http://www.ivyunion.org/index.php/ajpsr

Review

Why No One Knew that Hillary Clinton Would Lose the 2016 Election

Donald L. Buresh, PhD[1] and Theresa Pavone, PhD[2]

[1]Information Science and Systems, Department of Information Science and Systems, Earl Graves School of Business and Management, Morgan State University, Baltimore, United States

[2]School of Business, Embry-Riddle Aeronautical University, Daytona Beach Campus, Florida, United States

Abstract: The paper argues that a blind reliance on the use and results of predictive models led to the major pollsters never knowing that Hillary Rodham Clinton would lose the 2016 Presidental election. The reason is that history is not necessarily a good predictor of the future and that predictive models employed by the pollsters assumed the existence of a steady-state which was not present at the time. The paper introduces predictive models by providing a short history, followed by explaining how predictive models work. The paper lists the fundamental assumptions of predictive models, highlighting the advantages and disadvantages. The paper then analyzes in some detail why various predictive models incorrectly predicted that Hillary Clinton would win the 2016 presidential election, when in fact Donald Trump was elected President of the United States. The thesis is that when a predictive model experiences an exogenous shock or a superseding intervening cause, the dependence on data before the shock occurred is unwarranted. In fact, it is argued that all data before the exogenous shock or superseding intervening cause should have been ignored, and only the data that appeared after the shock should have been used in making predictions. The paper concludes that the predictive models used by the pollsters during the 2016 Presidential election were incapable of recognizing an exogenous shock or a superseding intervening cause and that human intervention was needed to correct for the limitations of the predictive models that were employed.

*Correspondence to: 3115 Enoch Avenue, Zion, Illinois 60099, United States

Email: LoganSquareDon@sbcglobal.net

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Purpose of the Paper

    The goal of this paper is to investigate why many of the predictive models that were used in the 2016 Presidential election failed to demonstrate the Hillary Rodham Clinton would lose. A prediction market using a predictive model is an exchange-traded market which is created for trading event outcomes. [1] The modeling can be utilized for any unknown event, regardless of the time in which it occurred. The idea of predictive modeling is to estimate the probability of an outcome, given a set or collection of input data. Predictive modeling overlaps with the field of machine learning. When predictive models are employed commercially, predictive models are

---

[1] SEMOUR GEISSER, PREDICTIVE INFERENCE: AN INTRODUCTION (Chapman & Hall 2016).

often called prediction analytics.[2] Predictive analytics use a variety of statistical techniques employing the principles of machine learning and data mining to analyze current and historical facts to predict unknown future events.[3]

Issue Statement

The blind reliance on the use and results of predictive models leads to spectacular failures. The reason is that history is not necessarily a good predictor of the future.[4] A predictive model assumes the existence of a steady-state.[5] This assumption is easily violated when a system involves human beings.

Significance of the Issue

In the modern world we live in today, it is essential to predict future events with a relative degree of accuracy and precision. Predictive models are used extensively in the fields of actuarial science[6], financial services[7], telecommunications[8], pharmaceuticals[9], and a wide range of applications.

---

[2]Charles Nyce, *Predictive Analytics*, White Paper, n.d.,

http://www.hedgechatter.com/wp-content/uploads/2014/09/predictivemodelingwhitepaper.pdf (last visited July 20, 2017).

[3]Id.

[4]DAVID HUME, AN ENQUIRY CONCERNING HUMAN UNDERSTANDING (Jovian Press 2017) (1776).

[5]Geisser, *supra.*

[6]Nathan Conz, *Insurers Shift to Customer-focused Predictive Analytics Technologies*, Insurance and Technology,

September 08, 2008,

http://www.insurancetech.com/data-and-analytics/insurers-shift-to-customer-focused-predictive-analytics-technologies/d/d-id/1310535? (last visited July 20, 2017).

[7] Sue Korn, *The Opportunity for Predictive Analytics in Finance*, HPC, April 21, 2011,

https://www.hpcwire.com/2011/04/21/the_opportunity_for_predictive_analytics_in_finance/ (last visited July 20, 2017).

[8] Eric Barkin, *CRM + Predictive Analytics: Why It All Adds Up*, Destination CRM, May, 2011,

http://www.destinationcrm.com/Articles/Editorial/Magazine-Features/CRM---Predictive-Analytics-Why-It-All-Adds-Up-74700.aspx (last visited July 20, 2017).

[9] Lauren McKay, *The New Prescription for Pharma*, CRM.com, August, 2009,

http://www.destinationcrm.com/articles/Web-Exclusives/Web-Only-Bonus-Articles/The-New-Prescription-for-Pharma-55774.aspx (last visited July 20, 2017).

