Insurance Probability:

Comparison & Analysis of Consumer Psychology and Pricing Strategy in U.S. and China

By

Wenjun Dong, Wanyue Zhang, and Yan Dai

December 12, 2011

Background

After the global economic downturn, the economies of the United States and China demonstrate substantial growth opportunities. Among many enterprises, insurance is one hopeful industry as more and more people are taking initiatives to protect themselves against risks in terms of lives, properties, investments, etc. While the insurance industry has been well-established in the U.S., the insurance industry in China is only beginning to grow rapidly.

Prior to China's economic reforms in the late 1970's, the insurance industry in China was non-existing as most properties and businesses were government owned. [1]When the insurance industry was reinstated in the late 1970's and early 1980's, property insurance and especially life insurance were heavily promoted. According to the Swiss Re sigma study, the annual life insurance premium in 1999 totaled ten billion US dollars.[2] Upon joining the World Trade Organization in 2001, China loosened its regulations and allowed foreign companies to participate in the insurance industry. Life-insurance premiums in 2006 grew to forty-six billion US dollars. The latest Swiss Re sigma study reports that China's total insurance premium grew 15.79% in 2009; life insurance premium alone increased by 13.9%.[3]

Due to many issues including socio-political structures and economic circumstances, insurance in the U.S. and China is quite different in many ways. This project aims to explore and analyze their similarities and differences.

---

[1] "China Insurance Industry | Economy Watch." *Economy Watch*. 26 May 2010. Web. 11 Dec. 2011. <http://www.economywatch.com/insurance/china-insurance-industry.html>.

[2] *World Insurance in 2006*. Rep. Vol. 4. Zurich: Swiss Reinsurance. Ser. 2007. *Sigma*. Swiss Reinsurance Comapny Ltd. Web. 11 Dec. 2011. <http://media.swissre.com/documents/sigma4_2007_en.pdf>.

[3] *World Insurance in 2009*. Rep. Vol. 2. Zurich: Swiss Reinsurance. Ser. 2010. *Sigma*. Swiss Reinsurance Comapny Ltd. Web. 11 Dec. 2011. <http://media.swissre.com/documents/sigma2_2010_en.pdf>.

Introduction to Ratemaking

Insurance is a business. An insured, or a policyholder, pays premiums to transfer his/her risks to the insurance companies, or the insurer. The goal of the insurer is to collect enough premiums to cover for losses should they occur as well as to make a marginal profit. In order to do so, they must price accurately, and the pricing of insurance is called ratemaking.

To evaluate if the ratemaking is optimal, measurements like loss ratios are used. The most basic ones are[4]:

$$loss\ ratio = \frac{\text{loss adjustment expenses}}{\text{earned premium}}$$

This measures how much of the premiums received are used for settling claims.

$$expense\ ratio = \frac{\text{underwriting expenses}}{\text{earned premium}}$$

This calculates how much of the premiums received are used for issuing new policies.

$$combined\ loss\&expense\ ratio = \text{loss ratio} + \text{expense ratio}$$

The sum of the two helps an insurance company gauge its actual cost of writing new policies and covering losses for a given period. The lower it is indicates more profits for the company. However, there is not one ideal loss ratio. Depending on industry subsidies and company objectives, companies' target loss ratios vary greatly from 40% to sometimes over 100%.

When the actual loss ratio differs from the expected loss ratio, a rate change is needed to correct the difference. Though a simple percentage surcharge or discount can be applied to the overall premiums, a rate change usually involves evaluating the entire ratemaking algorithm and revising individual rating variables. This process requires actuarial skills and can take weeks or months of time.

---

[4] "Glossary of Insurance Terms." *A.M. Best Company*. Web. 12 Dec. 2011. <http://www.ambest.com/resource/glossary.html>.

To help understand how the insurance policy is priced and the interactions between different ratemaking factors, we created a simulation of a simple case on life insurance policy. We suppose there are a male and a female client, and both of them want to purchase a life insurance policy. The market research shows that the expected loss for a female client is $180 and $200 for a male. How should they get charged? Should we charge them the same price or differently? We suppose there are two insurance companies, LA Life and CountryFarm. LA Life insurance company does not take gender as one of their rating factors, thus they charge the average price $190 for both genders. On the other hand, CountryFarm thinks that it is fair to consider the difference between genders, so it charge female applicants $180 while male applicants $200. It is assumed that all other conditions are held the same for both companies. The Figure 1 shows the summary of the initial scenario:

LA LIFE	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	500	$190	$180	$ 95,000	$ 90,000	95%
Males	500	$190	$200	$ 95,000	$ 100,000	105%
Total	1,000	$190	$190	$ 190,000	$ 190,000	100%

COUNTRYFARM	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	500	$180	$180	$ 90,000	$ 90,000	100%
Males	500	$200	$200	$ 100,000	$ 100,000	100%
Total	1,000	$190	$190	$ 190,000	$ 190,000	100%

Figure 1

From the chart above, we can see that both companies have 1000 clients in total and different product prices. Nevertheless, both of them can achieve a balanced 100% loss ratio at the beginning even if they charge premiums differently. However, can both companies keep this balance all the time? The consumers will constantly look for better rates. After one year, some of female clients in LA Life find out that there is a better rate at another company, and the male clients in CountryFarm also discover that LA Life has less expensive policy. Therefore, the clients will switch out of their own company and go to the better one. Figure 2 below shows the change in customer flow after one year.

LA LIFE	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	250	$190	$180	$ 47,500	$ 45,000	95%
Males	750	$190	$200	$ 142,500	$ 150,000	105%
Total	1,000	$190	$190	$ 190,000	$ 195,000	103%

COUNTRYFARM	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	750	$180	$180	$ 135,000	$ 135,000	100%
Males	250	$200	$200	$ 50,000	$ 50,000	100%
Total	1,000	$190	$190	$ 185,000	$ 185,000	100%

Figure 2

The chart above illustrates the new scenario: half of female clients in LA Life go to CountryFarm and half of male clients in CountryFarm come to LA Life. Even though both of companies still have the same total number of clients, LA Life incurred a loss ratio of 103% while CountryFarm is still 100%. The 3% increase in loss ratio is due to that LA Life have more male clients, who have higher loss cost. To avoid the loss, LA Life pricing department decides to increase their premium up to 3%, to $196. A summary of new scheme is shown in Figure 3.

LA LIFE	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	250	$196	$180	$ 48,925	$ 45,000	92%
Males	750	$196	$200	$ 146,775	$ 150,000	102%
Total	1,000	$196	$190	$ 195,700	$ 195,000	100%

COUNTRYFARM	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	750	$180	$180	$ 135,000	$ 135,000	100%
Males	250	$200	$200	$ 50,000	$ 50,000	100%
Total	1,000	$190	$100	$ 185,000	$ 185,000	100%

Figure 3

After the price adjustment, LA Life can reduce the loss and recover back to 100% loss ratio. However, the rising premium of LA Life will only lead to losing more female clients and absorbing more male clients as Figure 4 shows.

