

A Model Must Be Wrong to be Useful: The Role of Linear Modeling and False Assumptions in Theoretical Explanation

Chong Ho Yu*

Applied Learning Technologies Institute, Arizona State University, PO Box 612, Tempe AZ 85280, USA

Abstract: It is true that many times relationships in the real world do not fall into a linear pattern. Nevertheless, even if the true causal structure of the phenomenon under study is not linear, it does not mean that the causal relationship cannot be detected using linear modeling. With the advanced use of non-linear modeling, especially in the field of business data mining, researchers feel the tension of choosing between linear and nonlinear models. It is the conviction of the author that the appropriateness of non-linear modeling and linear modeling depends on specific research purposes (prediction vs. explanation). While nonlinear models are suitable to illustrate a physical or mechanical process under a natural interpretation, researchers occasionally have to go beyond the natural interpretation to look for theoretical explanations of the relationships between attributes. Judging the efficacy of statistical modeling, which is essentially a scientific method, should be based upon the criteria developed throughout the history of science, rather than through observations from the business market and political events within one or two decades. Examples from the history of science, including Dalton's atomic model, Galileo's law of uniform acceleration, and the Titius-Bode Law, will be cited to illustrate the usefulness of linear models in terms of providing explanation with theoretical depth. It is not the case that explanatory models are still useful *in spite of* the fact that they are wrong to some degree. On the contrary, they are useful *because* they are wrong.

Keywords: Simplicity, linearity, explanation, prediction, data mining, history of science.

The financial tsunami that exploded near the end of 2008 caught most people by surprise. Before this turmoil, quite a few authors had questioned the validity of conventional business research and statistical modeling in general, for these over-simplified models failed to take many unpredictable variables into account. For example, Taleb [1, 2] coined the phrase "black swans" to illustrate how outliers and improbable events throw all linear models off-track, whereas Smick [3] introduced the phrase "the world is curved" to convey the message that the fluctuation of world events and their consequences are highly unpredictable. Nonetheless, it is important to point out that judging the efficacy of statistical modeling, which is essentially a scientific method, should be based upon the criteria developed throughout the history of science, rather than through the observations from the business market and political events within one or two decades. Thus, this article aims to put simple linear modeling in a historical perspective, and defend its value.

Indeed, the debate regarding simplicity and linearity emerged much earlier than the current financial crisis. The idea of simplicity was popularized by Zellner's phrase "Keep It Simple, Stupid" (KISS). Later, Zellner [4] refined the phrase to "Keep It Sophisticatedly Simple" and further argued for KISS by citing its ubiquity in business domains. For example, the slogan of the Honda Motor Company is, "We make it simple." The Dutch Schipol airport in its advertising claims that, "It excels because it is simple and convenient". The counter-example cited by Zellner is the "Vector

Autogressive Regression" (VAR) developed by the Federal Reserve Bank of Minneapolis. Due to its poor performance, Zellner mocked it as "Very Awful Regression."

However, with advances in computer technology and mathematics, researchers feel the tension of choosing conventional linear modeling and nonlinear modeling. In recent years use of these methods has reached a crescendo where many non-linear data mining techniques have been proven to be effective tools for unveiling hidden data patterns. The following are some examples of popular nonlinear or/and non-parametric models:

- Polynomial regression [5]
- Wavelet smoothing [6-8]
- Kernel regression and Kernel discriminant analysis [9-12]
- Local polynomial smoothing [13, 14]
- Smoothing splines [15]
- Tree-based models [16-18]
- Multivariate adaptive regression splines (MARS) [19, 20]
- Neural networks [21, 22]
- Projection pursuit [23]
- Various Bayesian methods [24]

Some critics of linear modeling concentrate on the problem of unrealistic assumptions. Freedman [25] lists many popular responses to doubts about the linearity assumption and offers criticism against linear modeling:

* Address correspondence to this author at the Applied Learning Technologies Institute, Arizona State University, PO Box 612, Tempe AZ 85280, USA; Tel: 480-727-6978; Fax: 480-727-6767; E-mail: chonghoyu@gmail.com

In the social sciences, however, statistical assumptions are rarely made explicit, let alone validated. Questions provoke reactions that cover the gamut from indignation to obscurantism. *We know all that. Nothing is perfect. Linearity has to be a good first approximation. The assumptions are reasonable. The assumptions do not matter. The assumptions are conservative. You cannot prove the assumptions are wrong. The bias will cancel. We can model the bias. We are only doing what everybody else does. Now we use more sophisticated techniques. What would you do? The decision-maker has to be better off with us than without us. We all have mental models; not using a model is still a model* (p.103). (Italics appear in the original text).

In the Fisherian research tradition, parameter estimators are said to be consistent in the sense that as the sample size increases, the difference between the estimated and the true parameters should be smaller and smaller, and eventually the estimator is identical to the true parameter. But if the true relationships among variables are non-linear, how could the estimated parameters based on a linear model have any hope of resembling the true parameters [26]?

