A COMPARISON OF DETERMINISTIC IBNR RESERVING TO STOCHASTIC IBNR RESERVING IN KENYA

OTAMBO WASHINGTON MISIKO

078096

Submitted in partial fulfillment of the requirements for the Degree of Bachelor of Business Science in Actuarial Science at Strathmore University

Strathmore Institute of Mathematical Science

Strathmore University

Nairobi, Kenya

July, 2016

DECLARATION

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the Research Proposal contains no material previously published or written by another person except where due reference is made in the Research Proposal itself.

..…………….. Otambo Washington

……………….. *Signature*

……………... [1st July, 2016]

This Research Proposal has been submitted for examination with my approval as the Supervisor.

..…………….. Supervisor

……………….. *Signature*

……………... [1st July, 2016]

School of Finance and Applied Economics

Strathmore University

Table of Contents

List of Abbreviations

IBNR – Incurred but not reported

IFoA- Institute and Faculty of Actuaries

List of Figures

ABSTRACT

Different methods of reserving for claim liabilities of an insurer and in particular IBNR have been advanced throughout time. These methods strive to provide an estimate enough to cater for the future liabilities. Therefore it is important that the methods developed be as accurate as possible since an under reserving will mean more risk exposer to the insured's while holding too much reserve is not maximizing the shareholders commitment of funds. A balance should be made to set aside a just enough amount of reserve, hence need for a good reserving technique.

1.0 INTRODUCTION.

1.1 Background

One of the major concerns of insurance companies is to make sure that they can pay up the claims that is are due to the insured's. A reason for insurance companies going under can be attributed to improper reserving for the expected claims. It is important that the insurance companies sets up reserves for the future, the future referring to days, months and years from their current position. In the future their cash inflows may be overweighed by their cash out flows, therefore they need a cushion to top up on the deficit that they will incur.

At any one point an insurer wants to be certain that she can meet up the obligations of paying for a claim. While the insured wants to have assurance that whenever he/she may incur a loss the insurer will be able to pay up for the loss.

This paper revolves around IBNR reserving. A reported claim is one that has already been processed to the extent that a central record on it is held (Ratemo & Weke, 2013). Besides the reported claims, other claims may have occurred but not yet reported. Meaning that its processing has not reached an extent of a central record on it held. Therefore it is prudent and of essence for an insurer to think about those claims that have occurred meaning that he may be obliged to pay, but have not yet been reported to the office. Hence the wording Incurred but Not Reported Claims IBNR and Incurred But Not Reported Reserves.

Delays between the occurrence of a claim and the payment can be a result of several factors. First it may not be clear whether a legal liability of the insurer exists or not. Therefore a considerable amount of time may be taken before the court decides if the liability exists or not. Also within the insurance company there may be processing delays. Delays in recording of the claims incurred, processing of the files recorded, authorization of the payment to be made and the dispatch of the payment for the claim.

In the case the insurer closes an eye to the IBNR claims, then she is sure that later she is going to have to pay for claims which no provision were made for. If the claims turn out to be huge then the possibility of the insurer leaving the business will also be high.

Much has not been explained on the way forward for Incurred But Not Reported Claims reserving. No particular proper guidelines has been set to direct the treating of Incurred But Not Reported Claims reserving. Not much work has been written on the best ways to treat this reserving.

Although the situation is changing in recent times, lack of articles, discussions and other related means of presenting the theory and practice of IBNR reserves leads one to conclude that the subject has suffered from neglect over the years and companies have not been allocating sufficient time and talent to this subject. (Bornhuetter & Fergusson, 1972)

On this note it is important for the insurer to find the “best” way to make provisions (reserve) for the incurred but not reported claims. The important point is that the insurer neither holds too little reserves, as the company will strain to fulfill its obligation of paying up for claims when they arise. Nor the fund should not also be unnecessarily too much as the company could use these capital for other use. Reserve estimates that are too low can lead to the conclusion that pricing is adequate when it is not, so we may fail to achieve our underwriting target in future periods, and we may experience unprofitable growth. Reserve estimates that are too high may lead to inflated prices, potentially limiting competitive opportunities. (The Progressive Corporation, 2010)

For example they could undertake an investment that would yield more earnings than laying the funds in the reserves. The question to ask is if the proposed methods are reliable and whether other stochastic methods can provide better estimates and more information concerning reserving for IBNR claims. If the methods are not as reliable as the stochastic approaches then it means that insurance companies are either not maximizing the shareholders interest or policy holder needs.

Several ways for calculation of IBNR reserves have been advanced, some of which take the deterministic approach while others the stochastic approach.

The papers comparers a simple technique under the deterministic methods with another under the stochastic approach to relate which way is better for reserving for IBNR.

1.2 Problem Statement

One of the ways to curb this risk, is by holding up an extra fund (reserving), to provide for the future liabilities. The issue narrows down to how insurers set up their reserves. What are the methods for reserving, and are they efficient in their estimation of the reserves?

The models used need to be accurate, since a mistake in the reserving for IBNR claims could result in huge amounts of losses for the insurer. The Faculty and Institute of Actuaries (2011) state that the reserve for the IBNR losses on the balance sheet of the insurer has the greatest liability hence attention needs to be given to the IBNR reserves.

The Kenyan law under the Insurance Act stipulate that every insurer should determine and disclose to the authority their value of claim reserves. Besides determining and disclosing of the claims reserve, the Act stipulates some methods by which the IBNR claim reserves should be valued.

All the methods proposed by the Act are deterministic in nature. If the stochastic methods prove to be better estimators for reserving than deterministic method. Then a stakeholder for the reserves is not getting his or her full value. If the less efficient methods are used the insurers are either holding more reserves hence not maximizing the shareholders interest or less therefore exposing policy holders to more risk?

1.3 Research Objectives

- To compare the stochastic IBNR claim reserving methods with deterministic methods with the aim to of bridging the gap between the two methods and its implementation in practice.

1.4 Research questions

1. Does stochastic methods for IBNR claim reserving provide better estimates than deterministic methods?

1.4 Significance of the research

The study of the research will be significant as it will benefit a wide range of persons who on one way or the other are affected by insurance. A huge part of it being the insurers. From the research better ways of reserving for IBNR claims will be found and this will improve practices in the industry on how best to provide for their liabilities.

