An agent-based system with temporal data mining for monitoring financial stability on insurance markets

Iqbal Owadally[a,*], Feng Zhou[b], Rasaq Otunba[c], Jessica Lin[c], Douglas Wright[a]

[a] *Cass Business School, City, University of London, UK*
[b] *Future of Humanity Institute, University of Oxford, Oxford, UK*
[c] *Department of Computer Science, George Mason University, Virginia, USA*

Abstract

We describe an expert system to monitor the stability of insurance markets. It consists of two components: an agent-based simulation component and a temporal data mining component. Like other financial markets, insurance markets experience destabilizing cycles and suffer episodic crises. The expert system assists market regulators by monitoring the financial position of individual insurers and of the overall market, and by forecasting cycles and impending insolvencies. The agent-based simulation component runs a forward simulation allowing for interaction among insurers in a competitive market, and between insurers and customers. The temporal data mining component extracts useful information for market regulators from the simulations. A prototype of the system is applied to the automobile insurance market. We show how the system may be used to forecast cycles, investigate stability, and analyze insurers' herding behavior on the market. A practical policy conclusion is that regulators should monitor individual insurers' pricing pattern because aggressive price undercutting creates a "winner's curse", with subsequent losses and market instability.

Keywords: Agents, Motif, Anomaly, Cycle, Crisis, Automobile insurance

*Corresponding author. *Address:* Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom. *Email:* `m.i.owadally@city.ac.uk`. *Phone:* +44 (0)20 7040 8478. *Fax:* +44 (0)20 7040 8572.

1. Introduction

The financial crisis of 2008 and the ensuing recession have been the latest illustration of the unstable nature of financial markets. Insurance markets experience similar turbulence, much of it endogenously generated and unrelated to other financial markets. Insurance suffers from cycles in prices, in coverage for customers, and in profitability for insurers. Crises occur when there are deep troughs in the cycles, with exorbitant insurance prices, unavailability of insurance cover, and slow recovery thereafter (Dionne, 2013). Market regulators face a difficult task in attempting to forestall potential problems by acting countercyclically to stabilize markets. In this article, we describe an expert system which can assist insurance market regulators in this task. This is, to the best of our knowledge, the only expert system of its kind to help regulators monitor insurance markets.

The expert system that we propose consists of two components: an agent-based simulation component and a temporal (or time series) data mining component. The expert system can assist any national or state-wide insurance market regulator, such as the Federal Insurance Office in the U.S. Department of Treasury or the Prudential Regulation Authority at the Bank of England. The agent-based sub-system constitutes the knowledge base of the expert system, in that it takes information about the structure of the market with up-to-date statistics about insurers and customers, and projects this forward. It has a visualization tool to help regulators understand how the market is evolving. The regulator can monitor and forecast the performance of individual insurers and of the market as a whole. Simulations can be re-run using different parameter values to determine the effect of policy actions taken by the regulator.

The temporal data mining component of the expert system is the inference engine of the expert system. It takes the information generated within the agent-based component and uncovers patterns about market behavior. This can assist the regulator in detecting insurer insolvency and systemic instability, and in formulating policy to mitigate risks. Agent-based models create bottom-up simulations, starting at an individual agent level, but enabling interactions between heterogeneous agents. Because of this, considerable amounts of simulated data are generated in the form of time series about individual insurers' prices, losses, sales, capital, customer base, underwriting results etc. The data mining component makes sense of streaming time series simulated data from the agent-based simulation component. It can detect anomalies in the pricing or reserving of any of the insurers. For example, insurers may loosen underwriting standards dangerously and cut prices excessively to attract customers, leading to subsequent insurance insolvencies, from which customers lose out. The data mining component can also keep track of the overall market, detecting cycles (periodicities) in various insurance metrics. Since the simulation component is running forward in time, the data mining component is in effect forecasting anomalous insurer behavior and cyclical activity in the market.

There are a number of expert systems which share features with ours. The one which is closest, in terms of motivation, is the agent-based expert system of Streit & Borenstein (2009, 2012). Their system is concerned with the regulation or governance of the overall financial system, particularly the banking sector. By contrast, our system is concerned with the insurance sector only. Their system consists of two components: an agent-based component at the micro level, and a traditional econometric model to capture macroeconomic variables. The system that we describe in this paper also consists of an agent-based component, but we combine it with a temporal data mining component to visualize and analyze the simulation output from the agent-based simulation model. The financial governance system of Streit & Borenstein (2009, 2012) incorporates the monetary authority (central bank) as an agent and seeks to predict interest rate decisions, whereas our system takes the viewpoint of an external market regulator who has oversight of insurance firms for macro-prudential stability and for customer protection. The agents in Streit & Borenstein's (2009) system use a set of if-then fuzzy logic rules, whereas ours behave in accordance with microfoundations in insurance.

Two other systems related to ours are the premium rating expert systems of Lin (2009) and Imriyas (2009). Lin's (2009) decision support tool, based on a back-propagation neural network, helps insurance underwriters set premiums for fire insurance policies. Imriyas's (2009) system uses fuzzy logic to calculate rates for workers' compensation insurance in replacement of traditional insurance rating systems based on the claims experience of individuals working in the construction industry. Neither of these systems looks at the insurance market as a whole, however. Underwriting and actuarial rate-making are part of insurance pricing and hence are central to our system too, since the stability of an insurance market comes down to supply and demand for insurance products.

Also related to insurance pricing is the work of Guelman & Guillén (2014) who develop a causal-inference framework to measure the price-elasticity of demand for automobile insurance. This is designed to supplement actuarial rate-making and help insurers determine the impact on sales of a change in premium rates. In our system, demand-based pricing is also combined with actuarial cost-based pricing. Our insurer-agents use an optimization heuristic for this purpose, as might be realistically employed by insurance underwriters.

Expert systems are also widely used in marketing to insurance customers. Hsu (2011) uses a neural network to classify insurance policyholders by financial risk attitude, and Kaishev et al. (2013) use actuarial credibility-related techniques to cross-sell products to insurance customers. Lin (2010) trains a nonlinear fuzzy neural network on data about insurance customers' personality and emotional traits to capture the relationship between customers and insurers, and predict customers' inclination to switch to another insurer.

To the best of our knowledge, our paper is the first to apply data mining to discover patterns and extract knowledge from agent-based simulations. Data mining algorithms and predictive models are of course used in several other areas of insurance, notably fraud detection and marketing. In marketing, decision trees are used by Wu et al. (2005) and clustering analysis is applied to customers' financial lifecycle decisions by Liao et al. (2009). To detect insurance fraud, Shin et al. (2012) employ a scoring model whereas Šubelj et al. (2011) use social network analysis. Kuo et al. (2007) use association rules mining and clustering analysis on health insurance data. Guelman (2012) proposes a statistical learning technique, gradient boosting trees, to predict insurance claims.

Unlike the earlier literature, this paper is concerned with the stability of insurance markets from the viewpoint of an official market regulator. (Financial stability in general is discussed by Tsomocos (2003), Samitas et al. (2018) and others, but we restrict ourselves to insurance here.) Insurance markets suffer from destabilizing cycles and crises so an expert system that can assist market regulators has the potential to ameliorate the operation of these markets, benefiting customers and the economy generally. In the next two sections, we describe our agent-based simulation of the insurance market and the associated data mining tools. We then validate the system on the automobile insurance market and we show how an insurance regulator can use the system to forecast cycles on the market, to investigate market stability, and to analyze herding behavior of insurers.

2. Agent-Based Simulation Component

2.1. System Structure

The system is executed by first running the agent-based simulation component, which comprises current data about the market and generates multiple simulation sample paths. The simulation data is then fed through to the data mining sub-system which performs pattern discovery. In this section, we provide a step-wise description of the agent-based simulations. The data mining part is described later in section 3.