PREDICTIVE AND EXPLANATORY MODELING

This short article is not a bold attempt to discuss all pros and cons of linear and nonlinear models. Rather, this modest proposal is to state the point that the appropriateness of non-linear modeling and linear modeling depends on specific research purposes (e.g. prediction vs. explanation). Shmueli [27], and Shmueli and Koppius [28] observed that many researchers are confused by the distinction between explanatory and predictive models. The purpose of explanatory modeling is, as the name implies, to provide an explanation, usually a causal one, of how and why certain empirical phenomena occur. In contrast, predictive modeling aims to predict new observations with high accuracy even if a sound theory is absent. In the terminology of philosophy of science, predictive modeling is concerned with regularity expressed by empirical observation and thus empirical adequacy is the primary criterion of judging the efficacy of a predictive model. On the other hand, an explanatory model intends to study the relationships between properties that cannot be directly observed, and thus the relationships must be established by theorization. Borrowing from the terminology of psychometrics, predictive modeling can be conceptualized as a tool for *fitting the model into the data* in order to illustrate the process of a phenomenon and to forecast the future events, whereas explanatory modeling is a tool for *fitting the data into the model*.

Nonlinear models, such as data mining techniques, are suitable for identifying patterns and predicting the future occurrences. Nonetheless, it is important to point out that pattern recognition has a strong exploratory characteristic. The so-called “patterns” revealed by non-linear data mining may not represent the true underlying structure of the phenomenon, as illustrated in the problems of induction [29-31]. Based on the notions that humans are designed by Mother Nature to be fooled and that history is driven by the highly improbable events, Taleb [cf. 2, 3] explicitly downplayed the usefulness of virtually all statistical models. However, he did not clearly distinguish predictive modeling from explanatory modeling, and thus at most his criticisms

apply to one model type only, predictive models. Another type of modeling, namely, explanatory, is still very useful. The statement “all models are wrong [but] some are useful” [32] has been repeatedly cited by numerous modelers to counteract criticisms against “unrealistic” modeling made by Freedman, Taleb, and many other skeptics. Actually, it is not the case that explanatory models are still useful *in spite of* the fact that they are wrong to some degree. In contrast, they are useful *because* they are wrong. Linear modeling, which is inherently *wrong* for its high degree of noise suppression and pattern simplification, is instrumental in uncovering theoretical insights. In other words, explanatory modeling is abstract and decontextualized and thus it must be wrong in a concrete context.

It is crucial to clarify that linear modeling should not be equated with explanatory modeling; by the same token, non-linear modeling is by no means synonymous with predictive modeling. Linear models can be used for prediction, of course, and nonlinear models can also yield insightful explanation. Rueger and Sharp [33] argued that nonlinear dynamics in physics is capable of fulfilling empirical adequacy in terms of data-model fit, yet still providing theoreticians with ample explanatory power. Rather, the main point here is about the suitability of linear modeling for explanation. In the following section, examples from the history of science will be cited to illustrate the usefulness of linear models in terms of theory construction. To provide a balanced view, several counter-arguments will be discussed.

EXAMPLES IN THE HISTORY OF SCIENCE

Today it does not take a well-trained scientist to learn that atomic weight is calculated using linear approximation. The atomic number of an element is determined by the number of protons the element has, and the atomic weight depends on the number of neutrons and protons together. According to Aston’s discovery of isotopes, the number of protons and that of neutrons are almost equal in each element. Thus, we know that the atomic weight is approximately linearly dependent on the atomic number. However, what is common sense today was a puzzle in the early 19th century. At that time there was no certain procedure for determining atomic weights. Dalton, the advocate of atomism, knew about the weight ratios in which different chemical substances combined with each other, but he had little information about the molecular formulas of the compounds or the atomic weights of the elements. Determination of one required knowledge of the other. In spite of this “catch-22,” Dalton moved ahead with the assumption that elements combined with each other in the simplest possible atomic configurations. Although this approach led him to some incorrect conclusions, such as his hypothesis that water was HO rather than H₂O, at least a simple model allowed him to *get off the ground* and eventually learned the *fundamental structure* of the physical world [34].

The same pattern could be observed in other great discoveries in the history of science. Galileo noticed that the frequency with which a pendulum goes back and forth does not depend on how far the pendulum swings. This is because as long as the magnitude of the oscillation is small enough, the pendulum is a linear system in the sense that the rates of change of its displacement and momentum are proportional

to its momentum and its displacement [35]. Similarly, Galileo's law of uniform acceleration is considered a simple linear equation because air resistance is considered too negligible to be put into the equation of describing the movement of bodies. It is true that Galileo's linear approximation failed to support the design and manufacture of parachutes, in which a parameter must be added into the equation for representing the air resistance on the body [36]. Nevertheless, without the simple and linear model of acceleration suggested by Galileo, it is doubtful whether the complex model for designing parachutes could have been developed at all.

The discovery of the Titius-Bode Law, a linear model that predicts and explains the spacing of the planets in the Solar System, is another example. The pattern was first pointed out by Johann Titius in 1766 and was formulated as a linear model by J.E. Bode in 1778. As this pattern is only a linear approximation, it does not hold for the Neptune and Pluto. More importantly, it had no clear theoretical explanation. Nevertheless, later it was found that the suggested pattern agreed with the orbit of the newly discovered planet, Uranus. Moreover, it found a missing planet between Mars and Jupiter, and shortly thereafter, the asteroids were found in very similar orbits [cf. 36]. Although Pluto was demoted to be a dwarf planet in 2006, its position is unaffected and thus the linear approximation of the solar system model is still valid.