IFoA (2011) states that the reserve for the IBNR losses on the balance sheet of the insurer has the greatest liability hence attention needs to be given to the IBNR reserves. Therefore if the losses related to IBNR claims have a high impact on the insurer's balance sheet then the best ways to calculate these reserves need to be employed.

To the insurer better methods of reserving will be unveiled since reserving affect almost the entirety of an insurance company.

IFoA (1997) highlighted the needs to calculation of reserves. Some of the reasons for proper calculation of reserves which include assessing the financial condition of an insurer. Shifts in the reserve are key to assessing the progress of an insurance company.

Secondly pricing in the business is going to be affected with better ways of estimating IBNR claims. This is because prices of the products are functions of the estimated future cost of claims. Solvency of the insurer also comes into play, assessing the ability of the insurer to meet its obligations would be affected by the reserve. Other factors such as the value on an insurer in times of mergers and acquisition need to be assessed, of which the reserves have a weight in the net worth of the insurer. Lastly the reserves will affect the decision of the insurer with relevance to taking out of reinsurance.

Therefore the study will be importance as it will provided more information about reserving to the insurers which directly trickles down to issues that greatly affect the insurers as briefly outlined in the paragraphs above.

Other parties will also benefit from the study. To the policy holder better ways of providing for reserves mean that the insurer will be more likely to pay up for the claims. To the regulator whose interest is to protect the public, better way of reserving informs him that it is more likely that the insurer will be solvent. To the shareholder better methods of reserving means that moneys are being used effectively hence maximization of profits.

2.0 LITERATURE REVIEW

2.1 IBNR in general

A total loss reserve is held by the insurer to provide for claim expenses. The reserve is composed of both reserve for the known claims and for the unknown which forms the IBNR. Hence, the overall loss reserve consists of two groups, i.e. the reported (known) claims and the claim losses believed to have been incurred but not yet reported and hence not yet known to the insurer. (Liu, 2008). The total loss reserve can be divided into the reserve for known claims and the incurred but not reported reserve (IBNR). The reserve for known claims represents the amount of paid loss that will be required to settle all reported claims not including payments already made on these claims. (David, 2000). Rastogi, (2003) Also outlines that for non-life insurers reserving is concerned with provision for mainly two liabilities. These are liability for Unexpired Risks, liability for unpaid claims. Includes outstanding claims and IBNR

In insurance a claim is a request filed for payment of the losses incurred due to a certain insured event. Reported claim is one that has already been processed to the extent that a central record on it is held (Ratemo & Weke, 2013). Incurred But Not Reported claims are those claims that the have occurred, therefore the insured has experienced a certain loss that was insured by the insurer. But the insured has not yet reported the claim to the insurer. The term refers to claims not yet known to the insurer, but for which a liability is believed to exist at the reserving date. (IFoA, 1997).

There are several reasons why the insurers takes key interest in the providence of reserves. Besides been required by the law, which is the regulator. The reserve has a considerable weight in their income statements. There are at least two reasons for interest in IBNR reserves. Farrokh, (1988) outlines that first, insurance statutes require a provision for IBNR reserves as part of the total liability of insurance companies. Second, these IBNR reserves are large. For some companies IBNR reserves may be hundreds of millions of dollars, so their accurate estimation is a matter of considerable concern. For group and, to a lesser extent, individual products, the IBNR reserves are often a large component of the policy liabilities or past claim costs and are therefore material to both valuation and pricing. (Life Insurance & Wealth Management Practice Committee, 2014). For insurance companies, the claims reserve is a very substantial balance sheet item, which can be large in relation to shareholders funds. (Tom, 2005)

2.2 Best Estimate

Where else best estimate can be confused for many meanings. In the process of estimation of the reserves one can be subject to serving many masters. Depending on where the priority or more need arise the best estimate can vary in meaning. One concept that appears frequently both in actuarial literature and in regulations pertaining to loss reserves is the concept of a "best estimate". The term might provide a good foundation for clear communication, but the profession has not put forth any consistent definition (Blum & Otto, 2008).

Different stakeholder have different interest in the values that the estimate will take. For a company management the reserve estimate should relate information that will help them in maximizing the company's profits and the company's viability. For the regulator the reserving estimates will be of interest to see that the company's solvency is enough therefore reserves will be directed towards conservativeness. This is to ensure the company does not go under. For a tax agent the reserve estimates should be near as possible to the exact payments to be paid. This is to ensure the reporting on the income that is earned is timely and almost correct. To the policy holder the main interest towards the reserve is that it is enough to meet up with the liabilities when they arise. That his/her claims will be paid on time at the same time the premium not been overcharged for the policy. It then goes beyond estimation of reserves but for actuaries to fully communicate the implications of their work to avoid misunderstanding. Therefore it is of contest that when actuaries are estimating the reserves whose priorities to satisfy. Would an actuary sacrifice one for another?

Blum & Otto (2008) State out that the best estimate should satisfy some objectives like soundness, preclude optimistic or pessimistic skewing be explainable to a wide audience, identify a target for a point estimate and define the target unambiguously. Further they state that the word “best” implies a particular point (i.e. better than all others) within the range of reasonable estimates. While different actuaries may produce different “best estimate” numbers, the range of best estimates among the actuaries should be considerably narrower than the range of all “reasonable” estimates.

Dependence Concepts

Several measures of dependence have been developed this include

- Concordance measure
- Quadrant dependence
- Tail dependence

Concordance measures

When considering a pair of random variables, they are considered to be concordant if small values of one variable tend to be related or associated with the variables of the other small variable. On the other hand large values of one variable are associated with the ones of the other.

DE Matteis, (2001) describes that given observations (*xi; yi*) and (*xj; yj*) are concordant, if *xi* < *xj* and *yi* < *yj*, or if *xi* > *xj* and *yi* > *yj*. Analogously, (*xi; yi*) and (*xj; yj*) are discordant if *xi* < *yj* and *yi* > *yj* or if *xi* > *xj* and *yi* < *yj*. Alternatively.

If $(X_1 - X_2)(Y_1 - Y_2) > 0$, then the pair (X_1, Y_1) and (X_2, Y_2) are said to be concordant. (1)

If $(X_1 - X_2)(Y_1 - Y_2) < 0$, then the pair (X_1, Y_1) and (X_2, Y_2) are said to be discordant. (2)

If $(X_1 - X_2)(Y_1 - Y_2) = 0$, then the pair (X_1, Y_1) and (X_2, Y_2) are said to be neither concordant nor discordant. (3)

The measures used to determine concordance between variables include

Pearson's correlation coefficient

The correlation coefficient is obtained by get the covariance between the two variables. This is divided by the standard deviation of both variables. The measure is simple in nature in that it measures the linear relationship between the two variables.