LA LIFE	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	150	$196	$180	$ 29,400	$ 27,000	92%
Males	850	$196	$200	$ 166,600	$ 170,000	102%
Total	1,000	$196	$190	$ 196,000	$ 197,000	101%
COUNTRYFARM	Exposure	Price	Loss Cost	Total Premium	Total Losses	Loss Ratio
Females	850	$180	$180	$ 153,000	$ 153,000	100%
Males	150	$200	$200	$ 30,000	$ 30,000	100%
Total	1,000	$100	$100	$ 183,000	$ 183,000	100%

Figure 4

Figure 4 shows that LA Life will still result in a loss of 101% although they raised up the premium while its competitor could keep a profitable loss ratio all the time. So far, the simulation describes a process called adverse selection. In insurance industry, it refers to a situation in which people only buy insurance when they expect high risks. The main mechanism insurance companies employ to use against adverse selection is pricing segmentation. The above simulation shows that it does not work out to charge the same price to people who actually have different loss cost. Therefore, we can see that gender should be included as one of the ratemaking factor; charging a uniform price towards groups have different loss costs will result in a loss very likely. We all know that there are not only one factor that affect the price of an insurance policy. When we do a quoting online, the insurance company website always asks about gender, age, and sometimes health condition for a common application. These are all possible factors that rule the pricing of the policy, but how the factors interact is also a key to pricing strategy. In the second part of simulation ,we will add another factor, smoking habit and discover the relation between two factors.

Gender	One-Level EP	Losses	Loss Ratio=Losses/EP	Indicated Adjustment
Females	1100000	800000	0.73	0.00
Males	1150000	1100000	0.96	0.32
Smoking	One-Level EP	Losses	Loss Ratio=Losses/EP	Indicated Adjustment
Non Smoker	1200000	800000	0.67	0.00
Smoker	1100000	1200000	1.09	0.64

Figure 5

Figure 5 summarizes the two factors we want to simulate. The first factor is still gender, and females have a lower loss than males since the females have longer life expectancy than males do. The one-level EP means the premiums earned only from one factor. Since females have lower loss, the males rate are adjusted based on the females' loss rate, the adjustment is calculated as (0.96/0.73-1)=0.32. The same adjustment is applied the smoking factor. To examine the relationship between two factors, we first assume they are independent of each other and no correlation exists. Thus, when we combine the two factors together to price the premium, we will just multiply the ratio independently as shown in Figure 6.

Current Premium			Proposed Preium		
Smoking/Gender	Females	Males	Smoking/Gender	Females	Males
Non Smoker	800000	400000	Non Smoker	800000	526087
Smoker	300000	750000	Smoker	490909	1614130

Figure 6

Since female and non-smokers have lower loss relatively, a female who does not smoke will be a base and have lowest loss cost. To calculate the proposed premium for other combinations, we use current premium* (1+adjusted rate). For example, for a male who smokes, we will take 75000*(1+0.32)*(1+0.64)=1614130. Then we will compare the proposed premium with the loss cost to see the new loss ratio, which is indicated below.

Gender	One-Level EP	Losses	Loss Ratio	Indicated Adjustment
Females	1290909	800000	0.62	0.00
Males	2140217	1100000	0.51	-0.17
Smoking/Gender	**One-Level EP**	**Losses**	**Loss Ratio**	**Indicated Adjustment**
Non Smoker	1326087	800000	0.60	0.00
Smoker	2105040	1200000	0.57	-0.06

Figure 7

The calculation with proposed premium results in a negative adjustment rate for males and smokers, which indicates that the two groups get surcharged by the proposed premium. The phenomenon tells us that there must be some correlation between the two factors and they are not independent of each other. Otherwise, we should get zeros for indicated adjustment. Intuitively, it also makes sense that males probably tend to smoke more than females. Thus, it is important to explore the relationships between different factors when pricing an insurance product. The simulation explains the basic logic in insurance pricing, and also demonstrates the importance of locating correct correlation between different rate factors, and which is what we are going to do in our comparison of life insurance in China and US.

As demonstrated in the simulations, ratemaking is complex involving many rating variables and their interactions. We wanted to examine the similarities and differences between ratemaking in China and in the U.S..

Exploration of Rating Variables through Linear Regression Analysis

Suspecting that ratemaking differs in China and the U.S., we wanted to answer the following questions:

1) Do the two countries employ different rating variables, i.e. savings rate?

2) What are the influential rating variable for each country?

Data & Methodology:

Since life insurance is the most mature line of business in China, we chose to focus on life insurance. We would have liked to use real policy and claim data, but they are proprietary and confidential information. Therefore, we created our own policies.

From research, we selected six common rating variables in life insurance: age, education, income, health, gender, and marital status.[5] We added a seventh variables— savings rate— because the Chinese have a much higher savings rate than the Americans; we believe that savings habit reflects if a person is risk-adverse or risk-loving, characteristics that are important to insurers.

In summary:

Rating Variable	**Type**	**Notes**
Age	Numerical	Age of the insured
Education	Numerical	Years of education of the insured
Income	Numerical	Annual income in USD
Health	Categorical	0=standard health status, 1=preferred health status
Gender	Categorical	0=male, 1=female
Marital Status	Categorical	0=single, 1=married
Savings Rate	Numerical	Percentage of disposable income

[5] "What Factors Are Considered in Individual Life Insurance Underwriting?" *FreeAdvice*. Web. 12 Dec. 2011. <http://law.freeadvice.com/insurance_law/life_insurance_law/individual_life_insurance_underwriting.htm>.

Moreover, sixteen policies were created for each country. To capture the true demographics of the country, the profiles of the policyholders were based on census data and survey reports[6]. In addition to the most typical profile consisting of the median values for the respective country, a variety of individuals were accounted for—migrant workers, blue-collar workers, white-collar workers, college-educated professionals, retirees, self-employed individuals, etc. See appendix for detailed policies.

