Economics and econophysics in the era of Big Data

Cheong, Siew Ann

2016

Cheong, S. A. (2016). Economics and econophysics in the era of Big Data. The European Physical Journal Special Topics, 225(17), 3159-3170.

<https://hdl.handle.net/10356/83348>

<https://doi.org/10.1140/epjst/e2016-60131-x>

© 2016 EDP Sciences, Springer-Verlag. This is the author created version of a work that has been peer reviewed and accepted for publication by The European Physical Journal Special Topics, EDP Sciences, Springer-Verlag. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at:
[<http://dx.doi.org/10.1140/epjst/e2016-60131-x>].

Downloaded on 26 Apr 2023 21:28:53 SGT

Economics and Econophysics in the Era of Big Data

Data, Methods, Models, and the Scientific Approach

Siew Ann Cheong^{1,2,a}

¹Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371

²Complexity Institute, Nanyang Technological University, Block 2 Innovation Centre, Level 2 Unit 245, 18 Nanyang Drive, Singapore 637723

Abstract. There is an undeniable disconnect between theory-heavy economics and the real world, and some cross polination of ideas with econophysics, which is more balanced between data and models, might help economics along the way to become a truly scientific enterprise. With the coming of the era of Big Data, this transformation of economics into a data-driven science is becoming more urgent. In this article, I use the story of Kepler's discovery of his three laws of planetary motion to enlarge the framework of the scientific approach, from one that focuses on experimental sciences, to one that accommodates observational sciences, to one that embraces data mining and machine learning. I distinguish between the ontological values of Kepler's Laws vis-a-vis Newton's Laws, and argue that the latter is more fundamental because it is able to explain the former. I then argue that the fundamental laws of economics lie not in mathematical equations, but in models of adaptive economic agents. With this shift in mind set, it becomes possible to think about how interactions between agents can lead to the emergence of multiple stable states and critical transitions, and complex adaptive policies and regulations that might actually work in the real world. Finally, I discuss how Big Data, exploratory agent-based modeling, and predictive agent-based modeling can come together in a unified framework to make economics a true science.

1 Introduction

In this special issue, we are asked to explore the question on whether economics can be a physical science? Other experts will no doubt ask deep questions on what it means for economics to be a science, what makes the physical sciences (as opposed to the life sciences and the social sciences) the ideal to strive towards, and what it means for economics to be a physical science.

Drawing from my experience to foster a regional collaboration on economics and econophysics, I will take a more practical approach in this paper. I will start with the story of Kepler's discovery of the laws that bear his name, and use that to highlight the inner workings of the *scientific approach* or the *scientific method* [1–3]. I will explain that economics, like astronomy and biology, have practices that do not conform to this narrow definition of the scientific approach. In this article, I will give a broader definition of the scientific approach to make it more encompassing, for systems where we are

^ae-mail: cheongsa@ntu.edu.sg

restricted to make observations only (like in astronomy), and where the systems undergo evolutionary changes (like in biology).

Finally, I will explain how this broader definition of the scientific approach is compatible with Big Data, and describe how economics and econophysics can be advanced by taking such a Big-Data approach.

2 Kepler's Laws

Let me start by telling the story of Johannes Kepler (1571–1630). Strongly influenced by Ptolemy's Theory of Harmonics, but convinced that Copernicus's heliocentric model of the Solar System is correct, Kepler came up with the idea of *Harmonices Mundi* (music of the spheres) in 1595 to explain why the orbital radii of the six known planets (Mercury, Venus, Earth, Mars, Jupiter, and Saturn) were what they were.¹ In this *Harmonices Mundi* picture (see Figure 1), Kepler had the idea that the planetary orbital radii are in ratios according to spheres inscribed by a selected nesting of the five Platonic solids (octahedron, icosahedron, dodecahedron, tetrahedron, cube). In modern science, such an idea would be considered ludicrous, but in Kepler's time, scientific ideas were as influenced by aesthetics as they were mechanistic principles.

