DECISION MAKING USING

FINANCIAL TIME SERIES DATA

KARTIKAY GUPTA

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY DELHI

April 2021

DECISION MAKING USING

FINANCIAL TIME SERIES DATA

by

KARTIKAY GUPTA

DEPARTMENT OF MATHEMATICS

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

April 2021

Dedicated to my family

Certificate

This is to certify that the thesis entitled ***Decision Making Using Financial Time Series Data*** submitted by ***Mr. Kartikay Gupta*** to the Indian Institute of Technology Delhi, for the award of the Degree of the Doctor of Philosophy, is a record of the original bona fide research work carried out by him under my supervision and guidance. The thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

Place: Delhi

Date: 26 April 2021

Prof. Niladri Chatterjee

Supervisor

Department of Mathematics

Indian Institute of Technology Delhi

Acknowledgements

Words are not sufficient to express my gratitude towards people who helped me during the PhD. I had never thought that it would bring so much positivity and openness in my nature.

This PhD would not have been possible without the guidance and encouragement from Prof Niladri Chatterjee, my PhD supervisor. His comments and ideas played a critical role in the PhD. I learned the skill of academic writing through him. He opened up several opportunities in front of me which provided the much-needed exposure.

I remain grateful to my parents Mrs Geeta and Mr Ganesh Parshad for their love and care.

I thank Dr Deepak Joshi for collaborating with me which was a great learning experience. I thank Prof Aparna Mehra for her valuable tips. I am grateful to Prof B.S. Panda, Course coordinator of the course Data Mining and all the students who took this course. I learned many important things regarding PhD research work while working as a TA in this course.

I thank Prof Parag Singla, from whom I learnt significantly by attending two of his courses, namely Machine Learning and Deep Learning. I thank my Student Review Committee faculty members Prof Dharmaraja, Prof Aparna Mehra and Dr Parag Singla for their valuable comments, suggestions and supportive nature.

I also remain filled with gratitude towards all faculty members, PhD scholars and UG students with whom I interacted during my research work. They made my PhD work to be an enjoyable experience.

Place: NIIT Campus, Neemrana

Date: 26 April 2021 **Kartikay Gupta**

ABSTRACT

Forecasting plays a vital role in the lives of human beings. Time series forecasts are useful for weather forecasting, stocks and derivatives price forecasting, GDP forecasting, electricity load forecasting, among many other things. Good forecasts in any field are immensely valuable. Achieving better forecasts/ predictive-analytics for a stock price-series is the most prominent objective of this work. Specifically, the thesis tries to improvise upon certain different aspects of finance like trading strategies, investment strategies and identifying stock pairs for pairs trading. It works upon techniques of temporal data-mining to propose novel techniques in financial-engineering. Such financial research work helps in extracting money from the market while making the market more efficient and streamlining of the stock prices.

Financial price-series data seldom contains patterns which repeat over time. In this work, first we investigate ways of discovering such patterns (motifs) in financial data and using it for profitable trading. The results indicate the utility of such strategies.

Then, for further improving the forecasts we use fundamental and technical features of companies to predict future prices of their stocks. The technique developed is designed to be scalable to large datasets. The results verify the superiority of the proposed technique over other techniques like simple-regression or random-walk model. The returns generated are much higher than the returns of benchmark indices like NIFTY 50 and BSE Sensex. Here, two large datasets have been used for experimentation.

Our senses can only distinguish between two substances if they are some distance apart in space or time. Measurement of distances, which quantifies farness or nearness between objects, plays an essential role in temporal data-mining or any machine-learning task in general. Thus, the thesis apart from critically reviewing some of the measures, also proposes a few novel distance measures applicable in financial domain.

In this context, we explore clustering for grouping stocks into buckets for quick analysis. Various distance measures have been proposed in the past which can optimally cluster financial time series data. It is generally observed that some stock pairs may not have high Pearson correlation coefficient between them, but they are highly correlated at certain lead-lag. This phenomenon is generally referred to as the lead-lag effect. We propose two new

dissimilarity measures for the above task i.e., Cross-Correlation Type (CCT) and Cross-Correlation Type-II (CCT-II) measure. These measures are able to effectively take into account the time-varying lead-lag relationship between two stocks while measuring their dissimilarity.

We also propose Dynamic Cross-Correlation Type (DCCT) measure which allows the lead-lag relationship between two time series to vary continuously with time. DCCT subtly combines the properties of Dynamic Time Warping (DTW) measure with the CCT measure. Then we showcase the applicability of DCCT measure in identifying pairs for pairs trading. The experiments are conducted to compare the performance of DCCT measure and CCT measure with other popular measures for such task i.e., correlation and SSD measures. The DCCT measure when clubbed with SSD measure i.e., when pairs are selected through optimizing both these measures, then the selected pairs consistently generate the best profit, as compared to all other measures.

Further comparisons show the superiority of the DCCT based lead-lag alignment path over Thermal optimal path (TOP) for determining the lead-lag relationship between the two time series. TOP has been extensively used for analysing lead-lag relationship in the past. This work presents a good alternative for TOP. DCCT measure provides an opportunity to subjectively analyse different price series pairs which have known important lead-lag relationships.