The ideal gas law and the Hooke's law were also postulated as linear approximations. In the ideal gas law, gas will break down if the pressure is too high or the temperature is too low, whereas in Hooke's law gas will also break down if the displacements are too large [37]. Nevertheless, the non-linear structure of these natural phenomena did not stop scientists from producing a usable model. In the ideal gas law there are several assumptions aside from linearity. Specifically, the gas law tells us that the temperature of a gas is a function of pressure and volume. This causal relationship can be explained by Newton's laws of motion based upon the assumptions that particles are solid bodies and that the gravitational force is the only force that is taken into account. However, atoms are only approximately perfectly hard bodies and the gravitational force is just one of many forces [38]. The significant point in this example is that scientific inquiry could still proceed with *false* assumptions.

Consider a more recent example: Linearity in classical physics is never exact and thus Weinberg [cf. 35] has a strong reason to suspect that in quantum mechanics the wave function may not be exactly linear after all. Although he came up with a slightly non-linear model to explain quantum mechanics that seemed to promise better accuracy, he admits that the foundation of building a better non-linear model is the linear precursor. In brief, linear approximation has been used successfully to advance scientific inquiry.

There is a common thread in the preceding examples from the history of science: Researchers have to make bold assumptions, and live with uncertainty and inaccuracy in order to develop a theoretical, causal explanation of the phenomenon under study. One might invoke the "file-drawer problem" to respond to the preceding idea. While great scientific discoveries have resulted from linear approximation, those were the successful stories that went into science his-

tory records. Failed attempts due to over-simplified modeling simply stayed in the "file drawers" of anonymous scientists, and thus, those failed theories are unknown to us. I do not dispute that the file-drawer problem exists; however, there is no guarantee that every ambitious endeavor will lead to fruitful scientific discovery. It does not matter whether only 50%, 30%, or even 10% of incomplete models survive in the long run; the scientific community corrects and augments incomplete theories in subsequent inquiries. To put strong demands on initial research, as Freedman did, would kill theory generation before a model gets a chance to develop and prove itself.

NATURAL INTERPRETATION AND THEORETICAL EXPLANATION

One may point out that in recent history nonlinear models also produce numerous fruitful research results. But, it is noteworthy that many recent examples of nonlinear models are mechanistic in the sense that there is a natural physical interpretation of the non-linear pattern [39, 40]. By "natural" it means that the interpretation can be taken at the face value or at the observational, empirical level. A typical example is the growth of biological entities over time. One may argue that certain nonlinear models, such as neural networks, are hard to interpret and the meaning of the conclusion is hardly natural or obvious. Interestingly enough, there are many successful applications of neural nets in the domains that involve mechanical or physical processes. Examples of applications of neural nets include signal processing [41], automation in manufacturing [42-44], navigation, speech recognition, and object recognition in robotics [45-47]. More precisely, the process of neural nets, such as use of hidden layers, makes the results hard to interpret, but the physical or mechanical process by itself can be intuitively conceptualized. Consider another non-linear method--the Fourier series approach. Examples of its applications are modeling hourly energy use in commercial buildings [48] and biological rhythms [49]. Again, these subjects do not urge researchers to go beyond a natural interpretation.

However, whether the underlying relationship is physical should not be the criterion for determining the appropriateness of linear or non-linear modeling. It is a well-known fact that stress and performance forms a curvilinear relationship. It is absurd to claim that because this is a psychological, not physical process, non-linear modeling is treated as being improper to explain this phenomenon. The main point here is about whether the researcher aims to draw a causal explanation, no matter if it is physical or psychological, from the data.

As a matter of fact, many phenomena are not embedded with a natural interpretation. A sociological process, which involves many variables and human agents, is an obvious example. No doubt the complexity of the financial market makes it qualified as a sociological process. Using neural nets, some financial forecasters correctly predicted the direction of the market change for the next day with a very small error rate and made wise decisions about investments on mortgages [50, 51]. However, unlike basic or theoretical research, these predictions do not shed light on the causal mechanism of market value. On the other hand, psychological research and many other types of research have a differ-

ent goal. In the following section let's consider several arguments against linear modeling.

ARGUMENTS AGAINST SIMPLICITY

One of the popular arguments in favor of linear models is that they are too simple to be true. But, how simple is simple? What is the meaning of simplicity? Simplicity is usually defined in terms of the numbers of parameters within a given representational system. However, simplicity in a different conceptual sense is much more difficult to define. Zellner [52] pointed out that a quantitative or a parametric definition of simplicity can be controversial because parametric simplicity implies a formal (mathematical) structure. But some people may regard mathematics as quite complex, and thus, it violates another meaning of simplicity: *Communicability*. While one kind of simplicity/complexity is parametric, our overall judgment about models may be influenced by other ways of being simple/complex, too. It is appropriate, though, to take parametric simplicity into account. Weinberg [cf. 35] equates simplicity with the simplicity of ideas, rather than simplicity measured by counting equations and symbols. However, a definition like this can lead to very ambiguous results.

