Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

AN AGENT-BASED COMPUTATIONAL MODEL OF THE INDIVIDUAL HEALTH INSURANCE MARKET

Kevin T. Comer

Modeling, Simulation, Experimentation, and Analysis (MSEA) Technical Center
MITRE Corporation
7515 Colshire Drive
McLean, VA 22102, USA

ABSTRACT

This paper focuses on constructing an abstract baseline model of agents, simulating buyers and sellers of health insurance, in the individual health insurance market. Each buyer assesses their expected medical expenses, surveys the insurance contracts available on the market, and decides each year whether to buy insurance, and which plan they will choose. Each seller, upon seeing the costs to provide medical care to its subscriber base, will update the premium price it offers to the market to remain profitable. The aim of this paper is two-fold – to produce a self-sustaining representation of the insurance market that qualitatively represents the health insurance market without any external policy implementations and to quantitatively replicate the large-scale metrics seen in the insurance market, most notably the price elasticity of demand for insurance.

1 INTRODUCTION

1.1 Empirical Studies of the Health Insurance Market

The most notable and widely cited study of the health insurance market is the RAND Health Insurance Experiment (HIE), consisting of 7,791 individuals over three years and six U.S. cities. At the time of this writing (2018), it remains the largest health policy study in U.S. History. The primary aim of the study was to determine how much more medical care people will use if it is provided free of charge – RAND Corporation (2006). As noted from the focus, the goal of the study was to assess moral hazard, or the concern of excessive use of medical care without the controlling factor of price.

One of the more difficult metrics to determine about the health insurance market is the price elasticity of demand – how much the quantity of a good sold on the market will change given a set change in the price of that good. It was noted in Feldstein (1971) and Cutler and Zeckhauser (2000) that, prior to the RAND HIE, medical care was assumed to be a need, and that the demand was completely inelastic. The RAND study, as well as analyses from Phelps and Newhouse (1972) and Feldstein (1977), determined this was not the case. Estimates vary widely on the price elasticity of demand, which is dependent on a wide variety of factors, including employment, income, and age. The RAND HIE results estimated an overall price elasticity of -0.60. In other words, when the price of insurance increases by 1.0%, the quantity sold is expected to decrease by 0.60%.

Estimates on this measure in the real world vary considerably, from -1.50 in Rosett and Huang (1973) to -0.14 in Phelps and Newhouse (1972). This range is only widened when observing the market-wide elasticity metric for supplementary health insurance, as assessed in Marquis and Phelps (1987), ranging from near 0 to over -2.0. An exhaustive review of price elasticity estimates is summarized in a table in Cutler and Zeckhauser (2000). Marquis and Long (1995) discuss the price elasticity of demand of the non-group market – that is, the market of insurance buyers not purchasing insurance through their employer which they estimate at -0.40, based on a combination of empirical survey data and constructed premium prices from companies that develop individual plans.

1.2 Theoretical Models of the Health Insurance Market

While the health insurance market is an enormous system, both complex and adaptive, many efforts have been published to model certain interactions between large bodies of actors in the system. Cutler and Zeckhauser (2000) discuss the health insurance market in broad, abstract terms, as a three-body problem – patient, payer, and provider. They then discuss the analysis of plans through a game theoretical model, between high and low risk patients, to determine employers' payment option strategies for their workforce. They then examine this system empirically in their case study of Harvard University and Massachusetts Group Insurance Commission (GIC).

Marquis and Holmer (1986) used much of the data from the RAND HIE to inform a discrete choice model for a RAND study. Work prior to Marquis and Holmer focused on simulating rational choice for patients in the insurance market. However, this demonstrated a significant sensitivity to price changes among the patient population when choosing health insurance plans that was not validated by subsequent research by Holmer (1984) and Marquis and Phelps (1987). This model uses non-stochastic terms derived from the survey, as well as stochastic error terms, to develop and simulate a computational model representing families choosing plans for supplemental health insurance. This model much more closely represented the price elasticity of demand witnessed in the market for health insurance, in contrast to previous expected utility models. Marquis and Holmer (1996) updated their discrete choice model to incorporate more recent models of uncertainty, including Prospect Theory from Tversky and Kahneman (1981) which is the precursor to the Cumulative Prospect Theory discussed in Tversky and Kahneman (1992) and used later in this paper.

More recent models have been developed in the form of microsimulations. Most notable are the Health Insurance Reform Simulation Model (HIRSM) in 2003, as discussed in Blumberg et al. (2003), and the Health Insurance Policy Simulation Model (HIPSM) in 2011, as discussed in Buettgens (2011). Both models focus on all significant insurance markets, including the non-group market, and produce a synthetic population of actors (patients, as well as employers) from Current Population Surveys and the National Medical Expenditure Surveys. Both models use current law as the baseline model, then assess changes to the law on the population and each simulated actor's behavior. These models have been used to predict coverage percentages, multi-year costs, and premiums, as examined by Urban Institute Health Policy Center (2014), The models have also been used to assess the impact of removing elements of the Affordable Care Act (ACA) – such as the Individual Mandate – as discussed in Blumberg et al. (2013).

