ELLIOTT PRECHTER

EWAVES Flash

Published by Qualitative Analytics

June 2015 Issue

LIES, DAMN LIES AND BACKTESTS

Econometricians fit algebraic functions of all possible shapes to essentially the same sets of data without being able to advance, in any perceptible way, a systematic understanding of the structure and the operations of a real economic system. -Wassily Leontief[1]

Once upon a time there was a freshman engineering student named Ben Bitdiddle. He believed computers were almost magical machines, which could—given enough time, data and brute processing power—answer any question, reveal meaning, and forecast the future. He was equal parts intelligent and naive.

Professor Alyssa P. Hacker noticed Ben's enthusiasm and enlisted his help with her research, namely forecasting the behavioral patterns of cats. In other words, Dr. Hacker had tenure.

The first step was to outfit several cats with GPS collars, heart rate monitors, mini cameras and other tracking devices. Bitdiddle then began to compile thousands of data points from each cat, as reported by the devices at 15-second intervals. *What is the cat's temperature and heart rate? Where is the cat located? What is the cat's velocity and acceleration? Is the cat asleep or awake?* And so on.

Bitdiddle established massive databases. He began to fancy himself a "Big Data" resource for cats. He firmly believed he had the time and data and processing power to model cat behavior.

Really, Bitdiddle? You think you can forecast me?

At last the time came to integrate the data into a model. Bitdiddle used his computer to test billions of equations. He sought the combination of equations that would best serve as his model for cat behavior.

An amazingly precise model did emerge. Bitdiddle told Dr. Hacker that most of each cat's behavior could indeed be forecast nearly 15 minutes into the future. *He had the perfect model*.

Dr. Hacker was more experienced and thus less excited. She wanted Ben to learn the hard way.

"Test your model in real-time," she said.

P.O. Box 3102 • Gainesville, GA 30503 USA • www.ewaves.com • info@ewaves.com

So Bitdiddle hooked up the cat monitors and began to gather data in real time. As live data streamed in from the cats, his model generated predictive signals about the behavior of each cat. To Bitdiddle's dismay, his model had no predictive value. The cats did whatever the hell they pleased.

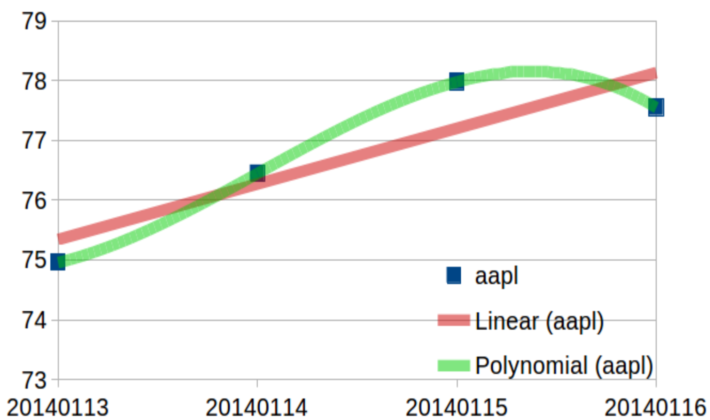
Befuddled, Bitdiddle thought: "I need more data." So he gathered and integrated more data and ran billions more equations to find a new model. But no matter how often he repeated the modeling process, he could not forecast the feline future. What Bitdiddle didn't realize is that he had trapped himself in an endless cycle of curve-fitting.

What is Curve-Fitting?

Ben Bitdiddle did not start by testing a theory or hypothesis. Rather, he forced the data into pre-existing mathematical models—i.e. he was curve-fitting. To understand what this means, let's try a simple example.

The chart to the left shows a few daily close prices for AAPL (Apple, Inc). To curve-fit to AAPL's price movements we need a model. The simplest possible model to depict this data set is a straight line (in red).

But, a straight line clearly does not *perfectly* reflect the data we want to show.