2.2. Agents and Simulation Landscape

The agent-based system simulates an insurance market, such as the automobile insurance market. Automobile insurance is compulsory in many jurisdictions, so we assume that there is a number M of customers who have to buy insurance at the beginning of every year. There are N insurers who sell automobile insurance policies. Insurers are heterogeneous in the sense that they have different business attributes. Customers are also heterogeneous because they have different preferences, or affinities, over the attributes of their insurance provider. Expert systems used for marketing to insurance customers, such as the ones of Kaishev et al. (2013), Hsu (2011) and Lin (2010), exploit this very heterogeneity. Customers and insurers are self-directing and interacting agents and they live on an n-dimensional simulation landscape, given n relevant business attributes. An example of a 2-d business attribute space is shown in Figure 1. The location of an insurer in this

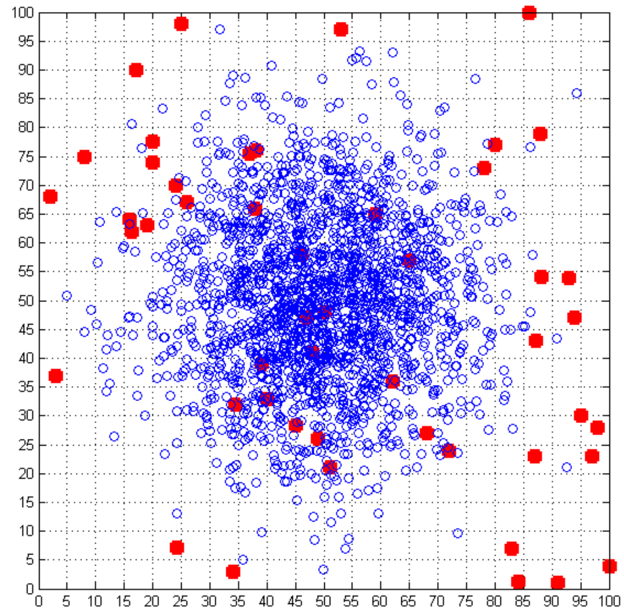

Figure 1: A simulation landscape, or business attribute space, in the agent-based system. Insurers and customers are represented by larger dots and smaller circles respectively. The insurer's business attributes determines its location on the space. The proximity of a customer to an insurer signifies greater affinity by the customer for the insurer's business attributes.

space is determined by its attributes. The proximity of a customer to an insurer on the space measures the affinity that the customer has for the insurer's attributes.

Business attributes may be quantifiable, or they may be non-measurable. Examples of the former include expenditure on advertising, expenditure on internet advertising, number of followers on Twitter, median age of insurer's policyholders, median market value of car or other property insured with the insurer, customer retention rate, number of customer complaints filed with regulator or ombudsman etc. These variables define whether the insurer is targeting customers in a particular demographic, or targeting them using a particular channel (for example, direct selling or use of an aggregator website). These quantities may be normalized and represented as a coordinate of the simulation landscape, so that every insurer can be located directly on the space. Customers may be also be directly located on the space if the coordinate relates to a customer attribute, for example age, otherwise they are randomly distributed on the relevant dimension in the simulation space (following a normal or uniform distribution depending on the relevant characteristic).

On the other hand, business attributes may also be non-measurable, for example, perceived reputation, branding, and image. Such attributes are distinct from product quality: all rational customers prefer better quality, but their response to attributes such as corporate image is more subjective. If business attributes are non-measurable, then the relevant coordinate of the simulation landscape can be purely abstract. Both insurers and customers may then be located randomly on the business attribute space. This has the benefit of allowing for insurer heterogeneity in respect of these business characteristics, and also customer heterogeneity in respect of response to these characteristics.

2.3. Outline of Execution of Agent-Based Market System

The steps in the execution of the market system are outlined in Figure 2. The discrete time unit in the system is a policy year, which is assumed to be the same for all customers. Every insurer offers a price to every customer at the beginning of the year: insurer i offers a price P_{it}, at the start of year $(t, t{+}1)$.(The price P_{it} can be viewed as a price per unit of risk, as in the simplified insurance simulation model of Taylor (2008).) An $N \times M$ array is constructed containing the total cost TC_{ijt}, for every customer j, of buying insurance from insurer i, where the total cost is a linear combination of price and customer-insurer distance. (The precise specifications for P_{it} and TC_{ijt} are given below in sections 2.4 and 2.5 respectively.) Customers buy the least-cost available policy, based on a ranking by total cost.

During the year, insurance claims are randomly generated. Insurers pay out at the end of the year. Their financial position is then evaluated. Insurers decide on their pricing for the following year based on their claims experience during the preceding year as well as experience on the overall market. Insurers also revise their business strategy by changing location on the simulation landscape and moving in the direction of the most financially successful insurer within their local peer group of competitors, if they are not the most

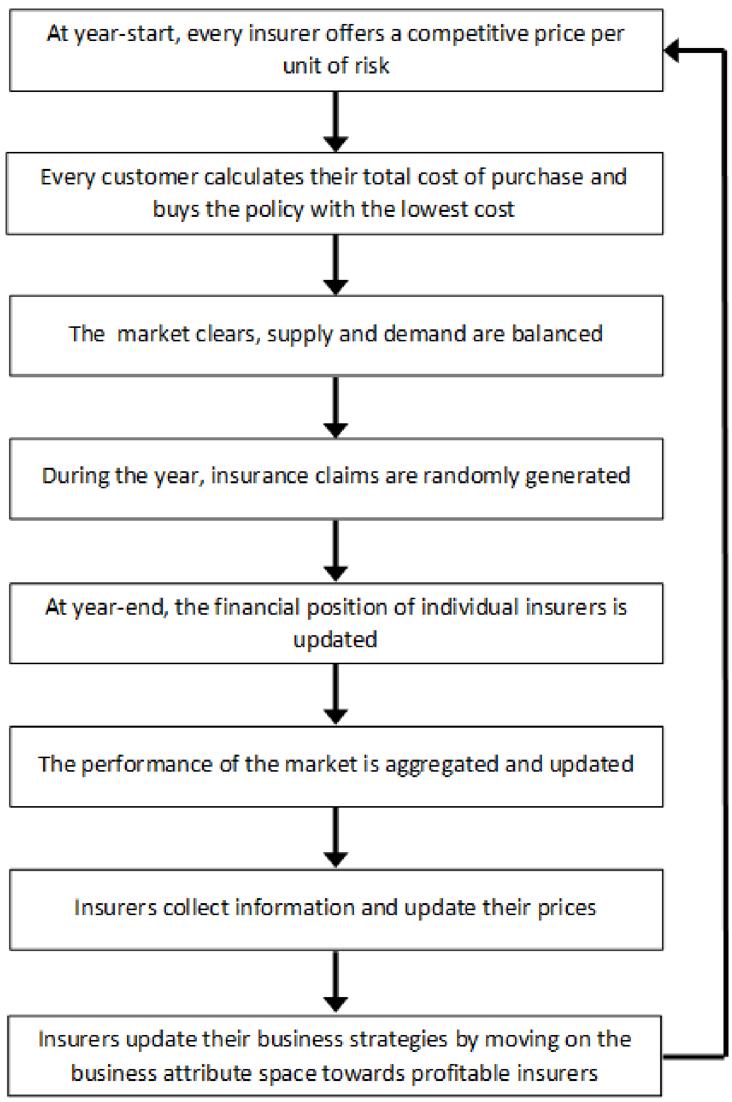

Figure 2: An outline of the execution of the agent-based market system.

successful themselves. The market process then resumes again at the start of the next policy year.

2.4. Pricing by Insurers

In order to price the insurance policies that they sell, insurers in the agent-based system closely replicate actual insurance pricing methods, as described by insurance experts such as Kaas et al. (2008), Hart et al. (2007) and Booth et al. (2005). A pure or statistical premium rate $\widetilde{P}_{it}$ is first calculated based on historic average claim amounts. (The specification for $\widetilde{P}_{it}$ is given below in section 2.7.) Two adjustments are then made.

The first adjustment allows for the riskiness or volatility of claims. A risk or safety loading is added to the pure premium based on the standard deviation F_{it} of claim amounts made against insurer i by time t. An actuarial premium rate is then calculated as $\widetilde{P}_{it}+\alpha F_{it}$, where $\alpha > 0$ is a safety (risk) loading factor. An insurer therefore charges a higher premium if claims are volatile and the insurer is exposed to greater risk. The safety loading factor α determines how much weight is given to the volatility or riskiness of claims when calculating premiums. See Kaas et al. (2008) for comprehensive details about safety loading and risk measures other than the standard deviation.

The second adjustment is a commercial adjustment to the premium rate, made by underwriters when they sell insurance policies, which gives the price P_{it} offered by insurer i. This depends on demand from customers and market pressure from competitors. In the agent-based system, this is implemented through a mark-up to the actuarial premium. The price offered by insurer i at time t is

$$P_{it} \;=\; \left(\widetilde{P}_{it} + \alpha F_{it}\right) e^{m_{it}}, \tag{1}$$

where m_{it} is a log mark-up. The mark-up depends on the price-elasticity of demand that the insurer estimates for its product. Insurer i estimates this by calculating an arc-elasticity of demand at time t over the previous two periods (Petersen & Lewis, 1999, p. 82):

$$\widehat{\epsilon}_{it} \;=\; \frac{(Q_{i,t-1} - Q_{i,t-2})/(Q_{i,t-1} + Q_{i,t-2})}{(P_{i,t-1} - P_{i,t-2})/(P_{i,t-1} + P_{i,t-2})}, \tag{2}$$

where $Q_{i,t}$ denotes the number of insurance policies sold by insurer i at time t and $P_{i,t}$ denotes the price of these policies. Standard economic theory states that a firm will maximize its profit if it charges a price $P^* = MC \times (1 + \epsilon^{-1})^{-1}$, where ϵ is the price-elasticity of demand and MC is marginal cost Petersen & Lewis (1999, p. 429). Comparing P^* with P_{it} in equation (1), and noting that ϵ^{-1} is typically small in a highly competitive market such as automobile insurance, suggests that insurers in the agent-based system use a first-order approximation for the mark-up as follows:

$$\widehat{m}_{it} \approx -\frac{1}{\widehat{\epsilon}_{it}} = -\frac{(P_{i,t-1} - P_{i,t-2})(Q_{i,t-1} + Q_{i,t-2})}{(Q_{i,t-1} - Q_{i,t-2})(P_{i,t-1} + P_{i,t-2})}, \tag{3}$$

and then estimate the profit-maximizing mark-up by using exponential smoothing (with $0 < \beta \leq 1$):

$$m_{it} = \beta \widehat{m}_{it} + (1 - \beta) m_{i,t-1}. \tag{4}$$

(There is a small probability that, purely by chance in the course of a stochastic simulation, $Q_{i,t-1} = Q_{i,t-2}$ in equation (3), which would lead to an infinite value of $\widehat{m}_{it}$. In this case, the insurer does not update his estimate in equation (4) and sets $\widehat{m}_{it} = m_{i,t-1}$.)

