ELDER RESEARCH
DATA SCIENCE · AI · MACHINE LEARNING

Technology Background
on Elder Research Long/Short Bond Strategies

INVESTMENT MODELING GROUNDED IN DATA SCIENCE

Elder Research is one of the world's leading Data Science consultancies. It has grown steadily since 1995 to over 60 scientists in four offices in or near Charlottesville VA, Washington DC, Baltimore MD, and Raleigh, NC. The Elder Research team has solved many challenging and previously unsolved technical problems in a wide variety of fields for Government, Commercial and Investment clients, including fraud prevention, insider threat discovery, image recognition, text mining, and oil and gas discovery. But Elder Research got its start with a hedge fund breakthrough (as described briefly in a couple of books[1,2]), and has remained active in that work, continuing to invent the underlying science necessary to address what is likely the hardest problem of all: accurately anticipating the enormous "ensemble model" of the markets.

It is extremely challenging to extract lasting and actionable patterns from extremely volatile and noisy market signals. In theory, timing the market is impossible—and in practice, that is a good first approximation. However, small but significant advances made over the past two decades in several contributing areas – briefly outlined below -- have combined to lead to the recent breakthrough reflected in the three related Long/Short Bond Strategies (Tactical, Opportunistic, and Steady), having high Sharpe ratios but varying degrees of return, risk, and market exposure, as detailed elsewhere.

LUCK, SKILL OR TORTURE? HOW TO TELL

Because of the power of modern analytic techniques, it is often possible to find apparent (but untrue) predictive correlations in the market due to *over-fit*—where the complexity of a model overwhelms the data or, even more dangerously, from *over-search*—where so many possible relationships are examined that one is found to work by chance. Wrestling with this serious problem over many years in many fields of applications, Dr. Elder developed a powerful resampling method, called *Target Shuffling* to measure the probability that an experimental finding could have occurred by chance.[3] It is far more accurate than t-tests and other formulaic statistical methods that don't take into account the vast search performed by modern inductive modeling algorithms. With this tool, Elder Research can much more accurately measure the "edge" (or lack thereof) of a proposed investment strategy.

Further, to more accurately measure the quality of market timing, or style-switching strategies, Dr. Elder created the DAPY (Days Ahead Per Year) criterion. It measures, in days of average-sized returns, the expected excess return for a timing strategy compared to a benchmark similarly exposed to the market. The Sharpe ratio of a strategy can be thought of as measuring the quality of its returns, whereas the DAPY measures the quality of its timing edge. Together, they are

Sharpe alone. Most importantly, ERI studies have shown DAPY to be better than Sharpe at predicting future performance.

SIMULATION, OPTIMIZATION, AND COMPLEXITY REGULARIZATION

Even the most modern data science tools most often attempt to minimize squared error, due to its optimization convenience, when forecasting or classifying. But this metric is not well-suited for obtaining market decisions, as the user's criteria of merit has much more to do with return, drawdown, volatility, exposure, etc., than with strict forecast accuracy. (If one gets the direction right, for instance, it is not bad to be wrong on magnitude, much less its square.) What one needs are optimization metrics that reflect the true criteria of interest, and an algorithm that can find the best values in a noisy, multi-modal, multi-dimensional space.

Knowing of this need from several years of experience (failing) applying analytics to the markets this was the focus of John Elder's PhD research. He created a global optimization algorithm GROPE (Global Rd Optimization when Probes are Expensive) which finds the global optimum value (within bounds) for the parameters of a strategy, using as few probes (experiments) as possible. By that criterion, it was for many years (and may still be) the world champion optimization algorithm. (Note below how it represents a nonlinear 2-dimensional surface as a set of interconnected triangular planes.)
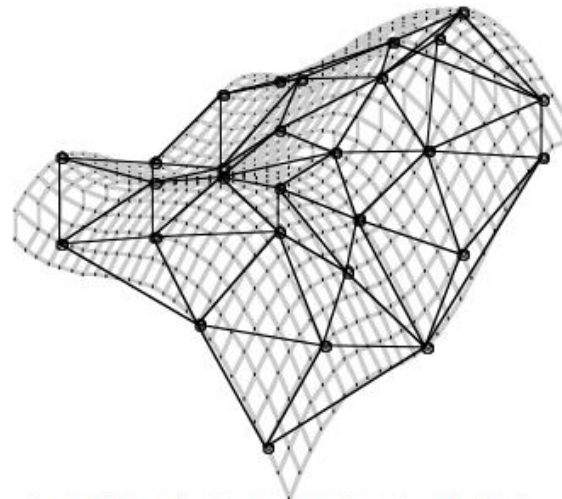