The equation below describes the correlation coefficient

$$\rho = \frac{cov(x, y)}{\delta x \delta y} \tag{4}$$

Kendall's tau

Considering $\binom{n}{k}$ distinct pairs of (Xi, Yi) and (Xj, Yj) random samples.

Then the Kendall's rank correlation can be obtained by

$$t = \frac{c - d}{c + d} = \frac{(c - d)}{\binom{n}{2}} \tag{5}$$

Or $$\rho t = \binom{n}{k}^{-1} \sum_{i<j}^{n} sign\ [(Xi - Xj)(Yi - Yj)] \quad (6)$$

Spearman's rho

The coefficient is used to measure the strength of monotonic relationship between the two variables

The estimator of the spearman's rank correlation $\rho S(X, Y)$ is defined by the following equation where X and Y are random variables with marginal distributions F_1 and F_2 and have a joint distribution F.

$$\rho s(X, Y) = \frac{12}{n(n^2 - 1)} \sum (\text{rank}(Xi) - \frac{n+1}{2})(\text{rank}(Yi) - \frac{n+1}{2}) \quad (7)$$

For a population

$$\rho s(X, Y) = \rho(F_1(X), F_2(Y)) \quad (8)$$

Positive Quadrant Dependence

This is a dependence structure which satisfies the following condition

$$Prob(X > x, Y > y) \geq Prob(X > x)\, Prob(Y > y) \quad (9)$$

In simple terms, this implies that the probability that two random variables are jointly large is greater than or equal to when they are looked at independently.

Tail dependence

The dependence structure is used to measure relationship between extreme values between the variables. Is basically explains the relationship on the tails.

Upper and lower tail dependence

Considering X and Y as random variables and F_1 and F_2 their respective distributions. The upper tail coefficient is obtained by

$$lim_{\propto -1-}\mathbb{P}[\mathrm{Y} > F_1^{-1}(\propto)|X > F_1^{-1}(\propto)] = \lambda u \tag{10}$$

Provided a limit $\lambda_u \epsilon[0,1\,]\ exists$ here

$$F_1^{-1}(\propto) = \inf\{x|F(x) \geq \alpha\}, \qquad \alpha\epsilon(0,1) \tag{11}$$

The lower tail dependence coefficient

$$lim_{\propto -1+}\mathbb{P}[\mathrm{Y} \leq F_2^{-1}(\propto)|X > F_1^{-1}(\propto)] = \lambda t \tag{12}$$

Provided $\lambda_t \epsilon[0,1\,]\ exists$

2.3 Deterministic approach

In the deterministic approach classical ways such as the Chain Ladder method, Inflation-Adjusted Chain Ladder, Accumulated Cost per Claim method, and the Bornhuetter-Fergusson method are used. (Ratemo & Weke, 2013) With these classical methods development patterns are established of which these patterns are used in the projection of the claim costs used in reserving. Development factors are determined and applied to the appropriate case of the incurred losses. The development factors may be used to show the ratios between cumulative claim amounts and claim numbers. (Subject CT6 Core Reading, 2011)A development factor may describe the ratio between cumulative claim amounts in consecutive years or between years over a longer period.

A typical way is to use the Run off triangle where data is presented inform of a triangle. Theresa, (2007) Highlights the purpose of the run off triangles been used as convenient way to present the data, easier to see patterns and relationships in historic data, easier to explain, logical and concise, can be for any data that demonstrates a reasonable growth pattern.

Claims are attributed to the year of which the insurance policy was taken. When the loss incurred is the accident years and the time (years) until the payment of the loss is referred to a development year. It may take some while when a loss occurs to be known or paid in full extent by the insurer.

The issue becomes to estimate the values of the lower triangle which represents the claims that are expected to be reported in the years to come until the policies written are run off. The lower triangle is used to estimate the reserve.

The table entries can include ultimate cost of claim or the number of claims depending on what the model desires to build its development factors.

All these is done with the assumption that the claims emerging from year to year follow a similar pattern in the accidents years.

In this paper the chain ladder method is used simply from the fact that the technique is simple to apply gives reasonable estimates and it is widely used. The basic chain-ladder is used as a benchmark, due to its generalized use and ease of application. The method is intuitively appealing and simple to calculate which often give reasonable results these attributes make the method popular (Gitonga, 2015). The reserving methods used in practice are frequently deterministic. A popular statistical method is the chain ladder method. (Bjorkwall, 2011)

However the deterministic method suffer a deficiency. This is because they provide the actuary with the point estimates leaving out the variance in the estimation which is an important statistic. Prakash & Thomas,(2000) States out that actuaries are often asked to provide a range or confidence level for the loss reserve along with a point estimate. Traditional methods of loss reserving do not provide an estimate of the variance of the estimated reserve.

The separation method is also used for later comparison with the stochastic methods because previous research on deterministic methods in short-term insurance contracts done by (Weke & Mureithi, 2006). Depicted that out of the deterministic methods used to estimate the best reserve estimates; the separation method proved to be the best. It gave the lowest mean, median, range and inter-quartile range for percentage residual errors.

2.3.1 The Chain Ladder Method

Mark (1993) State that the chain ladder method is probably the most popular method for estimating outstanding claim reserves. He further says that the reason for the chain ladder's popularity is because of its simplicity and does not rely on a distribution. (Schmidt & Wunsche, 1998) State that the reason for use of chain ladder method is because it exploits all data from the run-off triangle and provides simple estimate for the estimate total claims.

The model dates back to 1996. (Taylor, 2012) Expresses that the model coils from the chaining of a sequence of ratios into a ladder of factors (the sequence of rations are the age factors) from which projections of the ultimate claim value is estimated.

Incremental claims are expressed in the form of a run-off triangle. Development ratios for the accident years are then obtained from which projections of the cumulative claims are made. The reserve that needs to be held is the sum over all accident years from the projections made. Which is the difference of the cumulative amounts at the end of the last development year and the last entry in the development triangle in that accident year. The model is developed under the assumptions that payments from each accident year will develop in the same way. And that weighted average past inflation will be repeated in the future.