The first-order approximation in equation (3) is justified because demand is typically very sensitive to an individual insurer's offered price in a highly competitive market such as automobile insurance (so that $-\widehat{\epsilon}_{it}^{-1}$ is typically small in equation (3)). The exponential weighted moving average in equation (4) is also a pragmatic choice because insurance prices offered to individual customers are changed gradually by insurers. Furthermore, exponential smoothing is often an effective business forecasting tool (Hyndman et al., 2008).

In practice, it is difficult to evaluate an insurer's price-elasticity of demand precisely. It varies with time, with the changing customer base of the insurer, and with the dynamic competition experienced by the insurer. In the agent-based simulation, as in the real world, customers and insurers are interacting inter-temporally: customers move dynamically from one insurer to another, and insurers dynamically react to this by adjusting their prices to maximize profit. Some insurance analysts, such as Warthen & Somner (1996), refer to the elasticity of demand faced by individual insurers as a variable that cannot be estimated.

Expert systems based on causal inference have been developed recently to measure such price sensitivity: see Guelman & Guillén (2014). The premium rating expert systems of Lin (2009) and Imriyas (2009) are also noteworthy, as they remove imprecision and subjective judgment from the pricing process, but they do not incorporate a response to market competition.

2.5. Purchasing Decision by Customers

The agent-based simulation system provides for realistic behavior of customers. Clearly, customers respond to price but are also influenced by various characteristics of the insurer, for example its advertising methods, branding, reputation, perceived quality of service etc. Large amounts are spent on advertising and marketing precisely because of this. The simulation space, as exemplified in Figure 1, captures the fact that these non-price attributes matter. Location models such as this have been successfully used in a number of agent-based models, for example by Ladley (2013) in his model of contagion in banking.

The customer-agents in the insurance market simulation seek an insurance policy with a low price but also an insurer whose business attributes they prefer. At the start of year $(t, t+1)$, customer j chooses insurer i based on the lowest total cost TC_{ijt}, defined as

$$TC_{ijt} \;=\; P_{it} + \gamma\Delta_{ij}, \tag{5}$$

where Δ_{ij} is the shortest Euclidean distance between insurer-agent i and customer-agent j on the business attribute space. Here, γ is a cost per unit of distance, and it penalizes distance relative to price.

2.6. Insurers' Capital and Customers' Claims

During the policy year, random claims against insurance policies are generated. For every insurance policy sold, the timing of insurance claims follows a Bernoulli distribution with parameter b, whereas the amount of the claim follows a Gamma distribution with mean μ_G and standard deviation σ_G. We assume that claims occur as an independent and identically distributed process. If catastrophic events occur, such as natural disasters, then this assumption is clearly violated as many policyholders will claim at the same time and for the same kind of damage. As a first approximation, however, it is reasonable to disregard catastrophic events.

Insurers are endowed with capital from the outset. Their capital decreases as they pay out claims and increases as they earn premium income. Insurers endeavor to remain solvent by holding capital against every insurance policy that they sell. (This loosely mirrors solvency regulations. Insurers are required to hold capital so that they can pay out claims as and when the claims arise.) If they exhaust their capital, and cannot sell more policies, then some customers will not receive the least-cost policy. In this case, a random selection of customers buy from the insurer, while the remaining customers buy from their second-ranking insurer, and so forth, until all customers are insured. Finally, we note some implicit, simplifying assumptions in the agent-based simulation: customers cannot cancel their policies mid-year; there is no fraudulent insurance claim; there is zero inflation, administration expense, and tax on insurance premium for customers and on corporate profit for insurers.

2.7. Updating Prices

The agent-based simulation component of our expert system also enables insurer-agents to learn from their experience and from the market. At the end of every policy year, the pure premium $\widetilde{P}_{it}$ in equation (1) is calculated by every insurer as a credibility-weighted estimate based on its claim history and the market-collected average claim. This mirrors standard actuarial costing methods, as detailed by Kaas et al. (2008, p. 203–227) for example, as well as the claims information pooling activity of insurance rating bureaus or advisory organizations such as the Insurance Services Office in the U.S. or the Claims and Underwriting Exchange in the U.K. Note that credibility is also proposed by Kaishev et al. (2013) and Thuring et al. (2012) as a technique to market products to insurance customers.

Insurer i computes the pure premium

$$\widetilde{P}_{it} \;=\; z\overline{X}_{it} \,+\, (1-z)\lambda\mu, \tag{6}$$

where $\overline{X}_{it}$ is the average annual insurance loss of insurer i at time t. Also, λ and μ represent the average claim frequency per unit risk exposure and the average claim size respectively, as collected across all insurers during one simulation run, and $z \in [0, 1]$ is a credibility factor which weights the insurer's own claims experience against the market experience as a whole. Further, $\overline{X}_{it}$ is calculated as a weighted average:

$$\overline{X}_{it} \ = \ wX_{i,t-1} \ + \ (1 - w)\overline{X}_{i,t-1}, \qquad (7)$$

where $X_{i,t-1}$ is the insurance loss made by insurer i in the year to time $t - 1$. The weight $w \in [0, 1]$ determines how much weight should be given to past claims history.

2.8. Updating Business Strategy

The insurer-agents in the agent-based simulation do not have to be static on the simulation landscape or business attribute space. At the end of the year, insurers may change their business strategy, by directly modifying their attributes. For example, they can choose to increase expenditure on customer complaints handling, or they may devote more resources to augment their social media presence, or they may choose to focus on a particular demographic of the population (for example, older customers with luxury vehicles).

Changing business strategy means that insurers move on the simulation space. We define a speed parameter array $\mathbf{\Theta} = \{\theta_i \in [0, 1] : i = 1, \ldots, N\}$, where θ_i is the speed with which insurer i can move. We also define a scope parameter array $\mathbf{\Phi} = \{\phi_i \in [0, 1] : i = 1, \ldots, N\}$, where ϕ_i is the scope of insurer i's peer group.

More specifically, $\theta_i \in [0, 1]$ is the normalized Euclidean distance on the simulation space over which insurer i moves at the end of every year. The agent-based system assumes that insurers move towards strategies which are successful, by comparing themselves to their peers. Peer membership is determined as follows. At the end of the year, the underwriting result for every insurer is calculated in the form of the insurer's loss ratio, which is the insurance loss (total claims) for the year divided by premium income for the year. An $N \times N$ symmetric array containing the shortest Euclidean distance between every pair of insurers is updated. For every insurer i, we sort through this array to determine a proportion

$\phi_i \in [0, 1]$ of insurers who are close to the insurer on the simulation landscape. These insurers form the peer group. Insurer i then searches for the best-performing insurer in the peer group, this being the insurer with the lowest loss ratio in the past year. At the beginning of the next policy year, insurer i will then change its location to move over a distance of θ_i in the direction of the best-performing insurer. The peer group of insurer i includes insurer i, so insurer i does not move if it is the best-performing insurer in the group.

The fact that insurers compare themselves with some, and not necessarily all, of their competitors is also captured in the insurance model of Taylor (2008), where insurers compare themselves with insurers of about the same size as themselves. Depending on the attribute represented on a given axis of the simulation landscape, the landscape may be assumed to wrap around itself, to avoid boundary effects as the insurer-agents move about the simulation space.

2.9. Simulation Outputs

For every insurer-agent, the agent-based simulation system outputs a multivariate time series containing: the price charged by every insurer, the number of policies that it sells, the insurance loss faced by every insurer at the end of the year, and the location coordinates of the insurer on the simulation space. The insurer's premium income for the year is the price multiplied by the number of sales. Insurance analysts are usually concerned with the insurer's *loss ratio*, which is the annual insurance loss expressed as a proportion of the annual premium income. The insurance market regulator will also be concerned with the overall market loss ratio, computed by summing premium income and losses over all insurers in the market.