SHORT HISTORY OF PREDICTIVE MODELS

The theory behind prediction markets is associated with Friedrich Hayek[10] and Ludwig von Mises.[11] One of the older and more famous prediction markets using a predictive model is the University of Iowa Electronic Market which was introduced during the 1988 presidential election.[12] In 2001, Intrade.com created a prediction trading market program which was located in Ireland. The application permitted traders to use real currency to create contracts relating to business, current events, financial matters, etc.[13] The website stopped trading in 2013. In July 2003, the U.S. Department of Defense Advanced Research Projects Agency (DARPA) suggested creating a Policy Analysis Market based on the Net Exchange, a San Diego research company specializing in online prediction markets. The purpose of this prediction market was to create a market regarding the future of the Middle East, allowing investors to trade futures of political developments in that part of the world.[14]

In 2004, predictive models and markets were openly advocated by James Surowiecki.[15] Here, the author explored the simple idea that large groups of people are smarter than the elite. In other words, collective

---

[10]Friedrich Hayek, *The Use and Knowledge in Society*, American Economic Review, 1945,

http://www.econlib.org/library/Essays/hykKnw1.html (last visited July 20, 2017).

[11] Ludwig von Mises, *Economic Calculation in the Socialist Commonwealth*, Ludwig von Mises Institute, 2012, (1920),

https://mises.org/system/tdf/Economic%20Calculation%20in%20the%20Socialist%20Commonwealth_Vol_2_3.pdf?file=1 &type=document   (last visited July 20, 2017).

[12] Stanley W. August, *Iowa Market Takes Stock of Presidential Candidates*, (Reprinted with the permission of The Wall Street Journal) , 2012, https://tippie.biz.uiowa.edu/iem/media/wsj.html   (last visited July 20, 2017).

[13]*About Intrade*, Intrade.com, http://www.intrade.com/history.html (last visited July 20, 2017).

[14] Carl Hulse, *Threats and Responses: Plans and Criticisms: Pentagon Prepares A Futures Market On Terror Attacks*, New York Times, July 29, 2003,

http://www.nytimes.com/2003/07/29/us/threats-responses-plans-criticisms-pentagon-prepares-futures-market-terror.ht ml   (last visited July 20, 2017).

[15]JAMES SUROWIECKI, THE WISDOM OF CROWDS (Anchor Press 2005).

knowledge is greater than expert knowledge.[16] When people cooperate, they are better at solving problems, encouraging innovation, arriving at wise decisions, and predicting the future.[17] In 2005, Polgreen et al. observed that the pharmaceutical company Eli Lilly and Company was using prediction market models to figure out which drugs would be successful in progressing beyond clinical trials by employing internal company information regarding drug research development efforts.[18] In October 2007, various American and European firms formed the Prediction Market Industry Association, whose purpose is to promote awareness, education, and validation for predictive models and markets.[19]

Predictive models were copulating like rabbits. In 2008, the New Zealand Institute for the Study of Competition and Regulation and Victoria University of Wellington created the predictive model website entitled *iPredict*.[20] The site closed down on December 1, 2016. The *simExchange*, a predictive model, was introduced to predict the product lifetime sales of video games and software titles.[21] Hewlett-Packard, an innovator in sales forecasting, now employs predictive models in its internal business units.[22] Intel Corp. is currently managing its

---

[16] The statement that collective knowledge is greater than expert knowledge is one of the few contributions that we believe that we have made to the body of human knowledge. we am not aware how we formulated this statement. We only understand that as far as we know, we have never read nor have we ever heard anyone else make the statement. On the contrary, when we have stated that collective knowledge is greater than expert knowledge, many of our colleagues have disagreed with us. When hearing their opposition, on many occasions we have pointed to Wikipedia.org as a case on point.

[17]Id.

[18]Philip M. Polgreen, Forrest D. Nelson, & George R. Neumann, *Use of Prediction Markets to Forecast Infectious Disease Activity, Healthcare Epidmiology*, Oxford University Press, January 15, 2007, https://www.jstor.org/stable/4463972?seq=1#page_scan_tab_contents (last visited July 20, 2017).

[19]Jed Christiansen, *Prediction Markets Industry Association*, Prediction Market Thoughts, October 25, 2007, https://blog.mercury-rac.com/prediction-markets-industry-association-17367a1528a9. (last visited July 20, 2017).

[20]*Institute for the Study of Competition and Regulation*, Victoria Business School, n.d., http://www.victoria.ac.nz/vbs/centresandinstitutes/institutes   (last visited July 20, 2017).

[21]*What Is This?*, simExchange, n.d., http://www.simexchange.com/frontpage.php   (last visited July 20, 2017).

[22] Eric Siegel, *Predictive Analytics: The privacy pickle – Hewlett-Packard's prediction of employee behavior*, Analytics Magazine, November/December 2013, http://analytics-magazine.org/predictive-analytics-the-privacy-pickle-hewlett-packards-prediction-of-employee-behavior/ (last visited July 20, 2017).

sales efforts using predictive models.[23] Google.com offers a predictive model as part of its *BigQuery* cloud application.[24] General Electric Corp. uses predictive analytics to monitor it production optimizations solutions for onshore and offshore oil and gas pipeline operations.[25] The Hollywood Stock Exchange (HSX) built and now operates a virtual stock market, the Interactive Music Exchange for Fuse Networks Fuse TV, using a predictive model where a television audience can trade virtual stocks of artists, videos, and songs.[26]

This short history is by no means comprehensive. There are many organizations, for-profit and non-profit alike, that are actively engaged in employing predictive models for the benefit of the organization as well as their customers and clients. In fact, predictive models are now so pervasive that it would be unreasonable to assume that any Fortune 500 companies do not use predictive models in some way or another. It is probably correct to say that most major corporations have either developed their predictive model or subscribe to a predictive model from a third party.[27]

HOW DO PREDICTIVE MODELS WORK?

