

DOCTOR OF PHILOSOPHY IN MANAGEMENT

METHODS OF ANALYZING STRUCTURAL BREAKS IN
MULTIVARIATE TIME SERIES: APPLICATIONS TO FINANCIAL
DATA

By

ANCHAL SONI



भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

2023

DOCTOR OF PHILOSOPHY IN MANAGEMENT

**METHODS OF ANALYZING STRUCTURAL BREAKS IN
MULTIVARIATE TIME SERIES: APPLICATIONS TO FINANCIAL
DATA**

By

Anchal Soni

A Dissertation submitted in Partial Fulfillment of the Requirements for the degree of
Doctor of Philosophy at

INDIAN INSTITUTE OF MANAGEMENT BANGALORE

2023

Prof. Ananth Krishnamurthy
Chairperson
Doctoral Programme

Prof. Soudeep Deb
Chairperson
Dissertation Advisory Committee

Members of the Dissertation Advisory Committee

- | | |
|------------------------|-------------|
| 1. Prof. Soudeep Deb | Chairperson |
| 2. Prof. Rishideep Roy | Member |
| 3. Prof. Debojyoti Das | Member |

Copyright ©2023 by Anchal Soni
All rights reserved.

*To my parents - the stationary in the
volatile world*

ACKNOWLEDGEMENTS

I am grateful to God that I got an opportunity to write this part of my thesis which I have been longing for. The journey of a doctoral student is very different from that of any other student. There are many highs and lows in this voyage to become a PhD. I would like to express my gratitude to all the people who have cooperated and tolerated me in this path of completing my research.

To start with, I would forever be in debt to my DAC chair, Prof Soudeep Deb. His structured and organized methodology of working helped me complete my work on time. He has been extremely patient with me when times were rough. His continued motivation and timely guidance assisted me through this programme. His unwavering support and belief in me are the reasons behind my successful and wide research. I have enjoyed working with him and deeply appreciate his constant encouragement. I would also like to express my special thanks to the members of my DAC – Prof Rishideep Roy and Prof Debojyoti Das for their continuous support and mentoring. Their valuable insights have helped me look into my research from different perspectives. The immense knowledge and rich experience of my DAC members have encouraged me every day in my professional and personal life as well. My DAC have been extremely kind with their quick feedback always.

My gratitude to the current Doctoral programme chairperson, Prof Ananth Krishnamurthy for his enduring support during my thesis. I would also like to acknowledge the continuous support of the member of the doctoral programme office at IIMB for being always available and facilitating smooth administrative formalities.

I must thank Prof Haritha Saranga, the former Doctoral programme chairperson who showed her trust in me always and made me believe how I could complete my research efficiently.

I became a mother during my PhD and that made this research journey even more wonderful. Ever since the start of the course, I wanted to get my degree in the presence of my child. God has been kind enough to fulfil my dream. My daughter Ishanvi, as her name suggests, fills me with so much strength. I gathered all my energy to finish my

research after I embraced her. Her smile excites me to work more and more.

I would like to mention my husband, Harshit, who has been so patient with me in this entire journey. His faith in me that I could do this was unshakable. He has been my source of inspiration and support from the day I applied for this programme. He supported me continuously in the middle of every crisis. I am obligated to him.

Now requires the special mention of my parents (Babita-Surendra Agarwal) who have always been my support from the day I was born. I worked so comfortably while they took care of my daughter. They spent sleepless nights with me so that I could work efficiently. No amount of words and sentences can sum up the contribution and sacrifices made by my parents (P), my husband (H) and my daughter (D). Without the P, H and D in my life, this degree would have been impossible. I know how desperate they are to see me succeed in life.

I would like to mention the support of my siblings and friends who have always encouraged me to complete my work better and faster.