For example, Kepler, who discovered that the orbit of earth around the sun is elliptical rather than circular, claimed that his conjecture was based on the principle of the economy of nature. To him, nature is the paradigm of an efficient machine designed by an Intelligent Being, and thus, nature should be explained by simpler ideas. However, this simplicity of ideas did not result in simplicity of modeling [53]. Popper [54] argued that the hypothesis that the orbit of a planet is a circle is simpler than the hypothesis that it is an ellipse. This leads to another contentious point: *Representation*. In 2006, Creath asserted in a personal communication that defining a circle in an equation is more complex than doing the same for an ellipse because a circle can be defined as a special case of an ellipse (circle = ellipse + (radius x = radius y)). Let's go back to the example of linear modeling. In Cartesian coordinates, it takes fewer parameters to define a line than a curve. However, the opposite is true when curves are represented in a polar coordinate system [55]. In a polar coordinate system, a curve can be defined by just the distance from the center and the angle. Even if a single representational system, such as the Cartesian plane, is adopted, simplicity is still a contextually-dependent concept. For example, when comparing regression models with different sets of predictors, model comparison and variable selection procedures can be employed to determine whether the increase of R^2 can justify the increased complexity of the model. However, linear models in SEM are not necessarily simpler than non-linear models in regression analysis, and there is no objective way to tell whether one is simpler than another.

GENERALIZABILITY AND USEFUL INFORMATION

One might argue that a non-linear model with more terms can fit the data better than a linear model. However, this type of "good" fit tends to be very sample-dependent and hence this model may not be transferable to other samples. While a linear function that fits a dataset with a particular coefficient may have a different weight in another setting, whether the

relationship is positive or negative may remain the same in the new sample. On the contrary, it is very likely that a curve in a scatter plot that passes through as many points as possible does not hold the same quadratic pattern in another data set. It is important to note that a non-linear model does not only indicate whether the relationship between variables is positive or negative; it also indicates how many "turns" are in the functional curve. In another sample, the data points would not fall along the "turns" of that curve. Jeffreys [56] explicitly points out that including too many terms in a relation can improve fit but "the conclusion is that including too many terms will lose accuracy in prediction instead of gaining it" (p. 46). The aim of model selection should also include the ability of a model to generalize to predictions in a different domain [57]. In a similar vein, Sober [58] asserted that more adjustable parameters increase its risk of overfitting the data. In contrast, simpler hypotheses have higher likelihoods.

One may use the Item Characteristic Curve (ICC) in psychometrics to argue for the usefulness of a non-linear model (the red curve in Fig. 1). ICC indicates the relationship between the probability of answering an item correctly and the student skill level (the Y-axis is the probability of answering an item correctly and the X-axis is the skill level). For example, if the skill level of a student is +5, then the probability of passing the item is near 100%. If the skill level is 0, the probability is .5. But the probability drops to zero if the student is very unskilled (-5). It is important to point out that the relationship between the probability and the skill level is not proportional; rather, the slope of the ICC is less steep as the skill level increases from +2 to +5. The same pattern is observed with the range of -2 to -5, but the curve takes a sharp turn from -2 to +2. This non-linear pattern is called "ogive." At first glance, this seems to be a sound argument against linearity.

However, when the item information function (IIF, indicated by the blue curve) is superimposed on the graph, it is not clear that this is the case. The IIF indicates how much precise information about the student ability the test developer can obtain if the item is included in the test. The curve height indicates the amount of precise information at different skill levels. In this example, most information concentrates on the skill levels ranging from -1 to +1 (the green shade). Within the green region, the ICC is almost linear. In other words, a linear model in this case can yield most of the useful information. As mentioned before, proper use of non-linear and linear models depends on the research question. Theoretically speaking, if the skill level (theta estimation) of a student is +5, the probability of answering an item correctly is .99. However, in the empirical data there might not be any students who achieve such a high skill level, and thus, interpretation of the item attribute should be restricted within the linear range of the ICC. Nonetheless, if one is interested in counterfactual reasoning (what-if), it will be helpful to go beyond the linear range to the non-linear model. In short, even when the relationships are nonlinear, the researchers are often only interested in a limited range of the phenomenon, which can be sufficiently approximated by a linear model.

The range for proper interpretation is context-specific. Suppose that $Y=a+bX+e$ denotes an invariant relationship