Figure 1. Kepler's *music of the spheres* idea, where the orbits of the six known planets (Mercury, Venus, Earth, Mars, Jupiter, and Saturn) lie on spheres that circumscribe/inscribe the five Platonic solids (octahedron, icosahedron, dodecahedron, tetrahedron, cube). (This public domain figure originally from Kepler's *Mysterium Cosmographicum* (1596) is taken from Wikipedia.)

I tell this story not because Kepler eventually sold the world his *Harmonices Mundi* hypothesis. Certainly, had he tried to do it, there will be no lack of followers. But Kepler was somehow convinced, unlike others in his time, that beauty is not enough to validate a hypothesis. He sought data to confirm

¹Through computer simulations, we now understand that the Solar System started out as a ball of intergalactic gas [4]. Because of its spin, the ball of gas flattened to a disk. Nonlinear gravitational interactions between different parts of the spinning disk then caused mass accumulation to occur at various orbital radii ranges called *islands of stability*. At this stage of its evolution, the Solar System resembled a magnified version of Saturn's rings. Over billions of years, gravitationally-induced collisions between objects in each ring further aggregated into the planets we see today. Interestingly, we now understand from nonlinear dynamics and chaos that nonlinear resonance occurs when the orbital period T_1 of a small object around a massive body is a simple ratio (like $1/2$, $3/5$, ...) of the orbital period T_2 of a larger object around the same massive body, which eventually ejects the small body from the three-body system. The small object is in an island of stability if T_1/T_2 is irrational, i.e. not a simple fraction. This is exactly opposite to what Kepler had proposed, i.e. the ratios of orbital radii of the planets being simple fractions.

his hypothesis, and was aware that the most precise observation data of planetary motions were those collected by Tycho Brahe (1546–1601). A Danish nobleman, he persuaded the Danish king to finance the building of an observatory on the island on Hven. Brahe believed in a hybrid heliocentric model of the Solar System where the Sun and the Moon orbited a stationary Earth, but all other planets orbited the Sun. To prove this, Brahe tirelessly charted the angular positions of every luminous object in the night sky using his naked eye. By the time he fell out with the new King of Denmark in 1597, he had amassed a huge amount of high-precision astronomical data. According to modern astronomers, Brahe was able to measure the angular position of stars and planets down to between $1'$ and $3'$, using a large mural quadrant [5].

After leaving Hven, Brahe built a new observatory in the outskirts of Prague, but died shortly after it was completed. Kepler, being Brahe's assistant in the last years of the latter's life, inherited Brahe's rich trove of data. Keen to demonstrate his *Harmonices Mundi* idea, Kepler wasted no time fitting Brahe's data to circular planetary orbits. This was not a straight forward task: all observations were made in the reference frame of the Earth, and therefore Kepler needed to convert Brahe's readings on a planet to its position (x, y, z) in the reference frame of the Sun, taking care of retrograde caused by the motion of the Earth relative to the planet. Kepler started with Mars, because its data was the most complete and accurate. In doing so he was in part fortunate, because the Earth's orbit around the Sun is very close to being circular, with an eccentricity of $0.0034 < e < 0.058$. He had a different kind of fortune with Mars, however, because it has an eccentricity of $e = 0.0934$, the second largest among the six planets (Mercury's orbit is the most eccentric, with $e = 0.206$).

Once Kepler applied himself to fitting a circular orbit to the data of Mars, he quickly noticed systematic deviations. Had Kepler worked with a less precise data set, he would have dismissed these as observational errors. However, the deviations were significantly larger than the precision of Brahe's data. Kepler trusted Brahe's observations, and therefore he had to find a different explanation for these deviations. At this point in time, he was no longer concerned with the *Harmonices Mundi*, because he had found a far more interesting problem. After more trials and errors, Kepler fitted Mar's orbit to an ellipse. This was totally unconventional at that time, but Kepler fitted all planetary orbits to ellipses, and obtained better fits than to circles. He therefore concluded without further mechanistic insights that planetary orbits are ellipses, with the Sun at one of the two foci. These has come to be known as *Kepler's First Law*.