CONTENTS

1	Introduction	3
1.1	Preliminary definitions	8
1.1.1	Fourier series	8
1.1.2	Periodogram	9
1.1.3	Kullback-Leibler (KL) divergence	11
1.1.4	Standard models of financial data	11
1.1.5	Residual distributions	15
1.2	Final Remarks	17
2	Review of existing structural break detection algorithms	19
2.1	Typology	19
2.1.1	Methods for change in parameters of VAR	20
2.1.2	Methods to detect changes in mean, covariance or distribution	26
2.2	Conclusion	40
3	t-SNE based method of changepoint detection	41
3.1	t-SNE	43
3.2	Empirical justification of t-SNE	46
3.3	Proposed method of changepoint detection	49
3.4	Simulation	53
3.4.1	Data Generating processes	53
3.4.2	Evaluation criteria	60
3.4.3	Results and discussion for structure type A and B	62
3.4.4	Results of Simulation structure type C and D	70
3.5	Conclusion	80
4	Application to financial data	82
4.1	Data and method	83
4.1.1	Principal findings	84

4.2	Empirical analysis	85
4.2.1	Descriptive statistics	85
4.2.2	Statistical models	87
4.2.3	VaR forecasting and model evaluation	90
4.2.4	Group transfer entropy	92
4.3	Results and discussion	95
4.3.1	Structural breaks	95
4.3.2	Results from GARCH models	96
4.3.3	Results from GAS models	99
4.3.4	Volatility spillover between cryptocurrencies	101
4.4	Conclusion	104
5	Structural break detection using ICA	116
5.1	Principal Component Analysis	119
5.1.1	Optimization problem	121
5.1.2	Lagrangian method of calculating PCs	122
5.1.3	ICA as a successor of PCA	123
5.2	Independent Component Analysis	125
5.2.1	Measures of non-Gaussianity	126
5.2.2	Measures of Independence	129
5.2.3	Algorithms to solve ICA	131
5.3	Empirical justification of ICA	133
5.4	Proposed method of changepoint detection	137
5.5	Simulation	141
5.5.1	Results and discussion	143
5.6	Application to real datasets	153
5.6.1	Finding structural breaks	154
5.6.2	Transfer Entropy	157
5.7	Conclusion	161
6	Conclusion	162
	References	165

LIST OF TABLES

2.1	Some existing methods of changepoint detection in a multivariate series and their application in real data.	21
3.1	Parameters of ARCH and GARCH	48
3.2	<i>R</i> packages: This table lists down the <i>R</i> packages used to conduct simulation.	53
3.3	Basic simulation structure which applies to both bivariate as well as five variate cases. For example, DGP_{0A} indicate positively correlated Gaussian variables with zero changepoints.	54
3.4	Parameters of ARCH and GARCH	55
3.5	Parameters of ARCH and GARCH	57
3.6	Performance of different changepoint detection methods in bivariate simulated data, for $N_1 = 100$	64
3.7	Performance of different changepoint detection methods in bivariate simulated data, for $N_2 = 500$	65
3.8	Performance of different changepoint detection methods in bivariate simulated data, for $N_3 = 1000$	66
3.9	Performance of different changepoint detection methods in five variate simulated data, for $N_1 = 100$	68
3.10	Performance of different changepoint detection methods in five variate simulated data, for $N_2 = 500$	69
3.11	Performance of different changepoint detection methods in five variate simulated data, for $N_3 = 1000$	70
3.12	Performance of different changepoint detection methods in bivariate simulated data (type C and D), for $N_1 = 100$	72
3.13	Performance of different changepoint detection methods in bivariate simulated data (type C and D), for $N_2 = 500$	73
3.14	Performance of different changepoint detection methods in bivariate simulated data (type C and D), for $N_3 = 1000$	75