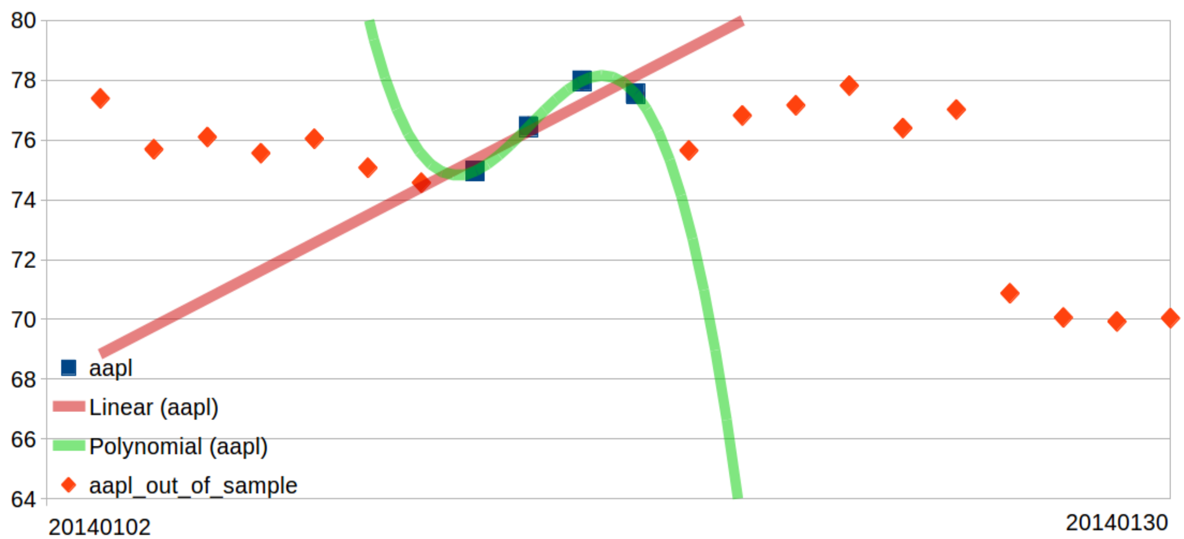


What other model can we use to curve-fit? The next level of complexity would be a polynomial. A polynomial can be forced (curve-fitted) to perfectly reproduce the AAPL data (in green).

If our polynomial fits this set of historical AAPL data, can we use it to forecast prices?

To answer this question we must look away from the data we used to curve-fit the model, which is *in-sample data* (IS). Instead we look at data which was *not* part of the curve-fitting process, which is *out-of-sample data* (OOS).

Consider the chart below. It shows that while our equations *seem* to work IS, they miserably fail OOS. Curve-fitting let us build a model that fit the data set perfectly but which offered *zero* knowledge or predictive value.



These examples are simplified, but the core error of curve-fitting is easy to inflict on more complex models. The most complex models are *Turing-complete*, meaning they can execute arbitrary logic. But a model's theoretical capability is no assurance that curve-fitting can bestow real predictive ability. Let's dig into why.

Why Curve-Fitting Fails

#1: Lack of a theory or hypothesis

The absence of a theory or hypothesis is the main reason curve-fitters fail to find viable solutions. Data is morphed into models which users treat as black boxes. They have no idea *why* the model should or should not function. They simply know that it *appears* to work…in the past.

But would you trust an auto mechanic who said, "I don't know how we fixed your car. We just tried random stuff till it quit making that weird noise." Maybe the car starts and drives down the street just fine. But two hours later you discover that the gas line passes through the cigarette lighter. Good thing you're not a smoker! And don't blame the mechanic: he never claimed to know how it worked.

If this analogy sounds far-fetched, take a closer look at the finance industry today. How many funds have been backed by finance whizzes, amazing marketing, and fantastic short-term performance—only to blow up with devastating losses to shareholders? Many hedge funds, commodity pools, mutual funds and junk bond funds did just this in the 2008 crisis, despite soaring in previous years. Even though they back-tested well and initially performed well, if they reflected no true theoretical understanding of financial markets, they were ultimately doomed.

In contrast, we're developing EWAVES 2 from the ground up with the Wave Principle as its foundation. While the software is proprietary, the Wave Principle itself is public knowledge: It's the difference between a black box vs. a glass box. For our development purposes, the elements of the Wave Principle are *invariant* to the underlying data—i.e., we have a theory that prevents us from curve-fitting.