Predictive models assess the likelihood that the data in a different sample will exhibit the same characteristics of the sample that was used in estimating the coefficients in the predictive model.[28] The units of data with known attributes are referred to as the "training sample," whereas the units in other samples are known as "out of [training] samples."[29] For example, if one is studying soups, the training sample would be the soups

[23] Elnat Ronen, Kenneth Burns, & Ajay Chandramouly, *Using Big Data Predictive Analytics to Optimize Sales*, Intel IT, August 2013,

https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/using-big-data-predictive-analytics-to-optimize-sales-paper.html (last visited July 20, 2017).

[24] *Google Analytics 360 + Google BigQuery for Predictive Digital Marketing*, Google Cloud, n.d.,

https://cloud.google.com/solutions/google-analytics-bigquery (last visited July 20, 2017).

[25] *What Does Digital Transformation Mean for Oil & Gas?*, General Electric, n.d.,

https://www.ge.com/digital/industries/oil-gas   (last visited July 20, 2017).

[26] *The Entertainment Market*, Hollywood Stock Exchange, n.d., https://www.hsx.com (last visited July 20, 2017).

[27] Wayne W. Erickson, *Making Predictive Analytics Pervasive*, Erickson Group, February 14, 2014,

https://www.eckerson.com/articles/making-predictive-analytics-pervasive-deployment-trends (last visited July 20, 2017).

[28] Hal Kalechofsky, *A Simple Framework for Building Predictive Models*, M-Squared Consulting, September, 2016,

http://www.msquared.com/wp-content/uploads/2017/01/A-Simple-Framework-for-Building-Predictive-Models.pdf (last visited July 20, 2017).

[29] *Subject: What are the population, sample, training set, design set, validation set, and test set?*, Neural Networks FAQ,

from known producers, such as soups from Campbell's Soup, etc., while the out of [training] sample would be soups from companies that are entering the market, such as firms that produce organic soups.

Types of Models

Predictive analytics is an area of statistics that is concerned with obtaining information from data and then exploiting it to predict trends and behavior patterns. There are three types of models employed by predictive analytics – descriptive models, decision models, and predictive models.[30] Predictive analytics can be used to analyze an unknown that occurred in the past, is occurring now, or will occur in the future.[31] For example, predictive analytics can be employed in identifying a suspect to a crime after the crime has been committed.[32]

Descriptive Models

Descriptive models quantify relationships in data that classify individuals into groups.[33] Descriptive models do not predict the behavior of a single person, nor do descriptive models rank-order member of a group by the likelihood of acting in a specific way.[34] Descriptive models are used to simulate the behavior of a vast number of individuals, predicting their behavior as a group.[35]

Decision Models

Decision models describe the relationship between the known data and the elements of the decision.[36] Decision models are used to predict the results of decisions when there are many variables to consider.[37] Decision models

---

n.d., ftp://ftp.sas.com/pub/neural/FAQ.html#A_data (last visited July 20, 2017).

[30]STEVEN FINLEY, PREDICTIVE ANALYTICS, DATA MINING AND BIG DATA.MYTHS, MISCONCEPTIONS AND METHODS (Palgrave Macmillan, 2014).

[31]Id.

[32]Id.

[33]Margaret Rouse, *Descriptive Modeling*, WhatIs.com, June, 2012,

http://whatis.techtarget.com/definition/descriptive-modeling (last visited July 20, 2017).

[34]Id.

[35]Id.

[36] Barbara von Halle & Larry Goldberg, *What is the Decision Model?*, IT Performance Improvement, n.d.,

http://www.ittoday.info/ITPerformanceImprovement/Articles/2011-03VonHalleGoldberg.html (last visited July 20, 2017).

[37]Id.

are typically used in developing decision logic trees so that a company knows the outcome for every customer or circumstance.[38]

Predictive Models

Predictive models are used when the amount of data is so vast and complex that employing 4GL database management tools (e.g., Structured Query Language (SQL)) are both awkward and time-consuming.[39] The size of the data set can be in hundreds, if not thousands or even hundreds of thousands of terabytes, where the sheer volume of information is overwhelming, boggling the imagination.[40]

Techniques Employed

Regression Analysis

Predictive models use two methods or approaches – regression analysis or machine learning technologies. The regression techniques are as varied as statisticians. There are linear regressions, weighted regressions, principles of components, discrete choice models, logistic regressions, logit and probit regressions, time-series regressions in all of its variations (e.g., Box-Jenkins), and maximum likelihood estimation, along with an analysis of variance and all of its mutations.[41] Predictive models may at times employ non-parametric techniques such as adaptive regression splines.[42]

Machine Learning

In contrast, machine learning techniques are a branch of artificial intelligence and werecreated to teach computers to learn.[43] Although machine learning techniques include statistical regression methods, in some applications, the technology can be employed to predict the value of the dependent variable without analyzing the relationships among the independent variables.[44] In essence, machine learning techniques mimic human cognition.[45] The methods involved in machine learning include neural networks, multilayer perceptron, radial

---

[38]Id.

