

## **CMCRC – New analytics improve accuracy of predicting daily stock returns**

26 September, 2013 (Sydney): Capital Markets Cooperative Research Centre (CMCRC), the Australian independent academic centre for capital market research, has found that models that aim to predict daily stock returns perform better when they combine text data (e.g. News announcements) and financial quantitative data (e.g. financial statements).

A study by Zhendong (Tony) Zhao, Nataliya Sokolovska and Professor Mark Johnson looks at combining quantitative and text data rather than treating them separately reducing errors by almost 3% compared to results when only quantitative data is examined.

Professor Mark Johnson said, "An almost 3% improvement doesn't seem large but it has a significant impact in finance because the new method's prediction is closer to the real value." Johnson adds, "This is of particular interest to traders, brokers and investors who want to make a decision on whether to hold, sell or buy after a certain event. For example, Company AAA issues an earnings statement, ASX issues an announcement on the earnings then the algorithm looks for key words in the announcement and compares these to current quantitative data and predicts the return at the end of next day's trading."

Tony Zhao explains, "By analysing the announcement and financial quantitative data the combination of these two different types of data gives the research far more variables to analyse, leading to more accurate predictions."

To examine the performance of these combinations the research uses 19,282 ASX announcements from the first half of 2010. The research uses 80% of the announcements to train the algorithm and 20% to test the different combinations.

The study compares the predictive performance of four different combinations of features including text data and quantitative data with various weighting schemes on the quantitative and text variables using advanced statistical techniques. The best performance was gained by applying different weightings between quantitative and qualitative text data, as this prevented the prediction model from over reacting to minor or random fluctuations in the data

This study is important as very little academic work has looked at combining quantitative and text data to predict daily stock returns. This research uses state of the art data mining techniques of key word weightings within the ASX announcements. Future research will include other advanced data mining techniques developed by CMCRC to analyse text data. Text data are things like company announcements, media news, and social media while financial quantitative data include factors like past daily returns, capital size, volatility, and the stock price.

Zhendong will back test his research results on CMCRC's Alluvial Backtesting Platform, which was launched to its PhD students in early 2013. If this back testing proves to be successful his research

will play a very important role in putting new science behind the methods of predicting daily stock returns, particularly as text based information is growing in quantity and importance.

#### **About CMCRC:**

The *Capital Markets Cooperative Research Centre* (CMCRC) is a world-leading research organisation that provides thought leadership and break-through technology solutions for capital and health markets ([www.cmcrc.com](http://www.cmcrc.com)).

CMCRC is funded equally by the Australian Government, an alliance of University partners, and industry partners including regulators, exchanges and market participants across 10 countries. Research is funded from pooled funding and not sponsored by individuals, companies or institutions. To address its prime goal of industry relevance, industry partners have a major role in the research questions the CMCRC addresses. Academics have primary control over the research design and therefore the results.

#### **About Zhendong (Tony) Zhao**

Zhao is a PhD student at Macquarie University. He has worked on a wide range of topics including embedded systems, digital signal processing, machine learning, large-scale data processing and natural language processing but his main research area is analysing the sentiment of financial announcements and evaluating topical collocation models.

#### **About Dr Matthew Clifton**

Dr Matthew Clifton is an experienced quantitative/algorithmic researcher and developer at CMCRC. He holds a PhD in Finance and a bachelor's degree in robotics engineering. His professional experience includes two years in London working in the area of optimal market design and market surveillance. He works with Alluvial Trading Group ([www.alluvialtrading.com](http://www.alluvialtrading.com)), a proprietary trading organisation specialising in the design, back-testing and implementation of automated trading strategies.

#### ***For more information and interviews:***

Kristin Westlake, The Continuum Partners  
[kwestlake@thecontinuumpartners.com](mailto:kwestlake@thecontinuumpartners.com) +61 416 219 358