In other words: We care as much about *how* EWAVES 2 does its analysis as we do about the system's external characteristics. Using the Wave Principle as a "guard rail" is critical, because the system's back tests matter less than the *reality*: Can that back test be maintained into the future? By sticking to the Wave Principle, EWAVES 2 will ensure that theory, testing, and real-time analysis are all congruent.

> "Even when all statistical precautions have been taken, a profitable back test that is not supported by sound theory is an isolated result and possibly still a lucky one. A successful back test always provokes the question: Will the rule continue to work in the future? Theory can be helpful here because a back test that is consistent with sound theory is less likely to be a statistical fluke. When a back test has theoretical foundation, it is no longer an isolated fact, but part of a larger cohesive picture in which theory explains fact and fact confirms theory.*"* -David Aronson[2]

#2: Insufficient OOS testing

Any model can be curve-fit to look good IS. But OOS is the arena that matters. When real money is on the line, everything is OOS.

Popular and academic literature is littered with IS-only financial analysis offered in support of "profitable" strategies, yet which are likely to fail OOS. It is possible for a researcher to properly use OOS data, but not possible to *prove* that the researcher did not cheat. The only reliable OOS method is the "hold out": Publish first, then report performance on live market data after the fact.[3]

The Wave Principle is the ultimate hold-out. Ralph Nelson Elliott published his structural model in 1938, and it remains unchanged to this day. He added a few minor observations in 1946. Therefore, all usage of EWP after that time is OOS. Since EWAVES is an implementation of EWP, it is intrinsically OOS past 1938. However, we still must employ OOS techniques heavily in our research and development to accommodate our unique implementation.

#3: Too many trials

Let's return to our friend Ben Bitdiddle. Suppose that Ben decides one day to be as good a marksman as Robin Hood. So he proceeds to shoot 1,000 arrows every day for three years. At long last he hits a bull's-eye and, on his next shot, splits the arrow in two. He immediately proclaims himself to be as skilled as the hero of Sherwood Forest!

What's wrong with his claim? Well, assuming he clears the target after every second shot, Bitdiddle had roughly *half a million* tries at duplicating Robin Hood's feat. If you make that many attempts—and succeed only once—it's a virtual certainty that this singular outcome is attributable not to skill but to random chance.

Bitdiddle's proclaiming his marksmanship to be on par with Robin Hood's (after one lucky shot) is the same as his claim to have the perfect behavioral model for cats after testing billions of models. Random chance, not underlying truth or accuracy, produced the bogus models that matched IS data but failed on the OOS data.

Professional statisticians understand the risks of using too many trials. Yet research and development teams often perform far more trials than they realize. Even when they do account for the number of trials when testing a given model in isolation, the *real* number of trials is rarely (if ever) accounted for. Ideally, each and every unique test performed by every researcher over the entire timespan of the project should be on the record. But it gets worse. It takes very few trials to accidentally lapse into large degrees of curve-fitting. The underlying math is beyond the scope of this newsletter, but you can read about it here: http://www.ams.org/notices/201405/rnoti-p458.pdf, p. 459.

The too-many-trials problem has huge implications. Plans have been made, funds have been launched, investors' money has been put at risk. I can't tell you how many times I've seen people who imagine they've finally built the perfect model only to have it blow up. Too many trials is one of the quickest ways to move away from true research and into ad-hoc, meaningless data-fitting.

As for EWAVES 1.1, our strategy included only *two* major trials. We didn't test an excessive number of strategy variations because we already knew how real traders use the Wave Principle. Its accuracy as a model is reflected by IS and OOS efficacies. More importantly, the limited number of trials ensured that we were not curve fitting, and greatly reduced the probability that random chance could explain the results.

#4: Too many degrees of freedom

Kim Peek, the autistic savant who inspired the movie "Rain Man," had amazing memory skills. He could recall hundreds of thousands of zip codes, phone numbers, historical events and random facts. Yet a more subtle cognitive test revealed something else. This test included a list of words such as "sugar," "candy" "caramel," and nearly a hundred more related to the theme of "sweet." Asked to recall the words on the list, he did so perfectly. But when asked what those words had in common, Peek drew a blank.[4]

Powerful and unique as Peek's mind was, he was unable to reason *inductively*—namely to use facts and evidence to reach general conclusions about rules, categories, and principles. Without inductive reasoning, only rote memorization remains, preventing the analysis of *previous* outcomes to forecast *future* outcomes.