[39]*Predictive Modelling*, Investopedia, n.d., http://www.investopedia.com/terms/p/predictive-modeling.asp (last visited July 22, 2017).

[40]Id.

[41]A. KOUTSOYIANNIS, THEORY OF ECONOMETRICS (Macmillan Press 2nd ed. 1977) (1973) and THOMAS J. SARGENT, MACROECONOMIC THEORY: ECONOMIC THEORY, ECONOMETRICS AND MATHEMATICAL ECONOMICS (Academic Press 1979).

[42]KENNETH P. MURPHY, MACHINE LEARNING: A PROBABILISTIC PERSPECTIVE(Adaptive Computation and Machine Learning series The MIT Press 2012).

[43]Id.

[44]TOM MITCHELL, MACHINE LEARNING (McGraw-Hill, 1997).

[45]Murphy, *supra.*

basis functions, support vector machines, naïve Bayes conditional probability rules, and nearest neighbors algorithms.[46]

WHAT ARE PREDICTIVE MODELS?

The *Foundation* series is a sequence science fiction books written by Isaac Asimov, an American author. The premise of the series is that a mathematician, Hari Seldon, developed a branch of mathematics called psychohistory.[47] By using the laws of mass action, psychohistory was able to predict the future on a large scale.[48] Although there is no such discipline as psychohistory, several mathematicians have had similar ideas in attempting to predict the future in the large by using Brownian motion from physics as underpinnings.[49] Predictive models are one such attempt but from a statistical perspective.

Predictive Model Assumptions

The fundamental premise of predictive models is the Efficient Market Hypothesis, which assumes that all of the information about a good is contained in its price.[50] Surowiecki observed that the three necessary conditions for collective wisdom include:[51]

- Informational diversity,
- Independence of decisions, and
- Decentralization of organization.

Informational diversity refers to differences in knowledge and perspectives held by individuals of a group.[52] Independence of decisions means that each person in a group makes up his or her mind.[53] Finally, decentralization of organization signifies that the decision-making power is the lowest level or the individual

---

[46]Mitchell, *supra.*

[47]ISAAC ASIMOV, FOUNDATION (Bantam Spectra Books, rev. ed. 1991).

[48]Id.

[49] Dan Crisen, et al., *A mathematician's view on Asimov's psychohistory*, n.d.,

https://rtraba.files.wordpress.com/2014/11/talkinaugural230113.pdf (last visited July 21, 2017).

[50]PAUL KRUGMAN & ROBIN WELLS, ECONOMICS (Worth Publishers 2nd ed. 2009).

[51]Surowiecki, *supra.*

[52] Wincang Zhou, *Is Informational Diversity Really Informational? An Investigation of What and When in*

*Entrepreneurial Team*s, Journal of Marketing Development and Competitiveness, 2013,

http://na-businesspress.homestead.com/JMDC/ZhouW_Web7_3_.pdf (last visited July 20, 2017).

[53]Surowiecki, *supra.*

within a group.[54] If one or more of these three assumptions are violated, then a predictive model is susceptible to erroneous predictions.

Advantages and Disadvantages

A predictive model has the following benefits:
- It can efficiently aggregate vast amounts of information, beliefs, and data;
- It can obtain material and relevant information through a variety of incentives, including financial incentives; and
- It can incorporate new information quickly that are difficult to manipulate.[55]

There are also disadvantages to predictive models. First, if the Efficient Marketing Hypothesis is violated, then the model will not accurately predict the future.[56] Second, if any one one or more of Surowiecki's three elements are broken, then the model will not yield the correct the result.[57] Third, if the steady-state assumption is violated, then a predictive model will forecast an erroneous outcome.[58]

It is this third possibility that needs additional clarification. In statistical regression analysis, when an exogenous shock to a model occurs, it forever changes the values of the independent variables.[59] In economics, an exogenous shock is an unexpected or unpredictable event that either positively or negatively affects an economic system.[60] It is an unpredictable change in a factor that cannot be explained by the economic system.[61] In legal parlance, an exogenous shock is similar to a superseding intervening cause. In law, a superseding intervening cause breaks the chain of causality between the old chain of causality and substitutes the new chain,

---

[54]Id.

[55]Id.

[56]Surowiecki, *supra.*

[57]Id.

[58]Id.

[59] Daniela Campello & Cesar Zucco, *Exogenous Shocks and Misattribution of Responsibility for Economic Performance: Results From Survey Experiments*, Prepared for the Annual Meeting for the American Political Science Association, September 03-06, 2015, http://www.fgv.br/professor/daniela.campello/files/CampelloZucco-APSA2015.pdf (last visited July 21, 2017).

[60]Id.

