling algorithms. *Mathematical modelling and analysis*. Vol. **12**, 4, pp. 409-418.

[3] Dacorogna R. et. al. (2001). *An Introduction of High-Frequency Finance*. Academic Press, San Diego.

[4] Kabašinskas A., Sakalauskas L., Sun W., Belovas I. (2012). Mixed-Stable Models for Analyzing High-Frequency Financial Data. *Journal of Computational Analysis and Applications*. Vol. **14**, 7, pp.1210-1226.

[5] Koutrouvelis I.A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *shape Biometrika*, Vol. **67**,1 , pp. 238-240.

[6] Rachev S., and Mittnik S. (2002). *Stable Paretian Models in Finance*. John Wiley and Sons, New York.

[7] Samorodnitsky G., Taqqu M.S. (2000). *Stable non-Gaussian random processes, stochastic models with infinite variance*. Chapman&Hall, New York-London.

[8] Sun W., Rachev S., Fabozzi, F. (2008). *Long-range dependence, fractal processes, and intra-daily data, in: Handbook of IT and Finance*, Springer, pp. 543–586.

[9] Wasserfallen W., Zimmermann H. (1985). The behavior of intradaily exchange rates. *Journal of Banking and Finance*, 9, pp. 55-72.

Table 2: Maximum likelihood estimates of mixed-stable parameters for 29 DAX financial data series with time step $\Delta t = 10$ sec, $\varepsilon_{ML} = 10^{-7}$, $\varepsilon_{pdf} = 10^{-12}$.

Company	r	α	β	μ	σ
Adidas AG	0.75	1.813224	0.004395	0.000002	0.000456
Deutsche Bank	0.47	1.822488	-0.013724	-0.000001	0.000272
BASF SE	0.58	1.798121	0.009887	0.000001	0.000292
BMW AG St	0.68	1.872899	-0.024771	0.000000	0.000405
Continental AG	0.66	1.704703	-0.007413	0.000000	0.000363
Deutsche Post	0.75	1.933809	0.030695	0.000001	0.000518
Deutsche Telekom	0.73	1.995341	-0.064854	0.000001	0.000572
Bayer AG O.N.	0.60	1.875897	0.023037	0.000001	0.000364
FMC AG	0.77	1.759640	-0.014327	0.000000	0.000484
Deutsche Börse	0.68	1.664209	0.015807	0.000003	0.000413
MAN SE St	0.68	1.669219	0.001631	0.000002	0.000443
Henkel AG	0.77	1.769518	-0.031952	0.000000	0.000511
Infineon Techn.	0.82	1.979982	-0.045985	-0.000004	0.000803
Linde AG	0.74	1.714015	-0.007574	0.000002	0.000367
Merck KGaA	0.77	1.612534	0.003719	0.000000	0.000442
RWE AG St	0.58	1.852625	0.025781	0.000001	0.000330
Daimler AG	0.49	1.853430	0.039817	0.000001	0.000322
SAP AG	0.58	1.919404	0.012302	0.000000	0.000384
Siemens AG	0.46	1.815861	-0.002928	0.000001	0.000276
METRO AG St	0.75	1.767016	0.044610	0.000004	0.000393
ThyssenKrupp	0.68	1.855014	-0.000734	0.000001	0.000461
Volkswagen AG St	0.59	1.744243	-0.004622	0.000002	0.000302
Deutsche Postbank	0.79	1.678412	-0.005400	0.000000	0.000519
HYPO RE	0.75	1.814879	0.027742	0.000000	0.000576
Commerzbank AG	0.66	1.901063	-0.007980	0.000000	0.000477
Deutsche Lufthansa	0.78	1.932395	-0.016547	0.000000	0.000587
Allianz SE	0.43	1.787790	-0.016081	0.000000	0.000255
Münchener Rück	0.59	1.779122	0.006533	0.000000	0.000271
TUI AG	0.80	1.903315	0.012532	0.000003	0.000699