[*JCH* 6.1–2 (2020–2021) 277–287]
https://doi.org/10.1558/jch.21156

*JCH* (print) ISSN 2051-9672
*JCH* (online) ISSN 2051-9680

# The Promise and Peril of the Data Deluge for Historians

Gary N. Smith[1]
Pomona College, CA
Email: gary_smith@pomona.edu

**Abstract:** Historical analyses are inevitably based on data – documents, fossils, drawings, oral traditions, artifacts, and more. Recently, historians have been urged to embrace the data deluge (Guldi and Armitage 2014) and teams are now systematically assembling large digital collections of historical data that can be used for rigorous statistical analysis (Slingerland and Sullivan 2017; Turchin *et al.* 2015; Whitehouse *et al.* 2019; Slingerland *et al.* 2018–2019).

The promise of large, widely accessible databases is the opportunity for rigorous statistical testing of plausible historical models. The peril is the temptation to ransack these databases for heretofore unknown statistical patterns. Statisticians bearing algorithms are a poor substitute for expertise.

**Keywords:** Big Data; data mining; HARKing; dimension reduction

## Models and Effects

Empirical models can often be cast as multiple regression equations,

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (1)$$

where $y$ is the dependent variable whose value is predicted from the $k$ explanatory variables $x_i$. Each coefficient $\beta_i$ represents the *ceteris paribus* effects of a change in the associated explanatory variable on $y$. The error term encompasses all of the determinants of $y$ that are not captured by the explanatory variables included in the model. The intercept $\alpha$ is literally the predicted value of $y$ if the value of each explanatory variable were equal to zero, but that situation often represents an incautious extrapolation in that the values of the explanatory variables might never equal 0 in practice. For example, a model for predicting aggregate household spending in England might use national household income and wealth as two explanatory variables, but would not be expected to predict spending during a doomsday scenario in which national income and wealth both go to zero. Instead, the estimated value of $\alpha$ is determined along with the estimates of the variable coefficients $\beta_i$ so as to give the most accurate predictions of $y$, which is commonly measured by the mean square error for the data at hand.

Multiple regression equations are quite general, in that the data might be numerical or categorical and the functional form can be nonlinear as well as linear, and are widely used by researchers in virtually all disciplines. Many machine learning algorithms are essentially multiple regression models.

The power of multiple regression models is that they offer the possibility of estimating the *ceteris paribus* effects of changes in individual explanatory variables without having to do controlled experiments. Economists might be able to estimate the effects of a drop in a nation's income on household spending without having to instigate an economic recession with other variables held constant. Historians might be able to estimate the effects of historical events that happened thousands of years ago and cannot possibly be replicated.

There are, however, several pitfalls to be avoided in the specification and estimation of empirical models (Smith and Cordes 2019). I will discuss one here that seems particularly relevant for historians considering the exploration of digital databases – data mining.

## HARKing

The scientific revolution was fueled by what has come to be known as the scientific method: specify a falsifiable conjecture and then collect data, ideally through a controlled experiment, to test this hypothesis. The development of powerful computers and the availability of vast amounts of data make it tempting to reverse the process by putting data before theory, by exploring data in order "to reveal hidden patterns and secret correlations" (Sagiroglu and Sinanc 2013). After a pattern is discovered, a theory can be concocted to explain the pattern, or it might be argued that theories are unnecessary (Fayyad et al. 1996; Cios et al. 2007; Begoli and Horsey, 2012).

This reversal of the scientific method goes by many names, including data mining, data exploration, knowledge discovery, and information harvesting. What they have in common is the belief that data come before theory. This belief is now gaining a foothold in historical analyses. Kohler (2018) has touted the "glory days" created by opportunities for mining large collections of historical data:

> "By so doing we find unanticipated features in these big-scale patterns with the capacity to surprise, delight, or terrify. What we are now learning suggests that the

> glory days of archaeology lie not with the Schliemanns of the nineteenth century and the gold of Troy, but right now and in the near future, as we begin to mine the riches in our rapidly accumulating data, turning them into knowledge."

The dubious assumption is that correlation supersedes causation, in that surprising statistical relationships represent knowledge discoveries, not temporary, fleeting coincidences.

The placement of data before theory has been labeled HARKing: Hypothesizing After the Results are Known. The harsh sound of the word reflects the dangers of HARKing: it is tempting to believe that patterns are unusual and their discovery meaningful; in large data sets, patterns are inevitable and generally meaningless.