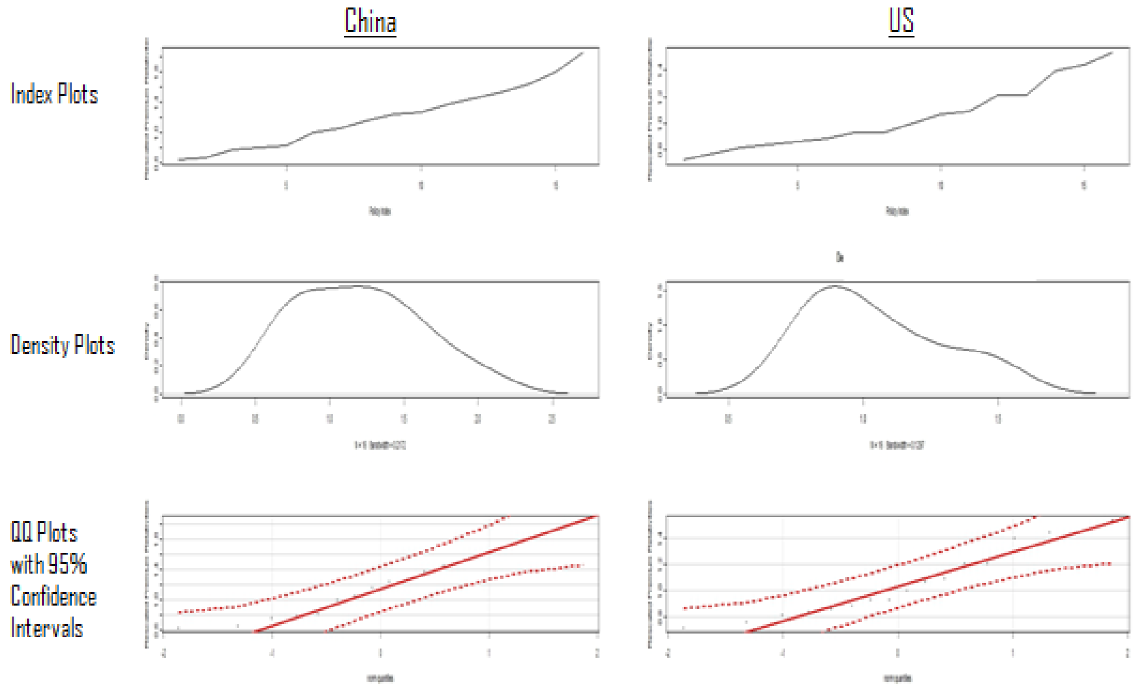
Next, the Chinese policies were rated by two life insurance analysts in China; the U.S. policies were rated by two life actuaries in the U.S. Essentially, each associate assigned a premium relativity to each policy considering the seven columns of information provided. Setting 1 as the base, a value smaller than 1 represents a discount for a preferred policy, and a value greater than 1 represents a surcharge for an undesirable policy. In the end, each policy received two ratings. We took a simple average of the two relativities. In order to compare the relativities across the two countries, we rebased the relativities to the median by setting the relativity of each country's median profile to be 1 (see appendix).

Finally, we regressed premium relativity on age, education, income, health, gender, marital status, and savings rate. We used anova tests to identify which variables are statistically significantly associated with the premium relativity. To simplify the regression, we assumed that the variables were independent and only modeled for main effects.

---

[6] "Compare China and United States." *NationMaster*. Web. 12 Dec. 2011. <http://www.nationmaster.com/compare/China/United-States>.

Exploratory Data Analysis:



Before running any regression, we used EDA to see if the response variable was appropriate for linear regression. As shown on the plots above, the values of response variable for both countries are about normally distributed and show no extreme skewness or multiple modes. Therefore, we proceeded to fitting linear models.

Regression Models:

U.S.

To begin, we try a model consisting of all seven main effects.

```
Anova Table (Type II tests)

Response: rescaled
           Sum Sq Df F value  Pr(>F)    VIF
age       0.024540  1  3.2134 0.11080  7.572308
education 0.000277  1  0.0362 0.85375  4.060945
income    0.046626  1  6.1055 0.03865 *7.968506
savings   0.001004  1  0.1315 0.72629  2.836824
health    0.027197  1  3.5614 0.09584 .2.319133
gender    0.001690  1  0.2212 0.65065  1.809767
marital   0.006403  1  0.8384 0.38662  2.833671
Residuals 0.061093  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

By the ANOVA type II test, only education is found to be statistically significant at the 5% level; health is found to be associated with premium relativity at the 10% level holding other variables constant. Immediately we suspect that income absorbed the effects of the other variables. To check possible collinearity, we look at the variance inflator factors. It is clear that the variance of income's estimated coefficients is inflated by a factor of 7.97 because it is highly correlated with at least one of the other explanatory variables in the model. For the same logic, age is problematic too. But since income is found to be statistically significant and has the highest VIF, we fit a second model excluding income but keeping age.

```
Anova Table (Type II tests)

Response: rescaled
           Sum Sq Df F value    Pr(>F)     VIF
education 0.05360  1  4.4786 0.0634208 .  1.677300
age       0.50419  1 42.1255 0.0001127 ***1.790982
gender    0.00104  1  0.0870 0.7746663    1.611019
marital   0.01591  1  1.3294 0.2786113    2.718490
savings   0.00127  1  0.1064 0.7517447    2.835855
health    0.00492  1  0.4110 0.5374278    1.788085
Residuals 0.10772  9
---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The VIF's look much more reasonable this time. According to anova type II test, education and age are statistically significantly associated with the response variable holding the other explanatory variables constant. In other words, deleting age or education from the model will result in a poorer fit.

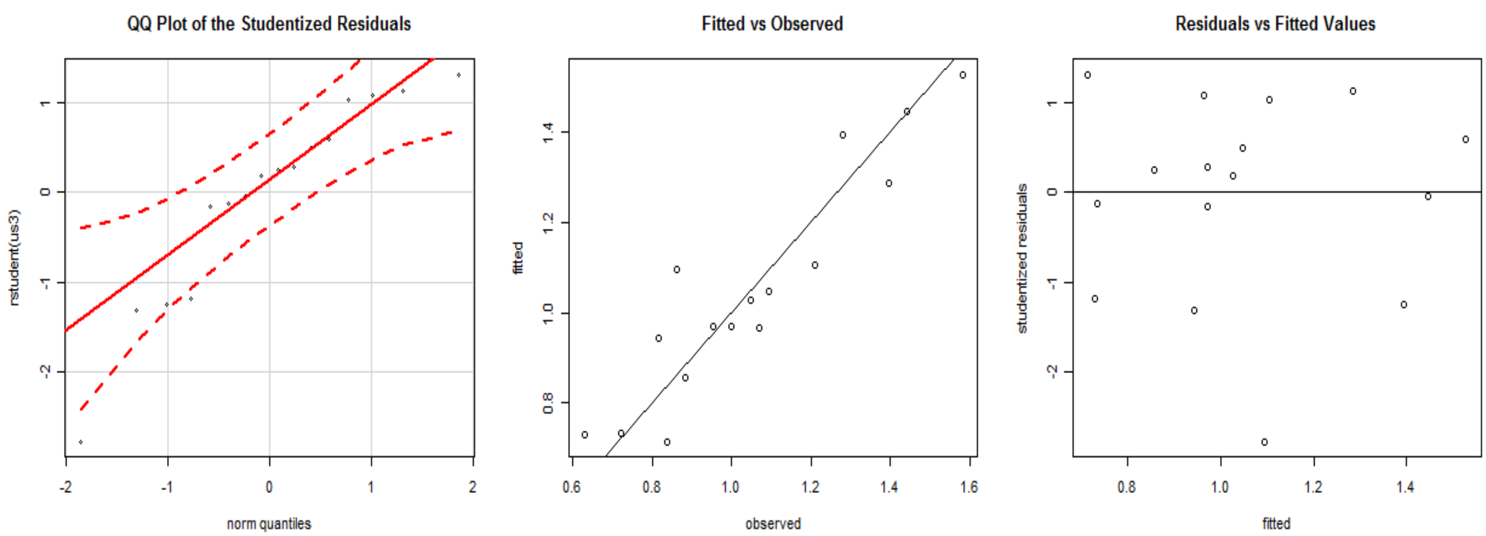
```
Call:
lm(formula = rescaled ~ education + age, data = us)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.23407 -0.03829  0.02255  0.06628  0.12239 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.876353   0.136639   6.414 2.29e-05 ***
education   -0.028224   0.007769  -3.633  0.00303 ** 
age          0.012375   0.001601   7.727 3.26e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1079 on 13 degrees of freedom
Multiple R-squared: 0.8623,     Adjusted R-squared: 0.8411 
F-statistic:  40.7 on 2 and 13 DF,  p-value: 2.532e-06
```

Therefore, the selected model for the U.S. data is:

*Premium relativity= 0.876 – 0.028*education +0.012*age*

To evaluate this model, we first note that the adjusted R-squared is high at 0.84, suggesting that 84% of the variability in the data is accounted for by the model.

