



Stockholms  
universitet

# Anpassning av copulamodeller för en villaförsäkring

Emma Södergren

Kandidatuppsats 2012:9  
Matematisk statistik  
December 2012

[www.math.su.se](http://www.math.su.se)

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm

# Anpassning av copulamodeller för en villaförsäkring

Emma Södergren\*

December 2012

## Sammanfattning

I denna uppsats studerar vi möjligheterna att beskriva resultatet från Trygg-Hansas villaförsäkring med hjälp av copulamodellering. Copulamodeller kan användas för att beskriva ett beroende mellan två eller flera variabler, grunderna för detta kommer även att beskrivas i detta arbete. Vi kommer att undersöka vilka datatyper som lämpar sig för denna typ av modellering och vilka justeringar som måste göras på data. Avslutningsvis kommer vi att utse den copulamodell som ger bästa anpassning till data och gå igenom hur man med hjälp av denna kan simulera fram nya observationer. För denna typ av försäkring visar det sig att inkrementella betalningar lämpar sig bra för copulamodellering samt att Gumbel copula beskriver detta beroende bäst av de testade modellerna.

---

\*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.  
E-post: [sodergren.emma@gmail.com](mailto:sodergren.emma@gmail.com). Handledare: Martin Sköld.

## **Abstract**

In this report we will study the possibilities to describe the result for stand alone house insurance with copula using data from Trygg-Hansa, a Swedish insurance company. Copulas can be used to describe a dependence between two or more variables, the basics of which will also be described in this report. We will investigate which type of data that is suitable for this type of modeling and which adjustments needs to be done on the basic data. Finally we will choose one copula model that gives the best fit to data and describe how new observations can be simulated from this copula model. The report suggests that for this type of insurance incremental payments is the most suitable for copula modeling and that Gumbel copula describes the dependence best of the models tested.

## **Förord och tack**

Detta är en kandidatuppsats på 15 hp vid Stockholms Universitet i samarbete med försäkringsbolaget Trygg-Hansa som leder till en Kandidatexamen i Matematisk Statistik.

Jag skulle vilja rikta ett stort tack till Lars Klingberg, chefsaktuarie på Trygg-Hansa samt till min handledare Martin Sköld för värdefull hjälp, rådgivning och förslag om nyttig litteratur i ämnet.

# Innehåll

<b>1</b>	<b>Introduktion</b>	<b>5</b>
<b>2</b>	<b>Copula</b>	<b>5</b>
2.1	Definition och några kända egenskaper . . . . .	5
2.2	Rangkorrelation . . . . .	7
2.2.1	Spearman's rho . . . . .	7
2.2.2	Kendall's tau . . . . .	8
2.3	Ett urval av copulamodeller . . . . .	8
2.3.1	Svansberoende . . . . .	9
2.3.2	Gumbel copula . . . . .	10
2.3.3	Clayton copula . . . . .	11
2.3.4	Ali-Mikhail-Haq copula . . . . .	11
2.3.5	Frank copula . . . . .	11
2.4	Slumpgenerering från tvådimensionella copulafunktioner . . .	11
2.5	Parameterskattningar . . . . .	12
<b>3</b>	<b>Termer och förklaring av data</b>	<b>13</b>
3.1	Datatyper . . . . .	14
3.1.1	Inkrementellt data . . . . .	14
3.1.2	Data baserat på skadedatum . . . . .	14
3.2	Årsrisk . . . . .	15
<b>4</b>	<b>Analys av data</b>	<b>15</b>
4.1	Förarbete med data . . . . .	15
4.2	Test av beroende . . . . .	17
4.3	Omvandlig till likformig fördelning . . . . .	22
4.4	Anpassning av copulamodeller . . . . .	23
<b>5</b>	<b>Diskussion</b>	<b>26</b>
<b>6</b>	<b>Appendix</b>	<b>30</b>

# 1 Introduktion

När en slumpvariabel är flerdimensionell kan det ibland vara viktigt att förstå hur de olika dimensionerna är beroende av varandra. Detta är speciellt nödvändigt då man önskar att simulera fram nya observationer baserat på historiska observationer. Om vi tänker oss att det handlar om en kostnad, då kan oftast den totala kostnaden delas upp i mindre segment där varje segment kan ses som en dimension av slumpvariabeln. Det gäller då alltså att totalsumman är summan av de värden slumpvariabeln ger i var dimension. När man ser på skadekostnader skulle segmentuppdelningen kunna vara baserad på vilken skadetyper kostnaden avser. För en villaförsäkring skulle sådana skadetyper kunna vara vatten, brand, stöld etc. Om det finns orsak att tro att dessa skadetyper inte är oberoende av varandra finns det fördelar att konstruera någon modell som beskriver beroendet. För detta finns det självklart fler än en väg att gå men i detta arbete så ska vi titta närmare på de modeller som går under benämningen copula. En orsak till att man skulle vilja se varje simulerad observation med denna uppdelning skulle kunna vara att man vill testa effekten av olika trender sett till skadetyperna. Detta skulle till exempel kunna röra sig om att försäkringsvillkoren förändras för någon viss skadetyper oberoende andra skadetyper och att man har en uppfattning om hur stor påverkan skulle vara sett till en bestämd procentsats.

## 2 Copula

Ordet copula är ett latinskt verb som betyder länk eller sammanknytning. Detta gäller även definitionen av begreppet i statistiska sammanhang. Inom statistiken används copula för att beskriva beroendet mellan två eller flera slumpvariabler.

### 2.1 Definition och några kända egenskaper

Ett sätt att beskriva beroendet mellan två eller flera slumpvariabler på är att uttrycka en fördelningsfunktion,  $H$ , enligt följande:

$$H(x_1, x_2, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n].$$

Förutsatt att slumpvariablerna  $X$  enbart kan anta reella tal är funktionen  $H$  definierad i rummet  $R^n$ . Funktionen  $H$  ger oss inte bara information om slumpvariablernas beroende utan även om var dimensions enskilda fördelningsfunktion. Det innebär att en och samma beroendestruktur kan få olika fördelningsfunktioner,  $H$ , om de enskilda variablerna i sig har andra fördelningsfunktioner. Det kan då vara av intresse att helt isolera beroendet från de enskilda fördelningsfunktionerna. Ett steg i riktning mot detta är att istället

betrakta den  $n$ -dimensionella slumpvariabeln  $(F_1(X_1), F_2(X_2), \dots, F_n(X_n))$  där  $F_i$  är fördelningsfunktionen för  $X_i$ . Om vi sätter  $U_i = F_i(X_i) \forall i \in [1, n]$  och låter funktionen  $C$  vara fördelningsfunktionen för  $(U_1, U_2, \dots, U_n)$ , så gäller det att:

$$\begin{aligned} C(u_1, u_2, \dots, u_n) &= P[U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n] \\ &= P[F_1^{-1}(U_1) \leq F_1^{-1}(u_1), F_2^{-1}(U_2) \leq F_2^{-1}(u_2), \dots, F_n^{-1}(U_n) \leq F_n^{-1}(u_n)] \\ &= P[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \dots, X_n \leq F_n^{-1}(u_n)], \end{aligned}$$

där

$$U_i \sim U(0, 1) \forall i \in [1, n].$$

Detta är även hur man definierar copulafunktionen,  $C$ , vilket kan sammanfattas:

$$C(u_1, u_2, \dots, u_n) = P[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \dots, X_n \leq F_n^{-1}(u_n)],$$

där

$$\begin{aligned} (X_1, X_2, \dots, X_n) &= (F_1^{-1}(U_1), F_2^{-1}(U_2), \dots, F_n^{-1}(U_n)), \\ U_i &\sim U(0, 1) \quad \forall i \in [1, n] \end{aligned}$$

och  $F_i$  fördelningsfunktionen för  $X_i$  samt  $F_i^{-1}$  den inverterade fördelningsfunktionen  $\forall i \in [1, n]$ . Det finns här även ett antal naturliga egenskaper hos copulafunktionen som bör nämnas.

- Den  $n$ -dimensionella copulan är definierad i rummet  $[0, 1]^n$  och antar enbart värden inom intervallet  $[0, 1]$ .
- Den  $n$ -dimensionella copulan har en likformig marginalfördelning för samtliga  $n$  dimensioner.
- Copulan antar värdet 0 om något av dess argument är 0, dvs.  
 $C(u_1, u_2, \dots, u_{i-1}, 0, u_{i+1}, u_{i+2}, \dots, u_n) = 0$ .
- Om ett av argumenten i copulafunktionen är  $u$  och övriga argument är 1 så antar copulan värdet  $u$ , givet att  $u \in [0, 1]$ ,  
 $C(1, 1, \dots, 1, u, 1, 1, \dots, 1) = u$ .
- $C = \prod_{i=1}^n u_i$  omm samtliga slumpvariabler är oberoende.

Utän att veta något om copulan i sig så kan man enbart med hjälp av dess argument ge ett minsta intervall för var det är möjligt att copulan kan anta värden. Det finns en sats för hur man konstruerar detta intervall som har baserats på arbeten av Maurice René Fréchet och Wassily Hoeffding, denna sats kallas Fréchet–Hoeffding copulagränser.



### Definition: Fréchet-Hoeffding copulagränser

För godtycklig Copula  $C:[0,1]^n \rightarrow [0,1]$  och godtycklig  $(u_1, u_2, \dots, u_n) \in [0,1]^n$  så gäller:

$$W(u_1, u_2, \dots, u_n) \leq C(u_1, u_2, \dots, u_n) \leq M(u_1, u_2, \dots, u_n),$$

där

$$W(u_1, u_2, \dots, u_n) = \max(1 - n + \sum_{i=1}^n u_i, 0)$$

$$M(u_1, u_2, \dots, u_n) = \min(u_1, u_2, \dots, u_n).$$

$W$  kallas Fréchet–Hoeffding nedre gräns och  $M$  kallas Fréchet–Hoeffding övre gräns.

## 2.2 Rangkorrelation

När man använder korrelationsmått såsom Pearsons korrelation mäter man hur pass bra man kan anpassa en rät linje till beroendet. Det är inte alltid ett lämpligt antagande att beroendet är linjärt. Antag slumpvariabeln  $(X, Y)$  där  $Y = \exp^X$ ,  $X$  och  $Y$  är alltså helt beroende av varandra samt givet att  $X$  är stort så kommer även  $Y$  vara stort och vice versa. Det är då önskvärt att använda ett korrelationsmått som indikerar detta beroende genom att returnera värdet 1. Givet att vi har fler än två observationer så kommer inte Pearsons korrelationsmått uppfylla det kravet. Det kan då vara en fördel att istället mäta korrelationen sett till rangen (Spearman's rho) alternativt antalet concordanta och discordanta par (Kendall's tau).

### 2.2.1 Spearman's rho

Detta mått är mycket likt Pearson's korrelation, vilken defineras som följer: Antag att vi har  $n$  utfall av slumpvariabeln  $(X, Y)$  och betecknar dessa med  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  så gäller:

$$\rho_{(X,Y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Skillnaden är att Spearman istället låter  $x_i$  samt  $y_i$  beteckna rangen av utfallen för  $X$  och  $Y$  separat.