1.3 Cost and Income Parameters

Many studies involved in estimating the cost of medical care have used a reductionist approach, establishing cost as a type of linear aggregation model of a combination of factors – prime example being age, location, and pre-existing health conditions. Few studies have attempted to arrive at a non-parametric approach – defining health care costs as a random variable with relatively few attributes. One such methodology, developed by Feenberg and Skinner (1992) and further explored by Rust and Phelan (1997) and French and Jones (2004), establishes both the cross-sectional distribution of health care costs, and the dependence of health care costs on previous cycles. Feenberg and Skinner (1992) analyze medical deductions from tax returns to establish an autoregressive, or AR, function to estimate the medical costs of an individual. Rust and Phelan (1997) and French and Jones (2004) build upon this in developing an autoregressive moving average function, or ARMA, incorporating both the AR function developed by Feenberg and Skinner, and adding a moving average, or MA, component to form a stationary stochastic process of two polynomials, for the AR and MA functions.

In addition to revising the process as an ARMA function, Rust and Phelan (1997) and French and Jones (2004) assess the heaviness of the right tail in the cross-sectional distribution of health care costs rather than a perfect lognormal distribution. This is particularly important in situations of the sickest individuals who are most likely to purchase health insurance. This model uses the parameters discussed in French and Jones (2004), as it is based on longitudinal medical expenditure surveys, multi-year data from the Health and Retirement Survey (HRS) by Juster and Suzman (1995), and the Assets and Health Dynamics of the Oldest Old (AHEAD) survey, as in Soldo et al. (1997), rather than expenses only deemed expensive enough to declare on tax returns. The downside of this data-set is the higher average age of the survey population, as the survey was centered on the costs of people prior to retirement (ages 51-65) and after retirement (65+) rather than a true representation of the entire population.

In a similar fashion to medical costs, income parameters are determined to similarly exhibit a heavy right tail, as depicted by Drăgulescu and Yakovenko (2001) and Clementi and Gallegati (2005). However, for the lower 95% of the population, the distribution can be depicted as a reasonably straight-forward exponential or lognormal function, depending on whether there exists a cut-off in the dataset, with a finite mean and variance. Drăgulescu and Yakovenko (2001) explicitly cite an exponential function with a mean of \$36.4K for the United States, and the right tail – representing the top 5% of the population – as having a power-law distribution, with an α value of 1.7. This data, like the medical care expenditure data presented in French and Jones (2004), is based on data from 1998, so the income and spending parameters are in similar year dollars.

2 METHODOLOGY

This paper is focused on the modeling of a functioning non-group health insurance market, using an agent-based computational model representing consumers (patients) and suppliers (payers) of health insurance. This model, coded in Python, attempts to replicate the interactions of patients and payers in a market using a small number of variables and parameters, to determine the fundamental elements of the individual health insurance market.

2.1 Patients

This model is run with 100,000 patient agents, choosing between ten plans over 20 years. In each simulated year, the patient agent assesses its current plan and all other plans being offered in the market. The patient agent can choose to keep its plan, move to a new plan, or forsake its plan to go uninsured. Similarly, each year, every patient agent earns an income, derived from the work of Drăgulescu and Yakovenko (2001), and has a medical expense, derived from the work of French and Jones (2004). Assessing these items, as well as the risk of a catastrophic event using a risk valuation assessment as presented in Tversky and Kahneman (1992), feeds into how each patient agent determines what to pay for health insurance.

The model presented within this paper initializes patient agents with a salary and an initial medical expense. As a distinction from many former models discussed in Section 1, this model does not consider age, gender, or other demographic information, and only looks at the medical expenditures that are expected as functions of the previous year, with the possibility of a catastrophic shock occurring.

The patient agents are initialized with a salary, determined by random variable from the probability distribution $Income \sim Exp(\lambda)$ where $\lambda = 36400$ represents the mean income of the lower-95% of the U.S. population, as was defined by Drăgulescu and Yakovenko (2001). As noted in Drăgulescu and Yakovenko (2001) and Clementi and Gallegati (2005), income distribution in the U.S. and in most developed nations has a right heavy tail which does not conform to an exponential function or even a log-normal function, but rather a power-law function. Given that the probability density function (PDF) of the exponential function provides the 95%-percentile at 109450, this becomes the x_{min} value for the conjoining power-law distribution, with an associated alpha value $\alpha = 1.7$.

The patient agents are also initialized with a starting medical expense, from the cross-sectional distribution determined by French and Jones (2004): $Care\ Cost \sim lognormal(6.69{,}2.11)$. This is the fitted lognormal cross-sectional distribution they determined fit the medical expenditure data they analyzed.