2.10. Visualization

The agent-based system incorporates simple visualization tools enabling the user to observe the evolution of the market in aggregate. Customers are fixed on the simulation space, but their distribution (possibly randomly allocated) can be visualized as in Figure 3.

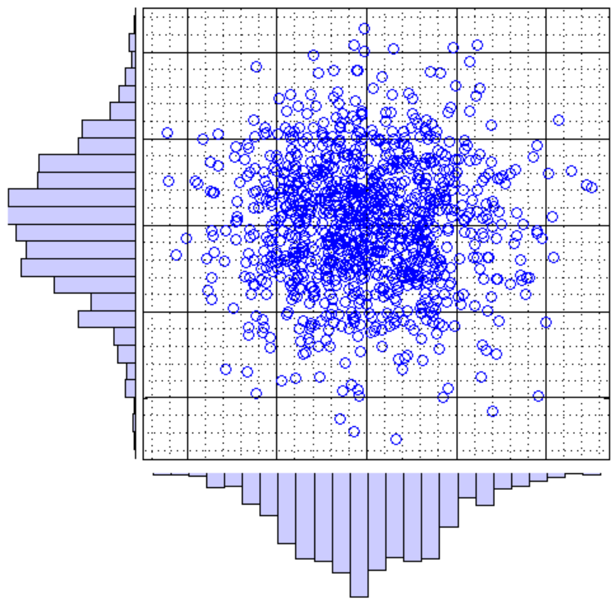

Figure 3: Visualization of customer distribution in the simulation space of the agent-based system.

The movement of insurers can also be followed, with the amount of capital held by the insurers (their financial 'size') and the network of peers against which insurers compare themselves being highlighted: see Figure 4. This can help the insurance market regulator observe market dominance and clustering or herding behavior by the insurers.

3. Temporal Data Mining Component

3.1. Data Mining Tools

The agent-based simulation component simulates the insurance market using a bottom-up approach. That is, every agent in the market, whether an insurer or customer, is modeled individually. Agents are also heterogeneous and they interact. Simulations therefore produce a large amount of data in every run. A user such as an insurance market regulator may run different simulations under different scenarios, for example to stress-test the market in the event of a major loss. Monte Carlo simulations may also be generated, by repeating simulations thousands of times, and then computing the probability that a particular event occurs (for example, the probability that the market average insurance premium rises by more than 25% in a year).

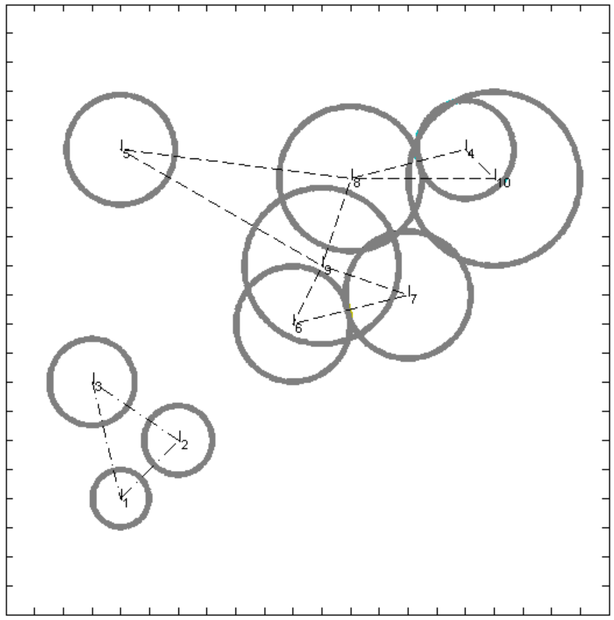

Figure 4: Visualization of insurers in the simulation space of the agent-based system. The radius of the circle drawn around the insurers represents their level of capital. The lines between the insurers show the peer group against which insurers are comparing themselves. Local clustering of the insurers is apparent.

There are two kinds of patterns that we may wish to uncover about the insurance market from the large amounts of simulated time series: *motifs* or repeated subsequences, and *anomalies* or unusual subsequences. Motifs indicate periodicity, which can help the insurance market regulator to forecast cycles in the market. Anomalies indicate unusual events such as a market crash, excessive volatility, or potential insurer insolvency.

The temporal data mining component of the system is used to uncover patterns in simulated data from the agent-based component. It consists of three tools: SAX, a time series discretization algorithm; Sequitur, a grammar induction algorithm; and GrammarViz, a motif and anomaly discovery tool.

3.2. Symbolic Representation with SAX

The simulated time series from the agent-based simulation system exhibit high dimensionality, high autocorrelation, and are also noisy. The SAX algorithm (Symbolic Aggregate ApproXimation) is an effective method to represent or approximate such data: it performs

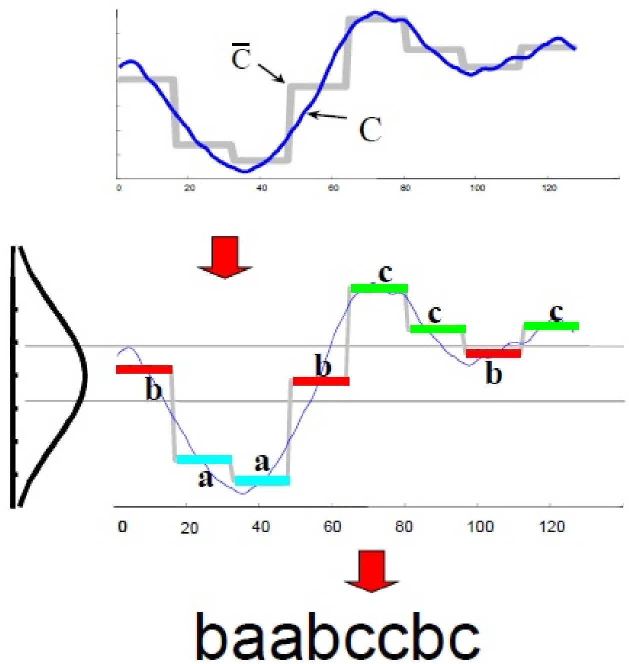

Figure 5: Symbolic Aggregate Approximation (SAX) discretizes a real-valued time series C resulting in the piecewise approximation $\overline{C}$ (top panel). The discrete values are then converted to a string of symbols, such that every symbol is equally likely to occur according to the standard Normal distribution, resulting in the symbolic sequence `baabccbc` (bottom panel). Source: Eamonn Keogh (reproduced with permission).

discretization and symbolic representation on real-valued data such as simulated market insurance data. SAX was developed by Lin et al. (2003).

The procedure within SAX is depicted in a simplified way in Figure 5. A real-valued time series is first z-normalized, then split into several equal-length subsequences. A piecewise linear approximation of the original time series is then created. The heights of the linear segments in the piecewise approximation equal the mean values of the subsequences that they are replacing. The heights are also approximately Normally distributed, and every segment is then mapped to a symbol in such a way that probabilistic bias is avoided and every symbol occurs with the same probability under the Normal distribution. Real-valued data are therefore converted to symbolic sequences. See Lin et al. (2003) for more details.

SAX is fast, allows lower bounding with various measures of dissimilarity and performs dimensionality and numerosity reduction. Although some information in the data is lost by discretization, it has been shown that SAX preserves and highlights important features such as cycles and anomalies (Ratanamahatana et al., 2010). The summarization that is achieved with SAX facilitates and speeds up data mining tasks such as classification, clustering, anomaly detection and periodicity mining.

3.3. Grammar Induction with Sequitur

Sequitur is an algorithm which performs grammar induction, or syntactic pattern recognition, on symbolic sequences. It was developed by Nevill-Manning & Witten (1997). Simulated time series data from the agent-based system are converted to symbolic sequences using SAX by first extracting subsequences via a sliding window. Grammar induction is then performed on these sequences by Sequitur in an incremental fashion as each new symbol arrives. The context-free grammar that is inferred is a compact generative representation of the symbolic sequence and Sequitur is, in effect, a string compression algorithm. Sequitur takes every word generated by SAX as an input token and recursively reduces every *digram*, or consecutive pair of tokens which repeat themselves, to a new symbol. Sequitur performs grammar inference under two constraints: a utility constraint which guarantees that all of the new symbols created correspond to recurring patterns in the SAX-generated symbolic sequence, and a digram uniqueness constraint which ensures that the digrams that are reduced to a new symbol do not repeat themselves.

The grammar rules that Sequitur generates therefore exploit recurring subsequences in the data, and hence highlight motifs (repeated patterns). The following example is given by Li et al. (2012). Consider the string $S_1 = abcd\,abcd\,eab$. Sequitur performs grammar induction and converts this into three grammar rules:

$$S_1 \to R_1 R_1 e R_2 \qquad R_1 \to R_2 cd \qquad R_2 \to ab$$

Observing the repetition of the rule R_1 in $S_1 \to R_1 R_1 e R_2$, we can tell that the sub-sequence *abcd* is a motif and is periodic in the data.