Calude and Longo (2017: 600) prove that large amounts of data necessarily contain a large number of patterns and correlations waiting to be discovered:

> "the more data, the more arbitrary, meaningless and useless (for future action) correlations will be found in them. Thus, paradoxically, the more information we have, the more difficult is to extract meaning from it. Too much information tends to behave like very little information.
>
> If there is a fixed set of true statistical relationships that are useful for making predictions, the data deluge necessarily increases the ratio of meaningless statistical relationships to true relationships."

From a Bayesian perspective, that is, taking into account how empirical evidence affects prior (pre-evidence) probabilities, suppose that 1 out of every 1,000 patterns that might be discovered in a large database is meaningful and the other 999 are meaningless coincidences, and that we use a reliable statistical test that will correctly identify a meaningful pattern as meaningful and a meaningless pattern as meaningless 95 percent of the time. If we use a data mining algorithm to discover patterns at random, the prior probability that a discovered pattern is meaningful is 0.001. If the discovered pattern is statistically significant at the 5 percent level, the posterior probability that it is meaningful is 0.0187. This is higher than 0.001, but hardly persuasive. Our discovered pattern is far more likely to be meaningless than meaningful. If, instead, only 1 of 10,000 patterns in a database is meaningful, the posterior probability that a randomly discovered, statistically significant pattern is meaningful falls to 0.0019.

We don't know how many useless patterns are in any given database, but we do know that with the data deluge, it is a very large number that is getting larger every day, which means that the probability that a randomly discovered, statistically significant pattern is truly meaningful is getting ever closer to 0.

This is the paradox of big data (Smith 2020):

> "It would seem that having data for a large number of variables will help us find more reliable patterns; however, the more variables we consider, the less likely it is that what we find will be useful."

Instead of unleashing a data-mining algorithm on hundreds or thousands or hundreds of thousands of unfiltered variables, it would be better to use human expertise to construct a model and narrow the list of explanatory variables. This is a corollary of the paradox of big data (Smith 2020):

> "The larger the number of possible explanatory variables, the more important is human expertise."

When laboratory experiments are possible, researchers can generate an essentially unlimited amount of fresh data to test data-mined theories. Observational data, in contrast, limit the opportunities for out-of-sample tests. After the 2016 United States presidential election, it was widely reported that a history professor had correctly predicted that Donald Trump would win the popular vote based on a model with 13 explanatory variables (Stevenson 2016). Data mining and overfitting were surely at work and the model's prediction turned out to be incorrect (Hillary Clinton won the popular vote by almost 2.9 million votes). The larger point is that out-of-sample testing is severely limited with only one observation every four years.

Even with more data, data-mining algorithms that go through a cycle of discovering and testing models will inevitably find a model that fits the in-sample and out-of-sample data. Discovering a model that fits 200 observations is more difficult than finding a model that fits 100 observations, but it is still data mining and still subject to the same pitfalls.

The perils of data mining are magnified by *black box* algorithms in which inputs are fed into a computer algorithm that provides output without the user knowing how the output was determined or having any way of assessing whether the results are tainted by logical mistakes, programming errors, or other problems.

Computers cannot distinguish between sensible conclusions and nonsense because algorithms do not have common sense, wisdom, or expertise. Computer algorithms are not intelligent in any meaningful sense of the word in that they literally do not know what words mean or what variables represent (Smith 2018b). If a computer algorithm found a highly statistically significant relationship between Trump tweeting the word *with* and the daily stock price of Urban Tea, a tea product distributer headquartered in China, it would have no way of assessing whether that relationship was meaningful or coincidental (Smith and Cordes 2020). Humans would.

Tosh, Ferguson, and Seoighe (2018) give the example of a complex algorithm that was used to generate missing values in a historical database. One imputation said that 100 CE Cuzco had 62 inhabitants, while its largest settlement had 17,856 inhabitants. Humans would know better.

## Long Waves

Turchin (2003) reported that he had discovered two interacting cycles that correlated with social unrest in Europe and Asia going back to 1000 BC. We are accustomed to seeing recurring cycles in our everyday lives – night and day, planting and harvesting, birth and death – and the idea that societies have long, regular cycles, too, has a seductive appeal to which many have succumbed. It is instructive to look at two examples where we can assess their success with fresh data.