[61]Id.

where the superseding intervening cause is the first link in the new chain.[62] Graphically, or a linear model of a single variable, would result in the following:
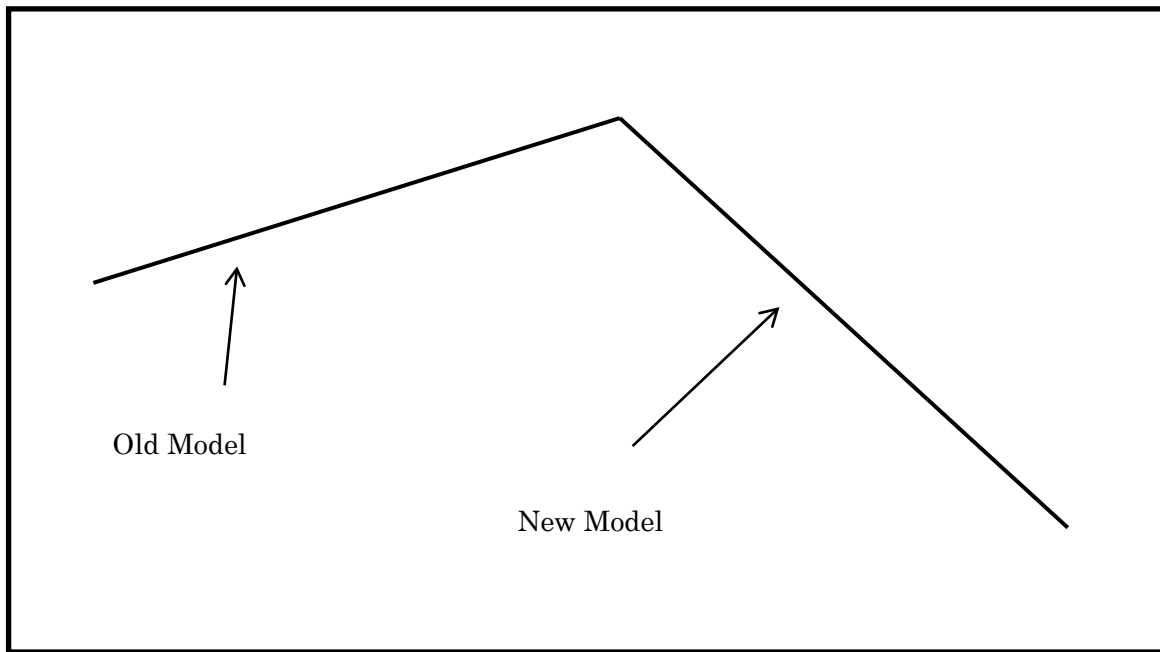

Figure 1 Graphical representation of structural change due to an exogenous shock to an independent or explanatory variable in a linear regression model.

There are statistical techniques to identify a fundamental shift in a regression model. First and foremost, a human researcher must determine the possible existence of a structuralchange in the model. Assuming that the

---

[62]VICTOR E. SCHWARTZ, KATHRYN KELLY, AND DAVID F. PARTLETT, PROSSER, WADE AND SCHWARTZ'S TORTS: CASES AND MATERIALS (Foundation Press 12th ed. 2010).

individual researcher believes that such a fundamentalshift has occurred, he or she can employ a Chow-Test.[63] By independently estimating the coefficients of the two models in *Figure 1*, one can calculate the ratio of the sum of the squares of the residuals, thereby forming an *F*-statistic. If the resulting *F*-statistic is statistically significant at the 95 percent confidence level, then the suspected exogenous shock has structurally changed the model.[64] Contra positively, if the *F*-statistic is not statistically significant at the 95 confidence level, then the suspected exogenous shock was insufficient to change the fundamental characteristics of the model.[65] It may also indicate that the exogenous shock did not exist, and that the researcher's belief was incorrect.[66]

The key factor here is that a researcher has to believe that an exogenous shock to the model has occurred. Left to its own devices, a predictive model is incapable of identifying the existence of an exogenous shock or a superseding intervening cause. This is extremely important because, in the example discussed below, it is the blind reliance on predictive models that resulted in the dramatic failure of the predictive models to predict the outcome of the 2016 United States presidential election.

2016 UNITED STATES PRESIDENTIAL ELECTION

On Tuesday, November 08, 2016, Donald John Trump was elected President of the United States of America. Donald Trump won 306 electoral votes while Hillary Rodham Clinton won 232 electoral votes.[67] Trump won 62,984,825 votes while Clinton won 65,853,516 votes.[68] Clinton won the popular vote by 2,868,691 votes but lost the election due to the constitutionally mandated Electoral College.

Various Predictive Model Results

According to Katz in July 2016, Clinton had a 76 percent chance of defeating Trump and becoming president.[69]Trump had as much chance of winning the presidency as a professional NBA basketball player has of missing a free throw.[70] In the article, the Katz wrote that the *Upshot* model used the voting history of each state

---

[63]Koutsoyiannis, *supra*.

[64]Id.

[65]Id.

[66]Id.

[67]*2016 Election Results*, CNN.Politics, n.d.,http://edition.cnn.com/election/results (last visited July 21, 2016).

[68]Id.