Figure 6: Screenshot of GrammarViz detecting motifs (recurring patterns) on test data.

The converse problem of detecting patterns which recur the least (or do not recur) is also of interest. Such patterns are indicative of structural irregularities or *anomalies*. The following example is given by Senin et al. (2015). Let $S_2 = abc\,abc\,cba\,xxx\,abc\,abc\,cba$. Sequitur converts this into three other grammar rules:

$$S_2 \rightarrow R_3 xxx R_3 \qquad R_3 \rightarrow R_4 cba \qquad R_4 \rightarrow abcabc$$

The subsequence xxx is non-repeated and is an anomaly which stands out in the rule $S_2 \rightarrow R_3 xxx R_3$.

3.4. Motif and Anomaly Discovery with GrammarViz

GrammarViz is a software tool which leverages SAX and Sequitur and implements algorithms to discover anomalies and motifs from time series data (Li et al., 2012; Senin et al., 2015). A screenshot of GrammarViz appears in Figure 6. Details of the motif and anomaly detection algorithms in GrammarViz are given by Li et al. (2012), Otunba et al. (2014) and Senin et al. (2015). We provide only a brief sketch of the algorithms here.

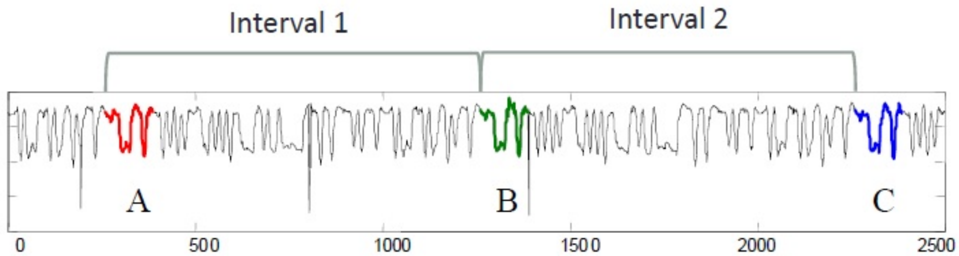

Figure 7: A motif or recurring pattern which repeats three times, in positions A, B and C, in test data. Source: Eamonn Keogh (reproduced with permission).

A simulated time series (for example containing insurer loss ratios) from the agent-based simulated market is passed to GrammarViz, which calls on SAX and Sequitur to convert the time series to a symbolic sequence and perform grammar inference. GrammarViz maps all grammar rules back on to the original simulated time series data. Rules can be ranked by their length and frequency, and may be visualized in a time plot of the data.

The motif-based periodicity detection algorithm in GrammarViz works as follows. For a motif to be regarded as valid, it must occur at least three times in the data. Suppose that GrammarViz detects m valid motifs in the simulated data. Three such motifs are shown in Figure 7. For the ith motif, GrammarViz records the intervals between each occurrence of the motif, and calculates the mean $\overline{\tau}_i$ and standard deviation s_i of the inter-motif time intervals. GrammarViz then estimates the periodicity in the data to be equal to $\overline{\tau}_j$ where $j = \arg\inf_i s_i$. That is, the periodicity is the average inter-motif interval for the motif which recurs with the least variability. If there is no valid motif, then GrammarViz will indicate that no periodicity is detectable in the output from the agent-based simulations.

GrammarViz also implements two anomaly detection algorithms (see Senin et al., 2015), and we describe one of them here. Suppose that the simulated time series input to GrammarViz from the agent-based simulation system contains n data points. Since GrammarViz maps all grammar rules created by Sequitur back on the simulated time series, every data point is encompassed by one or more grammar rule. GrammarViz creates an array of length n where each element contains a count of the grammar rules which span the corresponding data point. The array sequence is then plotted to give a rule density curve. This

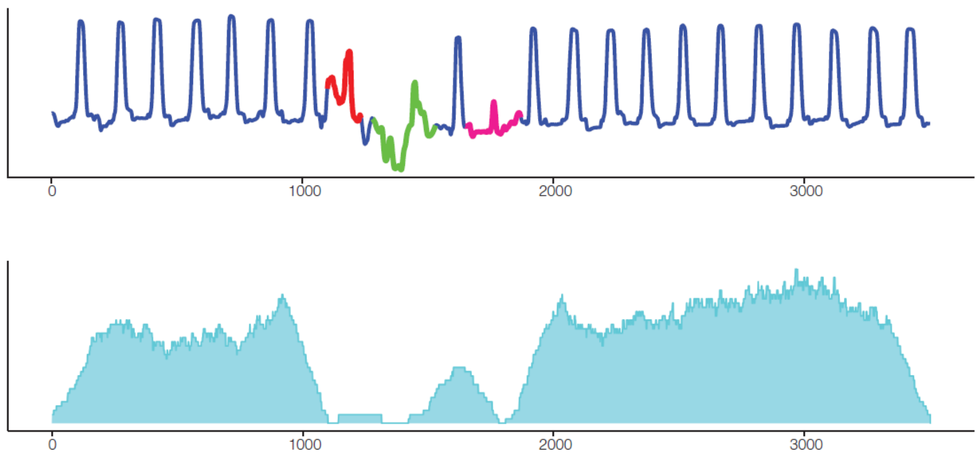

Figure 8: Top panel: test data containing anomalous or unusual subsequences. Bottom panel: the grammar rule density curve attains minima at the location of the anomalies. Source: Senin et al. (2015) (reproduced with permission).

is illustrated in Figure 8. Data points which are spanned by only a few grammar rules are unusual or anomalous, and merit further investigation. GrammarViz outputs the time points at which these anomalies occur.

4. Application to Automobile Insurance Market

4.1. Parameterization and Validation

In this section, we consider an application of our system to the automobile insurance market in the U.K. In order to parameterize and validate our system, we use data from the Association of British Insurers (ABI) from 1983 to 2011. We also consider a simplified agent-based market where the simulation space consists of an abstract circle with insurers located equidistantly, and customers located uniformly, on the circle. Based on the data, we obtain baseline parameter values as shown in Table 1.

Most of the parameter values are chosen to give a reasonable representation of the U.K. automobile insurance market, whilst avoiding long simulation runtimes. For example, about 97% of market share in the automobile insurance business was held by 20 insurers in

Parameter	Value	Definition
α	0.001	Safety loading factor in insurers' premium calculation, eq. (1)
β	0.3	Weight in insurers' mark-up calculation, eq. (4)
γ	0.05	Weight in customers' total cost calculation, eq. (5)
M	1000	Number of customers
N	20	Number of insurers
T	1000	Time horizon of simulation
μ_G	100	Mean of Gamma-distributed i.i.d. claims
σ_G	10	Standard deviation of Gamma-distributed i.i.d. claims
b	1	Parameter of claim frequency distribution
w	0.2	Weight in estimation of insurer's average claim size, eq. (7)
z	0.2	Credibility factor in pure risk premium, eq. (6)
$\mathbf{\Theta}$	$\mathbf{0}_N$	Speed of insurers' movement on business attribute space
$\mathbf{\Phi}$	$\mathbf{0}_N$	Scope of insurers' peer group

Table 1: Baseline parameter values in the agent-based simulation system. In the baseline case, insurers do not move in the business attribute space, so $\mathbf{\Theta} = \mathbf{\Phi} = \mathbf{0}_N$, an N-long vector of zeros.

2011, so we set the number of insurers $N = 20$. Numerical experimentation shows that the size of M, the number of customers, relative to N is more important than the absolute size of M, so we select $M = 1000$, to represent the fact that there are far more customer-agents than insurer-agents. The random claims process parameters μ_G, σ_G, and b are consistent with the simulation study of Taylor (2008), and experiments indicate that their only effect is on the scale and volatility of simulation results. The values that we use for w and z are based on parameters used by insurance experts on European insurance markets: see Kaas et al. (2008) and Booth et al. (2005). Since this is a baseline model, we also ignore business strategy evolution and populate the parameter arrays $\mathbf{\Theta}$ and $\mathbf{\Phi}$ with zeros, so that insurers do not move on the simulation landscape.

The key parameters which have the greatest influence on simulation results are the behavioral parameters α, β and γ related to the pricing of insurance by insurers and to customers' affinity to insurers' business attributes. We do not have micro-data of sufficient granularity on insurers and customers to estimate these parameters, so we estimate them by means of the method of moments combined with repeated Monte Carlo simulations on a grid of values of α, β and γ. The grid is searched for the parameter values which match the mean and standard deviation of the simulated market loss ratios to the corresponding statistics from the actual market data, and which minimize the squared difference between the lag-1 autocorrelation in the simulated loss ratio and the lag-1 autocorrelation in the actual loss ratio. Successive refinement and search are performed, with spot checks on surrounding grid values to verify that a globally optimal set of parameter values is obtained.