The First Law changed how Kepler looked at Brahe's data, leading him to discover further regularities. By drawing a line segment from the Sun to the planet, Kepler found that the area swept out by this line segment is the same for equal time intervals. Kepler had effectively discovered the Principle of Conservation of Angular Momentum, and called it *Kepler's Second Law*. Finally, Kepler also discovered that the square of the orbital period is proportional to the cube of the semi-major axis of the ellipse. This is known as *Kepler's Third Law*.

3 Economics and Astronomy

In many ways, Kepler's discovery of the laws that bore his name is typical in Science. First, Kepler came up with a beautiful idea (music of the spheres based on the five Platonic solids). Second, he seek empirical validation. To do this, he needed to acquire sufficient volume of high-precision data. Third, with enough attention to details he noticed discrepancies in the orbit of Mars. This failure to fit Mars' orbit to a circle was unexpected. But however enamoured to circular orbits in his nested Platonic solids idea, Kepler was honest enough to acknowledge that it was not compatible with empirical data, and magnimous enough to let it go. Fourth, the precise data of Brahe allowed Kepler to convince himself, and sufficiently many others, that planets go around the Sun in elliptical orbits, with the Sun at one of

the two foci. Finally, the discovery of the First Law made Kepler more alert to other possibilities that would otherwise not invite contemplation. This led Kepler to discover his Second and Third Laws.

In this section, let us consider how such a scientific process might play out in the economic realm. For example, many economists continue to talk about the Efficient Market Hypothesis (EMH) [6] because of the aesthetic appeal of (1) equilibrium markets, and (2) all information being incorporated into prices. This can be likened to the *Harmonices Mundi* hypothesis, and like Kepler, economists should aim at seriously testing the EMH against data. This is especially important, since the stylized facts discovered by economists and econophysicists about financial markets in recent years [7] cast serious doubts on the validity of the EMH. Indeed, this is what Malkiel and Fama have tried to do in their 1970 paper, where they tested autocorrelations at different time lags. The lack of autocorrelations is a prediction of the EMH. However, autocorrelations can also be weak for other reasons (for example, the stock market is run by monkeys playing dices). To convince other financial economists, Malkiel and Fama studied the impact of stock splits and other public announcements on stock prices to show that they produce only temporary price changes. Malkiel and Fama did this in the late 1960s, before stock markets were computerized. The data sets they used are small, and also have low temporal resolutions. Today we can easily get much more high-frequency data, but we must be clear which aspects of the EMH to test. In fact, the larger amount of data and higher data frequency have allowed us to discover that stock-market return autocorrelations are strongly momentum driven at intra-day time scales (see for example [8]), and weakly mean-reverting at time scales longer than a day (see for example [9]).

If we think of the EMH as the null hypothesis, then rejecting it after comparison with data means that we need to accept the alternate hypothesis that the market can sometimes be inefficient. After Kepler accepted his First Law as the alternate hypothesis to circular planetary orbits, he went on to discover the Second and Third Laws. In economics, it is possible we can also discover more interesting results after we open our minds to the possibility of inefficient markets. In fact, the EMH also predicts that no one can make money out of efficient markets, that efficient markets should never have asset bubbles, and hence never crashes. These predictions stand in stark contrast to reality, where we find bubbles, then crashes, and many funds making a killing trading in these highly-volatile times. Therefore, if we are to accept that the stock market has multiple stable states (which I will introduce in Section 4), then there are additional hypotheses we can make. First, within the stable states the market is close to being efficient, so that traders do not make much money. Second, we must assume that not all information on the market gets priced, so that the unpriced information can grow (an asset bubble) and eventually drive the market to a critical transition (market crash). Third, we may also ask where this unpriced information resides, and postulate that it is temporarily in the form of strategic interactions between traders. Here, let me clarify that I do not necessarily believe the above statements are true, but they are all plausible hypotheses that can come out from a multiple-stable-state market.

Finally, we know that Kepler's Laws were eventually explained in terms of Newton's Laws (the Principle of Conservation of Angular Momentum can also be thought of as a consequence of Newton's Laws) and Newton's Law of Universal Gravitation. Unlike Kepler's Laws, which are statements summarizing empirical observations (like the stylized facts in finance), Newton's Laws are more fundamental statements about motion. In particular, Newton's Second Law tells us how a massive body will respond to the presence of an external force. Therefore, provided an external force can be identified (in the case of Kepler's Laws this is the gravitational force imparted by the Sun), we will be able to calculate the trajectory of the body, i.e. the position of the body at each point in time. In general, when a set of laws is able to 'explain' another set of laws, the former is regarded as more fundamental than the latter.

