ISBN : 978-0-9742114-7-3

Credit Spread Forecasting using Higher Order Cumulants

Sanya Dudukovic
Franklin Collège Switzerland
Sdudukov@fc.edu

Abstrct :The aim of this article is to throw the light on financial time series modeling when invariants don't have Gaussian distribution. The first part of the article introduce the estimation method based on higher order cumulants and it's efficiency. The second part provides empirical analysis of the cumulants based method as applied to modeling Credit Spread between 10-year AAA corporate bond yields and 10-year Treasury bond yields for the period 1991:01-2007:07.Comparison with dynamical regression model is provided in the third part of the paper. Regression is based on the following explanatory variables : U.S. leading index ,Russell 2000 returns , BBB bond price changes , interest rate swaps, exchange rates EUR/ USD, Repo- rates , S&P 500 returns and S&P 500 volatility, Treasury bill changes , liquidity index-TRSW ,LIBOR rates , Moody's default rates, credit spread volatility and Treasury bills volatility, is also presented in the article. It is finally demonstrated that much of the information about variability of Credit Spread can be extracted from higher order cumulants.

I. INTRODUCTION

The predictability of credit spread has assumed a new importance since both fixed income investors and financial managers need reliable predictions to make more money.
Collin-Dufresne , Golden and Martin (2001) examined firm level related theoretical as determinants of credit spread changes, spot rate changes, changes in slope of yield curve, changes in firm leverage , changes in volatility, changes in the probability and magnitude of a large negative jumps. They concluded that the monthly changes in firm specific factors are not the driving force in credit spread changes .
Besides Merton (1974), Krishman , Ritchken and Thompson (2003) , showed that the predictability of credit spread, based on credit slope, very much depends on the maturity of the corporate bonds. Shorter dated future credit spreads are found to be influenced by market returns and market volatility . The analysis has not found evidence that those factors influence long dated credit spread.
Zhang ,Zhou and Zhu (2005), argued that the unsatisfactory performance of structural models may be in part attributed to the fact that the impact of volatility and jump risks are not treated seriously. They found strong volatility and jumps effects, which predict another 16% of credit spread.
Dudukovic (2006) investigated multiple lagged IAR-GARCH model for U.S. Credit Spread. The following explanatory macro-financial variables were investigated :U.S. Leading index , Russell 2000 returns , 5 year interest rate SWAPS , S&P500 returns, Treasury bill changes , liquidity index and Moody's default rates, S&P 500 volatility, credit spread volatility, treasury bill volatility, exchange rate EUR/USD , Repo rates and Libor rates. All the volatilities , Repo rates , Libor rate and exchange rates were not found to be causally related to credit spread changes.
However Credit Spread determinants are proven to be the following macro variables :
U.S. Leading index , Russell 2000 returns , interest rate SWAPS , S& P 500 returns, Treasury bill changes , liquidity index and Moody's default rates. The obtained model significantly improved predictability of credit spread changes .
Structured models based on micro independent variables ,default rate and recovery rate, have explanatory power which varies from 20% to 50% .The proposed model explained 73% of the credit spread variability.
Statistically speaking ,Credit Spreads Time Series is non Gaussian which means that its autocorrelation function does not provide sufficient statistics for its modeling and ARIMA-

ISBN : 978-0-9742114-7-3

GARCH parameter estimation. Cumulants (in frequency domain called polyspectra) have been received the attention of the statistics and signal processing . Gianninakis (19990) derived cumulant based ARIMA order determination method for communication signals . because second order cumulants for non Gaussian signals vanish , higher order cumulants are generally nonsymmetrical functions of their arguments , and as such carry phase information about ARMA parameters .If the process is not generated from ARMA filter (when noise is not present) , cumulate based order determination is more appropriate .

