The need for consistency and generalizability in tests of the weak-form efficient market hypothesis

Jerry K. Bilbrey, Jr.
Lander University

Michael Shurden
Lander University

Glenn Baigent
Lander University

ABSTRACT

The basis of scholarly research is to provide a new information for building the body of knowledge. However, in the field of financial modeling as it pertains to the efficient market hypothesis, advancement has been somewhat slowed due to a lack of sufficient modeling details. These details would allow for both use and advancement of previously published work. There is a technical aspect to the modeling of financial systems. This includes quantifying and testing various hypotheses often using empirical data. However, the literature has failed to require sufficient details for the replication of the modeling process. Without both the modeling methodology and details of the models used for testing, reproducing and advancing the field of technical analysis financial modeling has become slower than necessary. The goal of scholarly research is to provide a base upon which it can be used and/or added. Without a continually expanding base, the field of scholarly research is less than productive as reproducing methods (and subsequently adding to or improving the methods in some manner) becomes near impossible for most researchers and/or practitioners to achieve. This work demonstrates the need for explicit documentation for technical analysis models in the field of finance. Critical information needed for future modeling work is detailed.

Keywords: modeling, finance, verification, validation, efficient-market hypothesis

INTRODUCTION

Few topics in the social sciences have been studied to the extent of the efficient market hypothesis (EMH), first defined as a set of three testable hypotheses by Fama (1965). This was not the initial recognition of the importance of capital markets in the context of informational efficiency. Following the market crash of 1929, the Securities Exchange Act was passed in 1934 with the purpose of restoring confidence in the functioning of capital markets, and, in particular, with regulations that yielded better information for traders and investors. Part of the SEC's purview was delegated to the private auditing community and led to the establishment of GAAP in 1939.

Since 1934 and 1939 numerous attempts have been made by the SEC, members of Congress, and accountants to create regulations and reporting standards leading to better information and more transparent capital markets. For example, Graham-Leach Bliley and Dodd-Frank were pieces of legislation designed to protect market participants. Many market structures include circuit breakers so that markets in freefall can pause and process the available information. In fact, the general theme of regulations is to create markets that are as closely aligned with the definition of perfect competition as possible. That is, all agents have the same information, there is a homogeneous product, there are no barriers to exit or entry, and all participants are price takers. Financial economic theory has added to this list by suggesting that all individuals are rational. When empirical tests failed to confirm the EMH, researchers naturally questioned the assumption of rationality, thereby spawning research into behavioral economics.

The essence of the efficient market hypothesis is that market prices reflect all available information. In order for this to be true, markets must respond quickly and fully to the release of information (semi-strong form). Moreover, the information must be accurate, complete, and a surprise. If it is to be complete then there is no inside information (strong-form). If the semi-strong form and strong-form hold over time, it follows that the weak-form must also hold. That is, studying past price patterns to determine a pattern is useless because all available information was included and the next movement is related only by the next revelation in price.

Failure to reject the three null hypotheses (weak, semi-strong, and strong forms) is of critical importance to investors and the army of active money managers who claim to have the ability to consistently outperform the market as defined by a broad-based index such as the S&P500. If it is not possible to outperform the market, there is no need to pay a money manager. Ironically, the strongest proponent of efficient markets, Eugene Fama, is affiliated with Dimensional Fund Advisors, a fund with approximately $590B in assets under management.

MODELING TO TEST THE EMH

The era of modeling financial data using computers started in the 1960's. The interested reader can see Samuelson (1965) for further information on the state of finance research that had occurred up to that point in time. One aspect that was quite different then as compared to now is the speed of computers with which researchers have access to do their modeling. Testing a model then comprised of writing a program on "cards" and queueing their programs on a mainframe computer as the resource allowed. Now, virtually all researchers have access to at least one (and sometimes many) desktop computers that are more powerful than their room-sized mainframe counterparts from the 1960s.

There can always be errors in a given model that might never be discovered without ample information given. Herndon et al. (2014) presented an example of a model incorrectly represented. They found work published by Reinhart and Rogoff (2010a and 2010b) that had calculation errors in their model that distorted the final findings. This illustrates how uncaught errors can lead to serious misinterpretations of the model.