In order to validate the model, we test whether the distribution of loss ratios in the simulated agent-based market is the same as in the actual market. Before this, we test for stationarity in the loss ratios. The Augmented Dickey-Fuller (ADF) test comfortably rejects the null hypothesis of non-stationarity (unit root) at 5% in both the actual and simulated loss ratios (actual data: ADF test statistic = −6.075, 1% critical value = −3.696; simulated: ADF test statistic = −3.504, 5% critical value = 2.975). Testing the distributions, we find that the two-sample Kolmogorov-Smirnov test fails to reject the null hypothesis that the actual market loss ratios and the simulated agent-based market loss ratios have the same continuous probability distribution at the 5% significance level (p-value = 0.1997).

4.2. Cycles

An issue of great interest to insurance market regulators is whether the insurance market is cyclical or whether it performs a "random walk". Basic theory indicates that markets should move purely randomly, but market cycles are a stylized fact in insurance and in finance more generally. Cycles are defined here as a more or less regular succession of peaks and troughs in the overall profitability of insurers (see Dionne, 2013). Our agent-based simulation system can be used by a market regulator to determine the conditions under which cycles are present. The temporal data mining component of the system can be used to measure, and hence help the regulator forecast, these cycles.

Figure 9 shows the results of one simulation run of the agent-based insurance market system, with the baseline parameters in Table 1. The top panel of Figure 9 shows one sample path of loss ratios for a randomly selected individual insurer. The bottom panel of Figure 9 shows one sample path of loss ratios for all insurers combined, i.e. for the entire market. The central limit theorem might lead one to expect that, aggregated over the entire market, loss ratios would be noisy but stable and unstructured. However, the bottom panel of Figure 9 does appear to exhibit a cyclical pattern. This may be compared with Figure 10 which shows actual ratios from the U.K. automobile insurance market, which also appears cyclical. (The combined ratio is also shown on Figure 10 and allows for expenses, which are relatively stable here, so it is an upward shifted version of the loss ratio.)

The temporal data mining component can be used to measure cyclicality. On 10 simulation runs, GrammarViz returns an average periodicity of 6.6 years. This is consistent with the cycle lengths of 6–8 years estimated in econometric studies (Dionne, 2013; Boyer & Owadally, 2015). Interestingly, on the actual market data shown in Figure 10, GrammarViz returns a periodicity of 8 years in the loss ratio and 6 years in the combined ratio. This cycle length estimate enables market regulators effectively to forecast turning

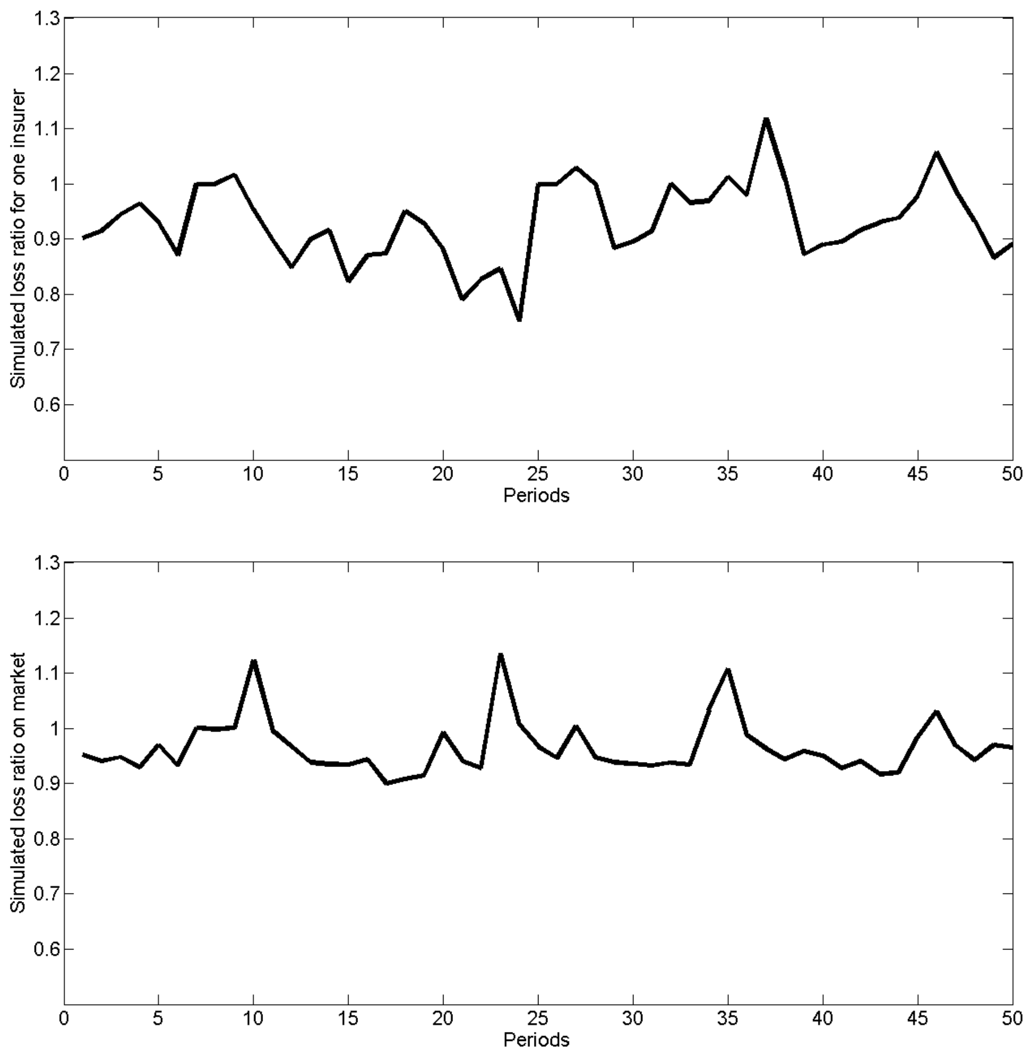

Figure 9: Sample paths of simulated loss ratios for one insurer (top panel) and for overall market (bottom panel).

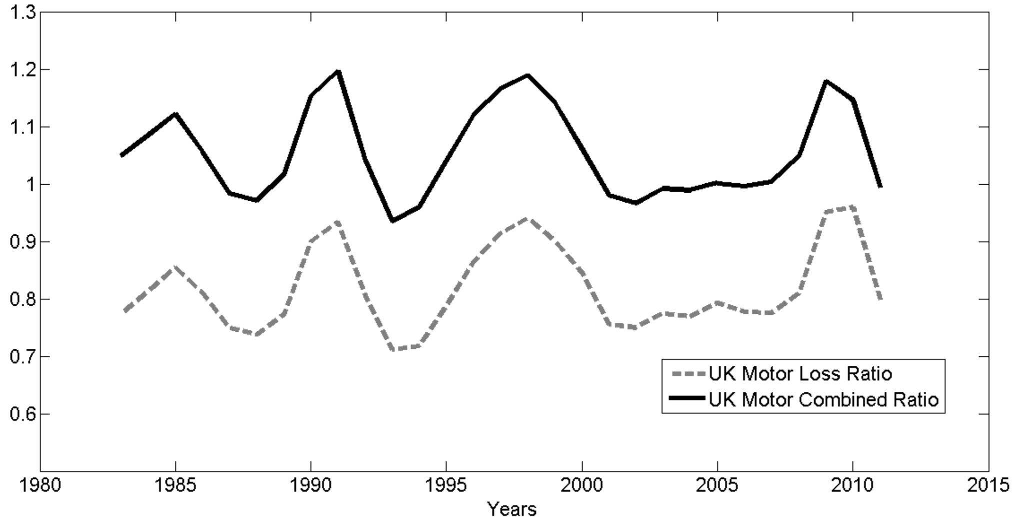

Figure 10: Loss ratios and combined ratios for the U.K. automobile (or 'motor') insurance market. Source: Association of British Insurers.

points in insurance market cycles, and recommend timely and appropriate policy action in response.

4.3. Volatility

Another issue of interest to insurance market regulators is how to dampen excessive volatility on the market and avoid a crash, known in the industry as an insurance crisis. When crises occur, insurers pull out of the market in a bid to stem accelerating losses, and insurance coverage for customers becomes either unavailable or very expensive.

It might be anticipated that, if insurance regulations encourage insurers to use a slightly higher safety loading factor α when they price insurance (see equation (1)), then customers might face slightly higher insurance bills, but would gain from a more stable market and would avoid steep premium rises or unavailable insurance during crisis episodes. The results from stochastic simulations show that, beyond a certain value, increasing the safety loading is counterproductive: loss ratios increase and become more volatile. This is shown in Figure 11.

