# PART IV

# MODELLING AND STATISTICAL ANALYSIS OF EMPIRICAL DATA

*Professor Bill McKelvey*

Phase 1 of complexity science began back in the 1950s when Ilia Prigogine started writing about phase transitions just beyond the '*edge of order*' (the *1st critical value*) in physical systems resulting from the reduction of imposed tensions (Prigogine 1955, 1962, 1997; Prigogine and Stengers 1984; Nicolis and Prigogine 1989). Starting in the late 1980s, scientists at the Santa Fe Institute of Complexity Sciences shifted complexity science to the study of emergent behaviour in living biological and social systems (Anderson 1972; Gell-Mann 1988, 2002; Holland 1988, 2002; Kauffman 1993; Arthur 1994).

Phase 2's focus was on heterogeneous agents interacting just before the '*edge of chaos*' (the *2nd critical value*) of imposed tensions. In between the '*edges*' of *order* and *chaos* is the *Region of Emergent Complexity*, what Kauffman terms the '*melting*' zone (1993, p. 109). Bak (1996) argued that to survive, organisms have to have a capability of staying within the melting zone, maintaining themselves in a state of '*self-organized criticality*,' that is, adaptive efficacy. Holland (2002) defined emergent phenomena as *multi-level hierarchies*, *intra- and inter-level causal processes*, and *nonlinearities*.

Phase 3 focuses on nonlinearity, scalability, power laws (PLs), the 'butterfly effect',[1] scalability and fractals. Though beginning many decades ago with Pareto (1897), Auerbach (1913) and Zipf (1949), Phase 3 re-focused attention to PL phenomena (Newman 2005; Andriani and McKelvey 2007, 2009), and eventually includes econophysics (West and Deering 1995, Mantegna and Stanley 2000). Econophysics began with Benoit Mandelbrot's focus on stock market crashes (1963). While crashes are negative extreme events, their showing of the PL signature indicates that the markets were free to go up or down without restraint. PLs often appear as indicators of self-organization, emergence-in-action, self-organizing economies (Krugman 1996), and the growth of firms, cities and economies (Stanley et al. 1996; Axtell 2001).

Various descriptions of how complexity science has been applied to organizations appear in Maguire et al. (2006) Hazy et al. (2007), Allen et al. (2011), Strathern and McGlade (2014). These existing views and theories about how complexity science thinking and concepts apply to organizations and management are more explicitly tested and elaborated in the chapters comprising this Section.

Crawford and McKelvey (Chapter 12) focus on statistical tests of whether power laws (PLs) actually exist in industry distributions. Especially important is their use of the $x_{min}$ concept to identify the point where normally-distributed phenomena shift into PL-distributed phenomena as some firms start growing out toward the 'stochastic frontier' (Aigner et al. 1977; Koop et al. 1999; Kumbhakar and Knox Lovell 2000; Lensink and Meesters 2014; Ravishankar and Stack 2014). Brown et al. (Chapter 13) focus on multifractal empirical analyses of 'web-traffic' connections to gender terms used in Wikipedia. Dister et al. (Chapter 14) conduct an empirical analysis of multiple cases (power-grid segments in the US state of Ohio) that uses both qualitative, quantitative, and statistical methods available in the SACS Toolkit created by Castellani and Rajaram (2012). Wolf-Branigin et al. (Chapter 16) use two NetLogo (Wilensky and Rand 2015) agent-based computational models (ABMs) to study the relative impacts of policy alternatives such as 'respite care, tax incentives, work-place policies and adult day-care services' on caregiver stress. Hazy and Wolenski (Chapter 15) use a canonical mathematical model – Goldstein et al.'s *Cusp of Change* model (2010) – to study the impact of degrees of freedom and internal and external complexity conditions on human interaction dynamics. Pourbohloul et al. (Chapter 17) focus on the use of complexity-relevant mathematical network modelling methods that would significantly improve the management of emerging disease outbreaks here and there around the world.