[69] Josh Katz, *Introducing The Upshot's Presidential Prediction Model*, New York Times, July 19, 2016,

https://www.nytimes.com/2016/07/20/upshot/hillary-clinton-has-a-76-percent-chance-to-win-the-presidency.html (last

visited July 21, 2017).

[70]Id.

coupled with approximately 300 national and state polls of the race conducted since mid-April, 2016.[71] Katz further observed that predictive model began with a weighted average in each state, where the polls with the larger sample size were given more weight. Also, the predictive model used an extendedperiod to calculate the averages, meaning that Upshot was stable, being less inclined to place, in their opinion, undue influence on the most recent poll.[72] Upshot believed that it was a mistake to rely too heavily on a week's worth of polling.[73]

On September 20, 2016, The Cthaeh wrote for *Prob(A)billistic World* using a machine learning methodology to predict that Clinton would win the election with a probability of 83.09 percent.[74] The predictive model forecasted that Clinton would win 281 electoral votes to Trump's 257 electoral votes.[75]

In October 2016, Jackson wrote for the *Huffington Post* that the news service uses the HuffPost Pollster database to predict who will become the next President of the United States.[76] The Huffington Post predictive model employed poll averaging by using Pollster's Bayesian Kalman filter model to average publicly accessible polls.[77] The critical word here is "Bayesian," because what it means is that the model used prior probabilities in the prediction of a future event.[78] The issue with using a Bayesian model is that the model assumes a steady-state and that trends exist. In other words, the future resembles the past, or more succinctly, there are no significant exogenous shocks or superseding intervening causes affecting the use and results of the predictive model.

On November 01, 2016, Moody's Analytics model forecasted that Clinton would receive 332 electoral votes, while Trump would get 206 electoral votes.[79] The economic factors in Moody's model include the

---

[71]Id.

[72]Id.

[73]Id.

[74]*Predicting The 2016 US Presidential Election*, The Cthaeh, September 20, 2016,

http://www.probabilisticworld.com/predicting-2016-us-presidential-election/ (last visited July 21, 2017).

[75]Id.

[76] Natalie Jackson, *How We're Forecasting The Presidential Election*, HuffPost, October 03, 2016,

http://www.huffingtonpost.com/entry/forecast-2016-president_us_57ee8eede4b0c2407cdd9155 (last visited July 21,

2017).

[77]Id.

[78]AMIR D. ACZEL & JAYAVEL SOUNDERPANDIAN, COMPLETE BUSINESS STATISTICS (Mc-Graw-Hill 5th ed. 2002) (1989).

[79] Heather Long, *Key model predicts big election win for Clinton*, CNN Money, November 01, 2016,

http://money.cnn.com/2016/11/01/news/economy/hillary-clinton-win-forecast-moodys-analytics/index.html (last visited

July 21, 2017).

two-year change in real household income, real home price growth and gas prices.[80] The political factors are the two-year change in the president's approval rating and political fatigue, where some states switch parties from one presidential election to another.[81] However, Long observed that the model did not account for personalities, only data.[82] In other words, the predictive model did not take into consideration any exogenous shocks or superseding intervening causes.

On election day, Tuesday, November 08, 2016, FiveThirtyEight gave Clinton a 71.4percent chance of winning the election with 302.2 electoral votes.[83] The FiveThirtyEight poll results aggregated from the Huffington Post, Pollster, RealClearPolitics, as well as other polling firms and news reports.[84] According to Silver, there were three versions of this predictive model:

- Polls-plus – It combines polls with an economic index.
- Polls-only – A simpler, what-you-see-is-what-you-get version of the model.
- Now-cast – A projection of what would happen in a hypothetical election held today.[85]

All three versions of the predictive model collected weight and averaged poll numbers, adjusted the poll numbers, combined the poll numbers with demographic and economic data, and accounted for uncertainty and simulated the election thousands of times.[86] All three models depended on data previously collected and ignored the exogenous shock of the Congressional hearings instigated by James Comey, then Director of the FBI, right before the election.[87]

---

[80]Id.

[81]Id.

[82]Id.

[83]*FiveThirtyEight 2016 Election Forecast*, ProjectsFiveThrityEight.com, n.d.,

https://projects.fivethirteght.com/2016-election-forecast/ (last visited July 21, 2017).

[84]Id.

[85] Ron Silver, *A User's Guide To FiveThirtyEight's 2016 General Election Forecast*, FiveThirtyEight, June 29, 2016,

https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/ (last visited ion

July 21, 2017).

[86]Id.

[87] Tim Rostan, *What motivated Comey to announce the reopening of the Clinton email probe after the Weiner laptop was found*, MarketWatch, April 23, 2017,

http://www.marketwatch.com/story/what-motivated-comey-to-announce-the-reopening-of-the-clinton-email-probe-after-the-weiner-laptop-was-found-2017-04-23   (last visited July 21, 2017).