Beräkningen för Spearman's  $\rho$  sker enligt följande schema:

1.) Sortera observationerna för  $X$  i växande ordning och låt  $x_i'$  beteckna ordningsnumret för  $x_i$ , dvs. för den observation med det högsta värdet ges den högsta rangen.

- 2.) Om något värde antas mer än en gång låt  $x_i'$  beteckna medel av de ordningsnummer som ges för det värde som  $x_i$  antar.
- 3.) På samma sätt för  $Y$
- 4.) Ersätt  $x_i$  med  $x_i'$  samt  $y_i$  med  $y_i'$  och beräkna  $\rho$  enligt definitionen ovan.

### 2.2.2 Kendall's tau

För att beräkna Kendalls tau behöver man veta antalet konkordanta och diskordanta par bland observationerna.

Låt oss betrakta paret  $(x_i, y_i)$  och  $(x_j, y_j)$  där  $i \neq j$  och sätter  $i$  och  $j$  så att  $x_i < x_j$ .

Om  $y_i < y_j$  så är paret är konkordant, är  $y_i > y_j$ , är paret diskordant.

Skulle det vara så att  $x_i = x_j$  eller  $y_i = y_j$  är paret varken konkordant eller diskordant.

Därefter beräknas Kendall's tau enligt följande där  $n$  är antalet observationer:

$$\tau = \frac{\#konkordanta\ par - \#diskordanta\ par}{\binom{n}{2}}$$

$\binom{n}{2}$  i nämnaren förklaras av att detta är det totala antalet par.

### 2.3 Ett urval av copulamodeller

Vi kommer enbart betrakta ett urval av de mer kända copulamodellerna som tillhör den arkimediska copulafamiljen. En arkimedisk copulafunktion är en funktion som kan defineras med hjälp av en generatorfunktion,  $\phi$ , enligt följande[7]:

$$C(u_1, u_2, \dots, u_d) = \phi^{-1}[\phi(u_1) + \phi(u_2) + \dots + \phi(u_d)]$$

där  $\phi^{-1}$  är inversen till  $\phi$  samt att  $\phi$  är en strikt växande funktion från  $[0,1]$  till  $[0,\infty]$ .

Orsaken till begränsningen att enbart inkludera den arkimediska copulafamiljen är att dessa copulamodeller är relativt enkla att konstruera samt att vi inom den familjen täcker många typer av beroenden. Inom den arkimediska copulafamiljen kommer vi att titta närmare på följande copulamodeller: *Gumbel*, *Clayton*, *Ali-Mikhail-Haq* samt *Frank*. Copulamodellernas funktioner kan uttryckas enligt följande där  $\theta$  är den obestämda parametern(se [2] samt [4]).

Gumbel copula för  $\theta \in [1, \infty)$ :

$$C^{Gu}(u_1, u_2, \theta) = \exp(-((- \log(u_1))^\theta + (- \log(u_2))^\theta)^{1/\theta})$$

Clayton copula för  $\theta \in (0, \infty)$ :

$$C^{Cl}(u_1, u_2, \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

Ali-Mikhail-Haq copula för  $\theta \in [-1, 1]$ :

$$C^{AMH}(u_1, u_2, \theta) = \frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}$$

Frank copula för  $\theta \in \mathbb{R} \setminus \{0\}$ :

$$C^{Fr}(u_1, u_2, \theta) = -\frac{1}{\theta} \log \left( 1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right)$$

Vi ska nu titta närmare på hur de olika modellerna ser ut samt hur de påverkas för olika val av parametern,  $\theta$ . Innan dess behöver vi dock en enklare förklaring av ett begrepp som fortsättningsvis kommer att användas nämligen *svansberoende*. En samling av slumpgenererade plottar för dessa modeller finns i fig. 1. För konturplottar samt fördelningsplottar för copulafunktionerna hänvisas till Appendix (fig. 8-9).

### 2.3.1 Svansberoende

Om vi har ett svansberoende så innebär detta att beroendet blir starkare då variablerna är väldigt höga och/eller låga. Ett sådant beroende känns igen genom att  $U(0, 1)^d$  observationerna lägger sig mer tätt i något/några hörn av rummet (eller enligt begränsningen för denna uppsats, ytan).

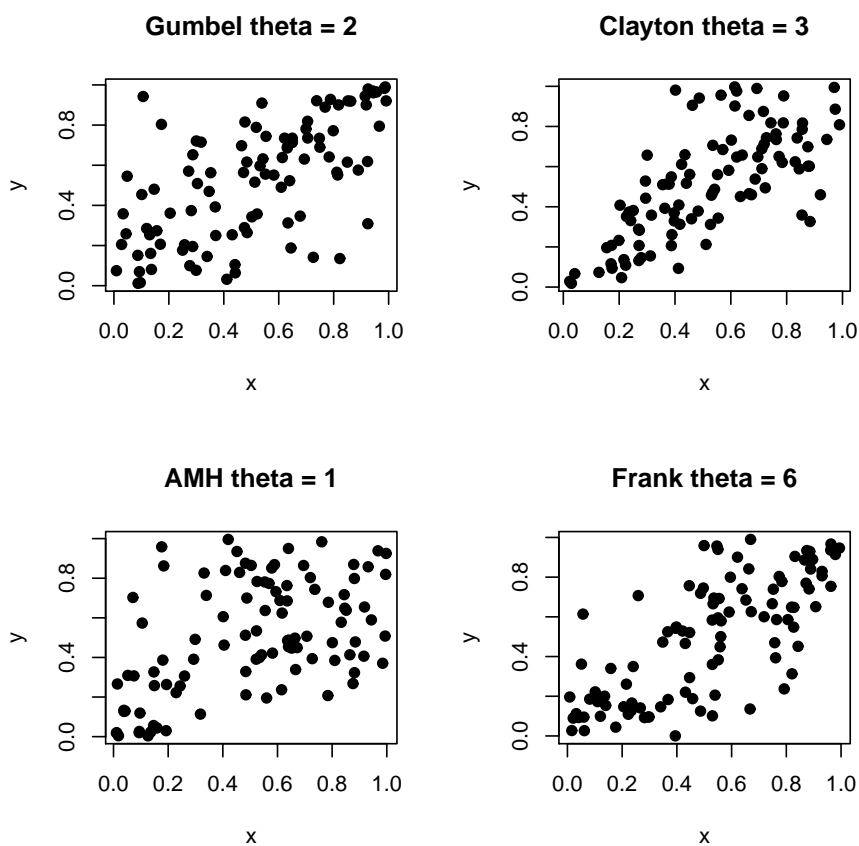
Detta kan även uttryckas med hjälp av en mer formell beskrivning. Betrakta den tvådimensionella copulafunktionen  $C(u_1, u_2)$  där:

$$\lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lambda_U.$$

Om  $\lambda_U$  existerar så har copulafunktionen ett övre svansberoende för  $\lambda_U \in (0, 1]$  men inget övre svansberoende för  $\lambda_U = 0$ .

På ett liknande sätt kan vi definiera det lägre svansberoendet.

$$\lim_{u \rightarrow 0} \frac{C(u, u)}{u} = \lambda_L$$



Figur 1: Scatterplot av slumpantal genererade från olika copulamodellerna för valt värde på  $\theta$

Om  $\lambda_L$  existerar så har copulafunktionen ett nedre svansberoende för  $\lambda_L \in (0,1]$  men inget nedre svansberoende för  $\lambda_L=0$ .

### 2.3.2 Gumbel copula

Copulafunktionen har i det valda parametervärdet ett svansberoende i den övre svansen, skulle vi öka värdet på  $\theta$  så skulle variablerna ha en mindre spridning i den nedre svansen samt ytterligare mindre spridning i den övre, även om detta inte är lika tydligt. Motsatt effekt gäller vid ett lägre värde av  $\theta$ , vad det gäller  $\theta=1$  kan vi förenkla dess copulafunktion till  $C(u_1, u_2, 1) = u_1 u_2$ , dvs. den copula funktion som gäller då variablerna är oberoende.

Något som bör observeras för denna copulamodell är att den kräver en positiv korrelation, skulle korrelationen vara negativ så kan man göra variabelutby-

tet för det tvådimensionella fallet  $u'_1=1 - u_1$ .

### 2.3.3 Clayton copula

Denna copula är mycket lik Gumbel copula med skillnaden att svansberoendet ligger i den nedre svansen. På samma sätt som för Gumbel gör ett högt parametervärde att observationerna lägger sig tätare efter diagonalen samt att ett lågt parametervärde ger en stor spridning samt ett oberoende mellan variablerna i sitt gränsvärde.

Precis som för Gumbel kräver modellen en positiv korrelation, varför en negativ sådan kräver ett variabelbyte som ges exempel på i beskrivningen av Gumbel.

### 2.3.4 Ali-Mikhail-Haq copula

För denna copulafunktion är det svårt att med hjälp av slumpsimuleringar se den påverkan som  $\theta$  ger. Det är visat av Kumar [4] att då  $\theta = 1$  finns ett nedre svansberoende samt att varken övre eller nedre svansberoenden finns då  $\theta \neq 1$ . Då  $\theta = 0$  föreligger oberoende mellan variablerna, vilket lätt kan förstås genom att betrakta dess copulafunktion. Hur funktionen i övrigt beter sig för olika värden för  $\theta$  krävs det att vi tittar närmare på konturplottar för olika värden på  $\theta$ .

I fig. 2 så kan vi se att konturerna för  $\theta=0,5$  samt  $\theta=-0,5$  är till synes spegelvända, det verkar alltså som att denna funktionen även fångar upp både positiva och negativa korrelationer, dock gäller inte detta för  $\theta= 1$  och  $\theta=-1$ . Detta gör att vi även för denna funktion bör överväga ett variabelutbyte enligt beskrivning ovan.

### 2.3.5 Frank copula

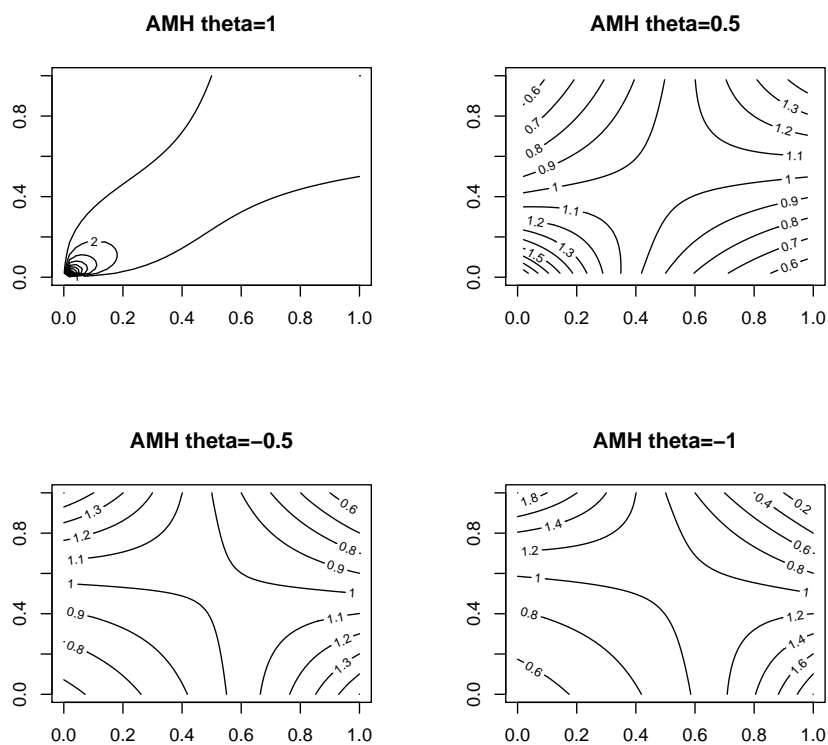
Desto närmare  $\theta$  är 0 både positiv som negativ så beter sig variablerna mer och mer som oberoende. Vad det gäller positiva  $\theta$  så ger detta en positiv korrelation samt negativa  $\theta$  en negativ korrelation, där korrelationen blir starkare desto längre avstånd  $\theta$  har från noll.