Furthermore, we looked at the residual analysis.



Based on the QQ plot, most of the residuals lie within the 95% confidence intervals. Two points lie slightly outside of the confidence intervals. The Shapiro-wilk test calculates a p-value of 0.09057, suggesting that there is not enough evidence to reject the assumption of normal residuals at the 5% level. For a small sample size of 16, we conclud the residuals are normally distributed.

To check the assumption of linearity, we plot the fitted values against the observed. There is no obvious pattern. For the most parts, the points are evenly scattered around the diagonal linear regression line. There is one exception at ~(0.85, 1.1). It deviates away from the regression line further from the rest. But again considering the small sample size, we believe that assumption of linearity is met.

Finally to check the assumptions of independence and homoscedascity of the errors, the studentized residuals against the fitted values are plotted. We notice the same outlier again at fitted value ~1.1. Ignoring this point, the residuals are randomly scatted around 0; the variance is constant. The residuals are indeed independent and homoscedastic.

<u>China:</u>

Similarly, a model with all seven main effects is tried first.

```
Anova Table (Type II tests)

Response: rescaled
            Sum Sq Df F value    Pr(>F)     VIF
age       0.198559  1 32.7545 0.0004424 ***5.521164
education 0.031309  1  5.1648 0.0526738 .  3.564620
income    0.025885  1  4.2701 0.0726326 .  6.724894
savings   0.050256  1  8.2902 0.0205342 *  2.515080
health    0.018544  1  3.0591 0.1184076    1.395376
gender    0.001626  1  0.2682 0.6185763    1.196773
marital   0.009468  1  1.5619 0.2467100    4.680743
Residuals 0.048496  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Income once again has the highest VIF and is thus deleted.

```
Anova Table (Type II tests)

Response: rescaled
           Sum Sq Df  F value    Pr(>F)     VIF
age       1.10272  1 133.4271 1.065e-06 ***1.682830
education 0.20105  1  24.3263 0.0008109 ***1.567788
savings   0.05524  1   6.6838 0.0294392 *  2.504643
health    0.01274  1   1.5417 0.2457543    1.362671
gender    0.00071  1   0.0865 0.7753807    1.188117
marital   0.05121  1   6.1957 0.0344720 *  3.211643
Residuals 0.07438  9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In the second model, the VIF's are all below 4—the general rule of thumb—suggesting collinearity is not an imminent problem. In this model, there is a statistically significant association between premium relativity and age, education, savings rate, and marital status. The model will be worse if any of these variables is deleted.

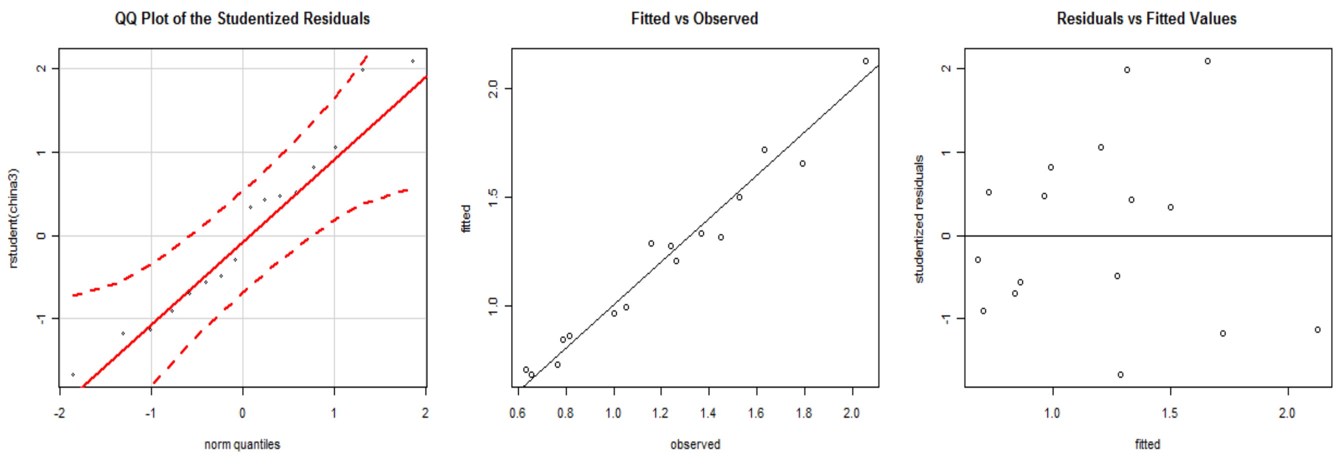
Therefore, a selected model regresses premium relativity on age, education, savings, and marital status:

```
Call:
lm(formula = rescaled ~ age + education + savings + marital,
    data = china)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.127975 -0.055593  0.001821  0.042573  0.131723 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.793705   0.092311   8.598 3.27e-06 ***
age          0.021621   0.001767  12.235 9.53e-08 ***
education   -0.028106   0.005329  -5.275 0.000262 ***
savings     -1.069449   0.409174  -2.614 0.024099 *  
marital     -0.205185   0.089543  -2.291 0.042668 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08944 on 11 degrees of freedom
Multiple R-squared: 0.9668,     Adjusted R-squared: 0.9547 
F-statistic: 80.06 on 4 and 11 DF,  p-value: 4.65e-08
```

The R-squared is very high, indicating that 95% of the variability in the data can be explained by the model shown below:

*Premium relativity= 0.794+0.022*age –0.028*education–1.069*savings– 0.205*marital status.*

To be consistent, residual analysis is also used to check the assumptions of this model.



All of the points are well within the confidence interval. The p-value from the Shapiro-wilk test is very high at 0.5673. There is strong evidence of normally distributed residuals.

Moreover, points in the fitted vs observed value plot are evenly scattered around the diagonal line, confirming linearity of the data.

Lastly, the residuals vs fitted values plot is plotted to check homoscedascity and independence. Though randomly scattered, the variance does seem to increase as fitted values increase. However, considering the small number of values at higher fitted values, this violation of constant variance is not as alarming to us.

Overall, the model does seem to be appropriate.

<u>Interpretation:</u>

<u>U.S.:</u>

*Premium relativity= 0.876 +0.012*age – 0.028*education*

Standardized Coefficients:

Age: 0.264 Education: -0.126

Based on the above linear regression analysis, the premium relativity in the U.S. rated by our American correspondents is highly associated with age and education. A year of increase in the insured's age corresponds to a 0.012 increase in the relativity; a year of decrease in the insured's year's education corresponds to a 0.028 decrease in his/her relativity. Intuitively this makes sense because an older insured has a higher chance of fatality; insurance company charges higher premiums to compensate for higher risks. Education allows people to take better care of themselves in terms of diets, lifestyles, activity levels; responsible habits likely decrease chance of death, leading to a discount in life insurance premiums for more years of education.