However, there is a heavy right tail, representing the most expensive medical care that might occur from a catastrophic event. The authors determine that the heavy tail has a Pareto distribution, with $x_{min} = 33800$ (the 99.5-percentile of the distribution) and $\alpha = 0.291$. As $\alpha < 1$, the distribution does not have a well-defined mean or expected value. It therefore does not make any sense to take an expected value of this distribution. Therefore, each patient agent is initialized with a randomly generated "estimated catastrophe" value, which is what the patient agent uses to assess the value of health insurance.

Each year, every patient agent is activated sequentially and asked to estimate its medical expenses for the coming year. They do this using a risk valuation function described in Cumulative Prospect Theory, from Tversky and Kahneman (1992), assessing the most likely medical expense as equivalent to the previous year. However, there is a small probability of a shock, estimated at 0.50% chance, that the patient agent might suffer a catastrophic event, which would cost an amount the patient agent estimated at initialization. As described in Cumulative Prospect Theory, the subjective estimation that each patient agent makes of the probability of a catastrophic event occurring is:

$$W(\rho, \delta) = \frac{\rho^{\delta}}{((1-\rho)^{\delta} + \rho^{\delta})^{1/\delta}} \tag{1}$$

Where $\rho = 0.005$, or the objective probability of a catastrophic event occurring, and $\delta = 0.69$, or the discounting factor established in Tversky and Kahneman (1992). Similarly, the estimated costs of the two possibilities are calculated as:

$$V_1(v_{t-1}, \lambda, \alpha) = \lambda v_{t-1}^{\alpha} \tag{2}$$

$$V_2(\chi, \lambda, \alpha) = \lambda \chi^{\alpha} \tag{3}$$

Where v_{t-1} is the actual cost from last year, χ is the cost of a predicted catastrophe, and $\lambda = 2.25$ and $\alpha = 0.88$ are parameters established by Tversky and Kahneman (1992). These three equations are then used to determine the estimated expenses in the next year:

$$\text{Estimated expenses} = (1 - W) \times V_1 + W \times V_2 \tag{4}$$

After the patient agent has calculated its expected expenses, it assesses the plans that are currently on the market, and will purchase the plan that provides the lowest expected expenditure. If the patient already has a plan, they will switch if the price of the new plan is less than the current expenditure divided by an *Inertia* parameter, defined by the author as 1.25. This is calculated using the premium price, the coinsurance rate, and the out of pocket maximum (OOP_Max) provided by the payer agent, as well as the estimated expenses of the patient agent.

$$\text{Insured Estimate} = \text{Premium} + \begin{cases} \text{Coinsurance} \times \text{Estimated Expenses if Estimated Expenses} \leq \text{Out of Pocket Maximum} \\ \text{Coinsurance} \times \text{OOP\_Max if Estimated Expenses} > \text{Out of Pocket Maximum} \end{cases} \tag{5}$$

After choosing a plan, the patient agent can drop its insurance plan, if the estimated costs of the plan are higher than the patient agent anticipates its medical expenses to be. The patient agent may also drop the plan if the costs of the plan are higher than the patient agent's salary.

After the estimations have been performed, the patient agent's actual medical expenditure is calculated, using an autoregressive moving average function, ARMA(1, 1) adapted from French and Jones (2004):

$$X_t = X_{t-1} + \epsilon_t + \theta\epsilon_{t-1} \tag{6}$$

$$\text{Care Cost} = e^{X_t} \tag{7}$$

Where X_{t-1} is the natural log of the previous year's medical expenditure, ε_t is a normally distributed error function $\varepsilon_t \sim N(0, 0.039)$, ε_{t-1} is the previous year's error term, and $\theta = 0.104$.

There is a 0.50% ($\rho = 0.005$) chance of a catastrophic event occurring, where the cost of care instead becomes a random variable with a Pareto distribution, "$Care\ Cost \sim Pareto(33800, 0.291)$". While this is a similar function to the one used to initialize each patient agent's estimation of a catastrophic event, there is no relation between the estimated cost and the actual cost of such an event.

Finally, the patient agent gets care. If the patient agent is not insured, then the care is paid for out of its wealth. If the patient agent has a plan to which it is a subscriber, then it pays the premium plus the care cost times the coinsurance rate, and the payer agent pays for the rest. If the care cost is greater than the Out of Pocket Maximum of the payer agent, than the patient agent simply pays the Out of Pocket Maximum times the coinsurance rate, plus the premium, as discussed in Cutler and Zeckhauser (2000).

2.2 Payers

This model is run with 10 payer agents over 20 years. Each simulated year, the payer agent chooses to accept or reject applicants to plans, and can drop subscribers who are too expensive to cover. It will pay for medical care for its subscribers and will then adjust its premium price to cover care for its subscriber base.