The sheer possibilities for accumulating wealth make the field of technical analysis research both fascinating, plausible and intriguing to the public. Standards are needed so that published work is useable. There are necessary and sufficient conditions for this. The work must perform reasonably well for a method to have minimum desirability (necessary condition). However, the work needs to be reproducible for the potential of comparing with other methods or used as a standalone method (sufficient condition). Without apples-to-apples comparisons, a method may not perform well enough for industry. Lastly, an erroneous model needs discovering and corrected in some manner.

Published work needs to be more than interesting, which is a presumptive condition for virtually all financial models. It needs to provide enough details that allows replication by other researchers (sufficient condition) and/or practitioners. Without both necessary and sufficient conditions realized, research output does not produce the kind of effect that is beneficial to the overall body of knowledge. After a sampling of research covering over 25 years of published work, it is clear that others need a standard so that work is actually usable to the fullest. This is the basic purpose of research in the first place.

In the context of financial models that fit into a small enough category, there is the field of technical analysis that enables the weak-form of the EMH to be tested. In fact, most of the models discussed in this paper are built under the assumption that the weak-form of efficient market hypothesis is not correct. This has been tested many times in the literature. For instance, Dow and Gorton (1993) investigated market response to new information. Similarly, in another paper, they (1994) investigated the effects of arbitrage chains based on private information and the effects of the markets as the information become reflected into the price of an asset. Balvers et al. (1990) used the efficient market hypothesis as a framework for predicting stock returns. For the interested reader, Dimson and Mussavian (1990) and Fama (1965) give discussions of the efficient market hypothesis and its implications.

In the end, bullish and bearish markets cannot be predicted with absolute certainty. However, any method should be directly comparable with another method for a given period. These types of comparisons are critical to the advancement of the literature. Whether it is risk, liquidity or any other parameter being investigated, a given method should produce results. Many scholars argue that poor results are important for direction so that further work in a given direction will not be pursued. However, industry is not so inclined to put value on work that does not produce positive results.

PURPOSE OF THIS WORK

Research leads to both applications and further research. A researcher and/or practitioner needs the ability to confirm that he/she is on the correct path. Building financial models is no different from other types of models. There is always the possibility that a model is incorrect. Here is an example of a simple statistical model because of its commonality and demonstration of applicability toward quantitative modeling. In virtually every business statistics textbook, there are examples for building confidence intervals based on small samples following Formula (1).

Formula (1): $\bar{x} \pm t(\frac{s}{\sqrt{n}})$

Where:
$\bar{x}$ is the sample mean
t is the critical value based on the t-distribution
s is the sample standard deviation
n is the given sample size

In the world of academics, it would never be acceptable to give such an example for creating a confidence interval without giving enough information for the recreation of the test by anyone wishing to use the function. Raw and/or summarized data (or both) are needed to plug into the above formula along with the output. However, without these, it becomes difficult at best for someone trying to determine whether they have been able to recreate the confidence interval correctly. Without such sufficient verification, someone could attempt to use the model even though it was not accurately implemented. It is exactly the same issue with more complex models.

For example, the price path for a security is frequently modeled as a Wiener process.

Formula (2): $W = W_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$

Where, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and are serially independent (ε_t are $i.i.d.$).

Given these assumptions, the expected value is W_0 and the standard deviation is $\sigma_\varepsilon\sqrt{n}$. Under this construct, the volatility is a function of time. The impact of time is demonstrated in DeBondt and Thaler (1985, 1987).

Still in the context of "time", a simple example makes the point. Suppose that security prices at times 0, 1, and 2 are \$20, \$21, and \$20. The return in period 1 is +5% and in period 2 is -4.76%. These returns are negatively related but if measured from period 0 to 2, there is no return. This affects the correlation and the volatility that is assumed to be constant (homoscedastic) over time. If a modeler somehow implements a published method but achieves inaccurate results, the error could easily be overlooked and the corresponding model could be misleading.