????????

Economists would of course like to make statements as strong and as fundamental as Newton's Laws. Some call this *physics envy*, and warn against it (see for example, [10]). For a long time, economists have tried to make economics more mathematical, but have not been able to replicate the success of physics in making quantifiable predictions that pass empirical tests. The reason why physics is able to predict the motions of stars and planets, down to atoms and electrons is that these entities do not have feelings! Even the great Richard Feynman acknowledged, "Imagine how much harder physics would be if electrons have feelings!" In contrast, economics deals with individuals making decisions. These individuals acquire, process, and thereafter act on information. When new information becomes available, particularly those about the decisions of other individuals, an individual can change its course of action. Therefore, the fundamental 'laws' of economics must be written in terms of entities that resemble intelligent computer programs, instead of mathematical equations as is done in physics. These autonomous *agents* can react to the environment, interact with other agents, and in contrast to atoms and molecules, set their own goals based on past, present, and future (expected) information [11–14]. The ability of agents to learn from the past and set up expectations of the future is particularly important.

4 Economics and Biology

As I have suggested in Section 3, the correct mechanistic description of economics should be in terms of agents, and how they make decisions. Indeed, Dooyne Farmer and Duncan Foley argued for this in their 2009 Nature opinion piece [15]. Farmer now leads an Institute of New Economic Thinking (INET) team in the EU-funded Complexity Research Initiatives for Systemic Instabilities (CRISIS) to develop a large-scale agent-based model of the European economy. As early as 1989, the Santa Fe artificial market model [16] has shown us that, when a market or an economy is framed in terms of adaptive agents, it can continually generate novel behavioural patterns that never settle into an equilibrium or a steady state. A market or economy of adaptive agents is a *complex adaptive system*. Like other complex adaptive systems, a market or economy of adaptive agents should also have multiple stable states [17]. For example, economist William Forster Lloyd wrote an essay in 1833 describing the problem of unregulated grazing on common land, where a common good can be depleted by excessive exploitation [18]. He called this a *tragedy of the commons*. A simple model of grazing is [19–21]

$$\frac{dN}{dt} = s + Nf(N) - Nv(N) - pN, \quad (1)$$

where s is the population-independent regeneration rate of the common good $N(t)$, $f(N) = aN + bN^2$ is the per capita population-dependent regeneration rate, $v(N) = c + dN^2$ the per capita population-dependent mortality rate, and p is the grazing rate. For a broad range of parameters a, b, c, d, s , $N(t)$ has two stable states, as shown in Figure 2.

Because of the natural parallel between adaptive agents in economics and adaptive agents in ecology, Robert May and Marten Scheffer have made repeated calls for economists to learn from ecologists [22, 23], and to give markets and economies their due respect as complex systems. Economists have thus far struggled with its inability to predict crises (and be forced to look for exogenous causes each and every time), or to recommend policies that led to unintended, catastrophic outcomes. Once we properly recognize markets and economies as complex adaptive systems with multiple stable states, we will be led naturally to think about endogenous critical transitions (which can be in response to slow external forcing) between stable states. Of course, ecologists are not the only ones who understand critical transitions in complex systems as bifurcations. These are rigorously studied by mathematicians in the past, and now routinely taught to undergraduate students [24–26]. In physics,

Figure 2. The two stable states (blue solid curves) and the unstable state (red dashed curve) of the grazing model in Equation (1).

critical transitions are also known as *phase transitions* or *critical phenomena*. These are also taught to students at the advanced undergraduate [27] or beginning graduate level [28–30].