The Real Rain Man

Mathematical models with too many degrees of freedom (DOF) face the same limitation as does Kim Peek's mind. DOF refers to how malleable a model is: The more DOF, the easier it is to curve-fit the model to the IS data (purposefully or accidentally).

Too many DOF in a model also means curve-fitting is less likely to induce *general* rules with predictive value, and more likely to simply *report* data.

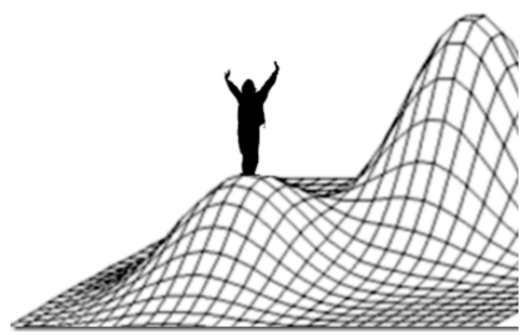
DOF is highly sensitive to OOS testing. If a model has memorized data, its OOS performance will be vastly inferior to its IS. This is why we emphasized OOS testing for our new EWAVES 1.1 strategy, with half of the markets omitted during initial development and testing and added only after the final strategy was selected. The performance was identical both IS and OOS, indicating that our implementation of EWP does not suffer from the too-many-DOF problem. We are taking the same precautions with EWAVES 2.

#5: Stuck in a local maximum

In sports ranging from tennis to baseball to golf, experienced players know that placing the right "spin on the ball" will produce advantageous outcomes. Yet beginners who try to use spin must go through a time when their performance gets worse before it gets better. The mathematical analog of this is called a *local maximum*: a point where short-term degradation is the only path towards long-term improvement.

Curve-fitting techniques have a strong tendency to get "stuck" in local maxima. This means the system's performance is non-optimal but *appears* optimal because any change to the system immediately degrades it. Only a major and radical change can improve it. But curve-fitters have no theoretical understanding of their systems, so they can't know which changes to try.

Fortunately, the "stuck in local maximum" problem is unique to black-box curve-fitting, where test results are the *only* thing available to guide modifications. Yet the Wave Principle, coupled with the Socionomic Hypothesis, is a theory based on the study of markets and human behavior. It may still have properties yet to be discovered, but we are not groping in the dark. We have concrete ideas about where to look that make logical sense within the rest of our ever-growing theoretical framework.



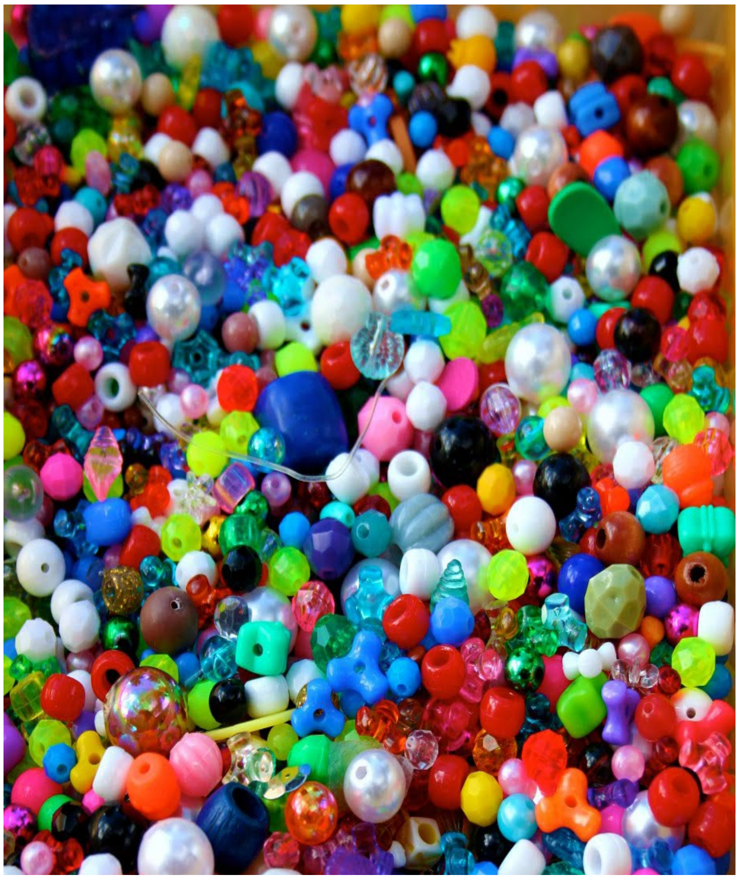
Sometimes you have to go down a hill to get to the highest point.