Below is a figure and an equation of the runoff triangle. Where C_{ij} represents the incremental claims.

$$C_{ij} = r_j . s_i . x_{i+j} + e_{ij} \tag{13}$$

R_j is the development factor for year j, representing the proportion of claim payments in year j. Each r_j is independent of the origin year i.

S_i is a parameter varying by origin year, *i*, representing the exposure, for example the number of claims incurred in the origin year *i*.

X_{i+j} is a parameter varying by calendar year, for example representing inflation

e_{ij} is an error term.

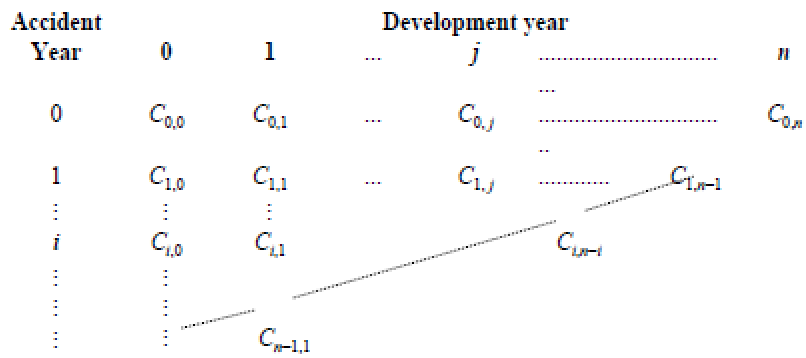

Figure 1 Runoff triangle

2.4 Stochastic Approach

Many studies have shown that there has been an increasing recognition in consideration of the random nature of the insurance loss of the insurance loss processes that leads to better predictions of ultimate losses. (Gitonga, 2015).The world is rapidly changing and therefore more sophisticated methods for reserving should be thought off. Methods that incorporate the randomness of liabilities. The global insurance industry is experiencing dramatic change as reserve and capital regulations transition from the traditional, prescriptive approach to stochastic, principles-based approaches. This fundamental change most likely portends the gradual global convergence of regulation across countries and insurance products. (Matthew, 2009)

Stochastic methods can also be employed in the estimation of IBNR reserving. Some of these methods include the Bootstrap, smoothing models and Archimedean Copulas. Throughout time many of the reserving in context of IBNR has been done from the deterministic approach. Perhaps it's because of its simplicity in understanding and applicability. But the question to ask is whether it is worthwhile to labor a bit more to understand the stochastic approach to reserving. Does the approach yield better estimates as compared to the deterministic approach.

In practice there is a long tradition of actuaries calculating reserve estimates according to deterministic methods without explicit reference to a stochastic model (Bjorkwall, 2011). But why is there little or no reference to the stochastic models in the calculation of reserves. Is it because there is little understanding or the deterministic models are enough hence no need for the stochastic models.

(England & Verrall, 2002) Provide some suggestions for little reference to stochastic methods. A number of reasons for this could be suggested, including a general lack of understanding of the methods; lack of flexibility in the methods; lack of suitable software. However the main reason is probably lack of need for the methods, when traditional methods suffice for the calculation of the best estimate of outstanding claims reserve.

However despite the little reference stochastic methods. The technique still bear some advantages over the deterministic methods. The stochastic methods can in many circumstances give more information as compared to the deterministic methods. This additional information can be used for better reserving practices.

Full distributions of the expected outcomes can be obtained from which measures such as percentiles and variability of claim reserve can be obtained, best estimates and worst case scenario of potential claim reserves. The primary advantage of stochastic reserving models is the availability of measures of precision of reserve estimates, and in this respect, attention is focused on the root mean squared error of prediction (England & Verrall, 2002).

In the stochastic methods estimation of IBNR reserves involves the development of probabilistic models. The reserve is viewed as a stochastic variable. The main parameters by which the IBNR reserve is estimated include the number of claims which is the claim frequency, claim amount forming the claim severity and the lag in reporting those claims. A distribution is developed using the variables from which the mean and variance of the reserve are obtained.

One of the shortcomings of the deterministic methods is that they don't pay much attention to the dependencies that exist between the cells. To estimate the reserve the dependencies between cells should be considered since the IBNR reserve is viewed as a summation of small univariate random variables which make the whole composition of the reserve. Most methods estimate the lower triangle cell-by-cell, and do not pay enough attention to the structure describing the dependencies between these cells. Indeed, each cell must be considered as a univariate random variable being part of the multivariate random variable describing the lower triangle. Hence, the IBNR reserve must be considered as a (univariate) random variable being the sum of the dependent components of the random vector describing the lower triangle (Mare, Jan, Eddy, & Hendrik, 2000)

More recently, greater interest has been expressed in estimating the downside potential of claim reserves, in addition to the best estimate. For that, it is necessary to be able to estimate the variability of claim reserve and ideally, to be able to estimate a full distribution of possible outcomes, from which percentiles (or other estimates) of that distribution can be obtained. Stochastic claim reserving methods extend traditional techniques to allow those additional methods to be estimated (England & Verrall, 2002). Mark,(1993) state that a confidence interval is of great interest for the practitioner because the estimated ultimate claim amount can never be an exact forecast of the true ultimate claims amount and therefore a confidence interval is of much greater information value. A confidence interval also automatically allows the inclusion of business policy into the claims reserving process by using a specific confidence probability

2.4.1 Bootstrap Method

This is one of the stochastic methods introduced by Efron in 1979 for calculation of loss reserving. The technique was a simple and intuitive method of making approximations to distributions which are hard or almost impossible to generate. It is a useful technique as it enables the assessment of the variability of the claim reserves alongside the predictions of the upper bounds at an adequate confidence interval.

Tom (2005) States out that its popularity is because of the combination of available computing power and theoretical development. He later points out that an advantage to the technique, that it can be applied to any data set without having to assume a distribution underlying the data set. Another advantage to that is that computer packages can handle large arrays of numbers with repeated sampling.

The data obtained is treated as the true reflection of the actual population and samples are obtained from it (bootstrap samples), this is referred to as re-sampling. Many samples are obtained from the data and this is done with replacement. From the sample the statistic of interest is calculated, this is referred to as the bootstrap statistic.

Tom (2005) the bootstrap technique is used to obtain prediction errors for different claims reserving methods, namely methods based on the chain-ladder technique and on generalized linear models.