Close to the critical point where the bifurcation occurs, we understand that the mathematical and physical descriptions are universal. Whenever the bifurcation is the result of a vanishing eigenvalue (of the Jacobian matrix), the critical transition will be preceded by *critical slowing down*. In his 2009 review, Marten Scheffer and his co-workers described how various critical slowing down indicators (growing autocorrelation, growing variance, spectral reddening, slowing recovery rates, ...) can be detected in very different complex systems [31]. They provided more details on how this early warning can be obtained in their 2012 review [32]. Following their prescription, my student and I mapped out the early warnings for a set of coupled critical transitions in the US housing market [33]. At this point, I would like to clarify that not all critical transitions are bad. For example, suppose a grassland have been overgrazed, and has gone into the low-density stable state (see Figure 2). Further suppose that the local government would like to bring the grassland back to the high-density stable state by regulating grazing. It must therefore reduce the grazing to $p < p_1$, where there is only the high-density stable state, and let the grassland recover. Once this is done, the local government can allow the grazing level to increase, and still keep the grassland in the high-density stable state, so long as $p < p_2$. Above p_2 , only the low-density stable state exists. Applying their understanding of multiple stable states and hysteresis, Marten Scheffer and his co-workers successfully brought about a critical transition from a turbid state to a clear state in shallow lakes in the Netherlands [34]. With the same kind of understanding, economists should think of what economic stimuli can be implemented to bring about a critical transition from recession to growth.

5 Economics and Big Data

In most areas of science (physical or biological), data is scarce. Therefore, the standard depiction of the scientific approach is for one or more scientists to contemplate the state of their field, identify knowledge gaps, and come up with hypotheses on what they think happens in these areas. In Kepler's case, his hypothesis was the planetary orbital radii being determined by nested Platonic solids. Once a hypothesis has been put forth, the scientists must extract a list of predictions from the hypothesis. They must then design a series of experiments to test some of these predictions, taking care to en-

????????

sure that their measurements are accurate and precise, and that their experiments are reproducible. Where more than one sample can be tested, reproducibility means that repeated measurements on the same sample produces similar outcomes, and also that measurements on different samples produce similar outcomes. Finally, the empirical results are compared against predictions from the hypothesis, sometimes with the aid of sophisticated statistics. If its predictions agree with the empirical results, the hypothesis is considered to be validated. Otherwise, it is refined or rejected. A hypothesis from which we can derive more and more sophisticated predictions becomes a theory if all or nearly all of its series of predictions are validated in succession. This process is shown schematically in Figure 3.

Figure 3. The scheme of the scientific approach for experimental sciences. Starting with the examination of (a) a small amount of initial data, we come up with a hypothesis (model in figure). We then (b) use the model to make predictions, which must later be tested in experiments. (c) The experiment generates more data, which is compared against predictions of the model, and thereafter (d) used to refine the model. We then use (e) the enhanced model to make more predictions, which (f) generates even more data, and the process of model refinement can be repeated *ad infinitum*.

However, if we go back to the Kepler story, we find that it did not progress according to the standard template described above. First of all, Kepler did not perform any experiments. Secondly, after he started suspecting that Mars’ orbit might be elliptical he no longer had the luxury to request Brahe to perform further observations. In fact, after he came up with the elliptical orbit hypothesis, he went back into Brahe’s data to test it. This is because all the data he needed to check for elliptical orbits has already been collected previously by Brahe. In fact, without any additional data, Kepler formulated and tested what were to become his Second and Third Laws. Because of how rich Tycho Brahe’s data was, Kepler did not have to design new observation procedures to collect new data. All he had to do was to dive into the data repeatedly to validate or reject his hypotheses. This process is shown schematically in Figure 4.

Figure 4. The scheme of a modified scientific approach for observational sciences. Suppose the initial model was created independently of the data, then the first step is to (a) use it to make predictions, and test these predictions against the rich data. If (b) the predictions are validated, then repeat with more predictions from the model. Otherwise, we reject the model. However, if (c) the predictions are partially validated, we refine the model, before testing its predictions against the rich data in (d). Sometimes, while (e) rejecting the model we come up with a radically new model. This represents a paradigm shift. After the paradigm shift, we must then (f) test the predictions of the new model against the rich data.