## 2.4 Slumpgenerering från tvådimensionella copulafunktioner

Givet en copulafunktion,  $C(u, v)$ , kan vi simulera fram nya observationer, vi kommer här att gå igenom en metod att göra detta på.

Generera slumpvariabeln  $u$ , där  $U \sim U(0, 1)$ .

Genom att ersätta det värdet som vi har för  $u$  i  $C(u, v)$  så har vi en funktion



Figur 2: Ali-Mikhail-Haq, konturplot för olika värden på  $\theta$

som enbart beror av  $v$ , vilken vi benämner  $F_u(v)$ .

Vi kan därefter simulera fram ett till värde från  $U(0, 1)$ , låt oss kalla detta värde  $z$ .

Vi löser därefter ut  $v$  genom att låta  $F_u(v) = z \Rightarrow v = F_u^{-1}(z)$ .

Därmed har vi fått fram en ny observation,  $(u, v)$ .

## 2.5 Parameterskattningar

Vi vill använda den parameter som ger bäst anpassning till data, metoden som vi kommer att använda för detta är maximum log-likelihood.

Denna skattning ges av:

$$\operatorname{argsup}_{\theta \in \Theta} \sum_{i=1}^n \ln(f_c(u_i, v_i, \theta))$$

där  $n$  är antalet observationer och

$$f_c(u, v, \theta) = \frac{\delta^2 C(u, v, \theta)}{\delta u \delta v}$$

### 3 Termer och förklaring av data

Data som vi bygger detta arbete på är samtliga skadeposter under perioden 1 Jan 2002 - 31 Maj 2012 för Trygg-Hansa's Villaförsäkring.

För att förstå betydelsen av en skadepost krävs lite grundläggande information om hur skadehanteringen för dessa skador ser ut.

När en skada rapporteras in tilldelas denna bland annat skadedatum, inrapporteringsdatum, estimerad kostnad för skadan (denna kostnad kommer fortsättningsvis benämnas taxationen) samt skadans skadetyper. Taxationsvärdet kan därefter förändras flera gånger, det behöver oftast göras närmare undersökningar av skadan innan man når den slutgiltiga skadekostnaden. Var gång som nytt estimat av taxationen ges ger det upphov till en ny skaderegleringspost med uppdaterade värden. Skulle det vara så att en skada i sig består av två eller flera skadetyper estimerar man hur mycket taxationen är för var skadetyper och betraktar dessa som separata poster. För varje utbetalning av skadeersättning så sker en uppdatering av skadans värden, vilket även detta genererar nya poster.

Vi kan nu betrakta ett fiktivt exempel för förtydligande runt hur registrering sker enligt följande förlopp:

- 1.) En brand uppstår i en försäkrad villa 1 jan 2012, skadan rapporteras in 20 jan 2012 och taxationen sätts till 2 000 000 kr.
- 2.) Den 23 jan 2012 rapporteras även att mindre delar av försäkringstagarens grannes egendom har tagit skada av branden och taxationen för detta sätts till 100 000 kr. Detta täcks av ansvarsförsäkringen som är tecknad på den villa där branden uppstod.
- 3.) Utbetalning av 1 200 000 kr för villan där branden uppstod sker den 13 mars 2012.
- 4.) Nytt estimat av taxationen för villan där branden uppstod sker den 7 april 2012, det nya estimatet är på 1 000 000 kr.
- 5.) Återbetalning sker för villan där branden uppstod med 200 000 kr den 16 april 2012
- 6.) Utbetalning av 75 000 kr för grannens egendom sker den 16 april 2012.

obs	skadedatum	obs.datum	taxation	betalning	skadetyper
1	01/01/12	20/01/12	2 000 000	0	Brand
2	01/01/12	23/01/12	100 000	0	Ansvar
3	01/01/12	13/03/12	2 000 000	1 200 000	Brand
4	01/01/12	07/04/12	1 000 000	1 200 000	Brand
5	01/01/12	16/04/12	1 000 000	1 000 000	Brand
6	01/01/12	16/04/12	100 000	75 000	Ansvar

### 3.1 Datatyper

De datatyper som vi kommer att titta närmare på är inkrementellt data samt data baserat på skadedatum. Dessa kommer att beskrivas sett till dagsnivå, för fallet månadsnivå är rader med samma år och månad hopsummerade.

#### 3.1.1 Inkrementellt data

Det inkrementella datamaterialet visar med hur stort belopp den totala taxationen/utbetalningen, oavsett skadedatum eller skada, ökar per dag separerat på olika skadetyper. Exemplet ovan skulle generera följande 2 tabeller beroende på om det är taxationerna eller utbetalningarna som man är intresserad av:

<i>Taxation</i>			<i>Utbetalning</i>		
datum	Brand	Ansvar	datum	Brand	Ansvar
20/01/12	2 000 000	0	13/03/12	1 200 000	0
23/01/12	0	100 000	16/04/12	-200 000	75 000
07/04/12	-1 000 000	0			

#### 3.1.2 Data baserat på skadedatum

För data av den här typen är enbart den senaste observationen per skademoment inkluderade och summerade på skadedatum. Exemplet ovan skulle generera följande data:

<i>Taxation</i>			<i>Utbetalning</i>		
datum	Brand	Ansvar	datum	Brand	Ansvar
01/01/12	1 000 000	100 000	01/01/12	1 000 000	75 000

Något som bör nämnas om denna datatyp är osäkerheten av de värden på skadedatum som ligger nära i tiden. Det kan dröja innan skador rapporteras in och regleringstiden för skadorna kan vara lång vilket orsakar denna osäkerhet. På Trygg-Hansa anser man att taxationen för skador från Villaförsäkringen kan betraktas som i stort sätt färdigutvecklade (98% av det



slutliga värdet) två år efter skadedatumet. Vi kommer i detta arbete betrakta dessa skador som helt färdigutvecklade. Betalningarna av denna datatyp kommer inte att användas då den efter tillräckligt lång tid kommer att vara identisk med taxationsdata.

### 3.2 Årsrisk

För en försäkring kan antalet försäkrade objekt förändras över tiden. Ett mått för att mäta denna förändring är årsrisken, årsrisken ger det genomsnittligt antal försäkrade under specifikt år. Om en försäkring tecknas 1 Juli 2010 - 31 Dec 2011 ger denna ett bidrag på ca 0,5 till årsrisken för 2010 och 1 till årsrisken för 2011. Orsaken till att det inte är exakt för 2010 beror på att det är 2-3 (beroende på skottår eller ej) fler dagar under andra halvåret jämfört med första halvåret.

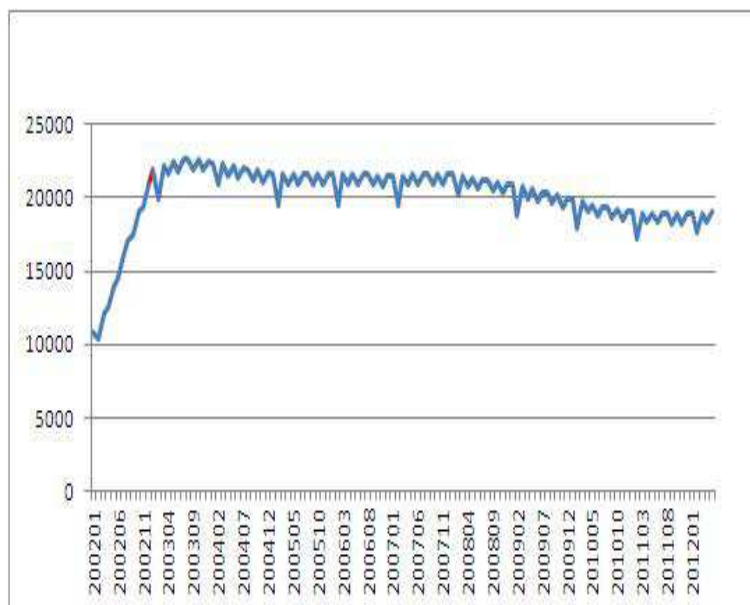
## 4 Analys av data

### 4.1 Förarbete med data

Då analysen bygger på historiskt data så behöver detta justeras med inflation för att få nuvärdet av observationerna. Vi antar en 2% årlig inflation och justerar till inflationsnivån för den senaste inträffade observationen för samtliga dataset. Detta medför att det inkrementella datasetet justeras till en inflationsnivå å 31 Maj 2012 och data baserat på skadedatum å 31 Maj 2010. Vi måste även välja mellan vilka tidsperioder vi anser data representativt. Något som påverkar data mycket utöver inflationen är förändringar av antalet försäkrade. Skulle vi under någon tidsperiod ha väldigt många försäkrade skulle detta troligtvis medföra att antalet skador inträffande under den perioden också ökade, vilket i sin tur påverkar summerad taxation samt betalningar. I fig. 3 illustreras var månads bidrag till respektive årsrisk.

Som vi kan se i fig. 3 så ökade bidraget till årsrisken kraftigt i början av perioden, efter Jan 2003 är exponeringen relativt stabil. Orsaken till denna kraftiga ökning beror på införandet av det skadesystem som data kommer ifrån, försäkringarna har successivt överförts till skadesystemet och med hjälp av denna plott kan vi rimligen anta att vi kan använda data från datumet 1 Jan 2003 sett till exponeringen. Exponeringsdata visar även systematiska hopp månad för månad, detta beror på att antal dagar skiljer sig åt mellan månaderna.

För att säkerställa att data som vi använder inte är alltför påverkat av trender i tiden utöver den antagna årliga inflationen på 2% så plottar vi



Figur 3: Exponering

variablerna för det inkrementella datasetet och data baserat på skadedatum, vilket återfinns i appendix (fig. 10-15) där extremobservationer är eliminerade för att ge en tydligare bild av datastrukturen.

För de inkrementella betalningarna kan vi se ett klart trendbrott för både Vatten och Brand i slutet av April 2003, därefter är det färre datum där kostnaden är 0 samt att de datum där kostnaden överstiger 0 generellt är lägre än för datum innan.

För de inkrementella taxationerna kan vi inte se ett lika tydligt trendbrott. Något som vi kan se är att datum fram till slutet av April har relativt få negativa kostnader om vi jämför med datum efter tidpunkten. Datasetet baserat på skadedatum verkar ha ett trendbrott som inträffar vid årsskiftet 2000/2001 för Vattenskador, det stora trendbrottet sett utifrån data för Brandskador inträffar innan dess. Då vi tidigare observerade att exponeringen var extremt låg för datum innan år 2003 betraktar vi tillsvidare enbart data från och med 2003 och framåt som representativt.