The notions of correctness and reproducibility of proposed models are the underlying issues here. In computer based financial modeling there are two concepts called verification and validation. Verification is the process of determining that a model performs as it is designed/specified to perform. For instance, if a model is set up to take an input and give a specified output, the model is verified if it produces the specified output given that input.

Seldom, however, is verification of a model so simplistic. For validation, the model is compared against the real-world system to show that the model closely represents the actual system. Robinson (1997) gives a good overview of computer model verification and validation. For a model to be reproducible and usable, it must be both verified and validated.

Fagiolo et al. (2005) and Dawid and Fagiolo (2007) identify validation as one of the major issues facing modeling approaches. However, validation can only begin after verification is complete. If a model is specified and/or built incorrectly, it is still possible for a model to be accepted as valid. However, it may or may not be an accurate representation of a system. For instance, in a queueing system, output (in units) can match the real world while queuing times may be far from real-world values. As such, output from a model like this would not render the appropriate "behavior" and consequent testing could yield results that are not reasonable. A contrived example can illustrate this behavior. Take a typical single teller queue in a bank with a service time of .99 minutes/customer and an arrival of 1.00 customer/minute. If a model were specified to perform on these averages, there would never be anyone in the queue. However, this is far from reality as the average length of the queue would be around 8 people for a given workday. Therefore, even though the output matches the real-world parameters, the queue length and associated waiting times in the model would be vastly different from real-world parameters.

Another issue in validation is accuracy. Robinson (1997) points out that accuracy is an issue of confidence and that models cannot completely be accurate. As a point of variation, it clearly illustrates the need to start at the same point of work. Subsequent researchers and/or practitioners cannot do this until a model's output is confirmed.

It is a reasonable assumption that researchers have made every effort to correctly build models that test their various theories. However, errors can be introduced by the most thorough and skilled researchers. Robinson (1999) discusses three common sources of variability (model, data and experimentation). Although he shows the concepts in a queuing system modeling scenario, the issues are the same with financial models. Herndon et al. (2014) gave one such documented case of a financial modeling error. Without proper documentation, it is impossible to find errors like this. Worse yet, an academic or industry person trying to replicate the work might become overly frustrated or even create a solution that has serious errors.

Without enough information to replicate a given method, it cannot be compared to other methods in an apples-to-apples comparison. For instance, virtually all research uses different data sets or different time ranges for input data. It is critical to be able to compare methods using the same data along with the same parameters chosen for the reported research. Furthermore, it is always possible for mistakes to be made during the modeling process that would otherwise never be discovered if a method were not properly reported.

Similar to the issues mentioned in this paper, Fagilio et al. (2007) examined empirical validation for agent-based economics models for the purpose of future modeling approaches. They identified a set of issues that are common to virtually all modelers engaged in empirical validation. They proposed a taxonomy that captures the relevant dimensions along which agent-based economics models differ. They mention four common issues associated with empirical validation. However, they failed to recognize the importance of the verification process. If a model is not built correctly, the validation process is flawed at best. They did point out the need for standards by saying that heterogeneity in empirical modeling techniques may be due to a lack of standards for constructing and analyzing agent-based models. Having a set of economics modeling standards is overdue.

Another aspect of comparisons is that of bull and bear markets. Some algorithms may work well on one type of data or another but may produce less than desirable results on a market other than which it was designed to operate. Also, an algorithm may perform less than desired on a given data set that no one would be interested in utilizing the algorithm for fear that the market could change – as it clearly does from time to time. When the world markets plummeted in 2007-09, many algorithms would have been worthless during that extended period. Similarly, an algorithm that performs well during bear markets might not perform well during bull markets. Most algorithms are designed for up trending markets as those occur (almost exclusively) over decades of time and therefore, are of interest to most people.

STUDY OF THE LITERATURE

An exhaustive survey of this vast ocean of research is beyond the scope of this paper. However, a look at representative portion of papers illustrates the need for standards. However, for the introduction of generic standards, necessary parameters need to be determined. Certainly, it is not reasonable to prescribe a given set of data or timeline and behaviors are almost certainly going to change over time. It is possible to make many different assumptions and factors investigated accordingly. However, it should be possible to compare quantitative methods over a static period that can be replicated.