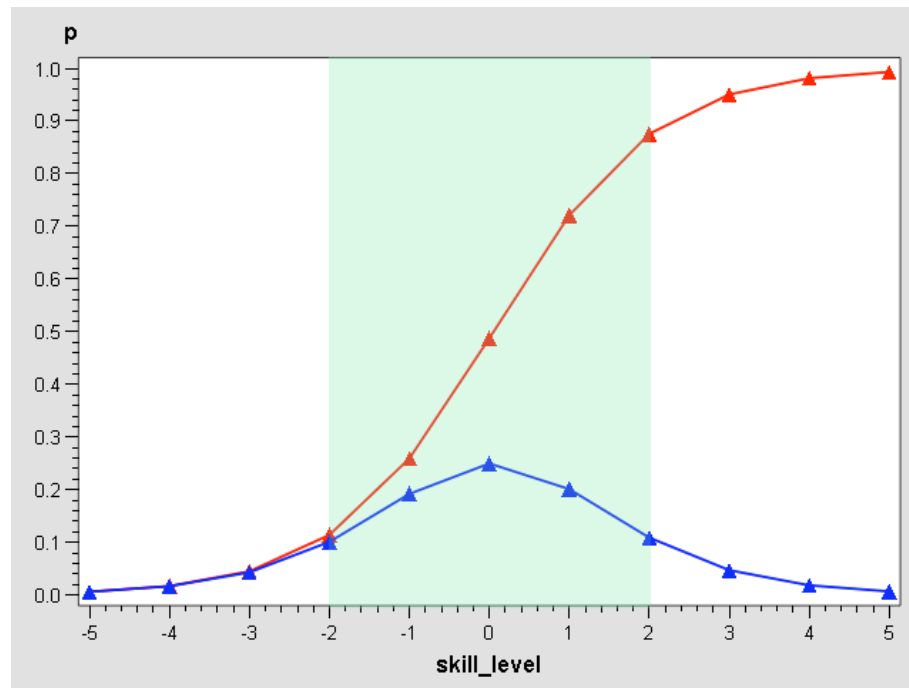


Fig. (1). ICC and IIF.

within a *reasonable range* of values of X and Y. However, if the values of any linear model are extended to the extreme, the model will certainly turn into a non-linear one. For example, if I water my plant with 1–3 liters of water, my plant would grow. If I pour more water, the plant will grow faster to some point. But if I water the same plant with 1,000 liters of water and the plant dies, it does not negate the linear relationship between watering and plant growth [59]. Like the preceding IRT model, the most useful information can be obtained within the linear region. Similar examples can be found in economics. Interventions that change the money supply may change the price level within some range of circumstances, but not in others [60].

FORCED LINEARITY IN FACTOR MODELING

The problem of Likert scaling is a less common, yet powerful argument against linearity. A Likert scale is commonly used for constructing answer options in psychological and sociological research. There is some controversy considering whether the Likert-scaled data can be properly used in Generalized Linear Model (GLM), such as ANOVA, because Likert-scaled data are essentially ordinal measures. Unlike an interval scale, in an ordinal scale the assumption of equal spacing is highly questionable. To be specific, the distance between “5” and “4” may not be equal to that between “4” and “3,” and so on. If the Likert Scale is mapped to a “true” interval scale (if there is one), the curve will be non-linear. For example, the “true” score of a particular construct between 1 and 100 would be assigned a Likert score of 1, 101–500 would be assigned to 2, and so on. It is argued that the non-linear nature of the data may lead the measures to be unsuitable for a linear model. Some critics go so far as to argue that as ordinal variables do not have origins or units of measurement, the reporting of their means, variances and covariances is meaningless [61, 62].

I agree that any data analysis on a single Likert-scaled item should avoid interval-based tools; rather, ordinal logistic regression is no doubt a better option. Interestingly enough, when many Likert-scaled items are loaded into a latent construct to formulate a factor structure, researchers seem to be more forgiving to the non-linear and non-additive nature of ranked order variables. Once a researcher explained his rationale to me by saying, “When ten five-point Likert scale items are summed together as a composite score, the range of the scores will expand from 0-5 to 0-50. It adds more variability to the data and thus parametric procedures such as Analysis of Variance can be appropriately employed.” Actually, this approach does not address the issue of non-linearity at all. My explanation is very straightforward. Researchers are simply forcing non-linear measures into linear modeling for convenience. After all, factor analysis is a technique of data reduction in an attempt to make modeling manageable. If one is critical of analyzing a single Likert scaled variable with parametric procedures, should one also reject use of parametric tests for all latent constructs yielded from a factor model consisting of Likert-scaled items? But this is not what is happening in psychological science. Examples of mental constructs using Likert scaling include *Attitudes Toward the Internet* [63], *Attributional Style Questionnaire* [64], *Curiosity and Exploration Inventory* [65], *Online Technologies Self-Efficacy Survey* [66], *Subjective Happiness Scale* [67], and *White Racial Identity Scale* [68]. In brief, most factor models of subjective cognitive constructs are measured by Likert scales that do not conform to linearity, yet psychologists are producing fruitful results.

CONCLUSION

It is not the intention of the author to downplay the merits of predictive modeling or nonlinear modeling. As a matter of

fact, data mining methods, which target on making predictions based upon nonlinear models, are helpful in exploratory data analysis and hypothesis generation [69]. Although the assumption of linear models is, in some sense, simpler than that of its nonlinear counterpart, on some occasions nonlinear modeling yields simpler conclusions than linear models. For example, in profiling college students who take online classes, Yu, *et al.* [70] employed nonlinear data mining techniques and found that only one or two variables are relevant to identify the attributes of the target audience. In addition, some subject matters in biology and engineering can lead to an effortless, natural interpretation, while in some situations, such as business applications, there is simply no need to substantiate the results with the merit of theoretical sophistication. In this case, nonlinear modeling is a natural fit to this type of research.