In contrast, Langer and de Jonge employed a multilevel regression using a post stratification technique (MRP).[88] The model consisted of 9,485 interviews from October 20, 2016 to November 06, 2016.[89] The model was based on pre-election polling. It correctly predicted the winner in 49 out of 50 states, including the District of Columbia.[90] The model predicted the wrong result in Michigan, where Trump won by 10,704 votes out of 4.8 million votes cast.[91] The authors postulated that the reasons the other polls were unreliable were because they relied on:

- Poorly executed or poorly timed state polls;
- Non-probability online sample; and
- Less effective pre-election polling techniques

Langer and de Jonge observed that they used a high-quality, random sample data set that was rated "A+" by FiveThirtyEight.[92] Note that Langer and de Jonge only hinted in the first reason that data employed by the other polling organizations was untimely. In other words, the polling organizations relied too heavily on past data, whereas Langer and de Jonge based their results on data that were collected as close to election day as possible.

Donald Trump's Victory and Hillary Clinton's Defeat

What was the exogenous school, or superseding intervening cause that occurred right before the 2016 election? It was the investigation of Clinton's use of a private email server by then FBI Director James Comey.

When Clinton was Secretary of State between 2009 and 2013, she used a private email server for government business.[93] After the news had broken that Clinton was employing a private email server while serving as Secretary of State, in September 2015, FBI investigators were engaged in sorting messages recovered

---

[88]Gary Langer and Chad Kiewiet de Jonge, *The polls didn't predict Trump's win in 2016, but this technique did*, Washington Post, June 13, 2017,

https://www.washingtonpost.com/news/monkey-cage/wp/2017/06/13/this-new-polling-method-predicted-trumps-win-while-we-were-testing-it/ (last visited July 21, 2017).

[89]Id.

[90]Id.

[91]Id.

[92]Id.

[93] Christi Parsons & Michael Memoli, *Hillary Clinton's use of private email not unusual, but still raises questions*, Los Angeles Times, March 03, 2015,

http://www.latimes.com/nation/politics/politicsnow/la-pn-hillary-clinton-emails-20150303-story.html (last visited July 21, 2017).

from the server.[94] The FBI investigation discovered 110 emails that should have been classified when they were sent, and approximately 2,000 emails which were not classified when they were sent, but was then classified after the emails were sent.[95] On July 05, 2016, Comey announced in a written statement that no charges would be filed against Clinton.[96]

In the early part of October 2016, FBI investigators were working on a criminal case involving former Congressman Anthony Wiener, who was being investigated for sending sexually explicit texts to a fifteen-year-old girl.[97] The investigators uncovered emails from his estranged wife, Huma Abedin, who was the vice chair of Clinton's campaign for president.[98] On October 28, 2016, eleven days before the election, Comey sent a letter to Congress, stating that he was re-opening the Clinton investigation.[99] On November 06, 2016, two days before the election, in another letter to Congress, Comey said that the FBI had completed the second inquiry, and found no wrongdoing by Clinton.[100]

---

[94] Katie Bo Williams, *FBI to begin transferring deleted Clinton emails Friday*, The Hill, July 18, 2016,

http://thehill.com/policy/national-security/288190-fbi-to-begin-transferring-deleted-clinton-emails-friday (last visited

July 21, 2017).

[95] Mike Levine, *What FBI Director James Comey Really Said About Hillary Clinton Email Probe*, ABC News August 02,

2016, http://abcnews.go.com/Politics/fbi-director-james-comey-hillary-clinton-email-probe/story?id=41044927 (last

visited July 21, 2017).

[96] Megehan Keneally & Geneva Sands, *FBI Recommends That No Charges Be Filed Against Hillary Clinton*, ABC News,

July 05, 2016,

http://abcnews.go.com/Politics/fbi-completes-investigation-hillary-clintons-personal-email-server/story?id=40346712

(last visited July 21, 2017).

[97] Evan Perez & Pamela Brown, *FBI discovered Clinton-related emails weeks ago*, CNN Politics, October 31, 2016,

http://www.cnn.com/2016/10/30/politics/clinton-emails-fbi-abedin/index.html (last visited July 21, 2017).

[98]Id.

[99] Evan Perez & Pamela Brown, *Comey notified Congress of email probe despite DOJ concerns*, CNN Politics, October

30, 2017,

http://www.cnn.com/2016/10/28/politics/fbi-reviewing-new-emails-in-clinton-probe-director-tells-senate-judiciary-committee/ (last visited July 21, 2017).

[100]Id.

The damage had been done. The exogenous shock or superseding intervening occurred, Clinton lost the 2016 presidential election.[101] Although Clinton won the popular vote by over 2.8 million votes,[102] Trump won 306 electoral votes, enough to win the Presidency, whereas Clinton won only 232 electoral votes.[103]

The Thesis Revisited

To the best of this author's knowledge, the only predictive model to correctly predict that Trump would win the 2016 presidential election was the Langer and de Jonge predictive model. In all of the other predictive models discussed above, the researcher employed data that were collectedbefore the email controversy. When additional data were gathered during the email server controversy, and then added to the predictive models discussed above, the pre-email server data dampened the effect that the controversy would have on the various models' predictions.