#6: Assuming Stationarity

In *Evidence Based Technical Analysis*, David Aronson imagines a box filled with different colored beads. The box has a constant ratio of each color, so if you reach into the box and grab a random handful of beads, you can predict with statistical significance what the color distribution will be. The box is *stationary* in the sense that once you know the ratio, the expected color distribution of each sample holds throughout time.

Now imagine that there's a hole in the box where old beads exit, even as new beads regularly pour into the box. The color ratios become an unknown; the box has become *non-stationary*. A statistical analysis of the color ratios at any point in time will not hold into the future.[2]

Financial markets are also non-stationary. So, even if a quantitative model is successful with IS *and* OOS data, there is no certainty that it succeed in the future. Some of the market's statistical properties may hold for indefinite periods, yet those properties can and do change over time or suddenly. Non-stationarity leads directly to model burnout, as discussed in the *first issue of EWAVES Flash*.



Stationary Beads

A great example of non-stationarity in the market is the demise of the Nifty Fifty. The Nifty Fifty were a group of growth stocks that outperformed the broad market averages in the 1968-73 period. Investors who bought these stocks assumed they would get higher returns without higher risk. Some even believed it was possible to generate risk-free returns by going long the Nifty Fifty and shorting the main stock indices! Yet, as financial markets always do, this property of the market suddenly reversed. The Nifty Fifty underperformed the broad averages for many years after 1973, neutering the viability of the strategy and punishing those who believed in its longevity.

Unlike transient artifacts such as the Nifty Fifty, the Elliott wave model is manifest in *all available broad market data*—from millennial waves to individual ticks. The Socionomic Hypothesis posits that these waves reflect fluctuations in social mood, or the unconscious herding impulse produced by primitive parts of the human brain. Unless the human species suddenly and radically evolves, the Wave Principle will continue to work into the future. (For more evidence of Socionomic Theory, please review our extensive and growing research at *www.socionomics.net*)

#7: Ignoring Tail Risk

Tail risk refers to the risk of extreme events. The crashes of 1929 and 1987, for example, exceed the risk boundaries of traditional statistical models. Market return data have *fat tails*, which means extreme events are less rare than they would be under a bell curve, and they can be devastating.

Alas, many popular financial analysis techniques are still based on the assumption that markets do *not* have fat tails. Such assumptions stand in direct contradiction to the available evidence.[5]

Our analysis does not consider the crashes of 1929 and 1987 as outliers—nor do we regard any market action as such. Both events are consistent with the Elliott wave model. Sometimes, as in 1987, we are prepared for an outlier, and sometimes not. But the model is not blind to them and in fact expects them. "Outliers" not only matter—they are arguably the most important market events of all.

Black Monday

In Search of the Idealized Elliott Wave

by Zach Allman

Portions of this article include excerpts and ideas from The Wave Principle of Human Social Behavior (1999) and the Elliott Wave Theorist (2000, 2004).

Ralph Nelson Elliott (1871-1948) discovered the first known fractal model of financial markets in the 1930s. Without computers and with limited data sources, Elliott's painstaking study of historic and real time stock market behavior described the building blocks of price trends, and how they combine to form more complex patterns at larger scales.