This would appear to be an instance of the "winner's curse" phenomenon, whereby

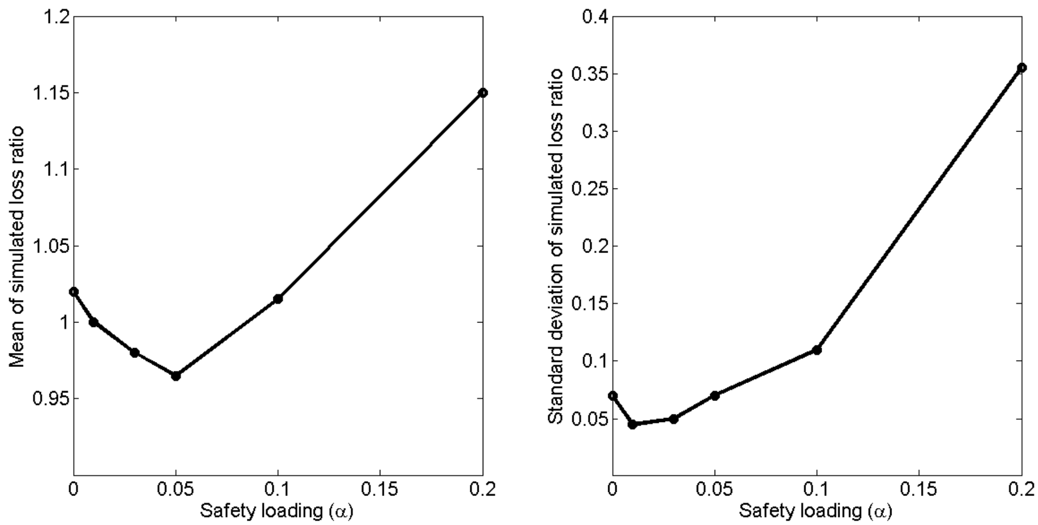

Figure 11: Mean loss ratios (left panel) and volatility of loss ratios (right panel) across the simulated agent-based market as safety loading α varies.

the highest bidder wins business but eventually incurs losses (Harrington et al., 2008). If insurers price risk too highly by applying a high safety loading, then this encourages customers to switch to other insurers as customers search for lower prices. Insurers which are under-pricing, possibly because of an initial favorable claims experience or because of low customer base, then acquire more and more customers. (Harrington et al. (2008) also suggest that insurers with poor forecasting ability may initiate price-undercutting.) This triggers a "price war" dragging other insurers into unaffordable price cuts in a bid to retain their customers. Unsustainably large losses eventually occur, hence the 'curse'. Losses are persistently higher and more volatile across the market, as shown graphically in Figure 11.

This result also shows that market regulators should monitor not just the capital held by insurers but also the level and dispersion of prices that they offer. Significant price undercutting should be a cause for concern by market regulators. Our expert system should enable market regulators to monitor the overall market as well as individual insurers and forestall potential crises.

4.4. Herding

The expert system can also help a market regulator observe how insurers' business strategies evolve over time. Experimentation is feasible within agent-based systems, so we adjust the baseline case, summarized in Table 1, by changing $\mathbf{\Theta} = \{\theta_i\}$ and $\mathbf{\Phi} = \{\phi_i\}$ to allow a proportion of the insurers to move on the business attribute space. Fixed insurers are insurers who do not change their business attributes. We set their speed and peer group scope parameters, respectively, to be $\theta_i = \phi_i = 0$ if i denotes a fixed insurer. Mobile insurers are insurers whose business strategy evolves as they move on the business attribute space towards financially successful insurers. We set $\phi_i = 0.2$ and $\theta_i = 0.01$ if i denotes a mobile insurer. Every mobile insurer compares itself to a peer group of insurers nearest to it, the peer group size being 20% of all insurers. The insurer then moves in the direction of the most financially successful insurer in the group (unless it is itself the most successful) by traversing 1% of the business attribute space every year.

Figure 12 shows the location of insurers over time, as the proportion of mobile insurers varies from 20% to 100%. Insurers with fixed strategies do not move and appear as the horizontal lines in Figure 12. Insurers with variable strategies evidently herd together: they either cluster around the same strategies followed by fixed insurers (visible in the 50% and 80% cases), or converge to the same strategy if there is no fixed insurer (100% case). This is consistent with Hotelling's Law, or the principle of minimum differentiation (see e.g. Petersen & Lewis, 1999, p. 574), which states that firms will seek to produce goods which are as similar as possible, under certain conditions. Further analysis of the output from the agent-based simulations show that customers benefit from lower insurance prices on average when herding occurs, thanks to insurers' aggressive competition on price. However, price under-cutting occurs as insurers compete on price only. This leads to large losses by individual insurers and destabilizes the insurance market.

4.5. Discussion and Practical Implications

The validity of the results that we obtain here depends of course on the modelling approach underpinning the expert system. A significant strength of our approach is that

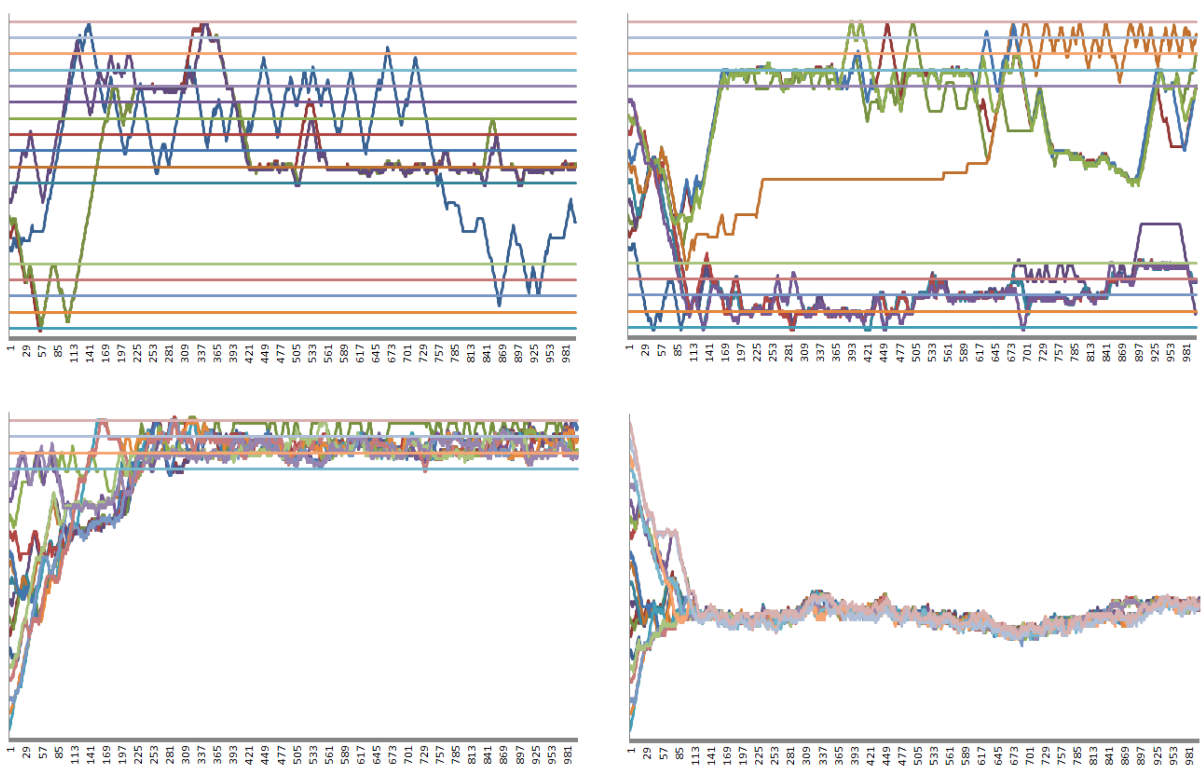

Figure 12: Evolution over time of business strategy of insurers. Vertical axes: position in business attribute space, horizontal axes: time (years). The proportion of 'mobile' insurers who can change strategy is 20% (top left panel), 50% (top right), 80% (bottom left), 100% (bottom right).

we do not make any assumption of linearity or Gaussianity. One or both of these is usually made in insurance econometric models. See Dionne (2013), Boyer & Owadally (2015) and Wang et al. (2011) for examples of ARIMA, cointegration, and regime-switching models. Nevertheless, the data mining sub-system makes subtle assumptions about the structure of market data, for example about its ergodicity. These assumptions deserve further investigation. Another strength is that the model is calibrated to actual market data, with a realistic representation of heterogeneous insurers interacting on the market. On the other hand, standard models, such as described by Dionne (2013) are theory-driven with a single representative insurer within the rational expectations framework. Dynamical simulation models, such as in Taylor (2008) and Warthen & Somner (1996), comprise a number of explicit difference equations and are therefore transparent. Our expert system is based on a market model which is constructed in a bottom-up fashion with implicit behavioral rules governing insurers and customers. This is a 'black box' approach, albeit closer to the industry models described by Mills (2010) and Ingram & Underwood (2010).