Each payer agent is initialized with a budget equal to the Out of Pocket Maximum of the plan, times the total number of patient agents in the model (set to 100K). Given that the baseline number of Out of Pocket Maximum (labeled OOP_Max in the simulation model) for all payer agents is 25,000 or 25K, this amount is set to $250M for all payer agents. The premium price is initialized as a random variable with a Gaussian distribution: $Premium \sim N(1000,200)$. The coinsurance rate is a discrete choice between 0.10, 0.25, and 0.50, each with equal probability, following the pattern laid out by Marquis and Holmer (1986). Lifetime Maximum (labeled Lifetime_Max in the model) is set to 100K. Elasticity is initially set to 0.

Each year, payer agents have limited actions they can perform while the patient agents decide between plans. When each patient agent decides to apply for a plan provided by a payer agent, the payer agent then performs medical underwriting. If the three (3) previous medical expenses have all been greater than a given value, calculated as "$OOP\_Max \times Coinsurance$", then the payer agent will reject the patient agent's application. Otherwise, the patient agent will be accepted and will be a subscriber for that year. Similarly, if a subscriber has a medical expense in any given year that is greater than the lifetime maximum established by the plan, then the payer agent can drop that patient agent as a subscriber.

After each patient agent has been activated, and their actual medical expenditures have been calculated and paid for, the payer agents are then activated sequentially to determine by how much they should raise their premium prices. First, the budget is updated, subtracting the total costs of all subscribers and adding the premium price multiplied by the number of subscribers currently on the books, as well as an administrative cost factor, defined at 1.15 by the author.

Next, the payer agent calculates the costs of the past five (5) years and the revenue generated by the past five (5) years, producing a simple autoregressive function of the five previous turns, AR(5) or ARMA(5,0) as there is no moving average component.

$$Cost_t = AdmitCostFactor \times \sum_{j=1}^{n} (1 - coinsurance) \times CareCost_j \tag{8}$$

$$Revenue_t = \sum_{j=0}^{n} Premium_t \tag{9}$$

$$\text{Profit} = \sum_{i=0}^{4} (Revenue_{t-i} - Cost_{t-i}) \tag{10}$$

Where n is the number of subscribers currently listed under the payer agent at year t, and $CareCost_j$ is the medical expenditure of subscribed patient agent j.

If the revenue is greater than the costs, the payer agent does nothing. If the costs are greater than the revenue, the payer agent then divides the difference by the number of subscribers, and adds this to the premium. If the difference is greater than the Premium Buffer times the current Premium value, then the payer agent increases the premium by the current premium times the premium buffer.

$$If\ \text{Profit} \geq 0, Premium_t = Premium_{t-1} \tag{11}$$

$$Else, Premium_t = Premium_{t-1} - \begin{cases} \dfrac{Profit}{Subscribers_{t-1}} \ if\ |Profit| < (Premium\ Buffer\ \times Premium_{t-1}) \\ \dfrac{Premium\ Buffer\ \times Premium_{t-1}}{Subscribers_{t-1}} \ if\ |Profit| > (Premium\ Buffer\ \times Premium_{t-1}) \end{cases} \tag{12}$$

If a plan has zero subscribers, the payer agent proceeds to try to make the plan more attractive to possible subscribers. The payer agent will decrease the premium by the premium buffer. If a payer agent ends up with a budget that is negative, then the payer agent will leave the market. All current subscribers will be dumped, and their insurance plan will be set to None. The payer agent will no longer be able to change their premium price, nor accept new patient agents.

Throughout this process of revising and changing their premium price offered to the market, the payer agents are learning and adapting to the circumstances of the market. If the price is too high, patient agents will not buy the payer agent's plan, and the payer agent will need to lower the premium to capture market share. If the price is too low, the payer agent risks losing money if the medical costs of their subscriber base is above the revenue collected. Rather than objectively determining the optimum price in the market, like what is performed in the neo-classical economic field, this model depicts a payer agent "groping" for a feasible solution through experimentation and feedback, in a process Herbert A. Simon termed "satisficing" (1956), as a form of learning and adaptation.

3 METRICS

3.1 Price Elasticity of Demand

The most fundamental output metric of this computational model, in terms of testing for the validity of the model, is the price elasticity of demand of the market. As premium prices increase, it is expected that the number of subscribers of any given plan will decrease. For every payer agent, this can be calculated each year of a simulation:

$$\text{Elasticity} = \frac{\text{Premium}}{\text{Subscribers}} \times \frac{\Delta\text{Subscribers}}{\Delta\text{Premium}} \tag{13}$$

where the subscribers and the premium are the number of subscribers and the premium price in a given year respectively, and the delta in subscribers is the number of subscribers that were gained or lost in the previous year, and the delta in the premium is the amount that the premium price changed in a given year. This becomes complicated when a premium does not change in a year (which is possible when the payer agent is operating at a net positive), so this value is calculated when the premium changes in a given year, rather than for every year of a simulation.