In the data-scarce version of the scientific approach, after coming up with a hypothesis we need to design experiments in a physical lab to measure specific variables. After testing this hypothesis we may be led to a more sophisticated hypothesis, requiring other variables to be measured. Thus new experiments have to be designed and carried out, when we want to test a new hypothesis. In Kepler's case, Brahe's notebooks were his laboratories, i.e. the rich body of data is some kind of virtual laboratory for hypothesis testing. To test a given hypothesis, Kepler did not need all the details in the records. He only needed part of the record to test one hypothesis, and another part of the record to test another hypothesis.

In this sense, what Brahe collected is Big Data [35–37]. Big Data is characterized by *volume* (how much there is), *velocity* (how quickly it comes), *variety* (how many different types), *variability* (how inconsistent it can be), *veracity* (how accurate it can be), the so-called '5V's'. By the standards of his time, Brahe's data has volume and variety, while Kepler trusts it because of the authority of Brahe. Although it was painstakingly collected, we can nevertheless consider it fast because astronomy theory is slow to digest the data. In systems where experimentation is out of the question, the preferred manner of data collection should be to collect whenever and wherever the observation opportunity arises, and to collect every bit of detail that we can and record them. Indeed, when stock markets were computerized in the 1980s, this is what was done, and economists and econophysicists ended up having tick-by-tick data to play with today. However, this level of detail in data collection is not uniform across all markets. In many markets, transactions remain primarily over the counter, and even though 'over-the-counter' today means they occur online, these transactions are not confined to central exchanges, and we will have to standardize these records later. There may also be missing

?????????

data, and it may be difficult to ascertain whether the transaction went through, because they can be last-minute cancellations offline.

Before I conclude, let me address two important questions. First, in Figure 4, we assume that the initial hypothesis was created independently of the data, and therefore could serve as a good starting point for the modified scientific approach for observational sciences. But what if we are not clever enough to come up with an initial hypothesis? Big Data appealed to businesses and organizations because they are not required to have ready-made hypotheses. Instead, data mining and machine learning techniques will be applied to the Big Data to identify patterns and correlations. Although they will initially be crude, these will be presented to economists and econophysicists as seed hypotheses. I expect that some of these hypotheses will sound plausible, and merit further exploration. Most hypotheses will not make any sense, and can be safely rejected by human experts. For the remaining plausible hypotheses, economists and econophysicists can then work out their predictions, so that we can go back to the Big Data for validation. Hypotheses that can be fully or partially validated at this stage will be refined, before we work out further predictions for a second round of validation. Hypotheses that cannot be validated will have to be rejected. This process is shown schematically in Figure 5. By repeating this procedure many times, the surviving hypotheses will become increasingly sophisticated, and at the same time strongly supported by data. These are then our Kepler's Laws (stylized facts) for economics. Also, in this Big-Data approach to science, the human experts are involved very early, when there are still a large number of simple hypotheses. In principle, we can also make computers learn how the human experts make decisions and generate predictions, so that these two processes can be automated. My vision is for economists and econophysicists to only be involved in this Big-Data scientific approach at a very high level, where decision criterion are not clear enough that we can teach them to machines.

Finally, we do not want to stop at the level of Kepler's Laws, but to proceed to develop Newton's Laws for economics. As I have explained in Section 4, to have this kind of explanatory powers we need to develop agent-based models. Unfortunately, the Big Data that I envisage we will collect reflects the decisions that individuals make, but not how they make them. To be able to specify what rules economic agents follow in our models, we need the framework shown in Figure 6. On the left of Figure 6, we see the empirical laws that we have distilled from repeatedly mining our rich Big Data. On the right of Figure 6, I show two classes of agent-based models (ABMs). The first class of agent-based models, developed using surveys and questionnaires to extract simple rules, are meant for exploration purposes. For these to be useful, they need only have qualitative agreement with all the empirical laws. Like the empirical laws, agent-based models that cannot agree with most of the empirical laws must be rejected. However, these models were built using surveys and questionnaires. If we have to reject an agent-based model, either the rich database is wrong, or the surveys and questionnaires are wrong. The exploratory agent-based models therefore suggest how to improve the surveys and questionnaires, or the rich database. With inputs from the empirical laws and the exploratory agent-based models, we can also develop predictive agent-based models. For these models, every rule used must be experimentally validated, i.e. it can be unambiguously shown that human experiment subjects use these rules. From Figure 6, we see that the designs of such economic experiments will be informed by the empirical laws, as well as the exploratory agent-based models.