डिसिशन मेकिंग युसिंग फाइनेंसियल टाइम सीरीज डाटा

शोध प्रबंध सार

पूर्वानुमान मनुष्य के जीवन में एक महत्वपूर्ण भूमिका निभाती है। मौसम की भविष्यवाणी, स्टॉक और डेरिवेटिव्स प्राइस फोरकास्टिंग, जीडीपी पूर्वानुमान, बिजली लोड पूर्वानुमान जैसी कई अन्य चीजों के लिए काल श्रेणी का विधिवत अध्ययन उपयोगी होता हैं। किसी भी क्षेत्र में अच्छे पूर्वानुमान अत्यधिक मूल्यवान होते हैं। शेयर की कीमत का पूर्वानुमान और र्विश्लेषिकी प्राप्त करना इस कार्य का सबसे प्रमुख उद्देश्य है। विशेष रूप से यह शोध प्रबंध व्यापार के कुछ पहलुओं जैसे निवेश रणनीतियों और जोड़े व्यापार के लिए स्टॉक जोड़े की पहचान करने की रणनीतियों में सुधार करने की कोशिश करता है। इस तरह के वित्तीय शोध कार्य बाजार से पैसे निकालने में मदद करते हैं और साध ही साध बाजार को अधिक कुशल बनाते हैं तथा स्टॉक की कीमतों को सुव्यवस्थित करते हैं।

वित्तीय मूल्य-श्रृंखला डेटा में कभी कभी ऐसे पैटर्न (रूपांकनों) होते हैं जो समय के साथ दोहराते हैं। इस काम में, पहले हम वित्तीय आंकड़ों में इस तरह के पैटर्न की खोज करने और लाभदायक व्यापार के लिए इसका उपयोग करने के तरीकों की जांच करते हैं। परिणाम ऐसी रणनीतियों की उपयोगिता का संकेत देते हैं।

फिर हम कंपनियों के मौलिक और तकनीकी विशेषताओं को उनके शेयरों की कीमतों की भविष्यवाणी करने के लिए उपयोग करते हैं। प्रस्तावित तकनीक को बड़े डेटासेट के लिए स्केलेबल बनाया गया है। परिणाम सरल-प्रतिगमन या रैंडम-वॉक मॉडल जैसी अन्य तकनीकों पर प्रस्तावित तकनीक की श्रेष्ठता को सत्यापित करते हैं। उत्पन्न रिटर्न निफ्टी और बीएसई सेंसेक्स जैसे बेंचमार्क सूचकांकों के रिटर्न से अधिक है। यहाँ, प्रयोग के लिए दो बड़े डेटासेट का उपयोग किया गया है।

हमारी इंद्रियां दो पदार्थों के बीच अंतर कर सकती हैं यदि वे अंतरिक्ष या समय में कुछ दूरी पर हैं। दूरियों का मापन, जो वस्तुओं के बीच की मात्रा निर्धारित करता है, सामान्य रूप से डेटा-खनन या किसी मशीन-शिक्षण कार्य में एक आवश्यक भूमिका निभाता है। थीसिस कुछ उपायों की गंभीर रूप से समीक्षा करने के अलावा, वित्तीय क्षेत्र में लागू कुछ दूरी उपायों का भी प्रस्ताव करता है।

इस संदर्भ में, हम विश्लेषण के लिए शेयरों को गुट में समूहित करने के लिए क्लस्टरिंग करते हैं। अतीत में विभिन्न दूरी के उपाय प्रस्तावित किए गए हैं जो वित्तीय समय श्रृंखला डेटा को बेहतर रूप से क्लस्टर कर सकते हैं। यह आमतौर पर देखा गया है कि कुछ स्टॉक जोड़े के बीच उच्च पियरसन सहसंबंध गुणांक नहीं होता है, लेकिन कुछ सीसा-अंतराल पर अत्यधिक सहसंबद्ध होता हैं। इस घटना को आम तौर पर लीड-लैग प्रभाव के रूप में जाना जाता है। हम उपरोक्त कार्य के लिए दो नए प्रसार दूरी उपायों का प्रस्ताव करते हैं अर्थात्, क्रॉस-सहसंबंध प्रकार और क्रॉस-सहसंबंध प्रकार-द्वितीय उपाय।

हम डायनेमिक क्रॉस-सहसंबंध प्रकार (डीसीसीटी) उपाय भी प्रस्तावित करते हैं जो दो समय श्रृंखला के बीच लीड-लैग संबंध को समय के साथ लगातार भिन्न करने की अनुमति देता है। फिर हम जोड़े व्यापार के लिए जोड़े की पहचान करने में डीसीसीटी उपाय की प्रयोज्यता का प्रदर्शन करते हैं। इस कार्य के लिए अन्य लोकप्रिय उपायों के साध तुलना करने के लिए प्रयोग किए गय हैं। डीसीसीटी उपाय जब एक अन्य उपाय के साथ जोड़ा जाता है यानी, जब जोड़े को इन दोनों उपायों को अनुकूलित करने के माध्यम से चुना जाता है, तो चयनित जोड़े लगातार अन्य सभी उपायों की तुलना में सबसे अच्छा लाभ उत्पन्न करते हैं।

आगे की तुलना दो समय श्रृंखला के बीच सीसा-अंतराल संबंध का निर्धारण करने के लिए थर्मल इष्टतम पथ (टीओपी) पर डीसीसीटी आधारित लीड-लैग संरेखण पथ की श्रेष्ठता दिखाती है। बीते समय में सीसा-लैग संबंधों के विश्लेषण के लिए टॉप का बड़े पैमाने पर इस्तेमाल किया गया है। प्रस्तावित काम टीओपी के लिए एक अच्छा विकल्प प्रस्तुत करता है। डीसीसीटी उपाय विभिन्न मूल्य श्रृंखला जोड़े, जिनमे महत्वपूर्ण लीड-लैग संबंध है, का विश्लेषण करने का अवसर प्रदान करता है।

कार्तिकेय गुप्ता

2015MAZ8144

Table of Contents

List of Tables

List of Figures

List of Notations

Symbol	Meaning
$\mathbb{R}$	Set of real numbers
$\mathbb{Z}$	Set of integers
$\mathbb{C}$	Set of complex numbers