Ett antagande som krävs vid användandet av copulamodeller är att samtliga observationer är oberoende och lika fördelade, vi antar därför att detta gäller. Det finns dock en viss osäkerhet i antagandet och det är då av fördel att inte låta vissa delar av året vara överrepresenterade. Därmed så kortar vi ner perioderna för de olika datatyperna ytterligare till helårsperioder, detta medför dock att några veckodagar blir överrepresenterade men vi antar att

den påverkan detta ger är minimal.

Dessa perioder kommer vi att fortsättningsvis använda för följande datatyper där vi har antagit en inflation på 2%:

- Inkrementella betalningar, 1 Juni 2003 - 31 Maj 2012
- Inkrementella taxationer, 1 Juni 2003 - 31 Maj 2012
- Taxationer baserat på skadedatum, 1 Juni 2003 - 31 Maj 2010

## 4.2 Test av beroende

Vi har tre stycken datatyper som vi vill testa på två olika nivåer. Datatyperna är inkrementella betalningar, inkrementella taxationer och taxation baserat på skadedatum, och nivåerna är dagsbasis samt månadsbasis, totalt utgör de sex stycken dataset. Då dataseten innehåller vissa extremobservationer så kan vi eliminera dessa och skapar därmed sex nya dataset. Detta görs genom följande elimineringar, för grafer plottar över samtliga observationer hänvisas till appendix(fig. 16-21).

typ\gräns	LB BRA	UB BRA	LB VAT	UB VAT
I bet dag	-15 000 000	20 000 000	-	-
I tax dag	-20 000 000	20 000 000	-4 000 000	7 000 000
SD dag	-	10 000 000	-	5 000 000
I bet mån	-4 000 000	-	-	40 000 000
I tax mån	-	-	-	50 000 000
SD mån	-	30 000 000	-	-

I tabellen ovan anges de övre(UB) samt de nedre(LB) gränserna som har använts vid eliminering. Observationer där vatten eller brand ligger utanför intervallet för datasetet elimineras. I nedanstående två tabeller anges antalet observationer i vart dataset före och efter eliminering, antalet eliminerade samt hur stor andel av det ursprungliga antalet observationer som eliminerats.

### Dagsbasis

Datatyp \Antal	Med extr.obs.	Utan extr.obs.	Elim.	Andel elim.
Inkr. bet.	3288	3285	2	0,06%
Inkr. tax.	3288	3276	12	0,36%
Skaddat tax.	2557	2549	8	0,31%

### Månadsbasis

Datatyp \Antal	Med extr.obs.	Utan extr.obs.	Elim.	Andel elim.
Inkr. bet.	108	105	3	2,78%
Inkr. tax.	108	107	1	0,93%
Skaddat tax.	84	81	3	3,57%

Vi kan här se att det är få observationer som har eliminerats, vi kan alltså tro att de dataseten där observationer har eliminerats ger en bra bild av verkligheten.

### Dagsbasis

Samtliga observationer

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,48801	<,0001	0,38520	<,0001
Inkr. tax.	0,47882	<,0001	0,36993	<,0001
Skaddat Tax.	0,06687	0,0007	0,04555	0,0006

Extremobservationer eliminerade

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,48719	<,0001	0,38445	<,0001
Inkr. tax.	0,48491	<,0001	0,37429	<,0001
Skaddat Tax.	0,06370	0,0013	0,04340	0,0011

### Månadsbasis

Samtliga observationer

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,44721	<,0001	0,31914	<,0001
Inkr. tax.	-0,11152	0,2505	-0,07754	0,2343
Skaddat Tax.	-0,04001	0,7178	-0,01606	0,8287

Extremobservationer eliminerade

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,47295	<,0001	0,34139	<,0001
Inkr. tax.	-0,10778	0,2692	-0,07565	0,2482
Skaddat Tax.	-0,09837	0,3823	-0,05494	0,4679

Vi kan till att börja med notera att korrelationen inte påverkas så mycket vid

eliminering av extremobservationer, detta beror dels på att det är väldigt få observationer som har eliminerats samt att vi använder rangkorrelation. Utan korrelationstest på rangerna så skulle några extremobservationer kunna ge en väldigt stor påverkan på den uppmätta korrelationen. Data som är baserat på skadedatum har en väldigt låg korrelation oavsett eliminering, dags eller månadsnivå. Det finns av den orsaken redan i detta skede starka skäl att ej använda detta dataset för copulamodellering. Vi kan även observera att de inkrementella taxationer på dagsnivå har en någorlunda hög korrelation medan data som är uppsummerad på månadsnivå är negativt. En orsak till detta skulle kunna vara att det finns många nollobserverationer, det vill säga att både variablerna vatten och brand är 0 för något datum, vi bör alltså testa nollobserverationernas påverkan på korrelationen. Det visar sig att det finns ett större antal observationer som antar värdet 0 för någon variabel i dataseten på dagsnivå, vad det gäller dataseten på månadsnivå existerar inga sådana variabler i dataseten. Värt att notera är att helgdagar påverkar nollobserverationerna mycket, utan helgdagar så hade vi troligtvis haft betydligt färre nollobserverationer. Nedan tabuleras antalet nollobserverationer samt antalet observationer med 0 i någon variabel för de olika dataseten på dagsnivå, andelarna ges i förhållande till grunddatasetet.

Datatyp	# Nollobs	Andel nollobs	# Nollobs Var	Andel nollobs Var
Inkr. tax	948	28,8%	1041	31,7%
Inkr. bet	961	29,2%	1045	31,8%
Skaddat Tax.	0	0%	188	7,4%

### Dagsbasis utan nollobserverationer

Samtliga observationer

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,34144	<,0001	0,23562	<,0001
Inkr. tax.	0,13783	<,0001	0,09421	<,0001

Extremobservationer eliminerade

datatyp\korr	Spearman	$P$ -värde $H_0:\rho=0$ $H_1:\rho \neq 0$	Kendall	$P$ -värde $H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,33974	<,0001	0,23430	<,0001
Inkr. tax.	0,14395	<,0001	0,09806	<,0001

### Dagsbasis, observationer utan 0 i någon variabel

Samtliga observationer

datatyp\korr	Spearman	$P$ -värde	Kendall	$P$ -värde
		$H_0:\rho=0$ $H_1:\rho \neq 0$		$H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,31654	<,0001	0,21745	<,0001
Inkr. tax.	0,08736	<,0001	0,05954	<,0001
Skaddat Tax.	0,06223	0,0025	0,04241	0,0020

Extremobservationer eliminerade

datatyp\korr	Spearman	$P$ -värde	Kendall	$P$ -värde
		$H_0:\rho=0$ $H_1:\rho \neq 0$		$H_0:\tau=0$ $H_1:\tau \neq 0$
Inkr. bet.	0,31471	<,0001	0,21605	<,0001
Inkr. tax.	0,0,09309	<,0001	0,06320	<,0001
Skaddat Tax.	0,06223	0,0025	0,04241	0,0020

Vi kan här se att korrelationen för de inkrementella taxationerna har minskat markant, även de inkrementella betalningarna har minskat men inte alls i lika hög grad.

Det kan nu vara bra att kontrollera att vi inte har något tydligt beroende som inte tas upp med hjälp av de Rangkorrelationstester som vi har gjort. Ett fall som exempelvis skulle kunna ge korrelationer nära 0 är om vatten betar sig som en andragsgradsfunktion gentemot brand. Vi kontrollerar detta genom att plotta upp data där elimineringar har gjorts, detta för att få en så bra bild som möjligt över beroendestrukturen, se appendix (för inkrementella betalningar se fig. 22-25, inkrementella taxationer se fig. 26-29, kostnader per skadedatum se fig. 30-33).

Vi kommer nu datatyp att för datatyp betrakta resultatet samt ge slutsatser kring beroendet mellan vatten och brandskador.

#### Data baserat på skadedatum

Vi har i korrelationstesterna fått resultatet svagt positiv korrelation för dagsnivå och svagt negativ på månadsnivå. Plotten av Brand mot Vatten på dagnivå ger intrycket av en negativ korrelation, dock kan vi se då vi observerar plotten över variabelernas ordningstal att det verkar som att strukturen som vi ser i data enbart är en följd av variabelernas individuella fördelningar. Vi kan efter dessa kontroller välja att tro på en korrelation nära 0. Vad det gäller data på månadsnivå ser vi ingen tydlig struktur mellan vatten och brandskadorna i någon av plottarna varför vi även här väljer att tro på en korrelation nära 0. Vi kan alltså nu ta beslutet att det inte är vettigt att försöka anpassa copulamodeller till data baserat på skadedatum oavsett dagsnivå eller månadsnivå.

### **Inkrementella taxationer**

Vi såg i korrelationstesterna att denna datatyp gav indikationer på en relativt hög korrelation på dagsnivå men att den till stor del berodde på nollobserverationerna. Då vi plottar Brandskador mot Vattensskador så ser vi en struktur i denna, då vi plottar deras ordningstal istället så kan vi även för detta dataset inse att det är variabelernas enskilda fördelningar som skapar den strukturen. Plotten av ordningstalen förstärker även bilden av att nollobserverationerna har stor betydelse för rangkorrelationen. Vi kan i den plotten se att observationer som för någon variabel är 0 ger en stor påverkan samt att en eliminering av dessa ger en korrelation nära noll. Då vi inte kan se någon form av beroendestruktur i data förutom att många variabler antar värdet 0 så verkar det inte rimligt att fortsätta att använda detta dataset på dagsnivå. På månadsnivå kan vi inte se någon tydlig struktur varken på scatterplotterna av variabelerna eller av variabelernas ranger, detta går i god linje med att de uppmätta korrelationsvärdena ligger nära 0. Vi anser därför att detta dataset inte heller går att använda på månadsnivå för framtida copulamodellering.

### **Inkrementella betalningar**

Det kan verka som att datat på dagsnivå har en visst beroende då vi ser på plotten av Brand mot Vatten. När vi ser på plotten av rangerna kan vi se att observationerna lägger sig något mer intensivare efter ena diagonalen, men är i övrigt väldigt spridda. Vi har dock observerat en uppmätt korrelation som är signifikant även då vi exkluderade nollobserverationerna. Vi ser i tabellen över korrelationer för alla observationer med 0 i någon variabel att korrelationen har minskat något jämfört med då enbart nollobserverationerna var eliminerade. Vad det gäller data på månadsnivå ser vi en ganska tydlig struktur i plotten av brand mot vatten som inte verkar bero på variabelernas individuella fördelningar, denna uppfattning förstärks när vi ser på plotten över variabelernas ranger, vi ser där att observationerna formar en diagonal som tyder på positiv korrelation. Denna uppfattning stärks även med resultatet av korrelationstesterna. Data på dagsnivå har vissa nackdelar vid jämförelse med detta data på månadsnivå, det har lägre uppmätt korrelation för båda testerna samt en svårhanterlig fördelning (detta då många observationer är noll för någon eller båda variabelerna). Med denna vetskap kommer vi inte att fortsätta använda data av denna datatyp på dagsnivå men däremot på månadsnivå.