Based on movements of various prices (especially copper and agricultural products), Kondratieff (1925) concluded that economies go through 50–60 year cycles (now called Kondratieff waves or K-waves). The statistical power of most cycle theories is bolstered by the flexibility of starting and ending dates and the co-existence of overlapping cycles, including Kitchin cycles of 40–59 months, Juglar cycles of 7–11 years, and Kuznets swings of 15–25 years (Mandel 1980; Modelski and Thompson 1996; Skwarek s.d.). Kondratieff and Stolper (1935) believed that Kondratieff waves co-existed with both intermediate (7–11 years) and shorter (about 3 1/2 years) cycles. This flexibility is useful for fitting cycles to historical data, but undermines the persuasiveness of the theory, as do specific predictions that turn out to be incorrect.

In the 1980s and 1990s, for example, some K-wave enthusiasts predicted a Third World War in the early 21st Century:

> "The probability of warfare among core states in the 2020s will be as high as 50/50." (Chase-Dunn and Podobnik 1995: 335)

> "I suggest the period around 2000 to 2030 as a "danger zone" for great power war." (Goldstein 1988: 353)

More recently, there have been several divergent, yet incorrect, K-wave economic forecasts:

> "In the second half of 2017 the United States and other more developed countries could experience a new recession… Then we expect the start of a new acceleration

> of global economic growth at the upswing phase of the 6th Kondratieff cycle (2018–2050)." (Akaev, Pantin, and Ayvazov 2009: 1)
>
> "In all probability we will be moving from a 'recession' to a 'depression' phase in the cycle about the year 2013 and it should last until approximately 2017–2020." (Quigley 2012)
>
> "The predicted crisis of the current Kondratieff cycle should take place between 2015 and 2030" (Salum and Vicente 2017)

The Elliot Wave theory is another example of how misleading patterns can be identified in historical data. In the 1930s an accountant named Ralph Nelson Elliot studied Fibonacci series and concluded that movements in stock prices are the complex result of several overlapping waves:

> Grand waves that last for centuries;
> Supercycles that last for decades;
> Regular cycles that last for years;
> Primary waves that last for months;
> Intermediate waves that last for weeks or months;
> Minor waves that last for weeks;
> Minute Waves that last for days;
> Minuette waves that last for hours;
> Subminuette waves that last for minutes.

Elliott (1938) proudly proclaimed that, "because man is subject to rhythmical procedure, calculations having to do with his activities can be projected far into the future with a justification and certainty heretofore unattainable." The theory's complexity gives it the flexibility to fit virtually any set of data, even random coin flips.

After the fact, wave believers can always come up with a wave interpretation, though different enthusiasts often have different explanations. Before the fact, wave theorists often disagree and are often wrong, though they are adept at explaining why (Smith 2018b).

In March 1986, *USA Today* called Elliot-wave enthusiast Robert Prechter the "hottest guru on Wall Street" after a bullish forecast made in September 1985 came true, and reported his forecast that the Dow Jones Industrial Average would rise to 3600–3700 by 1988. The highest level of the Dow in 1988 turned out to be 2184. In October 1987, Prechter said that, "The worst case [is] a drop to 2295," just days before the Dow collapsed to 1739. In 1993 the Dow hit 3600, just as Prechter predicted, but six years after he said it would.

Findings cyclical patterns in historical data is easy. Finding meaningful patterns that have a logical basis and can be used to make accurate predictions is elusive.

## Dimension Reduction

It is daunting to data mine a large digital database for an optimal combination of explanatory variables. Even with only 100 potential explanatory variables, there are more than 17 trillion possible combinations of 10 explanatory variables. With 1,000 potential explanatory variables, there are 2.6 x 1023 combinations of 10 explanatory variables.

Several statistical techniques can be used to reduce this dimensionality, but all are undermined by the pitfalls of data mining. I will give two examples of data-reduction procedures, both of which are used in Turchin *et al*. (2018).

Stepwise regression (Efroymson 1960) adds explanatory variables to a model like Equation 1 one at a time, in each step selecting the variable with the lowest $p$-value, as long as that $p$-value is less than a pre-specified level, such as 0.05. However, the standard statistical tests assume a single test of a pre-specified model and are not appropriate when a sequence of steps is used to choose the explanatory variables (Hurvich and Tsai 1990; Babyak 2004; Hendry and Krolzig 2001). The standard errors of the coefficient estimates are underestimated, which makes the confidence intervals too narrow, the $t$ statistics too high, and the $p$ values too low – which leads to overfitting and creates an unwarranted confidence in the final model. These issues once prompted an educational psychology journal to announce that authors should not bother submitting papers using stepwise regression (Thompson 1995). Yet, the data deluge has prompted many to turn to stepwise regression.