Schwert (1990) looked at stock returns over an entire century. One of the tests performed is to use “fresh” data to test stability of relations estimated earlier by Fama (1990). Another test is performed to compare the Miron-Romer (1989) index of industrial production with the Babson index of physical volume of business activity from Moore (1961). The data used are the nominal stock returns from the Center for Research in Security Prices (CRSP) and from the composite S&P portfolio over the time frame from 1889-1988. Without a published data set and with various changes in the data (dividends, stock splits, splicing of data, etc.), it would be virtually impossible to replicate the models given. However, Schwert does give ample outputs to understand the results obtained.

Dow and Gorton (1993) presented a model that studies two agents trading strategically and learning over time from historical prices. Their model was built to capture complex relationships that were difficult to otherwise interpret. They describe their trading strategy in detail. An example illustration is discussed but no information is given about inputs or the outputs from the model.

Gençay (1998) looked at the optimization of technical trading strategies and their profitability. This is done by measuring the profitability of technical trading rules based on nonparametric models. He used 25 years of data from the DJIA. The data was split into five subsamples to study the sensitivity of the results to sample variation. However, the DJIA changed components five times during that period. It is unclear how these changes were handled and would be virtually impossible to replicate.

Neural networks have perhaps received the most attention for estimation and classification techniques over the last twenty years. Jasic and Wood (2004) also present a case study where they test neural network predictions on the stock market. Yet Olson and Mossman (2003) present another AI example. They utilize neural networks for making predictions on the Canadien stock market. Thawornwong et al. (2003) looked at using neural networks for decision rules and their potential for predicting stock prices. There was so much activity using AI techniques for technical trading that Vellido et al. (1999) created a survey of the techniques and results. All of this happened with little are no mention of verification or giving practitioners information needed to use their methods.

Leung et al. (2000) also looked at forecasting stock indices by comparing classification and level estimation methods. They tested four classification models: linear discriminant analysis, logit, profit, and probabilistic neural networks. They compared these classification models with four level estimation models: exponential smoothing, multivariate transfer function, vector autoregression with Kalman filter, and multilayered feedforward neural network. They developed trading rules to estimate profit generated by the various forecasts. Furthermore, the classification models drive the trading rules. They mentioned the constrained set of input variables used in their testing as well as the output variables. Although they do give indirect model outputs (percentage of correct hits), it would be difficult, at best, to replicate these models.

Cichocki et al. (2005) studied variable selection in predicting a stock index. They used S&P 500 data, and their model was a neural network. Independent Component Analysis (ICA) was used to reduce the number of potential variables. They used a prepackaged Neural Network for Excel called "Predict." Even though their methods were reasonably well detailed, the paper did not provide enough information for another researcher to replicate their work.

Durlauf (2005) examines complex system modeling techniques and concludes that these methods (at that time) fail at representing the substantive properties of complex systems. He illustrates that financial models need a tight connection between empirical attempts and theoretical work. One of the issues brought up by Durlauf is later enhanced by Barde (2016). Barde conducted a full-scale comparison test for three different models to provide insight into how herding mechanisms explain the features of financial data. He compared model performance across a wide set of 24 empirical data series. The data used for the model comparisons come from 18 stock market indices and 6 exchange rate series covering the major time zones of Asia, Europe and the Americas. The extremely complex method described inevitably discards information and causes results to be randomly "close" but not exact. As such, replicating these results are not possible with certainty.

Besides Barde's attempt, many agent-based models have been presented to solve various types of problems. Dawid and Fagiolo (2008) contrast agent-based computational economic (ACE) approaches. They looked at creating models that enable better understanding of the characteristics of economic systems. They point out that large-scale simulation is becoming an attractive approach for modeling decentralized interactions of heterogeneous economic agents in systems such as markets. Kukacka and Barunik (2017) give one such approach. This approach uses logit regression to create early financial warnings for financial crises. They developed a model for testing the properties and behavior of an estimation framework while also gaging the model's ability to recover parameters consistently and efficiently. They do mention a published data set and give a mathematical representation of the problem as well as give parameters tested within their simulation model. This type of reporting is somewhat rare in the literature as enough information is given that a researcher could replicate the results.