### **Sammanfattning**

Vi har i detta stycke tagit beslutet att enbart arbeta med de inkrementella betalningarna på månadsnivå där extremobservationer har eliminerats och justeringar har gjorts enligt 4.1, vi kommer därför fortsättningsvis enbart att betrakta detta dataset.

### 4.3 Omvandlig till likformig fördelning

Vi vill nu undersöka variablernas individuella fördelning, detta kan lämpligen göras genom ett histogram samt normalplott för var variabel, se appendix (fig. 34-37). Vi kan konstatera att vårt antagande om normalfördelning för båda variablerna även stöds sett till dessa plottar. Vi kan nu skapa likformiga fördelningar givet antagandet om normalfördelning enligt följande:

$$U_{i,j} = \Phi^{-1} \left( \frac{X_{i,j} - \mu_j}{\sigma_j} \right)$$

där  $i$  betecknar observationsnumret,  $j$  skadeorsaken,  $n$  antal observationer samt  $\mu_j$  respektive  $\sigma_j$  skattas enligt:

$$\tilde{\mu}_j = \sum_{k=1}^n \frac{X_{k,j}}{n} \quad \tilde{\sigma}_j = \sqrt{\sum_{k=1}^n \frac{(X_{k,j} - \mu_j)^2}{n-1}}$$

Vi erhåller följande resultat för skattningarna:

Skadekod	$\tilde{\mu}$	$\tilde{\sigma}$
Brand	8 447 726	4 381 595
Vatten	11 928 529	6 843 598

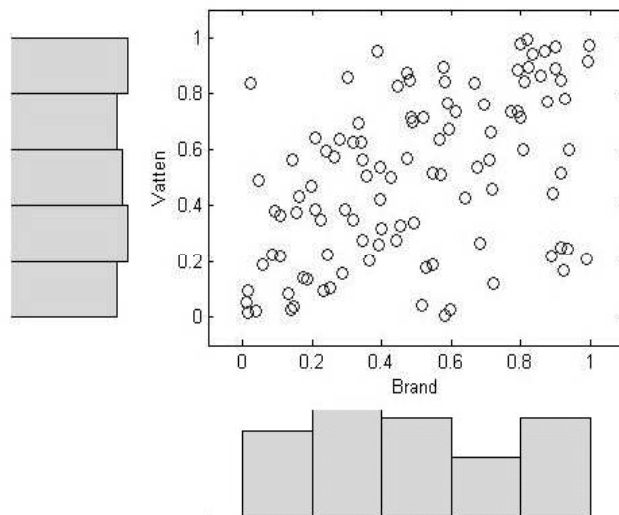
Detta resultat kan vi i ett senare läge använda för att omvandla resultat från en likformig fördelning till den fördelningen som vi har antagit att de individuella variablerna har i originaldatasetet. En viss extrakontroll av fördelningsantaganden kan göras, i fig.4 visualiseras det nya datasetet i ett histogram.

Vi kan komplettera fig. 4 med följande information om antalet observationer i vart intervall:

Skadetyper\Intervall nr.	1	2	3	4	5
Vatten	20	22	21	20	22
Brand	20	25	23	14	23

Som vi kan se verkar observationerna för vatten fallit väldigt jämnt i histogrammets olika intervall. Vad det gäller brand så kan vi se att vi har en ganska stor avvikelse från medel (21) i det fjärde intervallet. Sannolikheten att detta skulle inträffa i minst ett av intervallen givet att fördelningsantagandet för brand är korrekt kan ges med hjälp av simulering, vi finner att





Figur 4: Histogram för Vatten samt Brand med ranger

sannolikheten för detta är drygt 24,9%. Detta tyder på att även fördelningsantagandet vi har gjort för brand baserat på utfallet kan betraktas som fullt rimligt. Då det ändå finns en viss osäkerhet i antagandet runt variablernas enskilda fördelningar så skapar vi ytterligare ett dataset med hjälp av variablernas enskilda pseudofördelningar. Detta dataset skapar vi för att ge en känsla för hur stor påverkan fördelningsantagandet ger den slutligen valda copulamodellen. Variablerna i pseudodatasetet definierar vi på följande sätt:

$$U_{i,j} = \frac{\text{Rank}_j(X_{i,j})}{n + 1}$$

där  $i$  betecknar observationsnumret,  $j$  skadeorsaken samt  $n$  antalet observationer.

Observera att detta dataset inte har någon slump i variablernas enskilda fördelningar, detta då de enskilda fördelningarna enbart beror på antalet observationer, denna slump har nu istället tillförts beroendet mellan variablerna. Vi kan alltså inte förvänta oss att få exakt detta beroende även om variablernas individuella fördelningar hade varit kända.

#### 4.4 Anpassning av copulamodeller

Anpassning av copulafunktioner har gjorts med hjälp av inbyggda funktioner i copulapaketet för  $R$ . För mer information om dessa funktioner i  $R$  hänvisas till den dokumentation som finns tillgänglig [9]. Utöver test av faktiska obser-

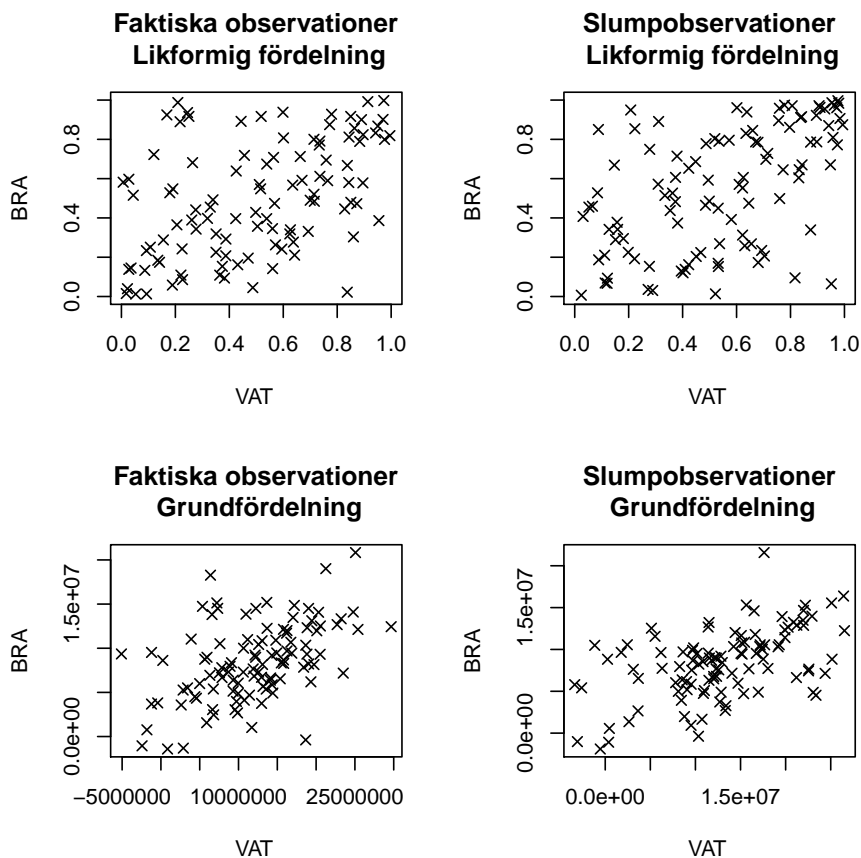
vationer och rangobservationer behöver även test av dessa observationer med omvänd svans göras, dvs. testa anpassningen på  $(1 - X_{BRA}), (1 - X_{VAT})$ . Detta behöver göras då vissa av modellerna ej är symmetriska, t.ex. kan modellen ha ett nedre svansberoende men ej ett övre. Vid anpassning av copulamodeller till data ges följande resultat i  $R$  där LL syftar på log-likelihood:

Copula	$\theta$	LL	$\theta$ rang	LL rang
Ej korr svans				
Gumbel	1,440223	12,83717	1,485027	13,60087
Clayton	0,745791	12,16181	0,8152281	12,72413
AMH	-	-	-	-
Frank	3,455084	14,07495	3.483184	13,96423
Omvänd svans				
Gumbel	1,470404	14,73003	1,501723	14,78974
Clayton	0,6440759	10,37874	0.740966	11.02126
AMH	-	-	-	-
Frank	3,455084	14,07495	3.483184	13,96423

Vi kan se här att vi för AMH ej får något resultat för varken datasetet eller rangerna av datasetet. Detta beror på att det finns ett inbyggt kontrollsystem i den valda funktionen där dataset med värden på Kendalls tau utanför intervallet  $[(5 - 8\log 2)/3, 1/3] = [-0.1817, 0.3333]$  ej accepteras, detta då AMH ej är teoretiskt definerad utanför detta intervall. Vi blir därför tvugna att anta att AMH inte ger den bästa anpassningen av de ovan testade copulafunktionerna. Vi ser bland de övriga tre copulafunktionerna att Gumbel ger det högsta värdet på log-likelihood både vad det gäller rangobservationerna och de fördelningsantagna observationerna. Då samtliga modeller har samma antal okända parametrar så är värdena på log-likelihood jämförbara. Vi kan även se att parametervärdena för rangobservationerna är ganska lika de parametrar som ges för det fördelningsantagna datasetet, vilket är bra då en stor skillnad tyder på ett felaktigt fördelningsantagande. Då Gumbel har ett övre svansberoende får vi i fallet för omvänd svans ett nedre svansberoende. Vi kan i fig. 5 se att de faktiska observationerna i det likformiga fallet ligger mer intensivt nära  $(0,0)$  jämfört med  $(1,1)$ , vilket är en förklaring till varför en modell med ett lägre svansberoende fick det högsta värdet på log-likelihooden. Det är dock svårt att uttala sig om ett lägre svansberoende faktiskt finns i data. Som kontroll av att modellen anpassar sig bra till data så kan vi skapa slumpstal från modellen och jämföra med de faktiska utfallet, detta görs i fig. 5. För plottar på rang mot slumpobservationer samt faktiska utfall mot dess rang hänvisas till appendix(fig. 38-39). Plottar i den nedre raden av fig. 5 illustrerar de transformerade observationerna till den antagda grundfördelningen, transformeringen görs genom

$$\widehat{X}_{i,j} = \Phi(U_{i,j}) \cdot \sigma_j + \mu_j$$

där  $j$  är skadeorsaken och  $i$  är observationsnumret,  $\sigma$  samt  $\mu$  definierade enligt skattningar i 4.3



Figur 5: Scatterplot av slumpantal genererade från anpassad Gumbel copula samt dataobservationer både för det likfomigafallet samt för den antagda grundfördelningen

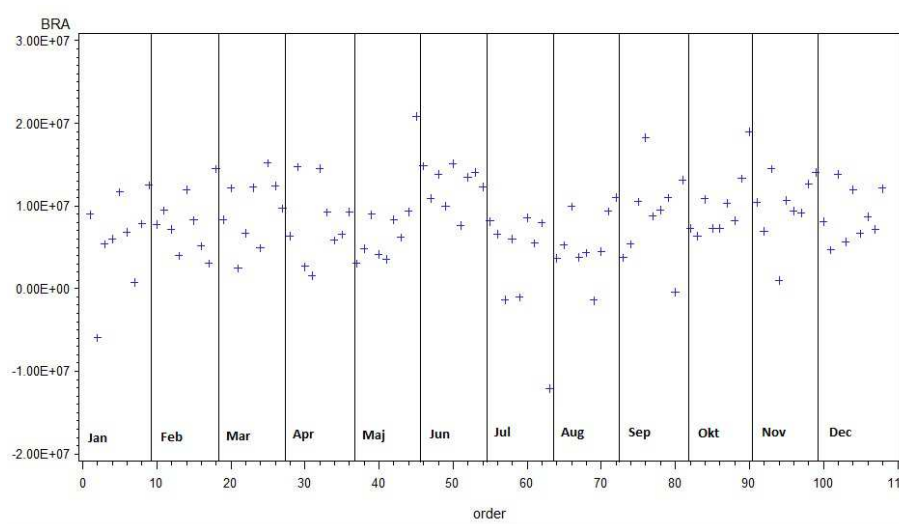
Vi kan utifrån fig. 5 se att de slumpgenererade observationerna från anpassad Gumbel copula inte verkar avvika sett till fördelning från de faktiska observationerna.