As discussed in the previous section, estimates on this measure in the real world vary wildly, from -1.50 in Rosett and Huang (1973) to -0.14 in Phelps and Newhouse (1972). An exhaustive review of price elasticity estimates in the literature is summarized in a table in Cutler and Zeckhauser (2000). The most notable, and often most cited, estimate of price elasticity of demand for health insurance comes from the previously cited RAND Health Insurance Experiment, as well as replicated by Marquis and Phelps (1987), which estimates an elasticity metric of -0.60. This will be the goal of the computational model, in terms of validation.

3.2 Number of Subscribers

The number of patient agents subscribed to a payer agent will be a secondary measure of performance for the computational model. In this model, the system should operate with a sizable portion of the population being insured, but with a significant portion of the population being uninsured. They can be uninsured for a variety of reasons, which will be captured by the model:

- **Rejected:** the patient agent's cost of care is too expensive for any insurers to insure against
- **Cheap:** the patient agent's salary is not enough to pay for the insurance it chooses
- **Volunteer:** the patient agent can afford to pay for a plan, but chose not to do so, because its estimated expenses are less than what it expected to pay for insurance
- **Dumped:** the patient agent's plan dumps it, either by leaving the market or not being able to pay for the patient agent's care.

4 ANALYSIS

4.1 Price Elasticity of Demand

The price elasticities of demand were calculated from those plans that had a change in both premium prices and subscribers. To be precise, this metric is a dataset of price point elasticities, rather than a global elasticity of demand. The graph in Figure 1 depicts all observations of elasticity – 794 observations in total – as a cumulative distribution function, depicting percentage of observations less than or equal to the price elasticity value observed.

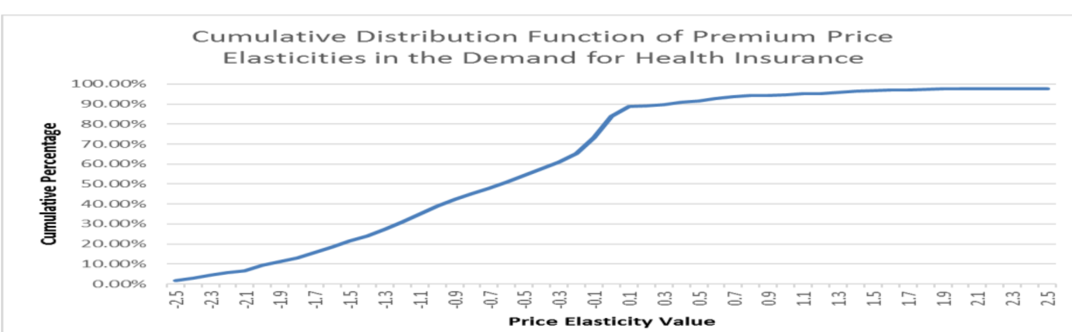

Figure 1: Cumulative Distribution Function of observations of Price Elasticity of Demand over 25 Runs.

The statistical measurements of the observations are misleading because there are some significant outliers, six values that are greater than 10, and one value less than -10. After removing these seven outliers, the metrics become much more descriptive, with mean (-0.68) and median (-0.65) values close to the price elasticity value determined by the RAND Study from empirical evidence (-0.60). Table 1 shows the statistics of both the raw dataset, and with the outliers removed.

Table 1: Descriptive Statistics for Price-Point Elasticity Values Dataset.

Statistic	Original Dataset	Outliers Removed
Mean	2.8871	-0.6783
Median	-0.6359	-0.6451
Variance	8696.57	1.2107
Standard Deviation	93.2554	1.1003
Skewness	27.9268	1.4906

One element of note is that the skewness is positive, rather than negative, which is at odds with what is seen in Figure 1 or the rest of the statistical data. It is important to note that skewness does not differentiate between fatness and length of the tail, and while only 15.4% of the observations are positive, they are much more uniform in their distribution, rather than concentrated. As a positive price elasticity value is odd enough in the real world, this may indicate the effects of a force other than the premium price being at play. Most notably, this could occur with the dropping of a patient agent by one plan and being picked up by another, regardless of the price of the premium, as what has happened in recent years with the Health Insurance Marketplace as discussed in Holahan et al. (2014).

While it can be shown graphically that the entire distribution is most likely not normally distributed, more quantitative measurements were calculated to confirm this. A Kolmogorov-Smirnov test versus a random variable distribution of equal size ($N = 795$), population mean and standard deviation computed a p-value of 2.2×10^{-16} ($D = 0.4817$), well below any reasonable value of α, leading to a rejection of the null hypothesis that this is a normally distributed population. Similarly, a Shapiro-Wilk test for normalcy also arrived at the same value for p as the Kolmogorov-Smirnov test, with a W-statistic of 0.0212.