However the bootstrap technique suffers from negative values, Mark & Olofsson (2006) outline that the bootstrap simulation is sensitive to negative values in the development triangle. The solution made here was to let the negative values in the development triangles equal zero.

2.4.2 Copula

Modelling of IBNR reserves consists of marginal distribution of the claim frequency claim amount and the delay time, and the dependency structure that is used to link these distributions. Therefore the basic problem become to understand the relationship or dependency among the multivariate outcomes.

Linear correlation is a commonly used measure of stochastic dependence. Its wide application is due to its ease of application and easiness to understand. However the linear correlation suffers a major drawback as it only works well for jointly elliptically distributed distributions meaning that the distributions should have a multivariate normal distribution. Linear and rank correlation measures do not explain well tail dependency.

In reality a lot of insurance classes are not adequately described by the normal distribution. Therefore a method to capture the tail dependency in the claims should be used. (Ratemo & Weke, 2013) State that general insurance contracts are either short or long tailed of the distribution function and that a multivariate simulation technique will only be possible if the whole dependency structure is known.

Regression analysis is another statistical methodology that is used to understand relationship between variables. Despite it capturing the effects of explanatory variables it requires the identification of one of the variables as the primary variable and another as the dependent variable. However it is also limited as in modelling of claims in insurance the main interest is to understand joint distributions as several variables are interacting simultaneously.

With all this deficiencies in the described methods copulas come in to fill in the gaps. Copulas combines individual marginal distribution into a multivariate joint distribution. It first describes each marginal distribution in isolation then then the dependency structure between the variables is established. Copulas thus are extremely helpful because they give a natural way of allowing dependency that is free from the drawbacks of correlation (Ratemo & Weke, 2013). Fortunately, the theory of copulas provides a flexible methodology for the general modelling of multivariate dependency (Zivot & Wang, 2007)

Sklar's Theorem

Let H be an n-dimensional distribution function with margins $F_1 \ldots F_n$. Then there exists an n-copula C such that for all **X** in $\mathbb{R}^n$

$$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \tag{14}$$

If C is an N-copula and $F_{1\ldots}$ F_n are the distributions functions making up the copulas, then it follows that H is an n-dimensional distribution function with $F_{1,\ldots\ldots,}F_n$ as the margins

Let H be an n-dimensional distribution function with continuous margins $F_{1\ \ldots}$ F_n and copula C (where C satisfies (2.1)). Then for any **u** in $[0, 1]^n$,

$$C(u_1, \ldots, u_n) = H(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) \tag{15}$$

Properties of copulas

1. $\forall(u_1, \ldots \ldots . u_n) in\ [0,1]^n$ if at least one component u_i is zero, then $C(u_1, \ldots \ldots . u_n)=0$ (16)

2. For $u_i \epsilon [0,1], C(1, \ldots .. ui \ldots \ldots ..1) = u_i\ \forall i \epsilon (1,2, \ldots \ldots , n)$ (17)

3. $\forall [u_{11}, u_{12}] * [u_{21}, u_{22}] * \ldots .* [u_{n1}, u_{n2}]\ n - dimensional\ rectangles\ in\ [0,1]^n$

$$\sum_{i1=1}^{2} \ldots \ldots \sum_{in=1}^{2} (-1)^{i1+i2+\cdots+1n} C(U1_{i1}, U2_{i2,\ldots\ldots,} Un_{in}) \geq 0 \tag{18}$$

Copulas and dependence measures

Copulas can be connected to the dependence measures discussed in the previous paragraphs.

Kendall's Tau: - Note that Kendall's tau is related to copulas in general in the following way;

$$\tau = 4 \int \int C(u,v) dC(u,v) - 1 \quad \text{or} \quad \tau = 4E[C(u,v)] - 1 \tag{19}$$

Spearman's *rho:* - Spearman's rho is related to copulas in the following ways;

$$\rho s = 12 \int \int uvdC\left(u, v\right) - 3 \tag{20}$$

$$\rho s = 12 \int \int C\left(u, v\right) dudv - 3 = 12E\left[UV\right] - 3$$
$$\rho s = \rho\left(u, v\right) = \rho\left(F\left(x\right), F\left(y\right)\right) \tag{21}$$

Positive Quadrant dependence: - In terms of copula; two variables are considered positive quadrant dependent if;

$$C\left(u, v\right) \geq uv \tag{22}$$

Upper tail dependence: - Consider the definition of upper tail dependence with regards to copula; if a bivariate copula *C* exists such that;

$$\lambda_U = \lim_{u->1^-} \frac{1 - 2u + C\left(u, u\right)}{1 - u} \tag{23}$$

If λ_U exists and is greater than 0 but less than or equal to 1 then *C* has upper tail dependence and if λ_U is 0 then *C* has upper tail independence.

Lower tail dependence: - If a bivariate copula *C* exists such that;

$$\lambda_L = \lim_{u->0^+} \frac{C(u,u)}{u} \tag{24}$$

If λ_L exists then C has lower tail dependence as long as λ_L is greater than 0 but less than or equal to 1 and has lower tail independence if λ_L is equal to 0.

3.0 RESEARCH METHODOLOGY

3.1 Research design

The research attempts to compare between the deterministic methods of IBNR claim reserving and stochastic IBNR claim reserving. The study examines the loss reserve estimation for the motor commercial insurance segment within one financial year. This simply because motor commercial has the highest contribution to gross premiums in the general insurance industry. According to the annual report of 2013 by the Association of Kenya Insurers. Motor commercial had the second largest contribution to the total gross premium in the insurance industry. Accounting for 22.6% of the gross premium.

On this given class of general insurance, IBNR claim reserves will be computed by both the stochastic and deterministic method. This research will be geared towards comparing deterministic reserving techniques and stochastic reserving technique for the motor commercial class of general insurance to ascertain the most effective.

Data required for the research will be obtained from an insurance company claims register. The data will mainly involve the claims information for the motor insurance line of business. Data on claims frequency, claims amount and delay time for the claims will also be collected for the application of the different models.

3.2 Population and Sampling

The sampling from the population will entail obtaining the claims data from a motor insurance line of business across the different years. Claims for one year will then be picked to model out both deterministic and stochastic IBNR reserving. Three hundred and twenty three claims will be used for the analysis. Data for one year will be preferred over another if it is complete, without missing entries.