Vi kan alltså med stöd av detta test som vi precis utfört påstå att den anpassade copulamodellen följer datasetets fördelning bra.

## 5 Diskussion

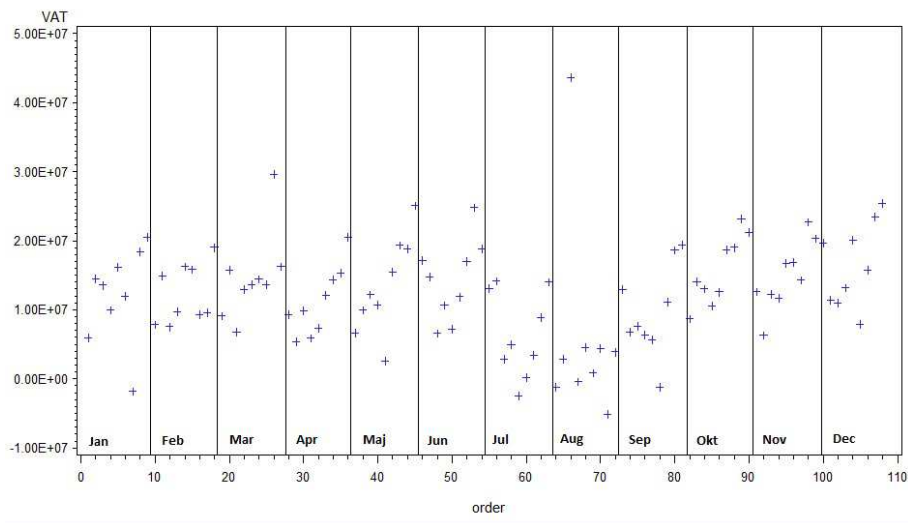
Det finns några antaganden som vi har gjort i det analytiska arbetet som kan ifrågasättas. Ett tidigt antagande som vi gjorde var att trenden i tiden kunde förklaras med en årlig inflation på 2%. Inga riktigt tydliga avvikelser från denna trend finns för betalningar eller data baserat på skadedatum. Vad det gäller de inkrementella taxationerna så kan vi för vatten se en tydlig trend som ej har eliminerats, se appendix fig. 12 - 13. Det skulle alltså behövas ett mer omfattande analytiskt arbete för att eliminera denna trend innan vi helt kan utesluta att detta data lämpar sig för copulamodellering.

Vi har även antagit att observationerna är oberoende och likafördelade (detta efter eliminering av inflationen på 2%). Detta antagande verkar dock inte rimligt då det är det är en stor skillnad mellan månader, se fig. 6 och 7.



Figur 6: Månadsbetalningar enligt inkrementellt data för brand, sorterat efter månad och år

Vi kan i fig. 6 och 7 se att det i juli och augusti generellt är lägre betalningar både för vatten och brand, vi har alltså orsak att ifrågasätta om den uppmätta korrelationen till stor del kan förklaras av skillnader mellan månaderna. För att simulera fram resultatet för en specifik månad skulle alltså en månadsjustering behöva göras och vi kan ställa oss tveksamma till om copula modellering verkligen är det bästa sättet att beskriva data på. Ytterligare en bra orsak att ifrågasätta beroendet för betalningar är fenomenet att taxationer. Något som även kan verka märkligt är att de uppmätta korrelationerna för taxationerna avviker en hel del från betalningarna på månadsnivå. Vi kan därför göra en likadan figur som i fig. 6 och 7 för taxationerna och jämföra



Figur 7: Månadsbetalningar enligt inkrementellt data för vatten, sorterat efter månad och år

resultaten, plottarna för taxationerna återfinns i appendix, fig. 40 och 41. Vi kan se i figurerna att de inte finns lika utmärkande perioder då taxationerna är låga som det fanns i betalningsdatamaterialet. Detta stärker även tron om att betalningarna bortsett från månadsskillnad är så starkt korrelerade som uppmätt.

Slutligen kan det vara intressant med månadsskillnaderna som bakgrund att undersöka om de elimineringar som vi gör är jämnt fördelade över månaderna, vilket vi gör i nedanstående tabell.

typ\ant	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
<b>Dag</b>												
I bet	0	0	0	0	0	2	0	0	0	0	0	0
I tax	0	1	0	0	0	0	3	6	1	1	0	0
SD	1	1	1	0	0	0	2	1	0	0	1	1
<b>Mån</b>												
I bet	1	0	0	0	0	0	0	1	1	0	0	0
I tax	0	0	0	0	0	0	0	1	0	0	0	0
SD	2	0	0	0	0	0	1	0	0	0	0	0

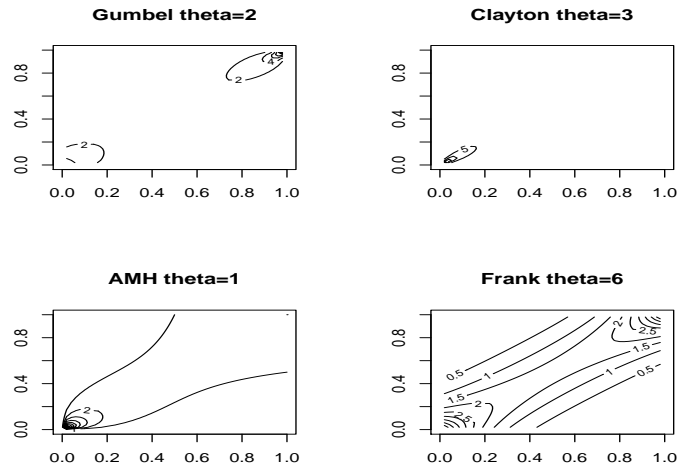
Då vi har åtta år av observationer är spridningen av elimineringen rimlig, dock avviker de inkrementella taxationerna på dagsnivå för augusti månad från detta. Det visar sig att 5 av dessa 6 observationer kommer ifrån år 2005 och samtliga av dessa 5 observationer är det vattenskador som är avsevärt över den övre elimineringssgränsen för vattenskador. Dessa skadedatum är dessutom på varandra följande vardagar, detta skulle alltså kunna vara ett

och samma event och eliminering är troligtvis därför rimlig. Då elimineringen verkar rimlig så tyder detta på att med ett större datamaterial skulle det kanske vara att föredra att undersökningen görs på årsnivå istället.

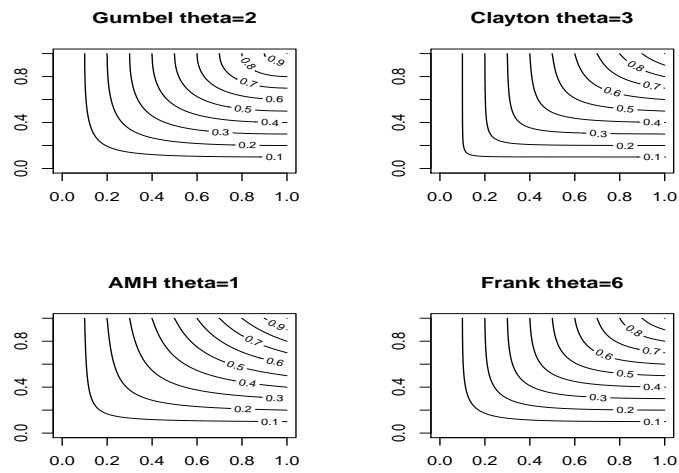
## Referenser

- [1] Roger B. Nelsen (2006), *An introduction to copulas*  
Springer
  
- [2] Stephan Haug, Claudia Klüppelberg och Liang Peng (2011), *Statistical models and methods for dependence in insurance data*  
Elsevier
  
- [3] Jun Yan (2007), *Enjoy the Joy of Copulas: With a Package copula*  
Journal of Statistical Software  
<http://www.jstatsoft.org/v21/i04/paper>
  
- [4] Pranesch Kumar (2010), *Probability Distribution and Estimation of Ali-Mikhail-Haq Copula*  
Applied Mathematical Sciences, Vol 4, 2010, no 14  
<http://www.m-hikari.com/ams/ams-2010/ams-13-16-2010/kumarAMS13-16-2010.pdf>
  
- [5] Paul C. Kettler (2008), *Fréchet - Hoeffding lower limit copulas in higher dimensions*  
<http://www.paulcarlislekettler.net/docs/FHcop.pdf?2bf7cb>
  
- [6] Henrik Hult och Filip Lindskog (2007), *Mathematical Modeling and Statistical Methods for Risk Management, Lecture Notes*  
KTH
  
- [7] Paul Embrechts, Filip Lindskog och Alexander McNeil (2001), *Modelling Dependence with Copulas and Applications to Risk Management*  
Department of Mathematics, ETHZ, Zürich
  
- [8] Paul Embrechts (2009), *Copulas: A personal view*  
Department of Mathematics, ETHZ, Zürich
  
- [9] Marius Hofert, Ivan Kojadinovic, Martin Maechler och Jun Yan (2012), *Cran R-project package Copula*  
<http://cran.r-project.org/web/packages/copula/copula.pdf>