To compare the effects of the two variables, we standardize the coefficients by multiplying them by their inter-quartile ranges. After adjustment, age has a larger effect than education over the central half of premium relativity observed in the data.

China:

*Single: Premium relativity=0.794 + 0.022*age – 0.028 *education -1.069 *savings*

*Married: Premium relativity=0.589 + 0.022*age – 0.028 *education -1.069 *savings*

Standardized Coefficients:

Age: 0.484 Education: -0.266 Savings: -0.160

On the other hand, the premium relativity in China rated by our Chinese correspondents is associated with not only age and education but also savings rate and marital status. A married insured's relativity is 0.205 lower than a single insured's. A year of increase in the insured's age corresponds to a 0.022 increase in the relativity; a year of decrease in the insured's year's education also corresponds to a 0.028 decrease in premium relativity like that in the U.S. model. Finally, a unit increase in savings rate is associated with 1.069 units of decrease in premium relativity. Again, these results make sense because an older insured is riskier. A married insured is more preferred because his/her spouse can care for each other. Higher education allows for self-care knowledge. High savings rate indicates more responsible behaviors and the ability to cover preventive cares and so on.

After standardizing the coefficients, we see that in this model age has the largest effect over the central half of observed data followed by education and savings rate.

U.S. VS China

Through our linear regression analysis, we find that ratemaking indeed differs in China and the U.S. Based on our samples, U.S. life insurance companies tend to rate based on age and education while Chinese companies tend to rate on age, education, savings rate, and marital status. In both models, age stands out to be the most influential rating variable followed by education.

However, recall that these regressions are extreme simplifications of ratemaking and rely heavily on the assumptions of independent variables. Moreover, the values of the response variable are the arbitrary judgments of a very small number of people. Any indication from these two models only prompts further examination and research.

To further understand the complications and implications behind different factors that affect two insurance markets, we conducted researches to study the insurance markets in China and United States and look for how consumer behaviors and cultural differences impact on two markets so that same factor can have different effects on the model in different country. Since there are many possible factors implementing on the insurance pricing model, we choose to emphasize a few major factors that are commonly asked in a life insurance application, and they are gender, age, income, family structure, and an overall examination on social and cultural differences between two countries that have crucial influences on these factors.

Gender

In US, although the overall statistics does not show an obvious difference or significance between the number of female life insurance policyholders and male life insurance policyholders, there is a recent trend that indicates more females are purchasing life insurance. According to the report from Insurance &Financial Advisor (IFA), sales of life insurance policies offering long-term care benefits obtained a 79% increase in 2010, and women accounted for about 60% of the sales. The research from the American Association for Long-Term Care Insurance shows that women between the ages of 55 and 64 bought more than one-third of the policies sold in 2010. [7] Also, male purchases tended to be slightly older. These evidences show that more women start to buy life insurance and they bought at earlier ages. One reason that women want to purchase life insurance is that women tend to have better rates than men have since women have a longer life expectancy. Another possible reason is that women will want to show more care and love toward their family so they would like to buy life insurance to protect their loved ones.

Whereas in China, in order to have a better idea of how various factors affect people's decisions of buying insurance, an insurance survey for Chinese life insurance consumers was conducted[8]. The survey was conducted by doing face-to-face interviews in three major cities of China, which were Shanghai, Shenzhen, and Chengdu. In total, 295 insurance buyers participated in the survey, in which 37 were not counted as they were used as testing samples. Therefore, the total sample size of the survey was 258. Out of the 258 participants, 45.74% are male, and 54.26% are female. The result clearly indicates the fact that female is the majority of

---

[7] "Women, especially older ones, buying into life-LTC policy hybrids." *Insurance and Finance Advisor*. June 1, 2010.
< http://ifawebnews.com/2011/06/01/women-especially-older-ones-buying-into-life-ltc-policy-hybrids/>.
[8] "Factors influencing consumers' life insurance purchasing decisions in China." Sep 22, 2010. < http://mspace.lib.umanitoba.ca/handle/1993/4235>.

insurance consumers in China. Therefore, under this category, we can conclude that female tend to be the majority of insurance consumers in both the U.S. and China.

<u>Age</u>

The table below is the data showing life insurance buyers in different ages in US in 2007, from the American Association for Long-Term Care Insurance website.[9] The data clearly illustrates that elderly people buy more life insurance in United State.

Ages When People Apply	
Under35	1%
35-44	6%
45-54	26%
55-64	50%
65-74	15%
75orOlder	2%

Table 1

Looking at the situation of China, the result of the same survey introduced under "Gender" category shows that more than 91% of the respondents are younger than 44 years old, among which 68% are between 25 and 44 years old[10], which demonstrates that compared to elderly, young people are the main insurance buyers in China. Therefore, two insurance markets differ under "age". While elderly people buy more life insurance in the United States, young people are the main consumers in China.

<u>Income</u>

Income is also an important factor in determining the price of insurance. When people have more extra money, they would more likely to buy insurance products. Oppositely, when the

---

[9] "Long-term Care Insurance Facts-Statistics." < http://www.aaltci.org/long-term-care-insurance/learning-center/fast-facts.php>.
[10] "Factors influencing consumers' life insurance purchasing decisions in China." Sep 22, 2010. < http://mspace.lib.umanitoba.ca/handle/1993/4235>.

economy is bad and unemployment rate is high, consumers will probably cut of the spending on life insurance. In 2010, the percentage of U.S. households with life insurance coverage is at its lowest in 50 years, leaving millions of families without a safety net, reported by USA Today.[11]According to the survey provided by LIMRA , an industry-sponsored group, only 44% of households have an individual life insurance policy, and 30% have no individual or employer-provided life insurance. About 11 million households with children younger than 18, who are seen as families with the greatest need for coverage, have no life insurance. One plausible explanation for this serious lack of life insurance coverage in US recently is the economic downturn. The survey conducted showed that people would not spend on life insurance when they suffer from reduced income. [12]

And China faces the similar situation. Based on the same survey of insurance consumers, nearly 64% of the respondents were making over RMB 4,000 (≈ US $503) per month per family unit, while the average monthly income at the time of the survey (2006) was RMB 1058.89 (≈ US $ 133) at exchange rate of US $1 = RMB 7.9602[13]. The huge difference between two monthly income shows that the majority of people buying life insurance in China has much more higher income than average. Therefore, comparing the situation in the United States and China, we conclude that people with higher income tend to buy more life insurance than others.

Family/Dependents

In United States, protecting family is one of main top reasons why people purchase life insurance, especially for primary income earners in a family. The primary income earners want

[11] "Households with life insurance hits lowest level in 50 years." *USA Today*. Dec 3, 2010. < http://www.usatoday.com/money/perfi/insurance/2010-12-03-1Alifeinsurance03_ST_N.htm>.

[12] Same as footnote 11.

[13] " National Monthly Income report in 2006" <http://www.stats.gov.cn/>.

to buy life insurance to cover for their family in case of any accidents happened to them and left their dependents vulnerable.