3.3 Data Collection.

Both primary and secondary data was used in the research. The primary data obtained from the an insurance company will involve getting data on the number of claims reported and settled, the year of origin of the events, the paid amounts on the claims and the year of payments will collected. Secondary data will mainly be obtained from Insurance regulatory Authority and the Association of Kenya Insurers. These data will involve the general insurance lines and their proportions in market share.

3.4 Data Analysis

In order to come up with conclusive findings and recommendations analysis of the data collected will be done. The data will first be sorted in one financial year by their dates. Missing entries, erroneous and inconsistent data will be removed. The data is then tabulated in the run-off triangle which depicts the accident year and development periods format in a Microsoft Excel sheet.

		year of development					
		0	1	2	3	4	5
	1	1002	1855	2423	3335	3485	3700
	2	1113	2774	3422	3844	3844	
year of Origin	3	1265	2433	3223	3977		
	4	1490	2873	3880			
	5	*1725*	3261				
	6	1889					

Figure 2. Claims per each development years

The chain ladder method of reserving is modelled on the data together with the stochastic model then compared with the actual IBNR claims revealed.

To model the stochastic IBNR reserve estimate. An appropriate copular model is chosen through a goodness of fit test. This can be easily done by an R code. The appropriate copula can also be determined using the VineCopula package in R. The package identifies the best copula to be fitted in the data. Marginal distributions for the claim frequency and claim amount are obtained then a joint multivariate distribution modelled out of it. The reserves are obtained by multiplying the estimated number of claims in development year with the average claim amount in the development year. With the help of Microsoft Excel and R software reserve estimate from the determinist methods and stochastic methods will then be compared to the actual amount to be reserved.

The copulas model is chosen because of its strength is modelling dependency structures between the parameters such as claim amount and delay time for the estimation of IBNR reserves. Secondly it is not affected by the negative claim amounts like the bootstrapping method. Mark & Olofsson, (2006) outline that the bootstrap simulation is sensitive to negative values in the development triangle. The solution made here was to let the negative values in the development triangles equal zero.

This analysis is done at a company level based on data from an insurance companies.

CHAPTER FOUR: RESULTS AND ANALYSIS

4.1Deterministic

Data was tabulated in a runoff triangle. Development factors calculated thereafter projections of the runoff claims made. IBNR was then estimated from the projected run off claims.

Below is a figure of the claims development triangle.

								CUMULATIVE CLAIM PAYMENTS											
ACCIDENT MONTH						DEVELOPMENT MONTH													
	0		1		2		3		4		5		6		7		8		9
1	8065675	1.001932	8081255	1.004851	8120455	1	8120455	1	8120455	1	8120455	1	8120455	1	8120455	1	8120455	1.003662	8150195
2	3003908	1.001664	3008908	1.004241	3021668	1	3021668	1	3021668	1	3021668	1	3021668	1	3021668	1	3021668	1	3021668
3	8009582	1.003808	8040082	1	8040082	1	8040082	1	8040082	1	8040082	1.018803	8191259	1	8191259	1	8191259	1	8191259
4	2927182	1.001879	2932682	1.006138	2950682	1.003389	2960682	1.006367	2979532	1	2979532	1	2979532	1	2979532	1	2979532	1.001221	2983169
5	5627460	1	5627460	1.003554	5647460	1	5647460	1.010438	5706410	1	5706410	1	5706410	1	5706410	1	5706410	1.001221	5713376
6	3880319	1.002835	3891319	1.008082	3922769	1.105997	4338569	1	4338569	1	4338569	1	4338569	1	4338569	1	4338569	1.001221	4343866
7	9029867	1.002384	9051397	1.001105	9061397	1.000441	9065397	1	9065397	1	9065397	1.003134	9093806	1	9093806	1	9093806	1.001221	9104907
8	6998988	1.002718	7018009	1	7018009	1.002824	7037829	1	7037829	1	7037829	1.003134	7059884	1	7059884	1	7059884	1.001221	7068503
9	2914240	1	2914240	1	2914240	1	2914240	1.002101	2920362	1	2920362	1.003134	2929514	1	2929514	1	2929514	1.001221	2933090
10	5518818	1.01879	5622518	1	5622518	1.012517	5692894	1.002101	5704853	1	5704853	1.003134	5722731	1	5722731	1	5722731	1.001221	5729717
11	4015776	1	4015776	1.002797	4027008	1.012517	4077413	1.002101	4085979	1	4085979	1.003134	4098783	1	4098783	1	4098783	1.001221	4103787
12	1144986	1.003274	1148734	1.002797	1151947	1.012517	1166366	1.002101	1168816	1	1168816	1.003134	1172479	1	1172479	1	1172479	1.001221	1173910

Figure 3 Runoff triangle with development factors and projected claims

5		6		7		8		9		10		11		
8120455	1	8120455	1	8120455	1	8120455	1.003662	8150195	1	8150195	1	8150195		8150194.9
3021668	1	3021668	1	3021668	1	3021668	1	3021668	1	3021668	1	3021668		0
8040082	1.018803	8191259	1	8191259	1	8191259	1	8191259	1	8191259	1	8191259		0
2979532	1	2979532	1	2979532	1	2979532	1.001221	2983169	1	2983169	1	2983169		3637.369366
5706410	1	5706410	1	5706410	1	5706410	1.001221	5713376	1	5713376	1	5713376		6966.302547
4338569	1	4338569	1	4338569	1	4338569	1.001221	4343866	1	4343866	1	4343866		4338569.08
9065397	1.003134	9093806	1	9093806	1	9093806	1.001221	9104907	1	9104907	1	9104907		39510.90306
7037829	1.003134	7059884	1	7059884	1	7059884	1.001221	7068503	1	7068503	1	7068503		30673.88978
2920362	1.003134	2929514	1	2929514	1	2929514	1.001221	2933090	1	2933090	1	2933090		18849.95531
5704853	1.003134	5722731	1	5722731	1	5722731	1.001221	5729717	1	5729717	1	5729717		107198.8403
4085979	1.003134	4098783	1	4098783	1	4098783	1.001221	4103787	1	4103787	1	4103787		88011.03829
1168816	1.003134	1172479	1	1172479	1	1172479	1.001221	1173910	1	1173910	1	1173910		28924.30623
													reserve	12,812,536.58

Figure 4 Runoff triangle with estimated IBNR reserve

A reserve estimate of 12,812,536 was estimated by the chain ladder technique.

4.2Stochastic.