In 1995, Elliott Wave International researcher Michael Buettner developed the first computer generated, mathematical expression of the Wave Principle. He applied three basic rules and a handful of easily programmable guidelines, depicted in the diagrams below (first published in the January 2000 *Elliott Wave Theorist*):

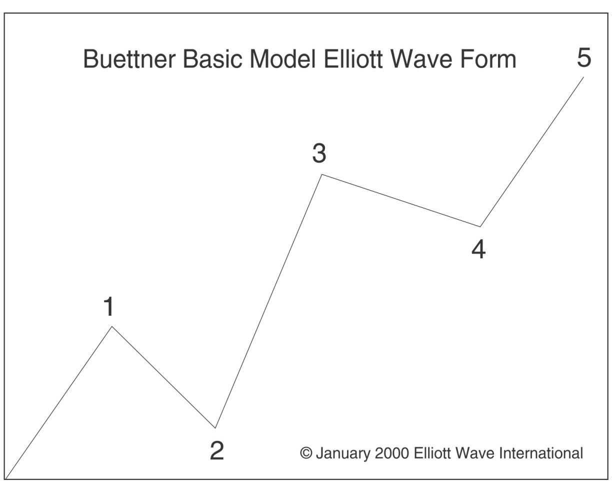


Figure 1

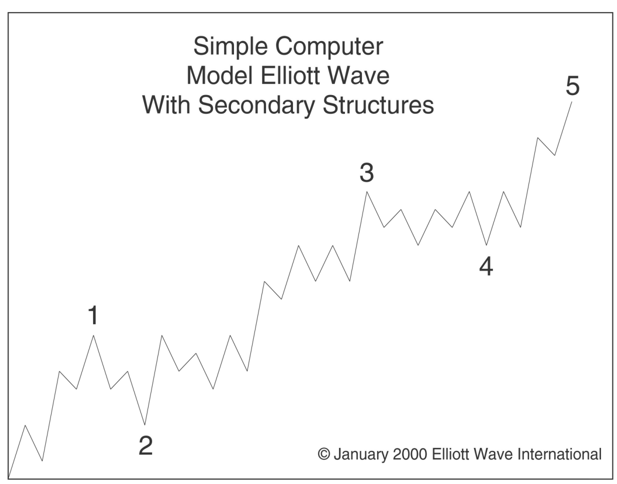


Figure 2

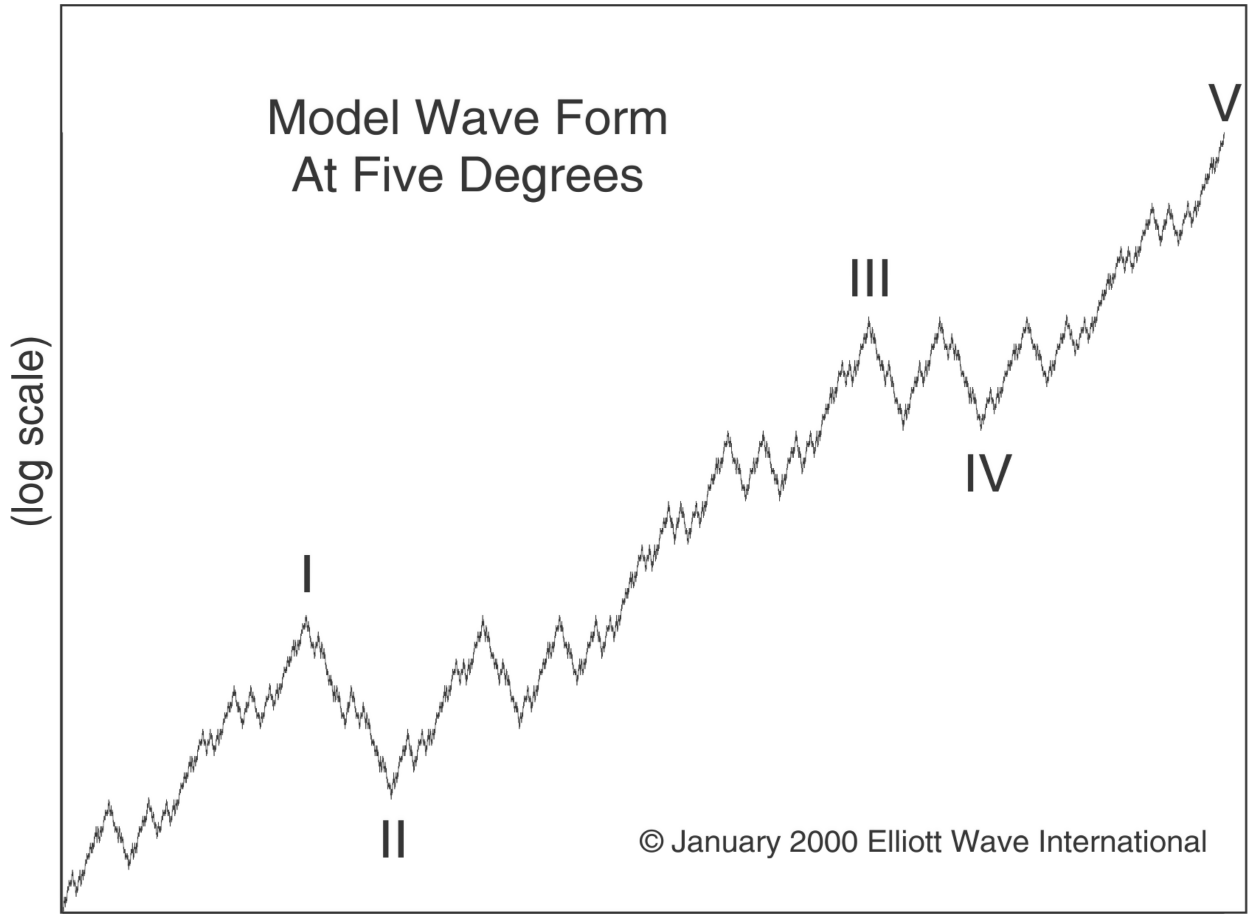


Figure 3

Figure 1 shows a rudimentary wave pattern with only six data points. Figure 2 shows the pattern's more complex subdivisions at the next degree of scale via 34 data points. Figure 3 shows the pattern at five degrees of scale and comprises more than 4,000 data points. Despite a +1,000-fold increase in complexity, the five-wave structure remains visually dominant at all degrees.

This representation of Elliott's model reflects real-world market behavior. As Buettner concluded, "The market model formulated in the 1930s, even when converted into a highly simplified mathematical idealization, reproduces to a remarkable degree the overall structure of the real market."

Buettner's idealization incorporated only the "zigzag" and the "flat" (two of the 13 types of corrective wave patterns Elliott identified) and could produce only identical wave forms and relationships at all degrees of scale, unlike the variable forms of the actual market. In 2009, the EWAVES development team resolved to go beyond these early limitations—thus was born the Idealized Elliott Wave (IEW) generator.

IEWs are designed to show the observable wave forms that R.N. Elliott discerned in actual market data. IEWs incorporate numerous WP guidelines as well as the variability that occurs in real-world markets.