## 6 Appendix

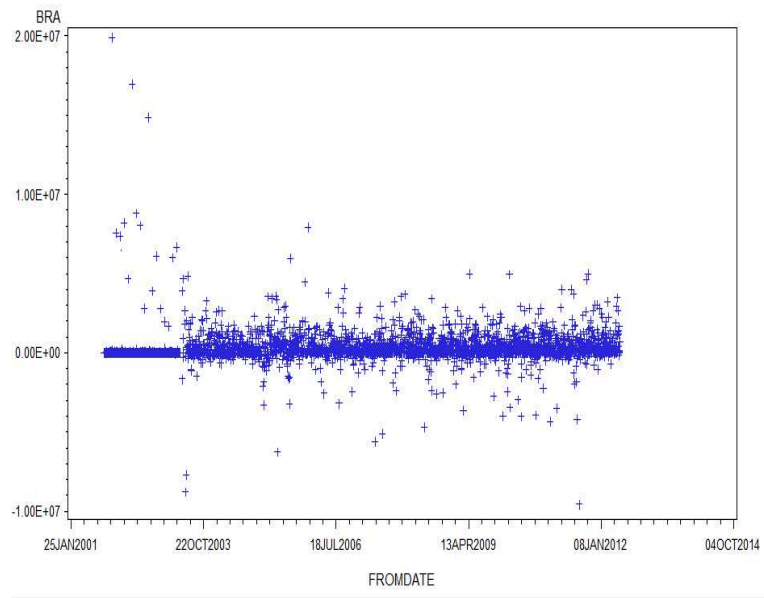


Figur 8: Konturplott över tätheten för respektive copula

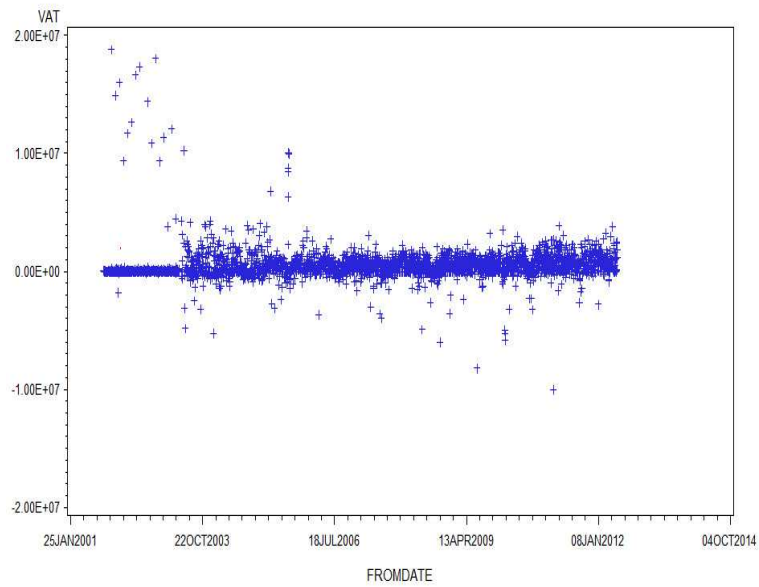


Figur 9: Konturplott över fördelningen för respektive copula

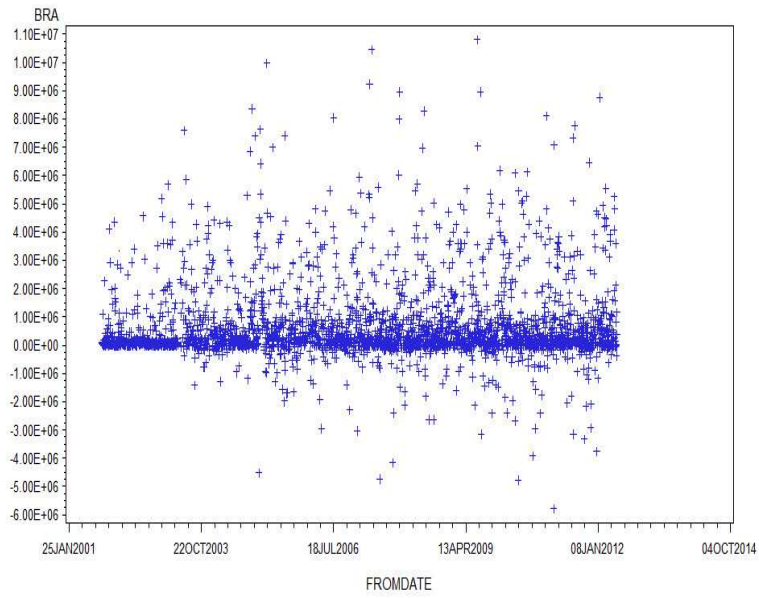




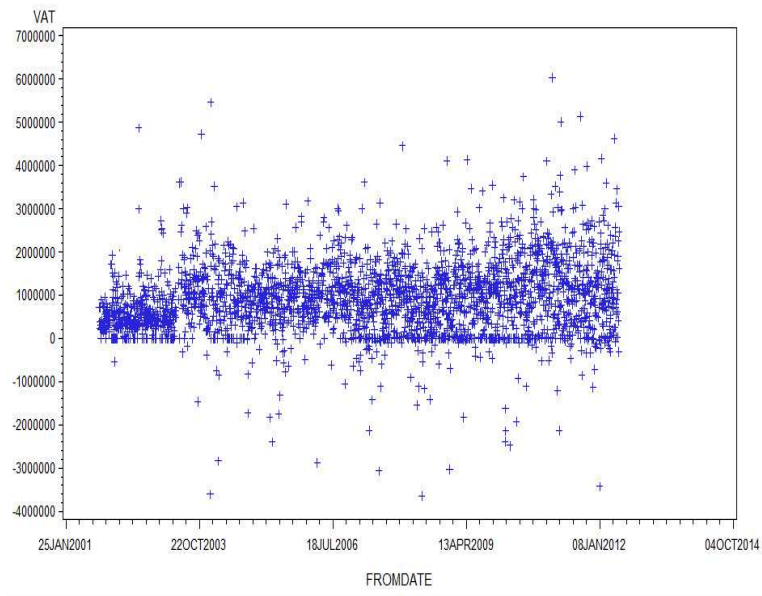
Figur 10: Inkrementella betalningar för brand med elimineringar mot datum



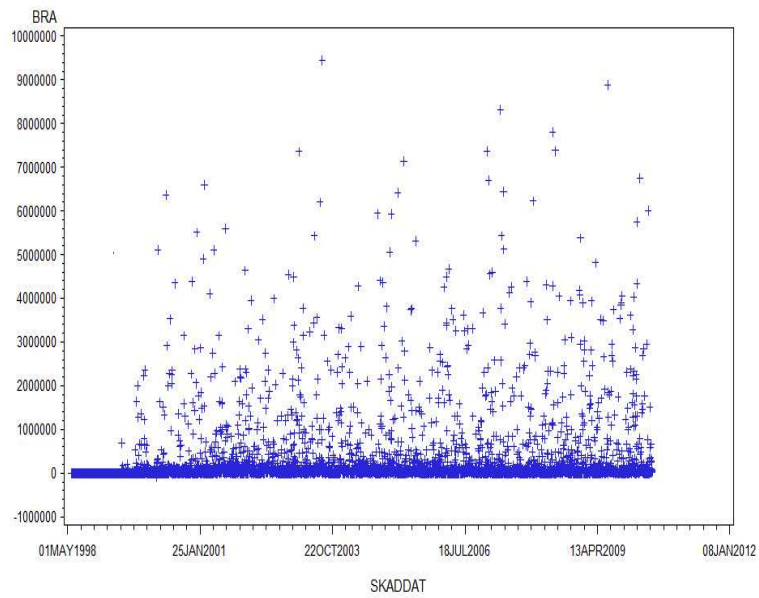
Figur 11: Inkrementella betalningar för vatten med elimineringar mot datum



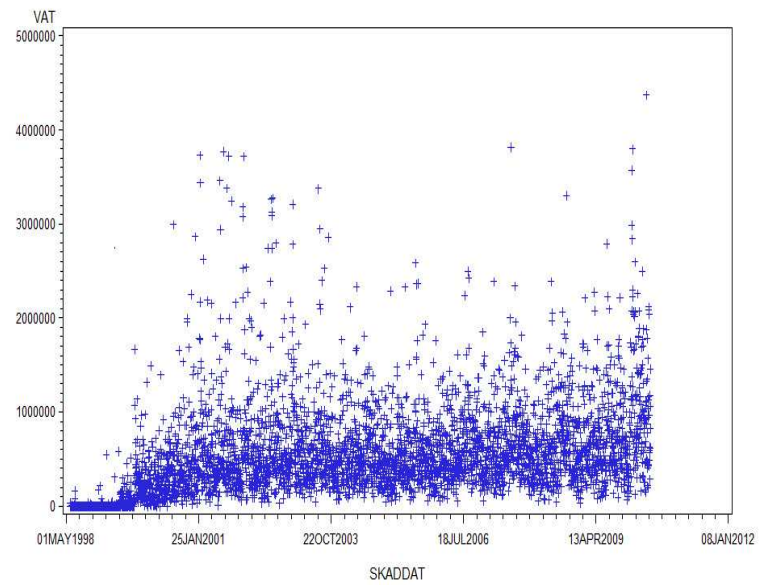
Figur 12: Inkrementella taxationer för brand med elimineringar mot datum



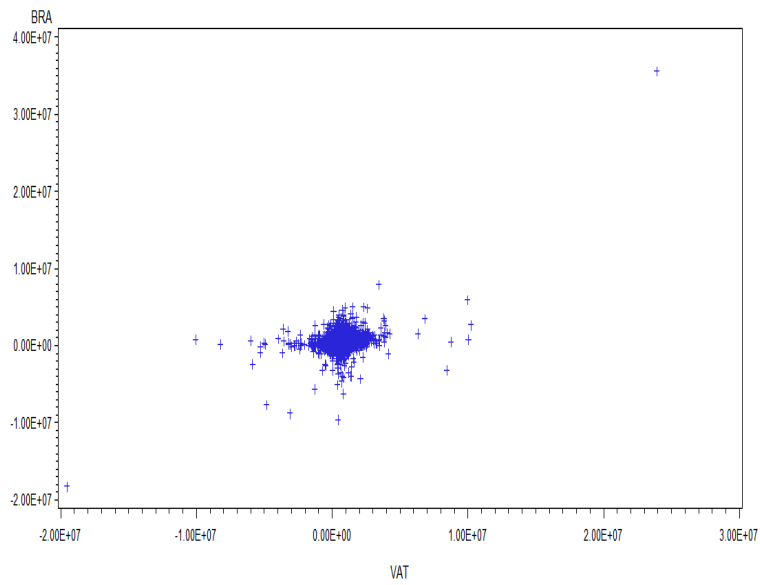
Figur 13: Inkrementella taxationer för vatten med elimineringar mot datum



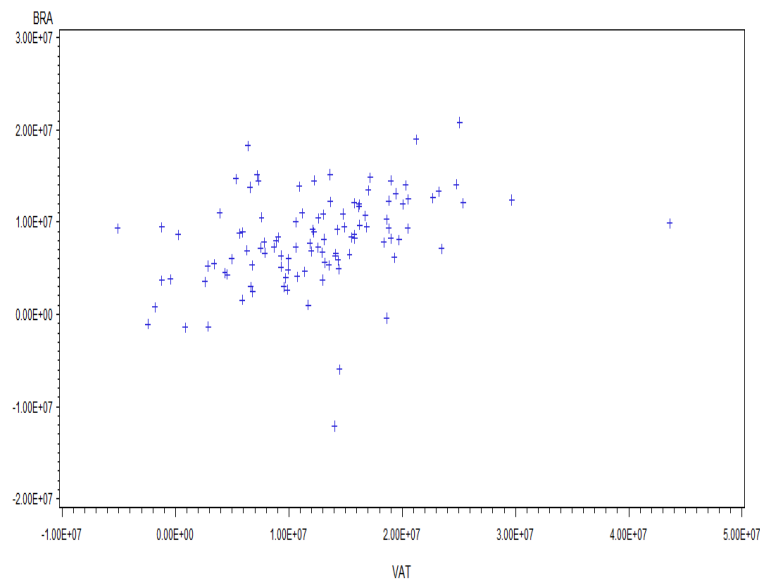
Figur 14: Data baserat på skadedatum, brand med elimineringar mot datum där historiken har förkortats