4.2.1Fitting Marginal Distribution

4.1.1 Report lag distribution

Data on report lag, which was the time between which an accident occurred and when it was reported, was considered to be discrete. That is in days. Discrete distributions were fitted on the data and the best distribution fit on the data picked. Sigma Magic package for excel was also used to identify which distribution fit was best for the data.

Below is a figure of the appropriate fit chosen for the data. The fit with the lowest chi-square value and the highest p-value is considered to be the best fit. The chi-square is a test statistic which is used to conduct a goodness of fit test. It is used to compare expected data with the actual observed data. The lower the chi-square value the better the model. While the P value shows the probability or likelihood that one will observe a sample statistic that will be as extreme as the test statistic. The higher the probability the better the fit.

Input Summary

Data type: Discrete
Method: Matching Moments
Fit: Discrete Uniform
Confidence level: 95%
Rows: 323

Goodness of Fit

Distribution: Discrete Uniform (0, 32)
Chi-Sq: 331.19, P Value: 0.062

Conclusion

⇒ The Discrete Uniform distribution was fit to the selected data

Figure 5 summary of report lag distribution fit

4.1.2 Claim Amount distribution

Data on the claim amount was considered to be continuous. Continuous distributions were fitted on the data and the best distribution fit on the data picked. The sigma magic package in excel was used to identify the best fit for the claim amounts.

Below is a figure of the best fit for the data.

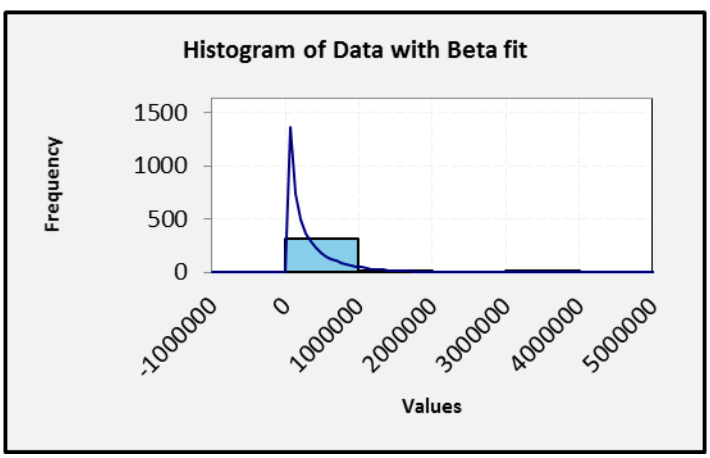

Figure 6 Claim amount histogram fit

Input Summary

Data type: Continuous
Method: Matching Moments
Fit: Best Distribution
Confidence level: 95%
Rows: 323

Analysis Results

Optimal distribution found: Beta
Distribution: Beta (0.309, 5.13, 0, 0.34E+7)

Distributi	Chi-Sq	P	Fit
Beta	19.13	0.321	OK
Cauchy	20.27	0.261	OK
Erlang	*	*	*
Extreme	4068.54	0.0	
Exponenti	209523.2	0.0	
Gamma	36.11	0.004	
Laplace	18317.16	0.0	
Log	*	*	*
Logistic	467452.2	0.0	
Log	*	*	*
Normal	20834.68	0.0	
Pareto	*	*	*
Power	41.46	< 0.001	
Rayleigh	1.74E+11	0.0	
Triangular	*	*	*
Uniform	269.01	< 0.001	
Weibull	89.92	< 0.001	

Conclusion

→ The Beta distribution best fits the selected data.

Figure 7 Summary statistics of the different fits on claim amount

4.2.2 Fitting the Copula distribution.

Using the VineCopula package in R the best copula was identified and fitted onto the data directly by a single code.

Below is a figure of the best copula chosen in R and the parameter specified.

```
> library(VineCopula)
> data<-read.csv(file.choose())
> var_a<-pobs(data)[,1]
> var_b<-pobs(data)[,2]
> selectedcopula<-BiCopSelect(var_a,var_b,familyset = TRUE)
> selectedcopula
$p.value.indeptest
[1] NA

$family
[1] 1

$par
[1] -0.03343665

$par2
[1] 0

attr(,"class")
[1] "BiCop"
> |
```

Figure 8 Copula family fitted on the data with the parameter

4.3 Estimating IBNR

Thereafter determining the best copula to use. For each development time unit the average claim size was found by simulation. The average number of claims in each time unit was determined from the number of claims distribution from the raw data. Finally reserve was calculated by multiplying average claim size with number of claims reported in each development time unit.

Below is a figure for the simulated claim amount and report delay data.

Console ~/

```
$family
[1] 1

$par
[1] -0.1837198

$par2
[1] 0

attr(,"class")
[1] "BiCop"
> cop<-BiCop(family=1,par=-0.1837198)
> simdata<-BiCopSim(200,cop)
> simdata
               [,1]       [,2]
  [1,] 0.248767158 0.57394903
  [2,] 0.792871338 0.46724913
  [3,] 0.669586610 0.14291616
  [4,] 0.588300342 0.78949463
  [5,] 0.601173507 0.05721714
  [6,] 0.596982289 0.62007668
  [7,] 0.368481077 0.84369123
  [8,] 0.617710644 0.84947866
  [9,] 0.763586808 0.21498412
 [10,] 0.349650110 0.42083640
 [11,] 0.110733193 0.99912101
 [12,] 0.341526116 0.35021860
 [13,] 0.968914531 0.01908652
 [14,] 0.911007310 0.76047976
 [15,] 0.637601228 0.27825135
 [16,] 0.081108870 0.40725968
 [17,] 0.447846441 0.43310700
 [18,] 0.549729916 0.59706393
 [19,] 0.918781986 0.43320040
 [20,] 0.066215515 0.23265715
 [21,] 0.466760190 0.58280563
 [22,] 0.884310253 0.79495694
 [23,] 0.521800337 0.34983595
 [24,] 0.542748782 0.91502574
 [25,] 0.678183211 0.20934464
```

Figure 9 simulated 200 data points from the copula fitted

CHAPTER FIVE: DISCUSSION AND CONCLUSION

5.1 Discussion

There are many ways of approaching the issue of reserving for losses and IBNR in particular. Throughout time methods are been refined and others achieved. If a method provides a better guide and a sounder estimate than another, it is prudent to prefer it over the other. Stochastic methods prove to be more daunting in their applications however they are handful and better in provision for loss reserving. The chain ladder model estimated an IBNR reserve of 12,812,536.58 while the copula model estimated an IBNR reserve of 5,293,021.79. Later on in the next year the actual amount to be reserved was 3,071,032.00. Though both models had over reserved as compared to the actual amount to be reserved. The copula model proved to be closer to the actual reserve amount than the chain ladder. Therefore better than the chain ladder.