China has slightly different situation under this category. The same life insurance survey shows that over 56% of the respondents do not have dependent children, while the rest have less and equal to 1 child. This indicates that people with fewer dependents buy more life insurance in China. And the main reason for this circumstance is that implementing “One family One child” policy has dramatically influenced people’s lives in China. When people start to have fewer children, the living expenses of a household will decline, thus people are able to earn more savings compared to before. Hence, the increase in people’s savings enables them to start to invest their money, and life insurance is one of the investments that people usually prefer.

Social and Cultural Difference between two countries

The comparison above demonstrates the differences between each country, and many of the differences can also be explained by understanding the social and cultural backgrounds. For example, more elderly purchase life insurance in US than old people do in China, and it could be explained by the societal evaluation on individualism. People in United States emphasize on the sense of individualism, thus, old people would like to buy life insurance for themselves to cover the funeral related expense occurred after their death and avoid shifting the burden to their kids. However, filial piety, a respect for the parents and ancestors, is one of the virtues to be held above all else in China. Holding a grand funeral is a traditional way for Chinese to show their love and respect to their parents. Since the social security system is still underdeveloped in China, most old people depend on their children after they retired. Thus, fewer old people in China purchase life insurance, since the policy is probably purchased by their kids. Another main difference in consumer behaviors between two country is due to the psychological barrier of facing death existing in US. The data shows that young people who are life insurance applicants are less than 50% of the total population. The procrastination is detected among young adults in US. Dr Ernest Becker, author of the Pulitzer Prize winning book, Denial of Death (1973) wrote that, "Denial of death pervades human culture. It is a subject we'd rather ignore than address."[14] This psychological refusal to face death leads to procrastination in purchasing life insurance. However in China, due to the growing awareness of risk and Chinese traditional way of showing love and respect to the elderly people, young people are the majority of life insurance consumers. Moreover, due to the difference of socio-demographic structure and traditional culture, while life insurance is a priority expense among all insurance products in China, it is not the same in the United States. Hence, people's preferences and demands of insurance products differ in two countries as well.

Conclusion

In conclusion, by conducting the simulation of two insurance companies' pricing strategies, we obtained a basic understanding of how pricing strategy works, and more importantly, how influential rating variables are in insurance markets. Then we proceeded to further analyze the rating variables in U.S. and China individually by our linear regression model, and compared the results jointly again between the two countries. Through these analysis and comparison, we observed that pricing strategy, influenced by variables including age, education, savings, and marital status, differs in China and the U.S.. However, age is observed to be the most influential rating variable in both countries.

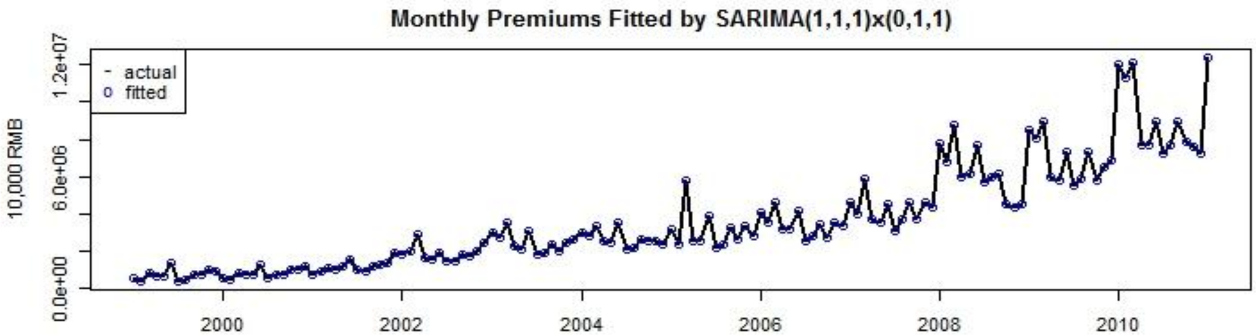
---

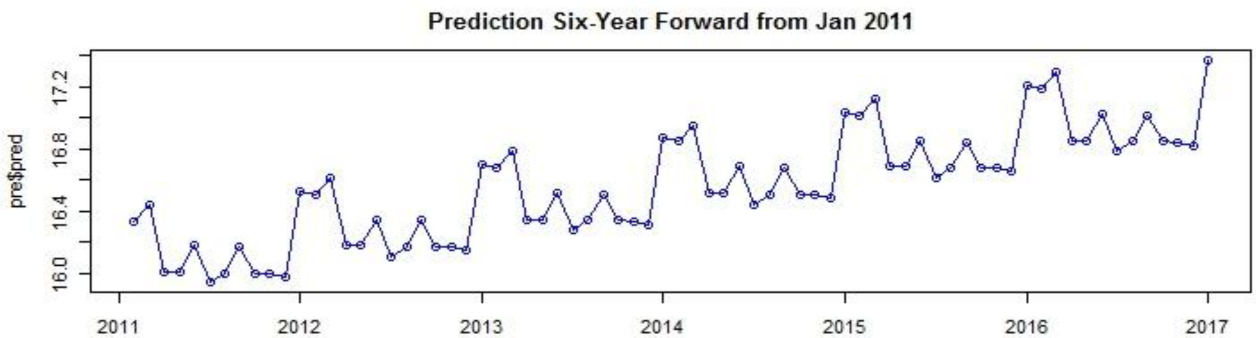
[14] "Reasons Behind Life Insurance Buy." *New York Life.* < http://www.newyorklife.com/nyl/v/index.jsp?contentId=16036&vgnextoid=11af1219a49d2210a2b3019d221024301cacRCRD >.

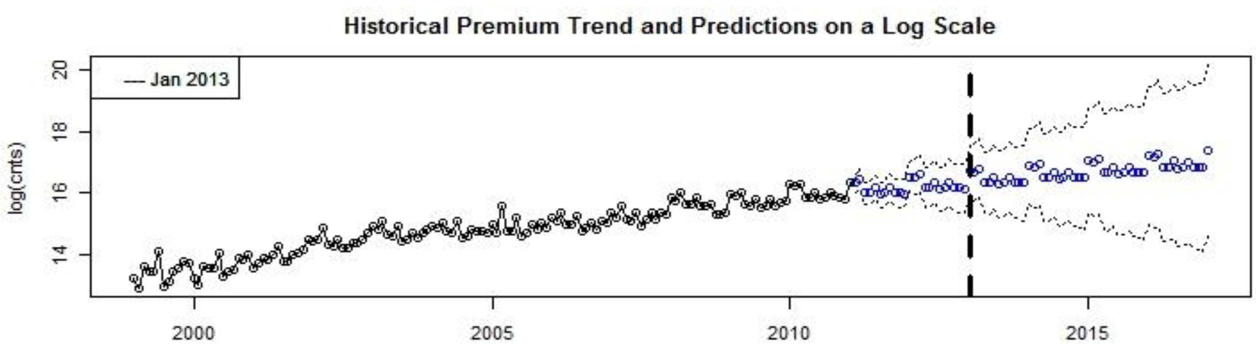
Furthermore, we examined the two insurance markets by considering more variables associated with social and cultural background. By doing so, we had a better understanding of the differences and similarities between these two insurance markets not only from statistical perspective, but also from the impact that social and cultural aspect has on consumer behavior as well as consumer psychology.