Crawford and McKelvey (Chapter 12) begin by explaining the basic causes of skew distributions. "These processes are generated by scale-free mechanisms – the same cause at multiple levels of analysis – that result in self-similar fractal structures, that is, power-law distributions (PLDs) within firms and other social entities". They then briefly describe how Bak's (1996) 'self-organized criticality' can be "used to explain how and why social and organizational entities maintain viable new-order creation as they coevolve" with competing firms. This discussion provides context for their data analysis and subsequent interpretation of the results. The authors use a bootstrapped maximum likelihood estimation (MLE) technique developed by Clauset et al. (2009) using three longitudinal datasets of entrepreneurial firms. They use "these variables and samples because any hypothesis about any mechanism driving the emergence of PLDs in a research domain requires a significant empirical finding (that is, the distribution is actually a PLD)". They "then draw conceptual links from the individual parameters of the MLE to the generative mechanisms". Most specifically, they identify the "critical point in a distribution where systems change from linear to nonlinear". They explain how each of the parameters generated by their statistical techniques can be interpreted via complexity science concepts.

Brown et al. (Chapter 13) begin by using gender studies – such as black versus white or male versus female – to illustrate how the traditional focus on binary examples supports the traditional deeply ingrained desire for logical formalisms and conceptually dynamic models of systems. The traditional way of avoiding math-driven simplifications was to use case studies so as to be able to describe the richness of human and social realities. In contrast, the authors focus on what is now termed 'intersectionality theory', which recognizes that many people have complex multiple memberships in various biological, behavioural, and social categories. They introduce multifractal analysis as a means for studying 'cascade dynamics', for example, defined as how DNA becomes more differentiated and complex as a family grows from two original parents to many distantly-related offspring. The authors suggest that "cascade dynamics and multifractal analysis can provide a

logical formalism and statistical framework to make intersectionality a quantitatively tractable" representation of gender development. They review recent cognitive-science advances in which multifractal analysis identified key features of the cascades affecting cognitive performance. Using "web-traffic data for gender terms on Wikipedia", they demonstrate how similar cascade structures in gender dynamics develop by using multifractal analysis. They conclude by proposing that "cascade formalisms and multifractal analysis may provide new avenues for gender studies that balance both logical formalisms and dynamic concepts".

Dister et al. (Chapter 14) observe that "if one is to improve reliability and resilience in infrastructures . . . it is necessary to adopt a 'complex, smart territory' modelling strategy, particularly one that gives particular attention to the importance of social complexity. To test the veracity" of their argument, they "conduct a case study on a segment of the United States power grid". Their goal is simple: they "seek to create a first proof-of-concept sufficient to show, in the simplest of cases, how thinking about infrastructures in 'complex systems' terms, primarily in terms of their social aspects, can prove beneficial". For their case study, they use "the SACS Toolkit: a new method for quantitatively modeling complex social systems, based on a case-based, computational approach to data analysis" (quote taken from Castellani and Rajaram, 2012: p. 153), "which is part of the new approach to modelling complex systems, called *case-based complexity*. As a technique, the SACS Toolkit is a computationally grounded, case-comparative, mixed-methods platform for modelling complex systems as sets of cases." It integrates "case-based reasoning with complex critical realism and the latest developments in complexity theory". They provide a "basic overview of their research process, ending with a summary of novel insights".

Hazy and Wolenski (Chapter 15) advocate a mathematical research agenda to categorize a theoretic representational framework to 'connect all scientific approaches to human interaction dynamics (HID), which seek to "build a cumulative base of knowledge to inform individual choice and behavior". They illustrate their approach by using the *Cusp of Change* canonical mathematical model (Goldstein et al., 2010) that describes the potential for first order-phase transitions. They suggest that the "degrees of freedom that are active in the action orientation of a stable organizing state may be a useful order parameter and further that indexes reflecting the internal and external complexity conditions that are confronting the population can be used as control parameters". Their approach takes the position that action orientation gathers its potency via "gradients of differences in predictability and uncertainty regarding relevant events. As agents seek to reduce the unpredictability and thus their cognitive load by moving along this gradient, informational-influence forces set up a potential field under which organizing for action occurs among independent semi-autonomous agents. By using the information in socio-technical structure that can be recognized and decoded, agents can ease their cognitive load while maintaining a perceived level of predictability about events in the environment."