The price point elasticity values that are positive, 15.4% of the observations ($n = 131$), can be shown to come from a different distribution than those that are negative ($n = 666$), using the Mann-Whitney U Test, also known as a Wilcoxon rank-sum test, on the change in premium price over the previous year. A normal distribution cannot be assumed for this population, as many of the values of premium increases are tied at the PremiumBuffer global value (baseline value = 0.20). With 95% confidence level, the null hypothesis ($U_1 = U_2$) can be rejected ($p = 7.27 \times 10^{-11}$), as changes in premium price resulting in negative elasticity are greater (one-sided test, $p = 3.64 \times 10^{-11}$) than those resulting in positive elasticity.

The statistics of the two points are assessed separately, using KS statistics from the Kolmogorov-Smirnov tests, and A-statistic from Anderson-Darling tests. The absolute values of the negative values of elasticity rejected every null hypothesis presented with regards to Kolmogorov-Smirnov Tests (prone to Type II errors, insensitive to larger values) and Anderson-Darling Tests (prone to Type I errors) – results are shown in Table 2.

Table 2: Test Statistics for Negative Elasticity Values.

Estimated Parameter 1	Estimated Parameter 2	KS-Test Statistic	AD-test Statistic
Lognormal	$\mu = -0.1463$	$\sigma = 0.8299$	0.1154 ($p = 0.0$)
Exponential	$\lambda = 1.0989$	N/A	0.0868 ($p = 0.0$)
Power Law	$\alpha = 4.6988$	$x_{min} = 1.5502$	0.5107 ($p = 0.0$)
Gamma	$k = 0.0088$	$\theta = 1.5663$	0.9508 ($p = 0.0$)

The distribution of the positive values of elasticity rejected every null hypothesis except for the lognormal distribution (via the Kolmogorov-Smirnov Test, given a confidence level of α =0.050), as demonstrated in Table 3.

Table 3: Test Statistics for Positive Elasticity Values.

Estimated Parameter 1	Estimated Parameter 2	KS-Test Statistic	AD-test Statistic
Lognormal	$\mu = -0.0002$	$\sigma = 0.3507$	$0.1061\ (p = 0.1044)$
Exponential	$\lambda = 28.3347$	N/A	$0.8009\ (p = 0.0)$
Power Law	$\alpha = 1.7887$	$x_{min} = 0.5076$	$0.4335\ (p = 0.0)$
Gamma	$k = 0.0508$	$\theta = 1570.95$	$0.5214\ (p = 0.0)$

4.2 Number of Subscribers

Shown in Figure 2 are the maximum, average, and minimum values of patient agent participation in the insurance market across the 25 runs of the model. What is interesting to note is the sinusoidal pattern of participation in the market, with a period of roughly ten years. This pattern remains consistent across all of the runs, which indicates an artifact of the system, and the patient agents responding to the changes in premiums, which respond in a delayed fashion to level of subscription across the patient agent population.

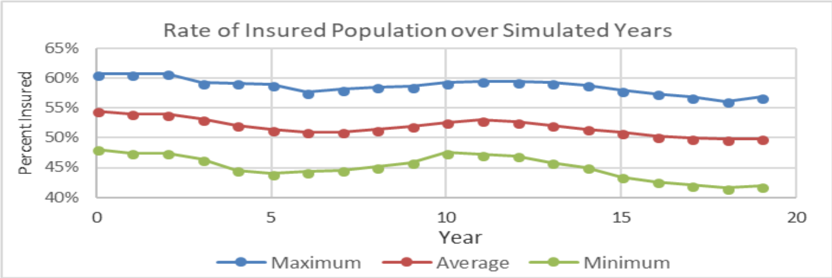

Figure 2: The maximum, average, and minimum values for proportion of total patient agents insured over 25 runs, by year.

With those that choose not to purchase insurance, there are a few reasons that patient agents may not be insured – being refused, having too little income to afford it, being dumped after having insurance, or choosing to forgo insurance because they do not believe it will benefit them. As can be shown in Figure 3, while the number of rejected and dumped stay reasonably constant throughout each run, the number of those who choose not to buy insurance, either not being able to afford it or making the rational decision to forgo insurance, rises slowly, which is consistent with the rising of premium prices through the simulation run.

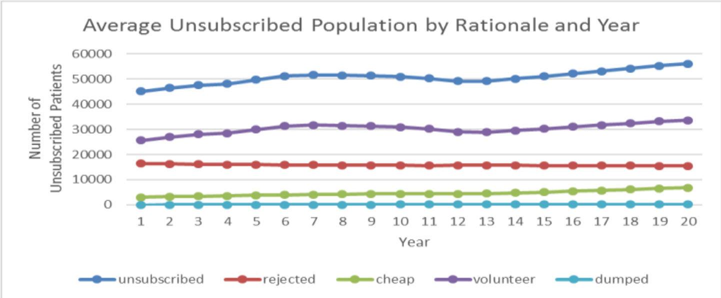

Figure 3: Average Unsubscribed Population by Rationale and Year Across 25 Runs of the Baseline Model.