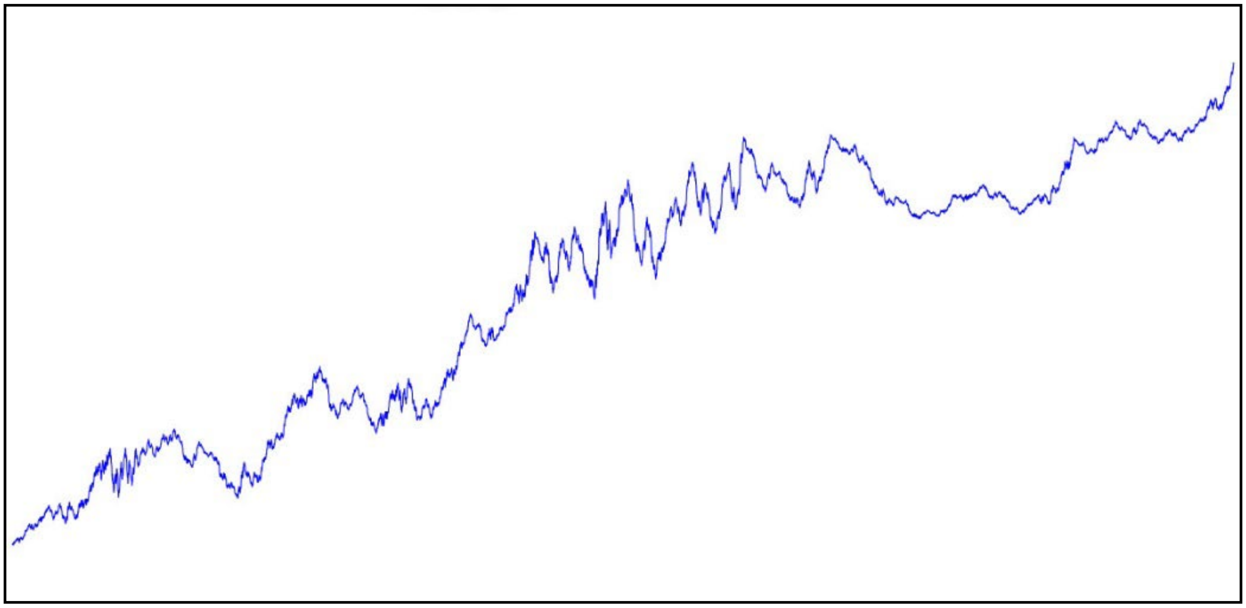


Figure 4

The 10,000 data points in Figure 4 above show one enhanced result.

On the academic front, the rules and guidelines in the IEW design also equip us to find out if a machine-generated data set would naturally include the stylized facts applicable to a general model of the stock market (e.g., volatility clustering, fat tails, skewness, leverage effects). Such a study could formally show that the Elliott wave forms first observed in the 1930s do indeed reflect what researchers have only recently discovered in stock market data.

That aspect of the project is still in the works. More immediately for Qualitative Analytics and EWAVES, we are upgrading the IEW generator. The early version of IEW was largely successful, but faster processors and more efficient programming languages now give the EWAVES team the ability to bring the IEW engine another decade into the future.

Beyond the improved performance of the existing model, the IEW generator now functions so that the unfolding waves have information about the larger pattern they comprise. (For instance, while generating wave 3, the program knows exactly about the structure of waves 1 and 2, and generally about the structure of waves 4 and 5.) We've also set up a series of linear equations—based on WP rules and guidelines—to generate patterns which remain well within the idealized range of wave relationships as described in *Elliott Wave Principle*, pp. 86-91.

In short, this is a very cool and useful testing tool. As the title of this newsletter (*Lies, Damn Lies, and Backtests*) indicates, it is often difficult to avoid curve fitting during the design and testing process. But EWAVES has EWP as a guard against curve fitting, and IEW is how we leverage EWP in our testing. IEW allows us to dynamically generate an infinite set of Elliott wave forms. In turn, we can use these forms to validate EWAVES 2's counting engine—without data-mining.

Elliott Wave International is also using IEWs to generate the charts for its new, fully-automated Certified Elliott Wave Analyst II (CEWA II) exam. If you've ever had to hand-grade labeled charts before, you can appreciate one more reason why this is exciting for us.

FREE TO SUBSCRIBE, FREE TO SHARE

The *EWAVES Flash* publication is open-access. Feel free to share the link ewaves.com/1506EWF for this issue, or visit ewaves.com and click on the "newsletter" tab to read other issues or to sign up for alerts when new issues are published. To subscribe to Flash Services, use the "services" tab or go directly to ewaves.com/flash-services

CITATIONS

[1]Leontief, W. (1982, July 9). Science Magazine, 104-107.

[2]Aronson, D. (2007). Evidence-based technical analysis: Applying the scientific method and statistical inference to trading signals (p. 332). Hoboken, N.J.: John Wiley & Sons.

[3]Bailey, D., Borwein, J., De Prado, M., & Zhu, Q. (2014, May 1). Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. Retrieved May 14, 2015, from http://www.ams.org/notices/201405/rnoti-p458.pdf

[4]The real rain man [Motion picture]. (2006). Five.

[5]Taleb, N. (2014, January 28). What We Learn From Firefighters. Retrieved May 15, 2015, from http://vezetoi-coaching.blog.hu/2014/01/28/mit_tanulunk_a_tuzoltoktol_what_we_learn_from_firefighters_n_n_taleb