6 Conclusions

To conclude, I explored in this article how economics can become a proper science, by adopting a scientific approach that is compatible with the character of the discipline itself, but always having data as the foundation. By telling the story of how Kepler discovered his three laws, I explained how

data, we (a) apply data as our hypotheses. (stops at (a)). For those against the database. If be eliminated (stops actions for the refined eliminated (stops at

astronomy, which
observational sciences,
hypothesis testing
ved on to describe
techniques are used
as play the role of
refine hypotheses
further predictions
. Ultimately, after
significant empirical

Models of economics
of complex adap-

????????

Figure 6. Applying data mining and machine learning techniques to the Big Data leads eventually to empirical laws, but yields no insights on the preferences of economic agents. To build exploratory agent-based models, we can rely on surveys and questionnaires from which we can distill simple rules. On the other hand, to build predictive agent-based models, we need to carry out small economics experiments.

tive agents will allow us to understand and expect the existence of multiple stable states, using which we can explain and predict regime shifts like market crashes and recessions. Closing the circle from data to methods to models, I describe a complementary framework for the development of exploratory agent-based models from surveys and questionnaires, as well as predictive agent-based models from small economics experiments, to identify those models capable of explaining all the empirical laws we distill from the data. Exploratory models and theories also have a part to play in the physical sciences, although economists and social scientists are more familiar with the most predictive models and theories in physics. For example, the Standard Model of high energy physics very successfully explains myriad phenomena associated with elementary particles, but very early on theoretical high energy physicists not satisfied with the Standard Model have explored the idea of supersymmetry, where bosons and fermions are treated on equal footing. Even though there is no experimental evidence of supersymmetry as yet, such theorizing has helped high energy physicists understand the Standard Model at a deeper level, partly because of all the experimental directions that supersymmetry suggests we look. I believe the same kind of synergy can also happen in economics, when it is approached from the Big Data, exploratory modeling, and predictive modeling directions.

References

- [1] B. Gower, *Scientific Method: A Historical and Philosophical Introduction* (Routledge, New York, USA, 1997).
- [2] H. G. Gauch, Jr., *Scientific Method in Practice*, (Cambridge University Press, Cambridge, UK, 2003).