5.2 Conclusion

Insurers should embrace better methods of reserving, so as to enable optimal utility of available funds. However a combination of both stochastic and deterministic methods can also be useful in informing the best reserve to set aside. This is because the world is dynamic and one cannot afford to rely entirely on one method of reserving. The relationships between variables should be explored. Vast access to technological capabilities should be taken advantage of in order to improve loss control.

Deterministic methods can be used as the basic underpinnings due to their simplicity, while the stochastic methods act as a supplement.

References

Bjorkwall, s. (2011). Stochastic Claim Reserving In Non-Life Insurance;Bootstrap and Smoothing models. *Department of Mathematics,Stockholm University*.

Blum, K. A., & Otto, D. J. (2008). Best estimate loss reserving: an actuarial perspective. *CAS Forum Fall, 1*.

Bornhuetter, R. L., & Fergusson, R. E. (1972). THE ACTUARY AND IBNR. *Proceedings of the Casualty Actuarial Society*.

David, S. (2000). A Survey of Loss Reserving Methods.

DE Matteis, R. (2001). FITTING COPULAS TO DATA. *Diploma thesis Institute of Mathematics of the University of Zurich*.

England, P., & Verrall, R. (2002). STOCHASTIC CLAIM RESERVING IN GENERAL INSURANCE. *Epsom: Institute of Actuaries.*

Faculty and Institute of Actuaries. (1997). *CLAIM RESERVING MANUAL.*

Farrokh, G. (1988). Assesing the Variability of IBNR Reserves: A Simulation Approach. *The Journal of Insurance Issues and Practices, 11*.

Gitonga, A. (2015). *Analysis of Variability in IBNR Estimates with Run off Triangles.* College of Biological and Physical Sciences School of Mathematics.

Kenya,T.G. (2015). *Laws of Kenya.* The Insurance Act Nairobi: Insurance Regulatory Authority.

Life Insurance & Wealth Management Practice Committee. (2014). Discussion Note: IBNR.

Liu, H. (2008). Stochastic Claim Reseving for Methods which Combine Information from Multiple Data Sets.

Mare, J. G., Jan, D., Eddy, V. B., & Hendrik, R. (2000). Some remarks on IBNR evaluation techniques.

Mark, T. (1993). MEASURING THE VARIABILITY OF CAHIN LADDER RESERVE ESTIMATES. *Casualty Actuarial Society* .

Mark, T., & Olofsson, M. (2006). Stochastic Loss Reserving. Testing the new guidelines from the Australian Prudential Regulation Authority on Swedish portfolio data using a Bootstrap simulation and distribution- free method.

Matthew, C. (2009). Uncertainities, Challanges and Opportunities of Global Insurance Regulatory Convergence. *Society of Actuaries*.

Prakash, N., & Thomas, W. (2000). A Comparative Satudy of the Performance of Loss Reserving Methods through Simulation. *Journal of Actuarial Practice, 8*.

Rastogi, A. A. (2003). Reserving for General Insurance Liabilities: Adequacy of reserves of Indian Non-life Insurers. *5th Global Conference of Actuaries*.

Ratemo , C., & Weke, P. (2013). ESTIMATING IBNR CLAIM RESERVES FOR GENERAL INSURANCE USING ARCHIMEDEAN COPULAS. *School of Mathematics, University of Nairobi, Kenya*.

Renshaw, A. E. (1989). Chain Ladder and Interactive modelling (claims reserving and GLIM).

Schmidt, K. D., & Wunsche, A. (1998). Chain ladder, marginal sum and maximum likelihood estimation. *Blatter der DGVFM, 23(3)*, 267-277.

Taylor, G. (2012). Loss reserving: an actuarial perspective. *Springer Science & Business Media, 21.*

The Progressive Corporation. (2010). *Report on Loss Reserving Practices.* Ohio.

Theresa, B. (2007). Property and Casualty Reserving Techniques. *Valuation and Reserving Techniques*.

Tom, H. (2005). *Modern Reserving Techniques For the Insurance Business.* Leuven.

Weke, P. G., & Mureithi, A. T. (2006). DETERMINISTIC AND STOCHASTIC MODELLING OF TECHNICAL RESERVES IN SHORT-TERM INSURANCE CONTRACTS. *The East African Journal of Statistics, VOL. 1, Number 2*.

Zivot, E., & Wang, J. (2007). *Modelling Financial Time Series with S-PLUS* (Vol. 191). Springer Science & Business Media.

APPENDIX

R Code:

```
Marginal Distributions
library(fitdistrplus)
library(maxLik)

##### REPORT LAG ########
data<-read.csv(file.choose() header = TRUE)
hist(data$Report, prob=TRUE, xlab="Report delay in days", main="Fitting distribution")

summary(fitdist(data$Report, "nbinom"))
lines(x, dnbinom(x,size=0.47, mu=56.24), col="blue")
summary(fitdist(data$Report, "geom"))
lines(x, dgeom(x,prob=0.01), col="red")
ks.test(data$Report, "pnbinom",size=0.47, mu=56.24)
ks.test(data$Report, "pgeom" ,prob=0.01)

######## CLAIM ESTIMATES #########
data<-read.csv(file.choose() header = TRUE)
hist(data$Estimate, prob=TRUE, xlab="Claim Amount", main="Fitting log normal distribution
curve")
x<-seq(from=10000,to=1000000, by=1)
summary(fitdist(data$Estimate, "lnorm"))
lines(x, dlnorm(x,meanlog=10.706871, sdlog=1.41), col="black")
summary(fitdist(data$Estimate, "weibull"))
lines(x, dweibull(x,shape=0.870, scale=0.0009), col="blue")
```

```
########fitting the best copula######

library(copula)
library(VineCopula)
data<-read.csv(file.choose())
var_a<-pobs(data)[,1]
var_b<-pobs(data)[,2]
selectedcopula<-BiCopSelect(var_a,var_b,familyset = TRUE)
selectedcopula

#######simulation from the copula#####
simdata<-BiCopSim(200,cop)
simdata
```