5 DISCUSSION

5.1 Discussion of Findings

The price elasticities and subscribers have been shown to exhibit qualities like those witnessed in the real-world market, such as primarily negative elasticity values centered around -0.60, as discussed in RAND Corporation (2006), and several insured individuals that remains constant throughout the results of the simulations. Several observations of price elasticity were shown to be positive, which is contrary to the behavior of most consumer goods. However, at points when patient agents moved to a plan despite the premium price increase, the patient agent was often leaving a plan that had an even greater proportionate price increase. As the availability of substitute goods increase, such as the case of multiple available insurance plans, so too would the price elasticity. It should be noted that, while the positive values of price elasticity followed a lognormal distribution, the negative values followed no such distribution. This indicates that the underlying behavior of the distributions may be caused by different phenomena.

It is an important finding that, even in the absence of any policy or significant exogenous effects, the subscriber base or number of insured persons remains stable and shows a similar cyclical pattern across all runs. This indicates a negative feedback in the model that is not explicitly coded into either the payer agent or patient agent behavior.

Many of the parameter values of the model, such as Inertia, Administrative Cost Factor, and Premium Buffer, were defined by the author. Performing sensitivity analysis using a Mann-Whitney U test on changing these values and their impact on the values of elasticity, there is a statistically significant difference in the price elasticity given 10% increase or decrease in these values, using a two-sides alternative hypothesis (H_1: $U_1 \neq U_2$). Further research would need to be done to identify what the empirical values of these parameter are in the real market, before this can be used for experimentation.

5.2 Discussion of Broader Implications

This model is a significant departure from many previous models on health insurance in several ways. While the model is driven by certain distributions based on empirical data, such as income and cost of medical care, the number of parameters needed to drive the model is rather small compared to many other microsimulation models of the subject matter, as in Blumberg et al. (2003) and Buettgens (2011). Even still, the model produces a result that is, to a reasonable degree, like what is observed in the real world, with regards to price elasticity of demand, number of insured, and premium prices.

A more notable departure from previous models of health insurance is the addition of heavy tails to the cost parameters, as in French and Jones (2004) and income parameters, as in Drăgulescu and Yakovenko (2001) and Clementi and Gallegati (2005). Both medical costs and household income have been shown to exhibit power law distributions at the highest values. For medical costs at least, this becomes enormously important to understanding the true costs of the system to the payers.

Finally, this model was run at true scale, representing one-for-one the non-group population one may observe in an average U.S. state. In 2013, the median number of non-group individuals per state was 169K, and 15 states had non-group populations of less than 100K, from Kaiser Family Foundation (2015). The scale of the model can impact the volatility and dynamics of the results, and so it was important to come as close to true scale as was feasible, given computing restraints.

5.3 Further Research

While this model clearly demonstrates the ability to replicate the qualitative patterns witnessed in the real-world health insurance market, further refinement of the global variables defining the model would be required to more thoroughly dock this quantitatively with other aspects of the health insurance market. Most notably, this would pertain to replicating the empirical data observed across the market for health insurance premium prices. If this baseline could replicate the distribution of insurance premiums pre-ACA, the output for any policy implementations of this model would have a much higher degree of quantitative fidelity.

This model uses 1998 data for both income and medical expenditures, so that spending patterns and the spending capacity of individuals remains chronologically consistent. A more comprehensive data set is available in the National Medical Expenditure Panel Survey (MEPS), which breaks out average medical expenditures across the population by a wide variety of subdivisions, based on age, sex, insured status, income, etc. However, one of the main reasons why these data were not used was the lack of granularity of the data available from survey queries, only calculating metrics for subsets of populations rather than data points of individual patients and households.

6 SUMMARY

This model and the subsequent results represent a considerable addition to the recent literature on modeling the health insurance market. While the current literature on computational modeling for the health insurance market is packed with discrete choice models and microsimulations using large data sets to formulate projected rates and population dynamics, this model has arrived at valid values of emergent properties of the health insurance market with only a handful of empirical parameter distributions, namely income and medical expenditure. The results of the simulations of the model can be shown to replicate the same price-point elasticity of demand that is seen in the health insurance market, given the findings of the RAND Health Insurance Experiment (2006), as well as the percentage of insured individuals across the simulation runs. This can be performed at a scale of one-to-one with current non-group markets for U.S. states, which is fundamental to understanding the volatility of the insurance market and prices. While admittedly abstract, this computational model can provide significant insight into the behavior of markets given certain elements of legislation affecting the purchase of health insurance in the individual market.

DISCLAIMER

The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author. Approved for Public Release ; Distribution Unlimited. Case Number 18-0507.

REFERENCES

Blumberg, L. J., J. Holahan, and M. Buettgens. 2013. “It's No Contest: The ACA's Employer Mandate Has Far Less Effect on Coverage and Costs Than The Individual Mandate.” The Urban Institute, Washington, D.C.