De Socio and Michelangeli (2017) give another risk-based approach. It assesses financial vulnerability for Italian firms over a two-year period. Both micro and macro-economic data are used input that takes into account demographic information and heterogeneity of firms. They did mention a 2-year period with baseline and stressed scenarios while mentioning a data set provided by Cerved. However, they complement this data set with macroeconomic forecasts, and estimated evolution of profitability, interest rates and debt levels, and do not show these complementary values. They did provide output results that would be comparable across models.

Li and Wong (2014) present a logit model for predicting early warning signals for a set of financial indicators. They used empirical data to show that their model gives a better predictability than previously published models. Their empirical data came from Chinese Company datasets. Although their modeling technique is described in detail and the outputs (in percentages) are given, without the ability to retrieve the input data, it is impossible for another researcher to verify the model.

Enke and Thawornwong (2005) give a neural network example where they attempt predicting stock market returns. A level change was used to capture monthly stock returns because these changes represent information that a neural network can learn. Data were used from 1976-1999 for 286 periods. They did an excellent job of listing each of the variables used along with definitions for them. In addition, they clearly define the data used for training and validation of the model. The performance measures are given and used that allows for comparisons of different configurations of their models. This approach is almost complete, except they do not give any sample output that would allow another practitioner to build and verify their model. Even though the authors are quite concerned with the validation of the model, there is no mention of verification. This tends to be the case across many modeling papers as the research has focused on ways of better modeling the real world. However, a model that does not give accurate results or is otherwise unreproducible is not beneficial to body of literature at all.

Kumar and Ravi (2007) compare over 100 published techniques for predicting bankruptcy in the financial industry. For neural network approaches, the authors mention the following variables: number of input variables, learning sample size, testing sample size, and the software used for each technique. They believe that Neural Network techniques are the most common form of model used for bankruptcy prediction. However, there is no mention of a common data set that would allow apples-to-apples comparisons of the models. Certainly, they mention some of the variables, but it is unclear, at best, that the models could be accurately replicated. Similarly, Gerke et al. (2013) compared multiple macroeconomic models that incorporate various financial frictions. The authors report that the tested models share qualitatively similar and interpretable features. However, even though they have “confidence” of common understanding of the mechanisms involved, they do not mention comparing the models with common data sets and common outputs. These modeling attempts have attempted to reproduce various models when there was not enough (documented) information to do so in the original work. It is certainly plausible that these authors contacted the original authors for guidance and/or model parameters. However, none of these papers explicitly provides sufficient details to begin the process of verification.

Similar to the issues mentioned in this paper, Fagilio et al. (2007) examined empirical validation for agent-based economics models. A set of issues is identified that are common to engagement in empirical validation. They propose a taxonomy that captures the relevant dimensions along which agent-based economics models differ. Four common issues are mentioned that are associated with empirical validation. However, they fail to recognize the importance of the verification process.

Some approaches compare different models to determine best performances. Ballings et al. (2015) present one such approach. The authors benchmark three ensemble methods (Random Forest, AdaBoost, and Kernel Factory) against four single classifier models (Neural Networks, Logistic Regression, Support Vector Machines, and K-Nearest Neighbor). They used data from 5767 publicly listed European companies and used a common measure of the area under the receiver operating characteristic curve as the performance measure to predict the stock price direction. They did an admirable job with their models. However, they give a practitioner no data to verify that they have successfully implemented the principles discussed.

Perez et al. (2015) present a similar approach for utilizing classifiers for binary financial data. They present a methodology that identifies data characteristics that allow for estimates of factor parameters. With simulation, they conclude that the method is appropriate for panel data sets that are often seen in finance. Their method is designed to overcome common problems seen with discrete factor models in data-rich environments common in finance. To evaluate their method, they developed a random Monte Carlo simulation under various scenarios. Using relatively large sample sizes, their results could reasonably be obtained by someone wanting to replicate their methods. The authors did give an excellent illustration in an example of a firms' decision to split their shares. However, the data used in their example uses data over a 44 year period with approximately 5000 companies in the data. It would be trying, at best, to replicate their example model. Researchers often fail to report enough information for the purpose of replication and/or verification. This paper is an excellent example of failing to think about others using their modeling techniques.