3.15	Performance of different changepoint detection methods in five variate simulated data (type C and D), for $N_1 = 100$.	77
3.16	Performance of different changepoint detection methods in five variate simulated data (type C and D), for $N_2 = 500$.	78
3.17	Performance of different changepoint detection methods in five variate simulated data (type C and D), for $N_3 = 1000$.	79
4.1	Descriptive statistics and time series properties of the five cryptocurrencies.	86
4.2	Structural breaks obtained using the standard univariate algorithm for the five cryptocurrencies.	96
4.3	Backtesting results from stage 1 for EGARCH models	98
4.4	Backtesting results from stage 1 for GAS models. Here, H_0 is rejected only in a few cases.	100
4.5	Backtesting results from stage 2 for EGARCH and GAS models: In bold are the lowest QPS values for each subsample across EGARCH and GAS, based on which the best model is chosen.	101
4.6	Values of effective group transfer entropy for VaR calculated under different residual distributions in SS2 and SS3. Information flow is towards each named cryptocurrency from the remaining four. Quantile choice considered here is 10%.	103
4.7	Descriptive Statistics of subsamples across cryptocurrencies.	106
4.8	In sample AIC values of EGARCH with different conditional distributions.	107
4.9	BTC GARCH coefficients.	108
4.10	ETH GARCH coefficients.	108
4.11	LTC GARCH coefficients.	109
4.12	XMR GARCH coefficients.	109
4.13	ETC GARCH coefficients.	110
4.14	RMSE values for out of sample forecast of EGARCH models.	111
4.15	In sample AIC values of GAS models with different conditional distributions.	112
4.16	RMSE values for out of sample forecast of GAS models.	113
5.1	Parameters of ARCH and GARCH	135
5.2	Parameters of ARCH and GARCH when conditional distribution is Gaussian.	142
5.3	Parameters of ARCH and GARCH when conditional distribution is student's t .	143

5.4	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data, for $N_1 = 100$	144
5.5	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data, for $N_2 = 500$	145
5.6	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data, for $N_3 = 1000$	146
5.7	Performance of the two proposed changepoint detection methods in a twenty variate simulated data, for $N_1 = 100$, $N_2 = 500$ and $N_3 = 1000$. .	147
5.8	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data (Type C and D), for $N_1 = 100$. .	149
5.9	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data (Type C and D), for $N_2 = 500$. .	150
5.10	Performance of the two proposed changepoint detection methods in both bivariate and five variate simulated data (Type C and D), for $N_3 = 1000$. .	151
5.11	Performance of the two proposed changepoint detection methods in a twenty variate simulated data (Type C and D), for $N_1 = 100$, $N_2 = 500$ and $N_3 = 1000$	152

LIST OF FIGURES

1.1	A multivariate time series representation of stock prices of four Indian banks – HDFC, SBI, ICICI, LIC Housing Finance and Kotak Mahindra bank.	7
3.1	Heatmap of the spectral density based distances between different series generated for 100 bivariate series.	48
3.2	Heatmap of the spectral density based distances between different series generated for 100 five variate series.	49
4.1	Multivariate correlation plot of the time series of the five cryptocurrencies (values are in hundreds).	87
4.2	Flow of the proposed methodology to analyze and forecast VaR dynamics in the cryptocurrency market.	94
4.3	Three subsamples are obtained with two structural breaks: τ_1 is at September 26, 2017 and τ_2 is at June 22, 2020. τ_2 coincides with the change in market conditions during the first wave of COVID-19.	96
4.4	Transfer entropy: The figures represent the values of EGTE and the confidence intervals for the five cryptocurrencies - BTC, ETH, LTC, XMR, ETC. Each row represents JSU and PSIV, respectively and the three columns denote the quantiles - $\{5,95\}$, $\{10,90\}$ and $\{15,85\}$. In every graph, the first set of five vertical lines speaks for SS2 and the second set for SS3.	102
4.5	Pairwise transfer entropy for the choices of JSU and PSIV in the three subsamples.	104
4.6	VaR forecasts for SS2 and SS3 of EGARCH model.	114
4.7	VaR forecasts for SS2 and SS3 of GAS model.	115
5.1	Heatmap of the spectral density based distances between different series generated for 100 bivariate series.	136