Blumberg, L. J., Y. C. Shen, L. Nichols, M. Buettgens, L. C. Dubay, and S. Mcmorrow. 2003. “The Health Insurance Reform Simulation Model (HIRSM): Methodological Detail and Prototypical Simulation Results.” The Urban Institute, Washington, D.C.

Buettgens, M. 2011. “Health Insurance Policy Simulation Model (HIPSM) Methodology Documentation.” The Urban Institute, Washington, D.C.

Clementi, F. and M. Gallegati. 2005. “Pareto's Law Of Income Distribution: Evidence For Germany, The United Kingdom, And The United States.” In *Econophysics For Wealth Distributions*, Edited By A. Chatterjee, S. Yarlagadda, B. K. Chakrabarti, 3-14 , Milano: Springer.

Cutler, D. M. and R. J. Zeckhauser. 2000. “The Anatomy of Health Insurance.” *Handbook of Health Economics* 1: 563-643

Drăgulescu, A. and V. Yakovenko. 2001. “Exponential and Power-Law Probability Distributions of Wealth And Income In The United Kingdom And The United States.” *Physica A: Statistical Mechanics and Its Applications* 299: 213-221.

Feenberg, D. and J. Skinner. 1992. "The Risk and Duration Of Catastrophic Health Care Expenditures." National Bureau Of Economic Research, Washington, D.C.
Feldstein, M. S. 1971. "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics." *The American Economic Review* 61(5): 853-872.
Feldstein, M. S. 1977. "The Rising Cost of Hospital Care." Information Resources Press, Washington, D.C.
French, E. and J. B. Jones. 2004. "On the Distribution and Dynamics of Health Care Costs." *Journal of Applied Econometrics* 19(6): 705-721.
Holahan, J., L. Blumberg, and M. Buettgens. 2014. "Will Those with Cancelled Insurance Policies Be Better Off In ACA Marketplaces?" The Urban Institute, Washington, D.C.
Holmer, M. 1984. "Tax Policy and The Demand for Health Insurance." *Journal of Health Economics* 3(3): 203-221.
Juster, F. T. and R. Suzman. 1995. "An Overview of The Health and Retirement Study." *The Journal of Human Resources* 30: 7-56.
Kaiser Family Foundation. 2015. Health Insurance Coverage of The Total Population, Accessed January 15th, 2017.
Marquis, M. S. and M. R. Holmer. 1986. "Choice Under Uncertainty and The Demand for Health Insurance." RAND Corporation, Santa Monica, CA.
Marquis, M. S. and M. R. Holmer. 1996. "Alternative Models Of Choice Under Uncertainty And Demand For Health Insurance." *The Review of Economics and Statistics* 78(3): 421-427.
Marquis, M. S. and S. H. Long. 1995. "Worker Demand for Health Insurance In The Non-Group Market." *Journal Of Health Economics* 14(1): 47-63.
Marquis, M. S. and C. E. Phelps. 1987. "Price Elasticity And Adverse Selection In The Demand For Supplementary Health Insurance." *Economic Inquiry* 25(2): 299-315.
Phelps, C. E. and J. P. Newhouse. 1972. "The Effects of Coinsurance on Demand For Physician Services." RAND Corporation, Santa Monica, CA.
RAND Corporation. 2006. "The Health Insurance Experiment." RAND Corporation, Santa Monica, CA.
Rosett, R. N., and Huang, L.-F. 1973. "The Effect of Health Insurance on The Demand For Medical Care." *Journal Of Political Economy* 81(2): 281-305.
Rust, J. and C. Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior In A World Of Incomplete Markets." *Econometrica Journal of The Econometric Society* 65(4): 781-831.
Simon, H. A. 1956. "Rational Choice and The Structure of The Environment." *Psychological Review* 63(2): 129-138.
Soldo, B. J., M. D. Hurd, W. L. Rodgers, and R. B. Wallace. 1997. "Asset And Health Dynamics Among The Oldest Old: An Overview Of The AHEAD Study." *The Journals Of Gerontology Series B: Psychological Sciences And Social Sciences* 52B: 1-20.
Tversky, A. and D. Kahneman. 1981. "The Framing Of Decisions And The Psychology of Choice." *Science* 211(30): 453-458.
Tversky, A. and D. Kahneman. 1992. "Advances In Prospect Theory: Cumulative Representation Of Uncertainty." *Journal Of Risk And Uncertainty* 5(4): 297-323.
Urban Institute Health Policy Center. 2014. "The Urban Institute's Health Microsimulation Capabilities." The Urban Institute, Washington, D.C.

AUTHOR BIOGRAPHIES

KEVIN COMER is a Simulation Modeling Engineer at the MITRE Corporation, a not-for-profit federally funded research and development center, located in McLean, Virginia. He holds a Ph.D. in computational social science from George Mason University. His research interests include behavioral economics, agent-based modeling, and quantitative policy analysis. His e-mail address is ktcomer@mitre.org.