Such higher-order techniques have typically been used less frequently than their second-order cousins *in* the signals world partly because of the significantly greater amount of processing power required to compute them. However recent1y such techniques have been the subject of increasing study and have found applications in such disciplines as image processing, biological monitoring, and the modeling of wave phenomena (Mendel 1991)

This article initially discusses some of the theory behind higher-order statistics, particularly as it applies to non Gaussian signal ARMA parameter estimation . It then details the steps taken towards constructing such an estimator. These steps are to determine the Credit Spread ARMA model order and sample length necessary to provide accurate cumulant estimates using Mahlab software , to examine the performance of cumulant based estimation for non-Gaussian Credit Spread model , and to examine the ability of the cumulant based model to outperform the classical regression model .

2. THEORY OF HIGHER ORDER STATISTICS

a. Moments

The nth-order moment of a real, stationary sequence X(t) with PDF $f_x(x)$ is given by $E\{X^n\} = \int x^n f_x(x)dx$ for $n = 1,2,3...$, where E{x.} denotes expectation and where *x,* as per probability convention, denotes a specific realization of the random variable X. For $n = 1$ (i.e., first-order), the moment is simply the expected value of *X* and is commonly known as the *mean,* denoted by $\mu x = E\{X\}$ .For $n = 2$ (i.e., second-order), the moment is the autocorrelation, which indicates how well the sequence correlates with a time-lagged version of itself.

Moments of arbitrary order may be calculated via a distribution's *moment- generating function* $M\,x(w) = E\{e^{XW}\} = \int e(xw) f\,x(x)dx$. The n-thorder moment of signal is generated by differentiating $Mx(w)$ *n* times, and setting *w* = O. Thus the mean of a signal may be calculated as $\mu x = E\{X\} = M_x'(0)$, where the single prime mark indicates the first differential of *Mx(w).* Moments of higher order may be generated similarly.

Because the concept of signal lag is frequently of interest in the signals world, however, it *is* frequently used to define second- and higher-order moments. Here, lags will be denoted by 'τn and the nth-order moment written as

$$m^x_n(\tau_1.\tau_2..\ \tau_{n-1})=E\{X(t)X(t+\tau_1)...X(+\tau_n)\} \quad (l)$$

b. Cumulants

Although moments of order greater than two are indeed higher-order statistics, frequent]y *cumulants,* rather than moments, are used. Cumulants may be ca1culated in a manner analogous to that of moments; the *cumulant-generating function* is produced by taking the natural logarithm of the moment-generating function:

Cx(w) = ln(Mx(w)). Differentiating Cx(w) *n* times and setting w = 0, as with the moment- generating function, produces a signal's n-th-order cumulant. However, if the study is limited to order four and below (as it is here), an equally va1id but more intuitive representation (for purposes of this article) is

$$C_n^x(\tau_1.\tau_2..\ \tau_{n-1}) = m_n^x(\tau_1.\tau_2..\ \tau_{n-1}) - m^G_n(\tau_1.\tau_2..\ \tau_{n-1}) \quad (2)$$

where $C_n^x(\tau_1.\tau_2..\ \tau_{n-1})$ is the nth-order cumulant of the random variable *x;* is the nth-order moment

ISBN : 978-0-9742114-7-3

ofthe random variable *x;* and $m_n^G(\tau_1.\tau_2.. \tau_{n-1})$- is the nth-order , moment of a Gaussian random variable that has the same first- and second-order moments as *x.* This notation makes it clear that cumulants may be used to measure the Gaussianity, or lack thereof, of a random variable. It is also a convenient way of illus trating a significant property of cumulants; namely that the cumulant of a Gaussian distributed signal is identica11y zero. (Note that even though the alternative representation above only applies to third- and fourth-order statistics, the nth-order cumulant of a Gaussian-distributed signal is zero for *any* value of *n* greater than two.)