Fig. 5: Triangular Facets Interpolate Function Surface

The global optimization works in a second stage after a smallish set (i.e., dozens) of useful inputs have been identified – in a completely quantitative and not qualitative manner – from thousands of candidate inputs. The winnowing is accomplished in a first stage through regularized model fitting, such as *Lasso Regression*, to filter out useless variables while allowing unexpected combinations to surface.

ENSEMBLE MODELS

Ensemble methods have been called "the most influential development in Data Mining and Machine Learning in the past decade." They combine multiple models into one usually more accurate than the best of its components. Ensembles have provided a critical boost to industrial challenges—from investment timing to drug discovery, and fraud detection to recommendation systems—where predictive accuracy is more vital than model interpretability. Dr. John Elder co-authored a key book on Ensembles in 2010, after being one of the early discoverers and promoters of the idea.[4] The investment system here uses an ensemble of separately-trained models to gain accuracy and robustness.

BIOS

Dr. Elder heads the quantitative modeling work, drawing on researchers at ERI often for assistance, ideas, and brainstorming. He collaborates on the work primarily with long-time colleague Dr. Jordan Barr. Their backgrounds are summarized below.

JOHN ELDER, PHD

22 years ago, John Elder founded Elder Research, America's most experienced analytics consultancy. With offices in Charlottesville VA, Baltimore MD, Raleigh, NC and Washington DC, they've solved hundreds of challenges for commercial and government clients by extracting actionable knowledge from all types of data. Dr. Elder co-authored three books -- on practical data mining, ensembles, and text mining -- two of which won "book of the year" awards. John has created data mining tools, was a discoverer of ensemble methods, chairs international conferences, and is a popular workshop and keynote speaker.

Dr. Elder earned Engineering degrees from Rice and UVA, where he's an Adjunct Professor in Engineering and a visiting executive lecturer in the Darden School of Business. He was named by President Bush to serve 5 years on a panel to guide technology for national security. Lastly, John is grateful to be a follower of Christ and the father of five.

JORDAN BARR, PHD

Data Scientist Dr. Jordan Barr has spent half his career (in two stints) at Elder Research and half as a Physical Scientist and Hydrologist with the Everglades National Park in Florida. Most recently at ERI, he has broken new ground in honing techniques with model ensembles, regularization, and visualization to avoid overfit and over-search. A decade ago at ERI he performed anti-fraud modeling for commercial and governmental clients, used text mining to help the Social Security Administration automatically identify claimants who would be approved for benefits, and was part of a team with Dr. Elder that debuted at 2nd place in the famous Netflix contest. In between for seven years, in the Everglades, Dr. Barr mixed high math and low physical data gathering to be the one of the first to create an accurate model of the carbon budgets and exchanges within the mangrove forest system. Jordan enjoys tennis, the outdoors, and time with his wife and four children.

FOOTNOTES

[1] See Chapter 1 of Dr. Eric Siegel's best-selling book, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*

[2] An excerpt from the book *Journeys to Data Mining; Experiences from 15 Renowned Researchers* briefly recounts the start: "The stock market project turned out, against all predictions of investment theory, to be very successful. We had stumbled across a persistent pricing inefficiency in a corner of the market. A slight pattern emerged from the overwhelming noise which, when followed fearlessly, led to roughly a decade of positive returns that were better than the market and had only two-thirds of its standard deviation —a home run as measured by risk-adjusted return. My slender share of the profits provided enough income to let me launch Elder Research, Inc. (ERI) in 1995 when my Rice fellowship ended, and I returned to Charlottesville for good. Elder Research was one of the first data mining consulting firms…"

[3] See explanation and video at www.elderresearch.com/resource/innovations/target-shuffling-process

[4] G. Seni and J. Elder (2010) *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions* (Synthesis Lectures on Data Mining and Knowledge Discovery), Morgan & Claypool.