On the other hand, a research agenda that requires explanation may demand the researcher to sacrifice accuracy and to make so-called false assumptions. Glymour, *et al.* [71] defend the adequacy of linearity by arguing that sciences have always proceeded by approximation and idealization. A linear approximation may not literally be true, however. The principal justification for use of a linear model is that it fits the correlation data very well and no alternative linear model is readily available which provides a comparably good explanation of the correlations. Many times relationships in the real world do not fall into a linear pattern. But even if the true causal structure of the phenomenon under study is not linear, it does not mean that the causal relationship cannot be detected using linear modeling. Natural sciences use a variety of techniques to show that the postulated causes are responsible for the observed effect in spite of using false assumptions to obtain them. Given that there are many success stories throughout the history of science by employing linear modeling and the criterion of simplicity in theory development, Glymour and colleagues' and Kincaid's notions might be right. On the other hand, possible counter-examples, such as Item Characteristic Curves and Likert-scaled measures, do not sufficiently negate the value of assuming linearity in modeling. In short, whether linear or non-linear modeling is more useful should be tied to specific research objectives. For discovering causal explanations, researchers should not be afraid of modeling based upon simplicity, linearity, and even false assumptions. Nonetheless, a causal explanation does not necessarily lead to a representation of the objective reality. Rather, for some philosophers of science a good explanatory model could be empirically adequate rather than literally true [72].

Further, it is important to emphasize that the above suggestion is a general guideline rather than an infallible, universal rule. There are always some exceptions. In other words, on some occasions a simple and linear model may not work well for explanatory purposes. For example, in a study aiming to explain why minority students have a low retention rate, the linear regression model yielded a so-called explanation by identifying low college GPA as the most crucial factor [73]. But saying poor performance, as manifested by low GPA, leads to attrition, is tautological. Further, while a model with false assumptions could lead to sophisticated theoretical explanations, it could also result in accurate predictions. The ancient Ptolemaic astronomical model, also

known as the geocentric model, postulates that the earth is the center of the universe, as well as many complicated notions, such as using "circles on circles" to predict retrograde planetary motions. Nonetheless, it turns out that this model still generated good predictions for the movement of heavenly bodies [74, 75]. Thus, researchers may need to make a case-by-case decision and be ready to defend the rationale.

In summary, as mentioned in the beginning, the usefulness of statistical modeling should be based upon the long term development throughout the history of science instead of the short term observation in the current market. Thus, the author sides with Zellner's [cf. 4] assertion that researchers in econometrics and statistics should follow the advice of natural scientists to keep analyses and models sophisticatedly simple.

ACKNOWLEDGEMENTS

Special thanks to Dr. Brad Armentdt, Dr. Bernard Kobes, Dr. Samuel Green, Mr. Charles Kaprolet, Mrs. Lori Long, and Mr. Chang Kim for their valuable input to this article.

REFERENCES

- [1] N. N. Taleb, *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*. New York: Random House, 2004.
- [2] N. N. Taleb, *The Black Swan: The Impact of Highly Improbable*. New York: Random House, 2008.
- [3] D. Smick, *The World is Curved: Hidden Dangers to the Global Economy*. New York, N.Y.: Portfolio, 2008.
- [4] A. Zellner, "Keep it sophisticatedly simple", 1998 [Online] Available: <http://faculty.chicagosb.edu/arnold.zellner/more/CURRENT-PAPERS/kiss.doc>. [Accessed February 16, 2009].
- [5] R.L. Eubank, *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker, 1999.
- [6] D.L. Donoho, and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", *Journal of the American Statistical Association*, vol. 90, pp.1200-1224, 1995.
- [7] D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Wavelet shrinkage: Asymptopia (with discussion)?" *Journal of the Royal Statistical Society, Series B*, vol. 57, pp. 301-369, 1995.
- [8] M. Jansen, and P. Oonin, *Second Generation Wavelets and Applications*. London: Springer, 2005.
- [9] D. J. Hand, *Discrimination and Classification*. New York: Wiley, 1981.
- [10] D. J. Hand, *Kernel Discriminant Analysis*, New York: Research Studies Press, 1982.
- [11] D. J. Hand, *Construction and Assessment of Classification Rules*. New York: Wiley, 1997.
- [12] J. R. Wolberg, *Data Analysis Using the Method of Least Squares*, New York: Springer, 2006.
- [13] M. P. Wand, and M. C. Jones, *Kernel Smoothing*, London: Chapman & Hall, 1995.
- [14] J. Fan, and I. Gijbels, "Data-driven bandwidth selection in local polynomial: Variable bandwidth and spatial adaptation", *Journal of the Royal Statistical Society, Series B*, vol. 57, pp.371-394, 1995.
- [15] W. Härdle, *Applied Nonparametric Regression*. Cambridge: Cambridge University Press, 1990.
- [16] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Monterey, CA: Wadsworth International Group, 1984.
- [17] J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, Boston, MA: Elsevier, 2006.
- [18] D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, NJ: Wiley-Interscience, 2005.
- [19] J. H. Friedman, "Multivariate adaptive regression splines", *Annals of Statistics*, vol. 19, pp.1-141, 1991.
- [20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, NY:

- Springer, 2001.
- [21] B. D. Ripley, "Neural networks and related methods for classification", *Journal of the Royal Statistical Society, Series B*, vol. 56, pp. 409-456, 1994.
- [22] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press, 1996.
- [23] J. H. Friedman, and W. Stuetzle, "Projection pursuit regression", *Journal of the American Statistical Association*, vol. 76, pp.817-823, 1981.
- [24] D. Dey, Ed., *Practical Nonparametric and Semiparametric Bayesian Statistics*, New York: Springer Verlag, 1998.
- [25] D. Freedman, "From association to causation via regression", *Advances in Applied Mathematics*, vol. 18, pp. 59-110, 1997.
- [26] D. Freedman, "Rejoinder to spirites and scheines", in: *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*, V. R. McKim, and S. P. Turner, Eds. Notre Dame, IN: University of Notre Dame Press, 1998, pp. 177-182.
- [27] G. Shmueli, "To explain or to predict?", 2009 [Online] Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1351252. [Accessed March 1, 2009].
- [28] G. Shmueli, and O. Koppius, "Contrasting predictive and explanatory modeling in IS Research." *Robert H. Smith School Research*, Paper No. RHS 06-058, 2008 [Online] Available: <http://ssrn.com/abstract=1112893>. [Accessed June 11, 2008].
- [29] N. Goodman, *Facts, Fictions, and Forecast*. Indianapolis, IN: Hackett, 1954.
- [30] D. Hume, *An Enquiry Concerning Human Understanding, and Selections from a Treatise of Human Nature*. Chicago: Open Court, 1777/1912.
- [31] C. H. Yu, "Is automated data mining an emerging paradigm that can supersede hypothesis testing?", *Theory and Science*, vol. 9, 2007 [Online] Available: <http://theoryandscience.icaap.org/content/vol9.2/Chong.html> [Accessed June 11 2008].
- [32] G. E. P. Box, and R. D. Norman, *Empirical Model-Building and Response Surfaces*. New York: Wiley, 1987.
- [33] A. Rueger, and W. D. Sharp, "Simple theories of a messy world: Truth and explanatory power in nonlinear dynamics", *British Journal for the Philosophy of Science*, vol. 47, pp. 93-112, 1996.
- [34] H. Chang, "How to take realism beyond foot-stamping", *Philosophy*, vol. 76, pp.5-30, 2001.
- [35] S. Weinberg, *Dreams of a Final Theory*, New York: Pantheon Books, 1992.
- [36] H. Simon, "Science seeks parsimony, not simplicity: Searching for pattern in phenomena", in: *Simplicity, Inference, and Modeling: Keeping it Sophisticatedly Simple*, A. Zellner, Ed. West Nyack, NY: Cambridge University Press, 2002, pp. 32-73.
- [37] G. Derry, *What Science is and How it Works*. Ewing, NJ: Princeton University Press, 1999.
- [38] Kincaid, H. "There are Laws in the Social Sciences", In: *Contemporary Debates in Philosophy of Science*, C. Hitchcock, Ed. MA: Blackwell Publishing, 2004, pp. 168-186.
- [39] S. Huet, A. Bouvier, M. A. Poursat, and E. Jolivet, *Statistical Tools for Nonlinear Regression*. New York: Springer, 2004.
- [40] J. C. Pinheiro, and D. M. Bates, *Mixed Effects Models in S and S-plus*. New York: Springer, 2004.
- [41] B. Kosko, *Neural Networks for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [42] K. Althoefer, B. Lara, and L. D. Seneviratne, "Monitoring of self-tapping screw fastenings using Artificial Neural Networks", *Journal of Manufacturing Science and Engineering*, vol. 127, pp. 236-243, 2005.
- [43] B. Karlik, M. Uzam, M. Cinsdikici, and A. H. Jones, "Neurovision-based logic control of an experimental manufacturing plant using neural net le-net5 and automation Petri nets", *Journal of Intelligent Manufacturing*, vol. 16, pp. 527-548, 2005.
- [44] W. L. Wang, J. G. Conley, Y. N. Yan, and J. Fuh, "Towards intelligent setting of process parameters for layered manufacturing", *Journal of Intelligent Manufacturing*, vol. 11, pp. 65-74, 2000.
- [45] F. Declercq, and R. DeKeyser, "Object recognition and selective handling by a robot", *Journal of Intelligent and Robotic Systems*, vol. 25, pp. 121-132, 1999.
- [46] T. Kitazoe, S. I. Kim, and T. Ichiki, "Speech recognition using a stereo vision neural network model", *Artificial Life and Robotics*, vol. 4, pp. 37-41, 2000.
- [47] D. T. Pham, and D. R. Parhi, "Navigation of multiple mobile robots using a neural network and a Petri Net model", *Robotica*, vol. 21, pp. 79-93, 2003.
- [48] A. Dhar, T.A. Reddy, and D. E. Claridge, "Generalization of the fourier series approach to model hourly energy use in commercial buildings", *Journal of Solar Energy Engineering*, vol. 121, pp. 54-62, 1999.
- [49] B. Benson, M. Gutierrez-Wing, and K. Rusch, "Application of the Fourier method to differentiate biological rhythms from stochastic processes in the growth of *Selenastrum Capricornutum* Printz: Implications for model development", *Journal of Applied Psychology*, vol. 20, pp. 103-111, 2008.