In this paper Credit Spread is modeled as ARIMA(p,q) model , (Box-Jenkins 1976) and GARCH model (Engel -Boleslev1995).. The model building, as usually consists of three steps: model identification (order determination using Akaike's information criterion - AIC), parameter estimation and model testing. The main premises in this methodology is that each stationary time series is treated as output from AR(p), MA(q) or ARIMA filter , which has as input uncorrelated non Gaussian shocks known as "white noise":

$$\mathbf{A(Z)^* DY_t= B2(Z)^* v_t}$$

Where **Z** is a backward shift operator :
$\mathbf{Y_{t-1}=ZY_t}$ **,** $\mathbf{Y_{t-k} =Z^kY_t}$, $\mathbf{A(Z) = 1\text{-}a_1Z\text{-}a_2Z^2 \text{-}a_pZ^p}$ **and**
$\mathbf{B(Z) = 1\text{-}b_1Z\text{-}b_2Z^2 - \ldots b_qZ^q}$ are AR and MA filters of orders p and q respectively , **D** is the first difference filter,

$$\mathbf{DY_t = Y_t - Y_{t-1}},\quad \mathbf{D^k}Y_t\mathbf{=Y_t - Y_{t-k}}$$

This type of models are not always identifiable on the abses of second order moments (Swami , Mendel 1989).
A new method of parameter estimation for non Gaussian processes ,using cumulants follows Yule-Walker system where autocorrelations are replaced by third order cumulants (Gianninakis -1990) :
$C^3_y(\tau1,\tau2)= (\sum(y(t)y(t+\tau_1)y(t+\tau_2))/n$,
where n is a number of observations.

$C^3(q+1\text{-}p,k)$ $C^3(q+2\text{-}p,k)$ $\ldots C^3(q,k)$ a(p) $C^3(q+1,k)$

a(p-1) = $C^3(q+2,k)$

$C_3(q+2\text{-}p,k)$ $C^3(q+3\text{-}p,k)$ $\ldots C^3(q+1,k)$

C3(q,k) C3(q+1,k) …C3(q+p-,k) a(1) $C^3(q+p,k)$

---

3. COMPUTER RESULTS

With the above theory in mind, the purpose of this article was to study third-order cumulants and their application for Credit Spread ARIMA modeling
The different factors discussed here were investigated using MA TLAB and its Higher-Order Spectral Analysis (HOSA) toolbox. MATLAB was used to calculate estimates of the data's third-order cumulants and as well as to estimate ARMA model parameters . Further residuals analysis is done using E-Views.

The results of the ARIMA estimations were compared to their the model regression model obtained in [6].
Data description of Credit Spread is presented in the table 1.

 ISBN : 978-0-9742114-7-3

Table 1:Summary statistics for Credit Spread

	DCRSP	AAA	TRE10Y
Mean	-0.000402	7.08017	5.781264
Median	0	7.17	5.81
Maximum	0.41	9.01	8.28
Minimum	-0.64	4.96	3.33
Std.Dev.	0.116715	0.96092	1.1809
Skewness	-0.405155	-0.2327	0.06616
Kurtosis	8.569198	2.45688	2.156968
Observations	174	174	174