- [3] S. S. Carey, *A Beginner's Guide to Scientific Method, 4th edition*, (Wadsworth, Boston, USA, 2011).
- [4] M. Duncan, T. Quinn, and S. Tremaine, "The formation and extent of the Solar System Comet Cloud", *The Astronomical Journal* **94**(5), 1330-1338 (1987).
- [5] W. G. Wesley, "The accuracy of Tycho Brahe's instruments", *Journal for the History of Astronomy* **9**, 42-53 (1978).
- [6] B. G. Malkiel and E. F. Fama, "Efficient capital markets: a review of theory and empirical work", *The Journal of Finance* **25**(2), 383-417 (1970).
- [7] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues", *Quantitative Finance* **1**(2), 223-236 (2001).
- [8] S. L. Heston, R. A. Korajczyk, and R. Sadka, "Intraday patterns in the cross-section of stock returns", *The Journal of Finance* **65**(4), 1369-1407 (2010).
- [9] J. M. Poterba and L. H. Summers, "Mean reversion in stock prices: evidence and implications", *Journal of Financial Economics* **22**(1), 27-59 (1988).
- [10] A. W. Lo and M. T. Mueller, "Warning: physics envy may be hazardous to your wealth!", SSRN preprint, 12 Mar 2010. Available at SSRN: <http://ssrn.com/abstract=1563882> or <http://dx.doi.org/10.2139/ssrn.1563882>.
- [11] M. W. Macy and R. Willer, "From factors to actors: computational sociology and agent-based modeling", *Annual Review of Sociology* **28**, 143-166 (2002).
- [12] E. Bonabeau, "Agent-based modeling: methods and techniques for simulating human systems", *Proceedings of the National Academy of Sciences of the United States of America* **99**(3), 7280-7287 (2002).
- [13] S. F. Railsback, S. L. Lytinen, and S. K. Jackson, "Agent-based simulation platforms: review and development recommendations", *Simulation* **82**(9), 609-623 (2006).
- [14] L. An, "Modeling human decisions in coupled human and natural systems: review of agent-based models", *Ecological Modeling* **229**, 25-36 (2012).
- [15] J. D. Farmer and D. Foley, "The economy needs agent-based modelling", *Nature* **460**, 685-686 (2009).
- [16] R. G. Palmer, W. B. Arthur, J. H. Holland, B. LeBaron, and P. Tayler, "Artificial economic life: a simple model of a stockmarket", *Physica A* **75**(1-3), 264-274 (1994).
- [17] P. Petratis, *Multiple Stable States in Natural Ecosystems*, (Oxford University Press, Oxford, UK, 2013).
- [18] W. F. Lloyd, *Two Lectures on the Checks on Population*, (Oxford University Press, Oxford, UK, 1833).
- [19] R. M. May, "Thresholds and breakpoints in ecosystems with a multiplicity of stable states", *Nature* **269**, 471-477 (1977).
- [20] I. Noy-Meir, "Stability of grazing systems: an application of predator-prey graphs", *Journal of Ecology* **63**(2), 459-481 (1975).
- [21] P. Petratis, "Multiple stable states and relationship between thresholds in processes and states", *Marine Ecology Progress Series* **413**, 189-200 (2010).
- [22] A. G. Haldane and R. M. May, "Systemic risk in banking ecosystems", *Nature* **469**, 351-355 (2011).
- [23] S. Battiston, J. D. Farmer, A. Flache, D. Garlaschelli, A. G. Haldane, H. Heesterbeek, C. Hommes, C. Jaeger, R. May, M. Scheffer, "Complexity theory and financial regulation", *Science* **351**, 818-819 (2016).

- [24] D. Ruelle, *Elements of Differentiable Dynamics and Bifurcation Theory*, (Academic Press, San Diego, USA, 1989).
- [25] J. K. Hale and H. Kocak, *Dynamics and Bifurcation*, (Springer-Verlag, New York, USA, 1991).
- [26] M. Demazure, *Bifurcations and Catastrophes: Geometry of Solutions to Nonlinear Problems*, (Springer-Verlag, Berlin-Heidelberg, Germany, 2000).
- [27] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*, (Oxford University Press, Oxford, UK, 1992).
- [28] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, (Oxford University Press, Oxford, UK, 1971).
- [29] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group*, (Addison-Wesley, Reading, USA, 1992).
- [30] H. Nishimori and G. Ortiz, *Elements of Phase Transitions and Critical Phenomena*, (Oxford University Press, Oxford, UK, 2011).
- [31] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara, “Early-warning signals for critical transitions”, *Nature* **461**, 53-59 (2009).
- [32] M. Scheffer, S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. van de Koppel, I. A. van de Leemput, S. A. Levin, E. H. van Nes, M. Pascual, and J. Vandermeer, “Anticipating critical transitions”, *Science* **338**, 344-348 (2012).
- [33] P. L. J. Tan and S. A. Cheong, “Critical slowing down associated with regime shifts in the US housing market”, *European Physical Journal B* **87**, 38 (2014).
- [34] M. Scheffer, S. H. Hosper, M.-L. Meijer, B. Moss, and E. Jeppesen, “Alternate equilibria in shallow lakes”, *Trends in Ecology & Evolution* **8(8)**, 275-279 (1993).
- [35] A. McAfee and E. Brynjolfsson, “Big Data: the management revolution”, *Harvard Business Review*, 61-68, October 2012.
- [36] S. Sagioglu and D. Sinanc, “Big Data: a review”, *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems* (San Diego, USA, 20-24 May 2013), 42-47 (2013).
- [37] M. Chen, S. Mao, and Y. Liu, “Big Data: a survey”, *Mobile Networks and Applications* **19(2)**, 171-209 (2014).