Lastly, we were very interested in the future development of these two insurance markets, especially in the undeveloped insurance market in China. While the insurance industry in the U.S. has rather stabilized, the growth of the insurance industry in China has very promising outlooks. Taking the historical (1999 to 2010) life insurance premiums in China published by China Insurance Regulatory Commission,[15] we fitted a Seasonal ARIMA (1,1,1)x (0,1, $1)_{12}$ model. Using this model to forecast the premiums for the next six years, a clear upward trend is observed. More importantly, it is predicted to increase exponentially. But the confidence interval of our prediction gets much wider as we predict further into the future. Therefore, we suggest that the predictions are reliable up to January 2013.

[15] "Tables of the Insurance Industry." *China Insurance Regulatory Commision*. Web. 12 Dec. 2011. <http://www.circ.gov.cn/web/site0/tab61/module438/more.htm>.


Monthly Premiums Fitted by SARIMA(1,1,1)x(0,1,1)
actual
fitted
10,000 RMB
0.0e+00
6.0e+06
1.2e+07
2000
2002
2004
2006
2008
2010


Prediction Six-Year Forward from Jan 2011
pre$pred
16.0
16.4
16.8
17.2
2011
2012
2013
2014
2015
2016
2017


Historical Premium Trend and Predictions on a Log Scale
Jan 2013
log(cnts)
14
16
18
20
2000
2005
2010
2015

Reference

"China Insurance Industry | Economy Watch." *Economy Watch*. 26 May 2010. Web. 11 Dec. 2011. <http://www.economywatch.com/insurance/china-insurance-industry.html>.

"Compare China and United States." *NationMaster*. Web. 12 Dec. 2011. <http://www.nationmaster.com/compare/China/United-States>.

“Factors influencing consumers’ life insurance purchasing decisions in China.” 22 Sep. 2010. < http://mspace.lib.umanitoba.ca/handle/1993/4235>.

"Glossary of Insurance Terms." *A.M. Best Company*. Web. 12 Dec. 2011. <http://www.ambest.com/resource/glossary.html>.

"Households with life insurance hits lowest level in 50 years." *USA Today*. Dec 3, 2010. < http://www.usatoday.com/money/perfi/insurance/2010-12-03-1Alifeinsurance03_ST_N.htm>.

"Long-term Care Insurance Facts-Statistics." < http://www.aaltci.org/long-term-care-insurance/learning-center/fast-facts.php>.

"Reasons Behind Life Insurance Buy." *New York Life*. <http://www.newyorklife.com/nyl/v/index.jsp?contentId=16036&vgnextoid=11af1219a49d2210a2b3019d221024301cacRCRD >.

"Tables of the Insurance Industry." *China Insurance Regulatory Commision*. Web. 12 Dec. 2011. <http://www.circ.gov.cn/web/site0/tab61/module438/more.htm>.

"What Factors Are Considered in Individual Life Insurance Underwriting?" *FreeAdvice*. Web. 12 Dec. 2011. <http://law.freeadvice.com/insurance_law/life_insurance_law/individual_life_insurance_underwriting.htm>.

"Women, especially older ones, buying into life-LTC policy hybrids." *Insurance and Finance Advisor*. June 1, 2010. < http://ifawebnews.com/2011/06/01/women-especially-older-ones-buying-into-life-ltc-policy-hybrids/>.

World Insurance in 2006. Rep. Vol. 4. Zurich: Swiss Reinsurance. Ser. 2007. *Sigma*. Swiss Reinsurance Comapny Ltd. Web. 11 Dec. 2011. <http://media.swissre.com/documents/sigma4_2007_en.pdf>.

World Insurance in 2009. Rep. Vol. 2. Zurich: Swiss Reinsurance. Ser. 2010. *Sigma*. Swiss Reinsurance Comapny Ltd. Web. 11 Dec. 2011. <http://media.swissre.com/documents/sigma2_2010_en.pdf>.

Appendix

Linear Regression Data: Policies for China

Policy	Age	Years of Education	Income	Savings	Health Status	Gender	Marrital Status	Description
China1	35	6	7500	0.2	0	1	1	median profile, male
China2	35	5	7600	0	0	0	1	median profile, female, no savings
China3	35	5	7600	0.2	0	0	1	median profile, female, good savings
China4	32	6	4000	0	0	1	0	young male, uneducated, no savings, below average income, standard health status, single
China5	30	6	3500	0	1	0	0	young female , uneducated, no savings, below average income, healthy, single
China6	35	16	9600	0.2	0	0	1	college educated female, median age, middle-upper class, good savings, standard health, married
China7	36	16	10000	0.2	1	1	1	college educated male, middle-upper class, good savings, health, married
China8	25	15	9000	0.05	1	1	0	young college educated male,middle-upper class, little savings, healthy, single
China9	28	18	10000	0.05	0	1	0	young male with graduate degree, middle-upper class, little savings, standard health, single
China10	42	16	10000	0.3	0	0	1	college educated married middle age female, middle-upper class, large savings, standard health
China11	45	6	8000	0.1	0	0	1	middle-age female with primary education, middle-class, moderate savings,
China12	60	2	3500	0.1	0	1	1	unedcated older male, blow average earnings, moderate savings
China13	55	12	5000	0.1	0	1	1	educated middle-age male, below average earnings, moderate savings
China14	70	12	4500	0.1	0	0	1	educated older female, below average earnings, moderate savings
China15	65	10	4500	0.2	0	0	1	educated older female with low income but high savings
China16	80	3	3000	0.1	0	0	1	uneducated elderly female with low income, moderate savings

Note: The “Description” column is intended for our reference. It was not sent to the actuaries.

Linear Regression Data: Policies for the U.S.

Policy	Age	Years of Education	Income	Savings	Health Status	Gender	Marrital Status	Description
US1	35	12	45000	0.2	0	1	1	median profile, male, good savings
US2	35	10	45600	0	0	0	1	median profile, female, no savings
US3	35	12	45600	0.2	0	0	1	median profile, female, good savings
US4	32	9	24000	0	0	1	0	young male, high school drop out, savings, lowincome, standard health status, single
US5	30	10	21000	0	1	0	0	young female , high school drop out, little savings, low income,healthy single
US6	35	16	57600	0.2	0	0	1	college educated female, median age, middle class, good savings, standard health, married
US7	36	18	60000	0.2	1	1	1	college educated male, middle class, good savings, health, married
US8	25	16	54000	0.05	1	1	0	young college educated male,middle class, little savings, healthy, single
US9	28	18	60000	0.05	0	1	0	young male with graduate degree, middle class, little savings, standard health, single
US10	42	16	60000	0.3	0	0	1	college educated married middle age female, middle class, large savings, standard health
US11	45	12	48000	0.1	0	0	1	middle-age female with high shcool education, middle-class, moderate savings, married
US12	60	9	21000	0.1	0	1	1	older male, middle school education, low income, moderate savings
US13	55	16	30000	0.1	0	1	1	educated middle-age male, low income, moderate savings
US14	70	16	27000	0.1	0	0	1	educated older female, low income, moderate savings
US15	65	16	27000	0.2	0	0	1	college educated older female with low income but high savings
US16	80	12	20000	0.1	0	0	1	high school educated elderly female with little income, moderate savings

Relativities Averaged and Rescaled.