Notwithstanding the above, our expert system has multiple potential benefits for a market regulator whose chief concern is the stability and smooth operation of the insurance market. The regulator can input the latest data on insurers and customers, with as much granularity as possible to maintain a close fit to the actual market, and project tens of thousands of Monte-Carlo scenarios about individual insurers over the next few years. The data mining component then detects anomalies in price and profitability, both at an individual insurer and a collective market level, in the generated sample paths. The regulator can therefore obtain a probabilistic forecast of variables such as insurer loss ratio, default rate, and price dispersion. Patterns such as significant price-undercutting by one or more insurers can be uncovered, as these can can trigger a "winner's curse" (section 4.3). The regulator can then warn individual insurers (and their shareholders and customers), similar to a bank stress test. The regulator can also estimate the probability of systemic instability, such as during prolonged troughs in the insurance market, as well as the incidence and timing of turning points in the cycle (section 4.2). Such instability can be costly for both corporate and retail customers leaving them uninsured or with unpaid insurance claims. The regulator can therefore enforce stricter capital reserves counter-cyclically and in advance to mitigate the risk of such events.

5. Conclusion

Insurance markets are unstable, experiencing cycles and crises regularly. We propose an agent-based system with temporal data mining to assist regulators in monitoring insurance markets. The agent-based simulation component runs a bottom-up simulation of every insurer and customer on the market. Agents employ simple optimization heuristics grounded in the microfoundations of the insurance business. Agent-based simulations generate a lot of data particularly when stochastic simulations are performed. The data mining component, powered by SAX, Sequitur and GrammarViz, enables market regulators to analyze the output of the agent-based simulation component. A prototype of the system is validated on an automobile insurance market. We show how it can be used to forecast the turning points of market cycles. We find that market regulators should be attentive to the level and dispersion of prices offered by insurers as price-undercutting can cause a "winner's curse", leading to unsustainably large losses and market instability. Finally, we show how regulators can use the system to analyze the evolution of insurers' business strategies, and to determine whether there are anti-competitive or destabilizing effects resulting from insurers' tendency to behave like a herd and converge upon the same strategies.

The research undertaken here will be extended in several directions. The expert system can be calibrated to insurance markets other than the automobile one. Specialist insurance markets, such as aviation or marine, may have specific features that can be captured. The business attribute space can be parameterized using different variables and with different topologies. Finally, a detailed comparison with econometric forecasts should also be illuminating.

References

Booth, P., Chadburn, R., Haberman, S., James, D., Khorasanee, Z., Plumb, R. H., & Rickayzen, B. (2005). *Modern Actuarial Theory and Practice, Second Edition*. (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Boyer, M. M., & Owadally, I. (2015). Underwriting Apophenia and Cryptids: Are Cycles Statistical Figments of our Imagination? *The Geneva Papers on Risk and Insurance Issues and Practice*, *40*, 232–255.

Dionne, G. (Ed.) (2013). *Handbook of insurance*. (2nd ed.). New York: Springer.

Guelman, L. (2012). Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Systems with Applications*, *39*, 3659–3667.

Guelman, L., & Guillén, M. (2014). A causal inference approach to measure price elasticity in Automobile Insurance. *Expert Systems with Applications*, *41*, 387–396.

Harrington, S. E., Danzon, P. M., & Epstein, A. J. (2008). "Crises" in medical malpractice insurance: Evidence of excessive price-cutting in the preceding soft market. *Journal of Banking & Finance*, *32*, 157–169.

Hart, D. G., Buchanan, R. A., & Howe, B. A. (2007). *The Actuarial Practice of General Insurance*. (7th ed.). Sydney, Australia: Institute of Actuaries of Australia.

Hsu, K.-T. (2011). Using a back propagation network combined with grey clustering to forecast policyholder decision to purchase investment-inked insurance. *Expert Systems with Applications*, *38*, 6736–6747.

Hyndman, R. J., Koehler, A. B., Ord, J., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: the State Space Approach*. Springer series in statistics. Berlin: Springer.

Imriyas, K. (2009). An expert system for strategic control of accidents and insurers' risks in building construction projects. *Expert Systems with Applications*, *36*, 4021–4034.

Ingram, D., & Underwood, A. (2010). *The human dynamics of the insurance cycle and implications for insurers: an introduction to the Theory of Plural Rationalities*. Technical Report Society of Actuaries Schaumburg, Illinois.

Kaas, R., Goovaerts, M. J., Dhaene, J., & Denuit, M. (2008). *Modern Actuarial Risk Theory Using R*. New York: Springer.

Kaishev, V. K., Nielsen, J. P., & Thuring, F. (2013). Optimal customer selection for cross-selling of financial services products. *Expert Systems with Applications*, *40*, 1748–1757.

Kuo, R., Lin, S., & Shih, C. (2007). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan. *Expert Systems with Applications*, *33*, 794–808.

Ladley, D. (2013). Contagion and risk-sharing on the inter-bank market. *Journal of Economic Dynamics and Control*, *37*, 1384–1400.

Li, Y., Lin, J., & Oates, T. (2012). Visualizing variable-length time series motifs. In *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 895–906). Anaheim, CA: SIAM.

Liao, S.-H., Chen, Y.-N., & Tseng, Y.-Y. (2009). Mining demand chain knowledge of life insurance market for new product development. *Expert Systems with Applications*, *36*, 9422–9437.

Lin, C. (2009). Using neural networks as a support tool in the decision making for insurance industry. *Expert Systems with Applications*, *36*, 6914–6917.

Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery* (pp. 2–11). ACM.

Lin, W.-B. (2010). Service failure and consumer switching behaviors: Evidence from the insurance industry. *Expert Systems with Applications*, *37*, 3209–3218.

Mills, A. (2010). *Complexity science. An introduction and invitation for actuaries*. Report for the Health Section Society of Actuaries Schaumburg, Illinois.

Nevill-Manning, C. G., & Witten, I. H. (1997). Identifying hierarchical structure in sequences: a linear-time algorithm. *Journal of Artificial Intelligence Research*, *7*, 67–82.

Otunba, R., Lin, J., & Senin, P. (2014). MBPD: Motif-based period detection. In *Proceedings of the 1st International Workshop on Pattern Mining and Application of Big Data* (pp. 793–804). Springer.

Petersen, H. C., & Lewis, W. C. (1999). *Managerial Economics*. (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., & Das, G. (2010). Mining time series data. In O. Maimon, & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook 2010* (pp. 1049–1077). Boston, MA: Springer US.

Samitas, A., Polyzos, S., & Siriopoulos, C. (2018). Brexit and financial stability: an agent-based simulation. *Economic Modelling*, *69*, 181–192.

Senin, P., Lin, J., Wang, X., Oates, T., Gandhi, S., Boedihardjo, A. P., Chen, C., & Frankenstein, S. (2015). Time series anomaly discovery with grammar-based compressions. In *Proceedings of the 18th International Conference on Extending Database Technology (EDBT)*. Brussels, Belgium. OCLC: 951739373.

Shin, H., Park, H., Lee, J., & Jhee, W. C. (2012). A scoring model to detect abusive billing patterns in health insurance claims. *Expert Systems with Applications*, *39*, 7441–7450.

Streit, R. E., & Borenstein, D. (2009). An agent-based simulation model for analyzing the governance of the Brazilian Financial System. *Expert Systems with Applications*, *36*, 11489–11501.

Streit, R. E., & Borenstein, D. (2012). Design and development of a fuzzy agent-based model to measure interest rate expectations. *Expert Systems with Applications*, *39*, 7391–7402.

Šubelj, L., Furlan, Š., & Bajec, M. (2011). An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications*, *38*, 1039–1052.

Taylor, G. (2008). A simple model of insurance market dynamics. *North American Actuarial Journal*, *12*, 242–262.

Thuring, F., Nielsen, J., Guillén, M., & Bolancé, C. (2012). Selecting prospects for cross-selling financial products using multivariate credibility. *Expert Systems with Applications*, *39*, 8809–8816.

Tsomocos, D. (2003). Equilibrium analysis, banking and financial instability. *Journal of Mathematical Economics*, *39*, 619–655.

Wang, S. S., Major, J. A., Pan, H. C., & Leong, J. W. (2011). US Property-Casualty: Underwriting Cycle Modeling and Risk Benchmarks. *Variance (Casualty Actuarial Society)*, *5*, 91–114.

Warthen, T., & Somner, D. (1996). Dynamic financial modeling–issues and approaches. *CAS Forum (Casualty Actuarial Society)*, *Spring 1996*, 291–328.

Wu, C., Kao, S., Su, Y., & Wu, C. (2005). Targeting customers via discovery knowledge for the insurance industry. *Expert Systems with Applications*, *29*, 291–299.