Wolf-Branigin et al. (Chapter 16) note that social planners "create innovative interventions to address the diverse and expanding needs of vulnerable populations. Over time, these social innovations require an increasing high level of flexibility and adaptability to remain effective in addressing the continually changing, disparate, and incompatible preferences of clients, funding sources, and other stakeholders". The authors apply

complexity science as a social programme evaluation methodology and then test pro-gramme and policy options using agent-based models. They then discuss how complex adaptive systems (CAS), and its related components are beneficial to social programme evaluators and researchers. Their use of complex adaptive systems (CAS) facilitates bridg-ing complexity science to social programme evaluation; several characteristics of complex adaptive systems have direct implications to programme evaluation, which "include non-linearity, emergence, being adaptive, having uncertainty in that estimating values cannot be exact, and being dynamical and co-evolutionary". The authors use ABMs to evaluate social programmes framed as complex adaptive systems. Foremost is the ability to fore-cast outcomes (represented as emergent behaviours) over time because the running of an ABM assumes the presence of an iterative process. They note that problems remain, however, in sufficiently matching modelling schemes and social realities. The ABM approach appears to function well when the individual agents are empowered and have the ability to make choices from the programmes from which they receive information and services.

Pourbohloul et al. (Chapter 17) use mathematical network models to describe various kinds of contacts among people that can cause quickly spreading disease outbreaks here and there around the world that are a special concern of health care providers and public health officials. They also note that further research is needed to better understand how infectious-disease spreads occur in hospitals, so as to better identify which individual(s) or group(s) of healthcare workers in hospitals who are most likely to foster the spreading of infectious diseases, for example, trainees, senior clinicians and so on. Strategies for more effectively requiring self-protection for these highly net-work-connected individuals (for example, wearing surgical masks during their working hours, wearing disposable gloves, and washing hands more often) may effectively reduce disease spread in hospitals. By incorporating the heterogeneous behaviour of individuals as well as the heterogeneity in disease transmissibility, complexity science tools allow them to accurately pinpoint optimal and economical intervention strategies, thus contributing to evidence-based public health research and decision-making, and also to the health of people worldwide. They emphasize one reassuring fact: the sooner control measures are implemented appropriately, the lower the rate of infection across many populations.

## NOTE

1. The so-called '*butterfly effect*' stems from Lorenz's 1972 paper: 'Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?' These are Holland's (2002) 'tiny initiating events' that scale up to extreme outcomes.

# REFERENCES

Aigner, D.J., Lovell, C.A.K. and Schmidt, P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* **6**(1): 21–37.

Allen, P., Maguire, S. and McKelvey, B. 2011. *The Sage Handbook of Complexity and Management*. London: Sage Publications.

Anderson, P.W. 1972. More is different. *Science* **177** 393–396.

Andriani, P. and McKelvey, B. 2007. Beyond Gaussian averages: Redirecting organization science toward extreme events and power laws. *Journal of International Business Studies* **38** 1212–1230.

Andriani, P. and McKelvey, B. 2009. From Gaussian to Paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, **20**(6): 1053–1071.

Arthur, W.B. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor, MI: University of Michigan Press.

Auerbach, F. 1913. Das Gesetz d er Bevolkerungskoncentration. *Petermanns Geographische Mitteilungen* **59** 74–76.

Axtell, R.L. 2001. Zipf distribution of US firm sizes. *Science* **293** 1818–1820.

Bak, P. 1996. *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus.

Castellani, B. and Rajaram, R. 2012. Case-based modeling and the SACS Toolkit: A mathematical outline. *Computational and Mathematical Organization Theory*, **18**(2): 153–174.

Clauset, A., Shalizi, C.R. and Newman, M.E.J. 2009. Power-law distributions in empirical data. *SIAM Review* **51**(4): 661–703.

Gell-Mann, M. 1988. The concept of the Institute. In D. Pines (ed.), *Emerging Synthesis in Science*. Boston, MA: Addison-Wesley, pp. 1–15.

Gell-Mann, M. 2002. What is complexity? In A.Q. Curzio and M. Fortis (eds), *Complexity and Industrial Clusters*. Heidelberg, Germany: Physica-Verlag, pp. 13–24.