A problem that is more fundamental than inappropriate statistical tests is – as with all data-mining models – the selection of variables on the basis of $p$-values, with no consideration of whether their inclusion makes sense. Models based on coincidences are inherently unreliable. Specifically, models constructed via stepwise regression are likely to include meaningless nuisance variables and to exclude truly meaningful variables, which creates a false sense of the model's usefulness. It might seem that stepwise regression is more useful, the larger the number of potential explanatory variables, but the reality is that the data deluge compounds the failings of stepwise regression (Smith 2018a).

Pearson (1901) and Hotelling (1933, 1936) independently developed principal component analysis, which can be used to create linear combinations of the explanatory variables in Equation 1 that are uncorrelated with each other. Hotelling (1957) and Kendall (1957) recommended replacing the original explanatory variables in a multiple regression model with their principal components. For dimension reduction, principal components regression discards the components with the smallest variances, presuming that components with small variances are of little use in explaining or predicting variations in the dependent variable (Hocking 1976; Mansfield et al., 1977; Mosteller and Tukey 1977). Even if a model like Equation 1 is not formally specified, the interpretation of the retained components implicitly assumes that they are related to something the researcher is trying to gauge.

The problem is that the principal-component weights on the explanatory variables depend solely on the correlations among these variables and are not related in any way to the variable that the components are intended to explain or predict. The most important explanatory variables may have smaller weights than less important variables and even nuisance variables that are entirely unrelated to the dependent variable. Important explanatory variables may even have the wrong signs. For example, a variable that has a positive effect on the dependent variable may, after being weighted in the construction of the retained principle components, be estimated to have a negative effect.

As with stepwise regression, principal components is less effective and more likely to be misleading, the larger is the number of potential explanatory variables (Artigue and Smith 2019).

## Conclusion

The promise of large digital collections of historical data is that an embrace of formal statistical tests can make history more scientific. The peril is the ill-founded idea that useful models can be discovered by unleashing advanced statistical procedures on large databases where meaningless patterns are endemic.

Data-mining algorithms that construct models with little or no human guidance are inherently suspect, and computer algorithms have no effective way of assessing whether the patterns they unearth are truly useful or just transitory coincidences.

While data mining sometimes discovers useful relationships, the data deluge has caused the number of possible patterns that can be spotted relative to the number that are genuinely useful to grow exponentially – which means that the more data that are mined, the more likely it is that what is found will be fortuitous, and of little or no use for understanding the past or predicting the future (cf. Ambasciano 2017; Spinney 2019).

Data are essential for the scientific testing of well-founded hypotheses, and should be welcomed by researchers in every field where reliable, relevant data can be collected. However, the ready availability of plentiful data should not be interpreted as an invitation to ransack data for patterns or to dispense with human expertise. The data deluge makes human common sense, wisdom, and expertise essential.

## Notes

1. Gary N. Smith is the Fletcher Jones Professor of Economics at Pomona College, Claremont, CA. Smith has a long history of research projects debunking dubious uses of data in statistical analysis. He is the author of eight textbooks, seven trade books, nearly 100 academic papers, and seven software programs on economics, finance and statistics. *The AI Delusion* (Oxford University Press, 2018), argues that, in the age of Big Data, the real danger is not that computers are smarter than us, but that we think computers are smarter than us and therefore trust computers to make important decisions for us. His most recent books are *The 9 Pitfalls of Data Science* (Oxford University Press, 2019, winner of the PROSE award for Excellence in Popular Science & Popular Mathematics) and *The Phantom Pattern Problem: The Mirage of Big Data* (Oxford University Press 2020), both co-authored with Jay Cordes.

## References

Akaev A. A., V. I. Pantin, and A. E. Ayvazov. 2009. "Analiz dinamiki dvizheniya mirovogo ekonomicheskogo krizisa na osnove teorii tsiklov." *Doklad na Pervom 31 Rossiyskom ekonomicheskom kongresse*, MGU im. M.V. Lomonosova ["Analysis of The Dynamics of Motion of the Global Economic Crisis on the Basis of the Theory of Cycles." Paper presented at the First Russian Economic Congress, Moscow State University].