Zhang et al. (2018) presented a data-driven price and trend prediction system using publicly available data such as price and technical indices. They predict both the category of direction (up, down, flat, and unknown) as well as the interval of growth (or decline). The authors applied the technique to transaction data taken from the Shenzhen Growth Enterprise Market. This approach is unique because they include pseudocode to illustrate the logic behind some of their methods. This is a much finer granularity of modeling than most researchers choose to present. They used 495 stocks from the Chinese market across a seven-year time span. However, there is no instance data and replicating a model this correctly, would be difficult at best.

Lastly, Henrique et al. (2018) utilized a machine-learning algorithm called the support vector machine (SVM). The authors compared this algorithm against a random-walk algorithm to look for abnormal gains on two years of Brazilian, Chinese and American small and large capitalization companies using 1-minute historical stock prices. They used a well-organized and relatively straightforward set of technical indicators such as simple moving average, weighted-moving average, and relative strength index. Similar to Enke and Thawornwong (2005) and Gençay (1998), they presented the measured root mean squared error for several configurations of a SVM. They presented table after table of errors for their various "optimal" parameters. In the end, with a small example showing the stock, starting and closing values for the stock, the results in the table could easily be followed to determine that a model had been successfully replicated. On the other hand, the results are quite thick and would be virtually impossible to determine if they were originally implemented correctly.

SUMMARY

There has been evidence provided of much interest in the field of predicting financial markets. However, there is one common denominator in the field. This is a lack of consistent reporting which would enable verification of the methods. Although the problem can be defined as an NP-Complete problem due to its infinite number of possible solutions, a standard is needed for the advancement of the field as a whole. Without consistent and methodological approaches, advancement in the field will be less efficient, at best.

Although the vast majority of research is likely modeled and represented correctly, there always exists the possibility for modeling errors. Assuming that all models are correct is not prudent to the field of research nor industry. Mistakes do happen and perhaps affect the outcomes as was illustrated by Herndon et al. (2014). Many papers are published without enough information to determine if the model was, in fact, represented correctly.

In the end, modeling techniques could be advanced both easier and with certainty if efforts were made to facilitate replication. It is no different from the one-sample t-test that is presented in virtually every introductory statistics textbook. Without the input data and known solution to the problem, it would be quite easy for even experienced modelers to make a mistake and never even realize it. Practitioners and/or researchers should not have to completely redevelop the methods presented in the literature while guessing/assuming that their replication is accurate. In fact, these methods should be verifiable in a minimal amount of time to allow a focus on improving or implementing the work instead of worrying with the correctness of the model.

With a given example that is verifiable, a practitioner/researcher could ensure that their model is built correctly even though perhaps the mentioned data set might have some issues. For instance, stock splits and adjustments happen in data sets from time to time. With beginning and ending data points, as well as model parameters, it would not be necessary to recreate each scenario. If a model is verified on one set of parameters and produces the same results on a separate model, it is extremely likely that the model is correct.

A category of financial models needs standardization to improve the overall quality of future work. The category is that of technical analysis modeling. These models are often predicated on the assumption of the weak-form EMH. This states that all public information is priced into an asset and no abnormal gains can be made from this information. However, there has been a plethora of work in this specific area. As these models are of high interest as well as quantitative, it is important for authors to present both necessary and sufficient information to make the work as useful as possible. For instance, it is necessary to describe a model and what the model is trying to accomplish. Without this information, there would be no hope of replicating the work. However, this does not meet the criteria of sufficiency. A model must be useful to others for the condition of sufficiency. Not only that, this is unfair to practitioners who might not have the contacts, time or other means of obtaining the needed information to replicate and/or expand on the modeling techniques.

With the sample of reviewed papers, there is a common theme. More effort is given toward presenting results rather than giving sufficient information to make use of the presented models. However, this is not consistent with the purpose of creating work that is usable by others. It is likely a combination of excitement for their method and oversight about what is needed for replication and usability. With minor changes in the process of thinking about research, advances will come in a timely fashion while producing work that is beneficial to researchers and practitioners alike.