Goldstein, J., Hazy, J.K. and Lichtenstein, B.B. (2010), *Complexity and the Nexus of Leadership: Leveraging Nonlinear Science to Create Ecologies of Innovation*. Englewood Cliffs, NJ: Palgrave Macmillan.

Hazy, J.K., Goldstein, J.A. and Lichtenstein, B.B. (eds) 2007. *Complex Systems Leadership Theory*. Mansfield, MA: ISCE Publishing.

Holland, J.H. 1988. The global economy as an adaptive system. In P.W. Anderson, K.J. Arrow and D. Pines (eds), *The Economy as an Evolving Complex System*, Vol. 5. Reading, MA: Addison-Wesley, pp. 117–124.

Holland, J.H. 2002. Complex adaptive systems and spontaneous emergence. In A.Q. Curzio and M. Fortis (eds), *Complexity and Industrial Clusters*. Heidelberg, Germany: Physica-Verlag, pp. 24–34.

Kauffman, S.A. 1993. *The Origins of Order*. Oxford, UK: Oxford University Press.

Koop, G, Osiewalski, J, and Steel, M. 1999. The components of output growth: A stochastic frontier analysis. *Oxford Bulletin of Economics and Statistics* **6**(1): 455–487.

Krugman, P. 1996. *The Self-Organizing Economy*. Malden, MA: Blackwell.

Kumbhakar, S.C. and Knox Lovell, C.A. 2000. *Stochastic Frontier Analysis*. New York: Cambridge University Press.

Lensink, R. and Meesters, A. 2014. Institutions and bank performance: A stochastic frontier analysis. *Oxford Bulletin of Economics and Statistics* **76**(1): 67–92.

Lorenz, E.N. 1972. Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? Paper presented at the 1972 meeting of the American Association for the Advancement of Science. Washington, DC.

Maguire. S., McKelvey, B., Mirabeau, L. and Öztas, N. 2006. Complexity science and organization studies. In S. Clegg, C. Hardy and T. Lawrence (eds), *Handbook of Organizational Studies* (2nd edition). London: Sage Publications, pp. 165–214.

Mandelbrot, B.B. 1963. The variation of certain speculative prices. *Journal of Business* **36**: 394–419.

Mantegna, R.N. and Stanley, H.E. 2000. *An Introduction to Econophysics*. Cambridge, UK: Cambridge University Press.

Newman, M.E.J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46** 323–351.

Nicolis, G. and Prigogine, I. 1989. *Exploring Complexity: An Introduction*. New York: Freeman.

Pareto, V. 1897. *Cours d'Economie Politique*. Paris: Rouge & Cie.

Prigogine, I. 1955. *An Introduction to Thermodynamics of Irreversible Processes*. Springfield, IL: Thomas.

Prigogine, I. 1962. *Non-Equilibrium Statistical Mechanics*. New York: Wiley-Interscience.

Prigogine, I. and Stengers, I. 1984. *Order Out of Chaos: Man's New Dialogue with Nature*. London: Fontana Paperbacks.

Prigogine, I. and Stengers, I. 1997. *The End of Certainty*. New York: Free Press.

Ravishankar, G. and Stack, M.M. 2014. The gravity model and trade efficiency: A stochastic frontier analysis of Eastern European countries' potential trade. *World Economy* **37**(5): 690–704.

Stanley, M.H.R., Amaral, L.A.N., Buldyrev, S.V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M.A. and Stanley, H.E. 1996. Scaling behaviour in the growth of companies. *Nature* **379** 804–806.
Strathern, M. and McGlade, J. (eds) 2014. *The Social Face of Complexity Science*. Litchfield Park, AZ: Emergent Publications.
West, G.B. and Deering, B. 1995. *The Lure of Modern Science: Fractal Thinking*. Singapore: World Scientific.
Wilensky, U. and Rand, W. 2015. *An Introduction to Agent-based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. Cambridge, MA: MIT Press.
Zipf, G.K. 1949. *Human Behavior and the Principle of Least Effort*. New York: Hafner.