Ambasciano, L. 2017. "Exiting the *Motel of the Mysteries*? How Historiographical Floccinaucinihilipilification Is Affecting CSR 2.0." In *Religion Explained? The Cognitive Science of Religion after Twenty-Five Years*, eds L. H. Martin and D. Wiebe,107–22. London and New York: Bloomsbury. https://doi.org/10.5040/9781350032491.ch-009

Artigue, H. M. and G. Smith. 2019. "The Principal Problem with Principal Components Regression," *Cogent Mathematics & Statistics* 6(1): 1622190. https://doi.org/10.1080/25742558.2019.1622190

Babyak, M. A. 2004. "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting In Regression-Type Models." *Psychosomatic Medicine* 66(3): 411–21. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Begoli, E. and J. Horsey. 2012. "Design Principles for Effective Knowledge Discovery From Big Data." Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference. https://doi.org/10.1109/WICSA-ECSA.212.32

Calude, C. S. and G. Longo. 2017. "The Deluge of Spurious Correlations in Big Data." *Foundations of Science* 22(3): 595–612. https://doi.org/10.1007/s10699-016-9489-4

Chase-Dunn, C. and B. Podobnik. 1995. "The Next World War: World-System Cycles and Trends." *Journal of World-Systems Research* 1(1): 1–47. https://doi.org/10.5195/jwsr.1995.40

Cios, K. J., W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan. 2007. *Data Mining: A Knowledge Discovery Approach*. New York: Springer.

Elliott, Ralph Nelson. 1938. *The Wave Principle*, New York: Elliott.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. 1996. "From Data Mining To Knowledge Discovery in Databases." *AI Magazine* 17(3): 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Goldstein, J. 1988. *Long Cycles: Prosperity and War in the Modern Age*. New Haven, CT: Yale University Press.

Guldi, J. and D. Armitage. 2014. *The History Manifesto*. Cambridge: Cambridge University Press. https://www.cambridge.org/core/what-we-publish/open-access/the-history-manifesto https://doi.org/10.1017/9781139923880

Hendry, D. F. and H. M. Krolzig. 2001. *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.

Hurvich, C. M. and C. L. Tsai. 1990. "The Impact of Model Selection on Inference in Linear Regression." *American Statistician* 44(3): 214 –17. https://doi.org/10.2307/2685338

Hocking, R. R. 1976. "The Analysis and Selection of Variables in Linear Regression." *Biometrics* 32(1): 1–49. https://doi.org/10.2307/2529336

Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24: 417–41, 498–520. https://psycnet.apa.org/doi/10.1037/h0071325 https://doi.org/10.1037/h0070888

Hotelling, H. 1936. "Relations Between Two Sets of Variates." *Biometrika* 28(3–4): 321–77. https://doi.org/10.2307/2333955

Hotelling, H. 1957. "The Relations of the Newer Multivariate Statistical Methods to Factor Analysis." *British Journal of Statistical Psychology* 10(2): 69–79. https://doi.org/10.1111/j.2044-8317.1957.tb00179.x

Kendall, M. G. 1957. *A Course in Multivariate Analysis*, London: Griffin.

Kohler, T. A. 2018. "Our Unfinished Agenda (What I Have Learned)." *The SAA Archaeological Record* 18(5): 37–42. http://onlinedigeditions.com/publication/?i=542220&article_id=3236418&view=articleBrowser

Kondratieff, N. D. 1925. *The Major Economic Cycles* (in Russian). Moscow. Translated and published in 1984 as *The Long Wave Cycle*. New York: Richardson & Snyder.

Kondratieff, N. D. and W. F. Stolper. 1935. "The Long Waves in Economic Life." *Review of Economic Statistics* 17(6): 105–15. https://doi.org/10.2307/1928486

Mandel, E. 1980. *Long Waves of Capitalist Development: The Marxist Interpretation. Based on The Marshall Lectures Given at the University of Cambridge, 1978*. Cambridge: Cambridge University Press.

Mansfield, E. R., J. T. Webster, and R. F. Gunst. 1977. "An Analytic Variable Selection Technique for Principal Component Regression." *Applied Statistics* 26(1): 34–40. https://doi.org/10.2307/2346865

Modelski, G., and W. R. Thompson. 1996. *Leading Sectors and World Powers: The Coevolution of Global Economics and Politics*. Columbia: University of South Carolina Press.