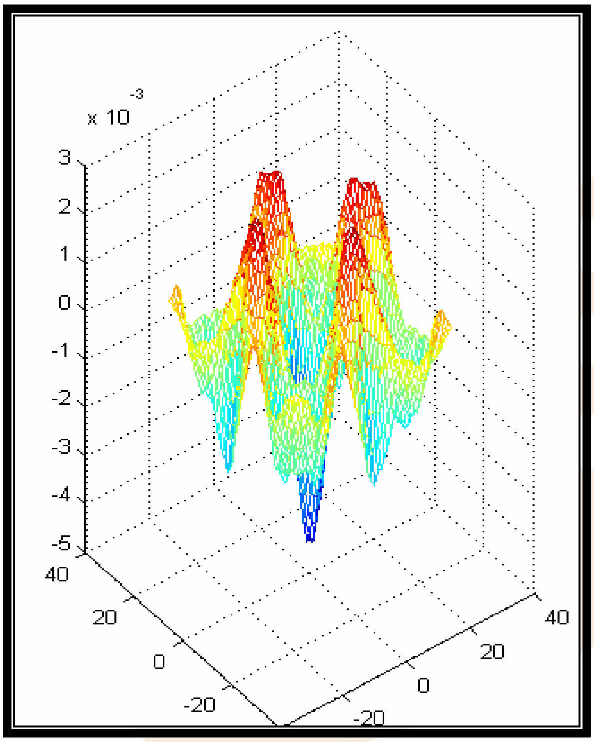
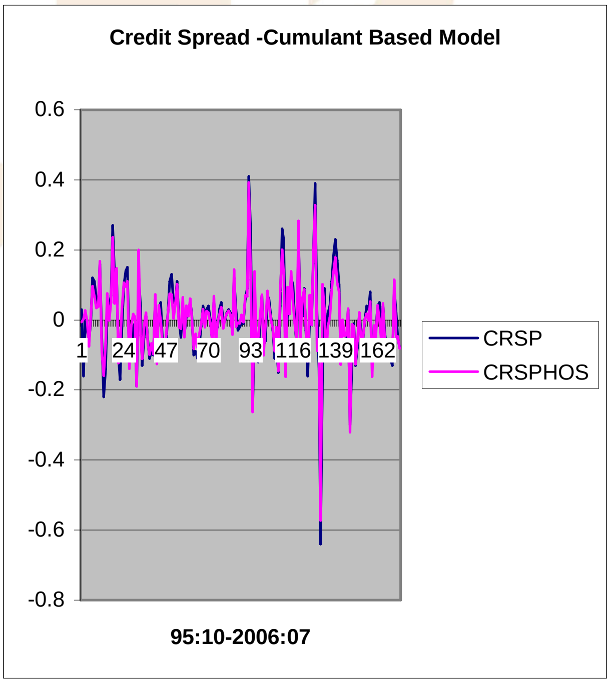


Figure 1: Third order cumulants for Credit Spread First difference

Obtained cumulants based parameters are presented in the table 2. The model and the real Credit Spread data are presented on the figure 2.
Table 2 : Cumulant based ARIMA estimates

Dependent Variable: DCRSP				
Method: Cumulant Based Least Squares				
Date: 12/02/07 Time: 18:57				
Sample(adjusted): 4 174				
Included observations: 171 after adjusting endpoints				
Convergence achieved after 30 iterations				
Backcast: 1 3				
Variable	Coefficien	Std. Error	t-Statistic	Prob.
AR(1)	0.939886	0.152151	6.177341	0
AR(2)	0.668867	0.217999	3.068214	0.0025
AR(3)	-0.718017	0.114241	-6.285113	0
MA(1)	-0.183302	0.103085	-1.778164	0.2911
MA(2)	-0.352788	0.167842	-2.101905	0.0371
MA(3)	0.166202	0.08135	2.043051	0.1616
R-squared	0.857758	Mean dependent v		0.007551
Adjusted R-s	0.853447	S.D. dependent va		0.27951
S.E. of regre	0.107003	Akaike info criterio		-1.59747
Sum squared	1.889176	Schwarz criterion		-1.487237
Log likelihoo	142.5837	Durbin-Watson sta		1.989511
Inverted AR	.89+.22i	.89 -.22i	-0.85	
Inverted MA	.43 -.23i	.43+.23i	-0.69	



ISBN : 978-0-9742114-7-3

Credits Spread model obtained by using Macro Financial Determinants is presented in the table 3.