Modeling work needs two items for future technical analysis modeling work. First, there is a need for an explicit attempt to describe a modeling attempt for replication purposes. With this mindset, future work can be successful (as intended) as useful to both practitioners and researchers. Second, there is a need for the following instance data and associated information for technical analysis modeling approaches that are attempting to disprove part of all of the EMH. These parameters are as follows:

- description of instance data
- detailed location of the instance data
- numerical (the actual calculation) output achieved using the instance data

With instance data, it is easy to verify whether the same results are achieved or not. For instance, if a simplistic buy and hold for two days strategy were applied to Yahoo's common stock (ticker symbol: YHOO) with a beginning price and date of $36.49 on 06/25/18 and a selling price at the close of $34.47 on 06/27/18 a loss of $2.02 per share would be the result. Even with a complex algorithm, with the instance data provided it would be straightforward for a practitioner to know if the method their model was correct as proposed in the research. As mentioned earlier, verification only requires one data point of instance data given it is a representative calculation generated by a given algorithm. Providing instance data is not intended as a constraint for modelers. It is actually quite the opposite. By providing sufficient information about instance data along with output, it should allow both researchers and practitioners to replicate work for use in industry or as a baseline for further research. After all, that is the basic purpose of research.

References

Barde, Silvain, (2016), "Direct Comparison of Agent-Based Models of Herding in Financial Markets," *Journal of Economic Dynamics and Control*, 73, pp. 329-353.

Ballings, Michel, Dirk Van den Poel, Nathalie Hespeels, and Ruben Gryp, (2015), "Evaluating Multiple Classifiers for Stock Price Direction," *Expert Systems with Applications*, 42(20), pp. 7046-7056, November.

Balvers, R., T. Cosimano and B. McDonald, (1990), "Predicting Stock Returns in an Efficient Market," *Journal of Finance*, 55(4), pp. 1177-1189.

Cichocki, A., S. Stansell, Z. Leonowicz, and J. Buck, (2005), "Independent Variable Selection: Application of Independent Component Analysis to Forecasting a Stock Index," *Journal of Asset Management*, 6(4), pp. 248-259.

Dawid, Herbert and Georgio Fagilio, (2008), "Agent-based models for Economic Policy Design: Introduction to the Special Issue," *Journal of Economic Behavior and Organization*, 67, pp. 351-354.

De Bondt, W. and R. Thaler, (1985), "Does the Stock Market Overact?" *Journal of Finance*, 40, pp. 793-808.

De Bondt, W. and R. Thaler, (1987), "Further evidence on investor overreaction and stock market seasonality," *Journal of Finance*, 42, pp. 557-581.

De Socio, Antonio, and Valentina Michelangeli, (2017), "A Model to Assess the Financial Vulnerability of Italian Firms," *Journal of Policy Modeling*, 39, pp. 147-168.

Dimson, Elroy and Massoud Mussavian, (1998), "A Brief History of Market Efficiency," *European Financial Management*, 4(1), pp. 91-193.

Dow, James and Gary Gorton, (1993), "Trading, communication and the response of price to new information," *The Economic Journal*, 103, pp. 639-646.

Dow, James and Gary Gorton, (1994), "Arbitrage chains," *Journal of Finance*, 49, pp. 819-849.

Durlauf, Steven N., (2005), "Complexity and Empirical Economics," *The Economic Journal*, 115(504), pp. F225-F243, June.

Enke, D. and S. Thawornwong (2005). "The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns." *Expert Systems with Applications*, 29(4), pp. 927-940.

Fagilio, Georgio, Paul Windrum, and Allesio Moneta, (2007), "A Critical Guide to Empirical Validation of Agent-Based Economics Models: Methodologies, Procedures and Open Problems," *Computational Economics*, 30, pp. 195-226.

Fama, Eugene, (1965), "The behavior of stock market prices," *Journal of Business*, 38, pp. 34-105.

Fama, Eugene, (1990), "Stock Returns, Expected Returns and Real-Activity," *Journal of Finance*, 45(4), pp. 1089-1108.