Mosteller, F., and J. W. Tukey. 1977. *Data Analysis and Regression: A Second Course in Statistics*. Reading, MA: Addison-Wesley.

Pearson, K. 1901. "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* 2(11): 559–72. https://doi.org/10.1080/14786440109462720

Quigley, C. 2012. "Kondratieff Waves and the Greater Depression of 2013–2020." https://www.financialsense.com/contributors/christopher-quigley/kondratieff-waves-and-the-greater-depression-of-2013-2020. Accessed 10 March 2020.

Sagiroglu S. and D. Sinanc. 2013. "Big Data: A Review." *Collaboration Technologies and Systems (CTS), 2013 International Conference*. https://doi.org/10.1109/CTS.2013.6567202

Salum, F. and P. Vicente. 2017. "The Next Cycle of Capitalism." INSEAD Knowledge. https://knowledge.insead.edu/strategy/the-next-cycle-of-capitalism-5226. Accessed 10 March 2020.

Skwarek, S. n.d. "Kondratieff Wave." CMT Association. Retrieved 2018–12–20. https://cmtassociation.org/kb/kondratieff-wave/.

Slingerland, E. and B. Sullivan. 2017. "Durkheim with Data: The Database of Religious History." *Journal of the American Academy of Religion* 85(2): 312–47. https://doi.org/10.1093/jaarel/lfw012

Slingerland, E., *et al.* 2019. 'Complex Societies Precede Moralizing Gods Throughout World History'". *Journal of Cognitive Historiography* 5(1–2): 124–41. https://doi.org/10.1558/jch.39393.

Slingerland, E. *et al*. 2018–2019. "Historians Respond to Whitehouse *et al.* (2019), 'Complex Societies Precede Moralizing Gods Throughout World History'". *Journal of Cognitive Historiography* 5(1-2): 124-41. https://doi.org/10.1558/jch.39393

Smith, G. 2018a. "Step Away From StepWise," *Journal of Big Data* 5: 32. https://doi.org/10.1186/s40537-018-0143-6

Smith, G. 2018b. *The AI Delusion*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198824305.001.0001

Smith, G. 2020. "Data Mining Fool's Gold.". *Journal of Information Technology*. https://doi.org/10.1177/0268396220915600

Smith, G. and J. Cordes. 2019. *The 9 Pitfalls of Data Science*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198844396.001.0001

Smith, G. and J. Cordes. 2020. *The Phantom Pattern Problem: The Mirage of Big Data*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198864165.001.0001

Spinney, L. 2019. "History as a Giant Data Set: How Analysing the Past Could Help Save the Future." *The Guardian*, 12 November. https://www.theguardian.com/technology/2019/nov/12/history-as-a-giant-data-set-how-analysing-the-past-could-help-save-the-future. Accessed 10 March 2020.

Stevenson, P. W. 2016. "Professor Who Predicted 30 Years of Presidential Elections Correctly Called a Trump Win in September." *The Washington Post*, 8 November.

Thompson B. 1995. "Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply Here: A Guidelines Editorial." *Educational and Psychological Measurement* 55: 525–34. https://doi.org/10.1177%2F0013164495055004001

Tosh, N., J. Ferguson, and C. Seoighe. 2018. "History by the numbers?" *Proceedings of the National Academy of Sciences* 115(26): E5840. www.pnas.org/cgi/doi/10.1073/pnas.1807023115. https://doi.org/10.1073/pnas.1807023115

Turchin, P. 2003. *Historical Dynamics: Why States Rise and Fall*. Princeton and Oxford: Princeton University Press. https://doi.org/10.1515/9781400889310

Turchin, P., et al. 2015. *Seshat: The Global History Databank*. *Cliodynamics* 6(1): 77–107. https://doi.org/10.21237/C7clio6127917

Turchin, P., et al. 2018. "Quantitative Historical Analysis Uncovers a Single Dimension of Complexity that Structures Global Variation in Human Social Organization." *Proceedings of the National Academy of Sciences* 115(2): E144- E144-E151. https://doi.org/10.1073/pnas.1708800115

Whitehouse, H. et al. 2019. "Complex Societies Precede Moralizing Gods Throughout World History." *Nature* 568: 226–29. https://doi.org/10.1038/s41586-019-1043-4.