5.2	Heatmap of the spectral density based distances between different series generated for 100 five variate series.	137
5.3	A multivariate time series representation of stock indices – S&P 500 and DAX.	154
5.4	A multivariate time series representation of stock indices – S&P 500, Dow Jones and DAX.	155
5.5	A multivariate time series representation of stock indices – S&P 500, DAX and Nikkei.	155
5.6	A multivariate time series representation of 19 American bluechip stock. This excludes the graphical representation of S&P 500.	157
5.7	Group Transfer entropy: The figures represent the values of EGTE and the confidence intervals for the three indices – S&P 500, Dow Jones and DAX. In every plot, the three set of vertical lines represent the different quantiles {5,95}, {10,90} and {15,85}.	158
5.8	Group Transfer entropy: The figures represent the values of EGTE and the confidence intervals for the three indices – S&P 500, DAX and Nikkei. In every plot, the three set of vertical lines represent the different quantiles {5,95}, {10,90} and {15,85}.	159
5.9	Group Transfer entropy: The figures represent the values of EGTE and the confidence intervals for the five banking stocks – HDFCBANK, SBIN, ICICIBANK, LICHSGFIN and KOTAKBANK. In every plot, the three set of vertical lines represent the different quantiles {5,95}, {10,90} and {15,85}.	159
5.10	Group Transfer entropy: The figures represent the values of EGTE and the confidence intervals for 20 US bluechip stocks which also includes the index S&P 500. In every plot, the three set of vertical lines represent the different quantiles {5,95}, {10,90} and {15,85}.	160

ABSTRACT

Our focus in this thesis is to contribute to the extant literature by proposing two methods of structural break detection in a multivariate time series based on two dimension reduction techniques: t-distributed Stochastic Neighbor Embedding (t-SNE) and Independent Component Analysis (ICA). We also apply cumulative sums based Binary Segmentation to detect the location of breakpoints. Both the methods proposed by us prove to be efficient when compared with a few existing techniques of changepoint detection. Performance of each of the methods is compared based on the values of Rand Index (RI), Adjusted Rand Index (ARI), average time take taken by the method (ATT) and the number of changepoints estimated (ACP) by each procedure.

Based on simulation studies, it is observed that the t-SNE based method is better than the several existing ones in terms of accuracy as well as the time taken. It rightly detects a break when there is a change in the generalized autoregressive conditional heteroskedastic (GARCH) structure of the time series. For generalization, we consider both Gaussian and student's t as conditional distributions for GARCH models. A change in the GARCH structure is very apparent in a financial time series and our method successfully detects the break with better precision. We administer the t-SNE based breakpoint location method on a real data set of five cryptocurrencies to fathom its volatility dynamics. First, our method rightly ascertains the location of changepoints which coincides with the cryptocurrency boom of 2017 and the COVID-19 pandemic. This divides the time series into three subsamples (SS1, SS2 and SS3). To add to that, we also model the conditional variances of this dataset using EGARCH models augmented by several heavy

tailed residual distributions. Residual distributions play an important role in modelling the data appropriately for better forecasting. It is observed that for SS2 which is longer in duration and has low volatility, Exponential GARCH (EGARCH) models with Johnson's S_U distribution (JSU) and Pearson Type IV distribution (PSIV) provide a better fit, while for SS3, which is more volatile and shorter in duration, both EGARCH and generalised autoregressive score (GAS) models perform well with skewed student's t -distribution (SST), skewed generalised error distribution (SGED). The estimation of Group Transfer Entropy breaks the myth that Bitcoin generates risk spillovers to other currencies.

Detection of changepoints based on ICA demonstrates competitive and even better performance sometimes when compared with the proposed t-SNE based procedure. The computation time taken by ICA is very less. Though t-SNE based method worked better for smaller samples sometimes, both the proposed methods are competitive. We verify our claims by extensive simulation and application on real datasets. We apply the ICA based technique of changepoint estimation on several multistock portfolio and it rightly detects the position of changepoints which coincides with the recession of 2008. The information flow is also observed to traverse from the west to east.