The dampening of the effect of new data is common to statistical models that are computing trends. For example, suppose that the data points 3, 5, 6, 4, 5 occur at time periods one through five respectively. The average value is 4.6. Assume that the number 3, the first data point, is removed. Then, assume that the new data point 8 takes its place at the end of the sequence. The new average becomes 5.3. Although the difference in the two averages has increased by 0.7, note that 5.3 is 2.7 away from 8. If all following points after the sixth period are between 7 and 11, it will take someiterations for the moving average to catch up with the exogenous change in the data. On the other hand, if data points 3, 5, 6, and 4are ignored, and only the number 8 is preserved, then 8 is a reasonably decent predictor of future data points between 7 and 11. In the same way, if an exogenous shock or a superseding intervening cause occurs, it best to ignore the previous, data and use only the data that manifests itself after the exogenous shock or superseding intervening cause. This is essentially what Langer and de Jonge did when they predicted the outcome of the 2016 presidential election, and their results speak for themselves.

Why were people so fixated on the predictions being released by the major pollsters that proudly proclaimed that Clinton would win the 2016 presidential election? Simply stated, the pundits et al. either did not appreciate the underlying assumptions of predictive models or did not want to understand the consequences of the multiple email server FBI investigations. After all, it appeared that Clinton would win the 2016 presidential election. Various predictive models proclaimed it to be an inescapable fact. There was apparently no need to visit the predictive model assumptions. We find a statement from Ayn Rand particularly appropriate here; namely, that one can ignore reality, but one cannot ignore the consequences of ignoring reality.[104]

---

[101] 2016 Election Results, *supra.*

[102]Id.

[103]Id.

[104]Ayn Rand, Atlas Shrugged (Signet Press, 1996) (1957).

THE AFTERMATH AND COMMENTARY

Shortly after Trump was sworn into office, he conducted a press conference where he boldly proclaimed that CNN and its mainstream media sisters to be "fake news."[105] It must have been rather embarrassing for mainstream media to eat crow. However, in our opinion, the mainstream media has only itself to blame. It blindly relied on the pollsters to get their predictions right without understanding that when the assumptions of a predictive model are violated, the results can vary dramatically.

Some government officials have suggested that a law should be passed preventing the main stream media from broadcasting so-called "fake news."[106] In other words, the forecasts from predictive models should comply with the actual result and with the law. The problem such a law is that it would probably violate a person's First Amendment rights as well as possibly being void for vagueness.[107] The definition of what constitutes "fake news" would be the issue. On its face, the rule of thumb is elementary – do not print anything that is not true. However, Pontius Pilate's famous question from the New Testament is particularly relevant here – "What is truth?"[108] In our experience, the truth is a challenging thing to ascertain. Truth, like the law, is filled with nuance, and then there is the First Amendment right of Free Speech to consider.

From a statistical perspective to ensure that such a blunder does not happen again. What a researcher has to do is to conduct what is known as a Chow-Test every time that new data is collected.[109] Under the Chow-Test, two regressions have to be calculated. The first calculation involves estimating the value of the coefficients of the regression equation(s) using only data from the past. The second calculation is concerned with estimating the value of the coefficients so of the regression equation employing only the new data. These two regression equations will generate their residual sums of squares. The next step is to form the F-statistic which is the ratio of the sums of the squares. If the value of the F-statistic is statistically significant at the 95 percent confidence level, then this means that the old data should be ignored when estimating the value of the coefficients. On the other hand, if the value of the F-statistic is statistically insignificant at the 95 percent confidence level, then this means that the new data should be aggregated with the old data.[110]

---

[105] Lizzie Dearden, *Donald Trump accuses CNN of 'fake news' after three reporters resign over retracted Russia story*,

Independent, June 27, 2017,

http://www.independent.co.uk/news/world/americas/us-politics/donald-trump-cnn-fake-news-russia-story-retract-three-r

eporters-resign-anthony-scaramucci-a7810071.html (last visited July 21, 2017).

[106] Jane Haskins, *Fake News: What Laws Are Designed to Protect*, Legal Zoom, May, 2017,

https://www.legalzoom.com/articles/fake-news-what-laws-are-designed-to-protect (last visited July 21, 2017).

[107]Id.

[108] John 18:38.

[109]Koutsoyiannis, *supra*.

[110]Id.

The problem is that it is unreasonable to mandate by law that forecasters be required to perform a Chow-Test. There may be more sensitive tests that will be developed by statisticians as time progresses. Also, even if forecasters were required by law to perform a Chow-Test before disclosing their findings, it is possible that the Chow-Test will indicate that the two data sets could be combined, but yet prediction could still be wrong.[111] There is no magic legal wand to be waived over predictive models, thereby ensuring that a modelprovides accurate forecasts.

In conclusion, we are not always sure what truth is, but we arecertain what truth is not. In our opinion, the truth is a balancing act. In logic, one weighs the truth of the premises against the truth of the conclusion. In statistics, one considers the value of the assumptions against the value of the predictions. In legal arguments, one balances the facts against the law. But most of all, in reality, things unanticipated and unexpected things can happen. As humans, we must be smart enough to recognize when an exogenous shock or a superseding intervening cause occurs and then make the appropriate behavioral adjustments. It is a human thing to do.

---

[111]Id.