Gençay, R. (1998). "Optimization of Technical Trading Strategies and the Profitability in Security Markets." *Economics Letters*, 59(2), pp. 249-254.

Gerke, R., M. Jonsson, M. Kolasa, P. Lafourcade, A. Locarno, K. Makarski, and P. McAdam, (2013), "Assessing Macro-Financial Linkages: A Model Comparison Exercise," *Economic Modelling*, 31, pp. 253-264.

Henrique, Bruno Miranda, Vinicius Amorim Sobreiro and Herbert Kimura, (2018), "Stock Price Prediction Using Support Vector Regression on Daily and Up to the Minute Prices," *The Journal of Finance and Data Science,* (accepted for print).

Herndon, Thomas, Michael Ash, and Robert Pollin, (2014), "Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff, *Cambridge Journal of Economics,* 38(2), pp. 257-279, March.

Jasic, T. and D. Wood, (2004), "The Profitability of Daily Stock Market Indices Trades Based on Neural Network Predictions: Case Study for the S&P 500, the DAX, the TOPIX, and the FTSE in the Period 1965-1999," *Applied Financial Economics*, 14, pp. 285-297.

Kukacka, Jiri, and Jozef Barunik, (2017), "Estimation of Financial Agent-Based Models with Simulated Maximum Likelihood," *Journal of Economic Dynamics and Control*, 85, pp. 21-45.

Kumar, P. Ravi and V. Ravi, (2007), "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques," *European Journal of Operational Research*, 180, pp. 1-18.

Leung, M., H. Daouk and A. Chen, (2000), "Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models," *International Journal of Forecasting*, 16 (2), pp. 173-190.

Li, Shuangjie and Shao Wang, (2014), "A Financial Early Warning Logit Model and Its Efficiency Verification Approach," *Knowledge-Based Systems*, 70, pp. 78-87.

Miron, Jeffrey, and Christina Romer, (1989), "A New Monthly Index of Industrial Production, 1884-1940," *Journal of Economic History*, 50(2), pp. 321-337.

Olson, D. and C. Mossman, (2003), "Neural Network Forecasts of Canadian Stock Returns using Accounting Ratios," *International Journal of Forecasting*, 19(3), pp. 453-467.

Perez, M. Fabricio, Andriy Shkilko and Konstantin Sokolov, (2015), “Factor Models for Binary Financial Data,” *Journal of Banking & Finance*, 61, pp. S177-S188.

Reinhart, C. M. and K. S. Rogo, (2010a), “Growth in a Time of Debt,” *American Economic Review: Papers & Proceedings*, 100, pp. 573-578.

Reinhart, C. M. and K.S. Rogo, (2010b), “Growth in a Time of Debt,” Working Paper 15639, National Bureau of Economic Research, http://www.nber.org/papers/w15639 [accessed 05/09/2018].

Robinson, Stewart, (1997), “Simulation Model Verification and Validation: Increasing the User’s Confidence,” *Proceedings of the Winter Simulation Conference*, pp. 53-59.

Robinson, Stewart, (1999), “Three Sources of Simulation Inaccuracy (and How to Overcome Them),” *Proceedings of the Winter Simulation Conference*, pp. 1701-1708.

Samuelson, Paul, (1965), “Proof That Properly Anticipated Prices Fluctuate Randomly,” *Industrial Management Review*, 6, pp. 41-49.

Schwert, W., (1990), “Stock Returns and Real Activity: A Century of Evidence,” *Journal of Finance*, 45, pp. 1237-1257.

Thawornwong, S., D. Enke, and C. Dagli, (2003), “Neural Networks as a Decision Maker for Stock Trading: A Technical Analysis Approach,” *International Journal of Smart Engineering System Design*, 5(4), pp. 313-325.

Vellido, A., P. Lisboa and J. Vaughan, (1999), “Neural Networks in Business: A Survey of Application (1992-1998),” *Expert Systems with Applications*, 17(1), pp. 51-70.

Zhang, Jing, Shicheng Cui, Yan Xu, Qianmu Li and Tao Li, (2018), “A Novel Data-Driven Stock Price Prediction System,” *Expert Systems with Applications*, 97, pp. 60-69.