- [50] C. Giles, S. Lawrence, and A. C. Tsoi, "Noisy time series prediction using recurrent neural networks and grammatical inference", *Machine Learning*, vol. 44, pp. 161-183, 2001.
- [51] D. Wilson, S. D. Paris, J. A. Ware, and D. H. Jenkins, "Residential property price time series forecasting with neural networks", *Knowledge-Based Systems*, vol. 15, pp. 335-341, 2002.
- [52] Zellner, Ed., *Simplicity, Inference, and Modeling: Keeping it Sophisticatedly Simple*, West Nyack, NY: Cambridge University Press, 2002.
- [53] J. Kozhamthadam, *The Discovery of Kepler's Laws: The Interaction of Science, Philosophy, and Religion*, Notre Dame: University of Notre Dame Press, 1994.
- [54] K. Popper, *Logic of Scientific Discovery*, New York: Routledge, 1959.
- [55] C. Hempel, *Philosophy of Natural Science*, Englewood Cliffs, N.J.: Prentice-Hall, 1966.
- [56] H. Jeffreys, *Theory of Probability*, Oxford: Clarendon Press, 1961.
- [57] M. Forster, "Key concepts in model selection: Performance and generalizability", *Journal of Mathematical Psychology*, vol. 44, pp. 205-231, 2000.
- [58] E. Sober, "What is the problem of simplicity?" In: *Simplicity, Inference, and Econometric Modeling*, H. Keuzenkamp, and M. McAleer, and A. Zellner, Eds. Cambridge, UK: Cambridge University Press, 2002, pp. 13-32.
- [59] J. Woodward, "Causal interpretation in systems of equations", *Synthese*, vol. 121, pp. 199-247, 1999.
- [60] J. Woodward, *Making Things Happen*, Oxford: Oxford University Press, 2003.
- [61] K. Joreskog, "On the estimation of polychoric correlations and their asymptotic covariance matrix", *Psychometrika*, vol. 59, pp. 381-389, 1994.
- [62] K. Joreskog, and A. M. Aish, *Structural Equation Modeling with Ordinal Variables*, Chicago: Scientific Software, 1993.
- [63] B. Ohlund, C. H. Yu, S. DiGangi, and A. Jannasch-Pennell, "Impact of asynchronous and synchronous Internet-based communication on collaboration and performance among K-12 teachers", *Journal of Educational Computing Research*, vol. 23, pp. 435-450, 2001.
- [64] G. Buchanan, and M. G. Seligman, Ed., *Explanatory Style*. Hillsdale, N.J.: Erlbaum, 1995.
- [65] T. Kashdan, "Social anxiety dimensions, neuroticism, and the contours of positive psychological functioning", *Cognitive Therapy and Research*, vol. 26, pp. 789-810, 2002.
- [66] M. Miltiadou, and C. H. Yu, *Validation of the Online Technologies Self-Efficacy Survey (OTSES)*, in Association for Educational Communications and Technology (AECT). International Convention, Denver CO, 2000.
- [67] S. Lyubomirsky, "Why are some people happier than others?: The role of cognitive and motivational processes in well-being", *American Psychologist*, vol. 56, pp. 239-249, 2001.
- [68] J. Behrens, C. H. Yu, and W. Rowe, *Meta-analysis of White Racial Identity Attitude Scale (RIAS-W)*, in Annual Meeting of American Psychological Association. New York, 1995.
- [69] C. H. Yu, A. Jannasch-Pennell, S. DiGangi, C. Kim, and S. Andrews, "A data visualization and data mining approach to response and non-response analysis in survey research". *Practical Assessment, Research and Evaluation*, vol. 12, no. 19, 2007 [Online] Available: <http://pareonline.net/getvn.asp?v=12&n=19> [Accessed Dec 20, 2007].
- [70] C. H. Yu, S. DiGangi, A. Jannasch-Pennell, and C. Kaprolet, "Profiling students who take online courses using data mining", *Online Journal of Distance Learning Administration*, vol. 11, no. 2, 2008 [Online] Available: <http://www.westga.edu/~distance/ojdl/summer112/yu112.html> [Accessed June 15, 2008].
- [71] C. Glymour, R. Scheines, P. Spirtes, and K. Kelly, *Discovering*

- Causal Structure: Artificial Intelligence, Philosophy of science, and Statistical Modeling*. Orlando, FL: Academic Press, 1987.
- [72] B. Van Fraassen, *The Scientific Image*, Oxford: Clarendon Press, 1980.
- [73] C. H. Yu, A. Jannasch-Pennell, S. DiGangi, and C. Kaprolet, "An examination of the relationship between high school technological-oriented courses and college retention using regression modeling", in Joint Statistical Meeting Conference, Denver, CO, 2008.
- [74] T. S. Kuhn, *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*, Cambridge: Harvard University Press, 1957.
- [75] G. Shmueli, "Good predictions by wrong model?", 2008 [Online] Available: <http://blog.bzst.com/2008/04/good-predictions-by-wrong-model.html>. [Accessed February 16, 2009].

Received: January 05, 2010

Revised: April 06, 2010

Accepted: April 18, 2010

© Chong Ho Yu; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.