Policy	CN1	CN2	Average CN	Rescaled
CN1	1.00	0.90	0.95	1.00
CN2	1.20	1.20	1.20	1.26
CN3	1.00	1.00	1.00	1.05
CN4	1.45	1.30	1.38	1.45
CN5	1.15	1.20	1.18	1.24
CN6	0.65	0.60	0.63	0.66
CN7	0.60	0.60	0.60	0.63
CN8	0.75	0.80	0.78	0.82
CN9	0.80	0.70	0.75	0.79
CN10	0.70	0.75	0.73	0.76
CN11	1.10	1.10	1.10	1.16
CN12	1.60	1.50	1.55	1.63
CN13	1.30	1.30	1.30	1.37
CN14	1.80	1.60	1.70	1.79
CN15	1.40	1.50	1.45	1.53
CN16	2.00	1.90	1.95	2.05

Policy	US 1	US 2	Average US	Rescaled
US1	1.05	1.10	1.08	1.00
US2	1.00	1.00	1.00	0.93
US3	1.00	1.00	1.00	0.93
US4	1.20	1.15	1.18	1.09
US5	1.20	1.10	1.15	1.07
US6	0.90	1.00	0.95	0.88
US7	0.80	0.85	0.83	0.77
US8	0.75	0.80	0.78	0.72
US9	0.90	0.90	0.90	0.84
US10	0.90	0.85	0.88	0.81
US11	0.95	0.90	0.93	0.86
US12	1.45	1.30	1.38	1.28
US13	1.30	1.30	1.30	1.21
US14	1.50	1.60	1.55	1.44
US15	1.50	1.50	1.50	1.40
US16	1.60	1.70	1.65	1.53

R-Code

```
library(car)
library(TSA)
#reading data

us=read.csv('us.csv',header=T,stringsAsFactors=F)
china=read.csv('china.csv',header=T,stringsAsFactors=F)
names(us)=c('policy','age','education','income','savings','health','gender','marital','factor1','factor2',
'avg','rescaled')
names(china)=c('policy','age','education','income','savings','health','gender','marital','factor1','facto
r2','avg','rescaled')
us= us[with(us, order(rescaled)), ]
china= china[with(china, order(rescaled)), ]

#EDA
#US
par(mfrow=c(3,1))
plot(us$rescaled,type='l',xlab='Policy Index',ylab='Rescaled Premium Relativities',main='Index
Plot of the Rescaled Premium Relativities-US')
plot(density(us$rescaled),main='Density Plot of Rescaled Premium Relativities-US')
qqPlot(us$rescaled,main='QQ Plot with 95% Confidence Intervals-US',ylab='Rescaled Premium
Relativities')

#China
par(mfrow=c(3,1))
plot(china$rescaled,type='l',xlab='Policy Index',ylab='Rescaled Premium
Relativites',main='Index Plot of the Rescaled Premium Relativities-China')
plot(density(china$rescaled),main='Density Plot of Rescaled Premium Relativities')
qqPlot(china$rescaled,main='QQ Plot with 95% Confidence Intervals',ylab='Rescaled Premium
Relativities')

#US regressions
us1=lm(rescaled~age+education+income+savings+health+gender+marital,data=us)
Anova(us1)
vif(us1)
# delete income
us2=lm(rescaled~education+age+gender+marital+savings+health,data=us)
Anova(us2)
vif(us2)
#final
us3=lm(rescaled~education+age,data=us)
summary(us3)
#residual analysis
par(mfrow=c(1,3))
qqPlot(rstudent(us3),main='QQ Plot of the Studentized Residuals')
plot(us$rescaled,us3$fitted,xlab='observed',ylab='fitted',main='Fitted vs Observed')
```

```
abline(lm(us$rescaled~us3$fitted))
plot(us3$fitted,rstudent(us3),xlab='fitted',ylab='studentized residuals',main='Residuals vs Fitted Values')
abline(h=0)
shapiro.test(rstudent(us3))
summary(us)

#China regression
china1=lm(rescaled~age+education+income+savings+health+gender+marital,data=china)
Anova(china1)
vif(china1)
#removes income
china2=lm(rescaled~age+education+savings+health+gender+marital,data=china)
Anova(china2)
vif(china2)
#final
china3=lm(rescaled~age+education+savings+marital,data=china)
summary(china3)
#residual analysis
shapiro.test(rstudent(china3))
par(mfrow=c(1,3))
qqPlot(rstudent(china3),main='QQ Plot of the Studentized Residuals')
plot(china$rescaled,china3$fitted,xlab='observed',ylab='fitted',main='Fitted vs Observed')
abline(lm(china$rescaled~china3$fitted))
plot(china3$fitted,rstudent(china3),xlab='fitted',ylab='studentized residuals',main='Residuals vs Fitted Values')
abline(h=0)
summary(china)

#Time Series Prediction
cnprem=read.csv('cn_premiums.csv',header=F,stringsAsFactors=F)
cnts=ts(cnprem[,2],start=c(1999,1),frequency=12)

par(mfrow=c(1,2))
plot(cnts,main='Time Series Plot: Monthly Life Insurance Premiums in China')
plot(log(cnts),main='Time Series Plot: Log (Premiums)')

par(mfrow=c(1,2))
plot(diff(log(cnts)),main='Difference of Log(Premiums)')
spec(diff(log(cnts)),spans=c(3,3),main='Smoothed Spectral Density of the Difference of Log(premiums)')
abline(v=c(1:5)/12,lty='dashed')
#clear seasonality

par(mfrow=c(2,2))
acf(diff(log(cnts)),main='ACF of the Difference of Log(premiums)',ci.type='ma',lag.max=48)
```

```
pacf(diff(log(cnts)),main='PACF of the Difference of Log(premiums)',lag.max=48)
acf(diff(diff(log(cnts)),lag=12),main='ACF of the First and Seasonal Difference of Log(premiums)',ci.type='ma',lag.max=48)
pacf(diff(diff(log(cnts)),lag=12),main='PACF of the First and Seasonal Difference of Log(premiums)',lag.max=48)

cn1=arima(log(cnts),order=c(1,1,1),seasonal=list(order=c(0,1,0),period=12))
#log likelihood = 54.82,  aic = -105.64

cn2=arima(log(cnts),order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12))
#log likelihood = 58.83,  aic = -111.66

#final
par(mfrow=c(3,1))
plot(cnts,lwd=2,main='Monthly Premiums Fitted by SARIMA(1,1,1)x(0,1,1)',ylab='10,000 RMB')
points(cnts-exp(cn2$resid),col='blue')
legend('topleft',c('actual','fitted'),col=c('black','blue'),pch=c('-','o'))
pre=predict(cn2,n.ahead=72)
plot(pre$pred,type='o',main='Prediction Six-Year Forward from Jan 2011',col='blue')
plot(cn2,n.ahead=72,main='Historical Premium Trend and Predictions on a Log Scale')
points(pre$pred,col='blue')
abline(v=c(2013,1),col='black',lty='dashed',lwd=3)
legend('topleft','--- Jan 2013')
```