Table 3

Dependent	Variable:	D(CRSP)		
Sample(adjusted):	1991:01	2005:07		
Included	observation	135		
Variable	Coefficient	Std.Error	t-Statistic	Prob.
USLEADR100(-1)	-0.040062	0.01227	-3.26593	0.001
USLEADR100(-3)	0.034588	0.01143	3.025645	0.003
D(SWAP5(-2))	0.03728	0.01973	1.889892	0.119
D(SWAP10(-1))	-0.284462	0.03068	-9.27221	0
D(TRE3M(-2))	-0.13943	0.03527	-3.95359	1E-04
D(CRSP(-1))	0.328051	0.06139	5.343578	0
D(TRE3M(-1))	0.147576	0.03435	4.295932	0
SP500R(-1)	0.214505	0.1246	1.721577	0.114
R2000R(-1)	-0.276917	0.109	-2.54059	0.012
D(MDEFR(-1))	-0.082142	0.04032	-2.03705	0.105
D(MDEFR(-2))	0.113739	0.05174	2.198339	0.03
D(TRSW(-1))	-0.301093	0.03959	-7.60514	0
	Variance Equation			
C	0.001933	0.00108	1.789954	0.074
ARCH(1)	-0.059636	0.06532	-0.91294	0.361
GARCH(1)	0.696361	0.16662	4.179337	0
R-squared	0.729435	Mean dependent v		0.002

4. CONCLUSION

The first part of the article introduces the estimation method based on higher order cumulants.. The second part provides empirical results obtained by using Matlab functions and E-Views software as applied to Credit Spread between 10-year AAA corporate bond yields and 10-year Treasury bond yields for the period 1995:01-2007:07.Comparison with dynamical regression model is also provided . Regression is based on the following explanatory variables : U.S. leading index ,Russell 2000 returns , BBB bond price changes , 5 year interest rate swaps, exchange rate EUR/ USD, Repo- rates , S& P 500 returns and S&P 500 volatility, Treasury bill changes , liquidity index-TRSW ,LIBOR rates , Moody's default rates, credit spread volatility and Treasury bills volatility, is also presented in the article. It is finally demonstrated that much of the information about variability of Credit Spread can be extracted from higher order cumulants.In fact the coefficient of determination obtained by regression for Credit Spread data is .7294 while the coefficient of determination obtained by using the third order cumulants and applying HOS method appears to be .857 .This demonstrates the fact that HOS based ARMA estimation achieves statistically higher coefficient of determination .

5. REFERENCES

[1] Box G.E.P., Jenkins G.: "Time Series Analysis, Forecasting and Control", Prentice Hall, 1995.

[2] Bolleslev .:"Generalized Autoregressive Conditional Heteroskedasticity ", Journal of Econometrics, 1986,31 307-327.

[3]Collin-Dufresne P. Goldstein S.R., Martin J.S. : "The Determinants of Credit Spread Changes", Journal of Finance ,56, 2177-2207,2001.

[4] Elton J. E, Martin J. Gruber. at all :"Explaining the Rate Spread on Corporate Bonds", Journal of Finance, Vol. LVI, 2001,No. 1, pp. 247-277.

[5] Huang J.,Huang M. : How much of the Corporate treasury Yield Spread is due to Credit Risk ,working Paper, Stanford University, http://www.stanford.edu/~mhuang/ papers/Huang-Huang.pdf ,2003.

[6] Dudukovic S.: "Dynamic Time Series Approach to Credit Spread Modeling", Proceedings of the Fourth Global Conference ion Business& Economics, Oxford University ,2005.

[7] Joutz F.,Mansi S.A., Maxwell W.F.: "The Dynamics of Corporate Credit Spread" , The George Washington University , Working paper,2001.

[8] Roger M. Stein: "Evidence on the Incompleteness of Merton-type Structural Models for Default Prediction" ,2005, http://www.moodyskmv.com/research/files/wp/EvidenceOnIncompletenessOfMerton_AK20050214.pdf.

[9]Huang J.,Kong W.: "Explaining Credit Spread Changes : New Evidence from Option-Adjusted Bond Indexes", The Journal of Derivatives ,Fall 2003,30-44.

[10] Lando D.: "Credit Spread Modeling - Theory and Applications " ,Princeton University Press , 2004 .

[11]Gianninakis , Mendel J.:Cumulant-Based Order Determinntion of Non-Gaussian ARMA Models.

ISBN : 978-0-9742114-7-3