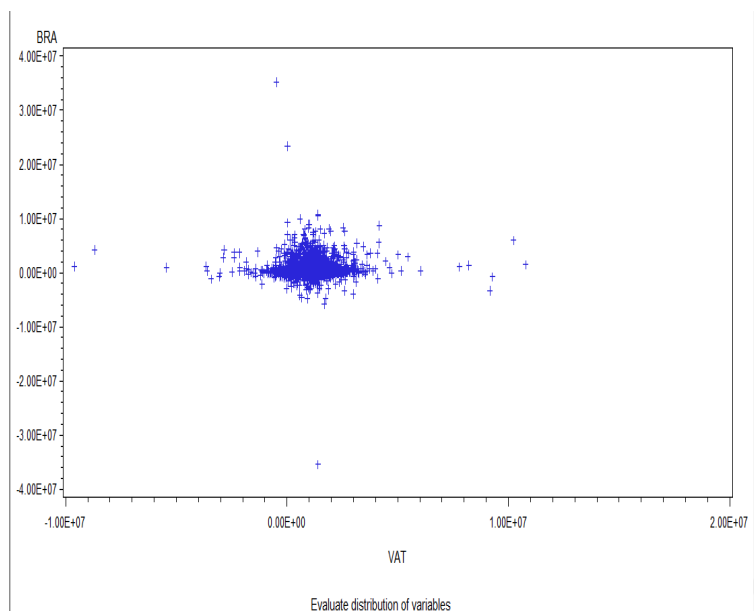
Figur 15: Data baserat på skadedatum, vatten med elimineringar mot datum där historiken har förkortats



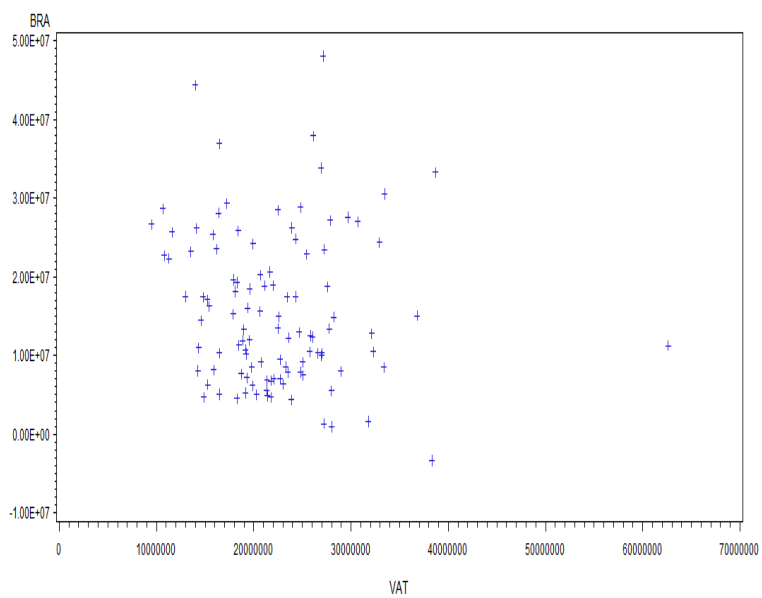
Figur 16: Inkrementella betalningar dagssnivå, samtliga observationer



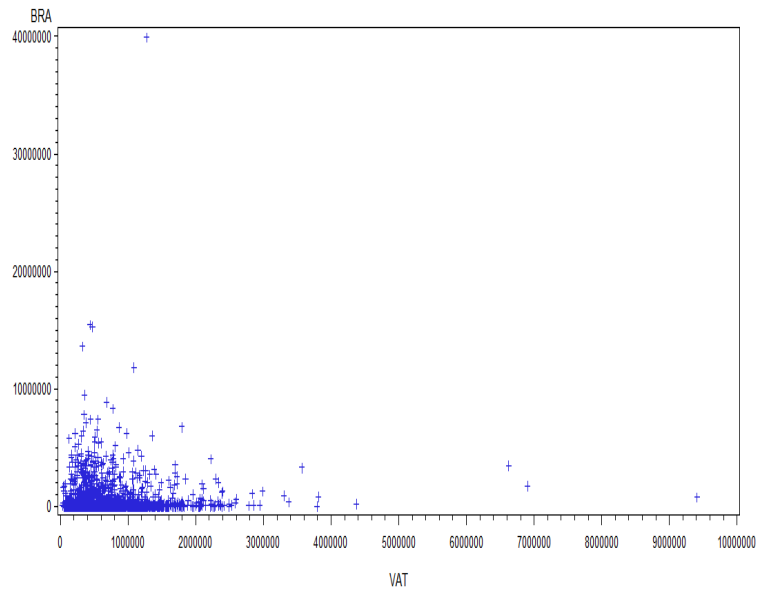
Figur 17: Inkrementella betalningar månadsnivå, samtliga observationer



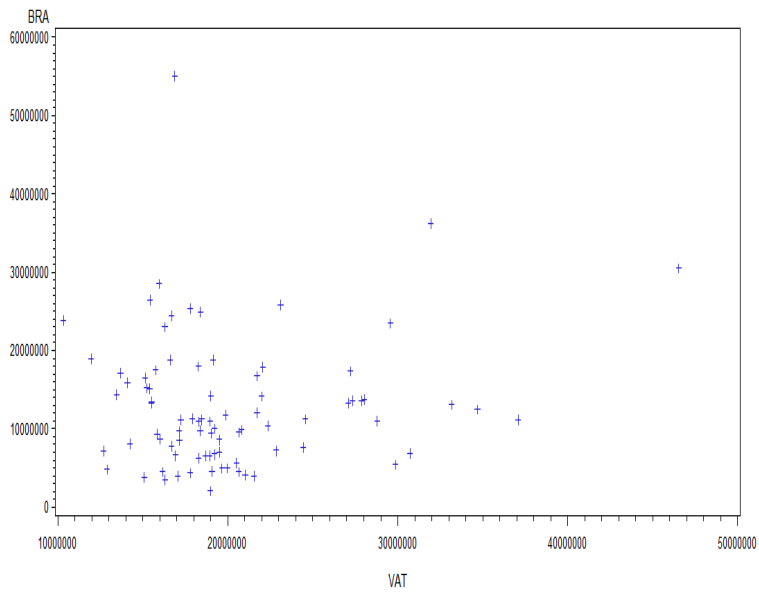
Figur 18: Inkrementella taxationer dagsnivå, samtliga observationer



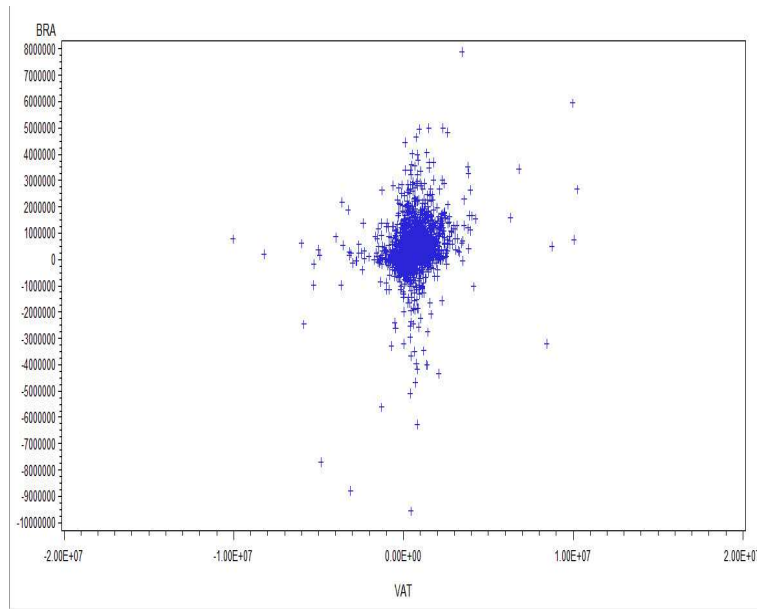
Figur 19: Inkrementella taxationer månadsnivå, samtliga observationer



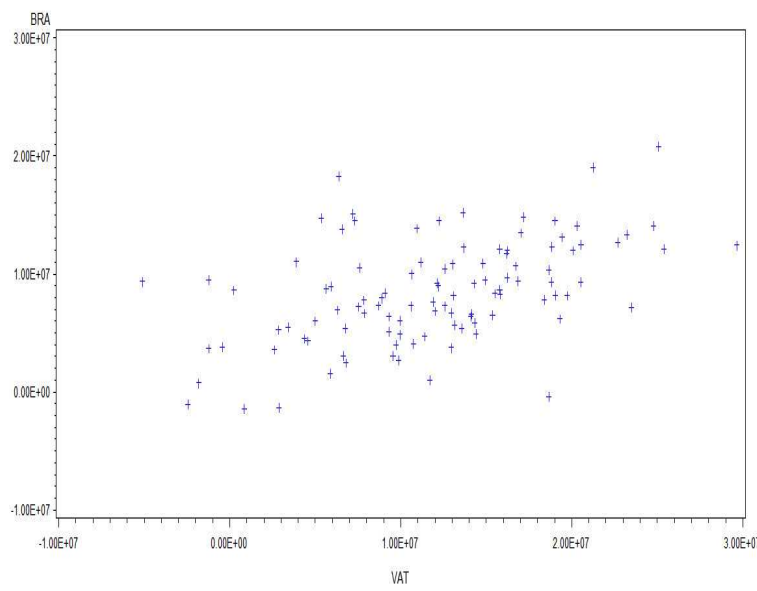
Figur 20: Kostnad per skadedatum dagsnivå, samtliga observationer



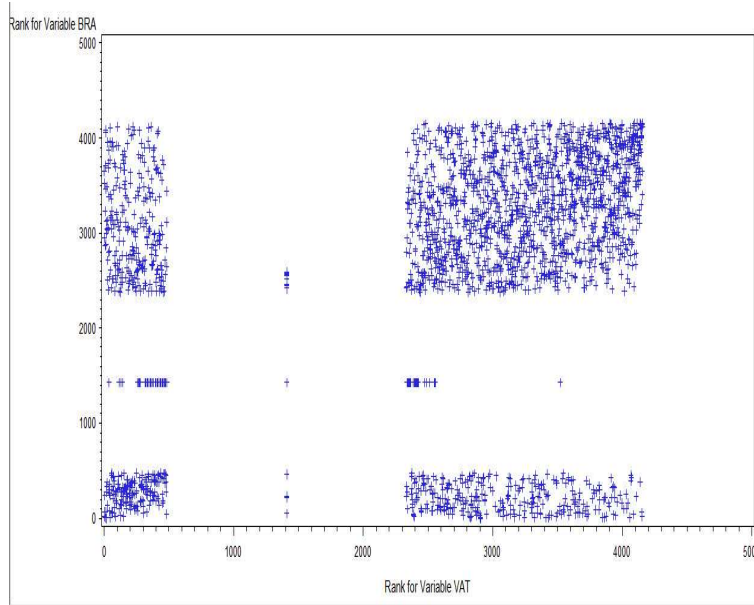
Figur 21: Kostnad per skadedatum månadsnivå, samtliga observationer



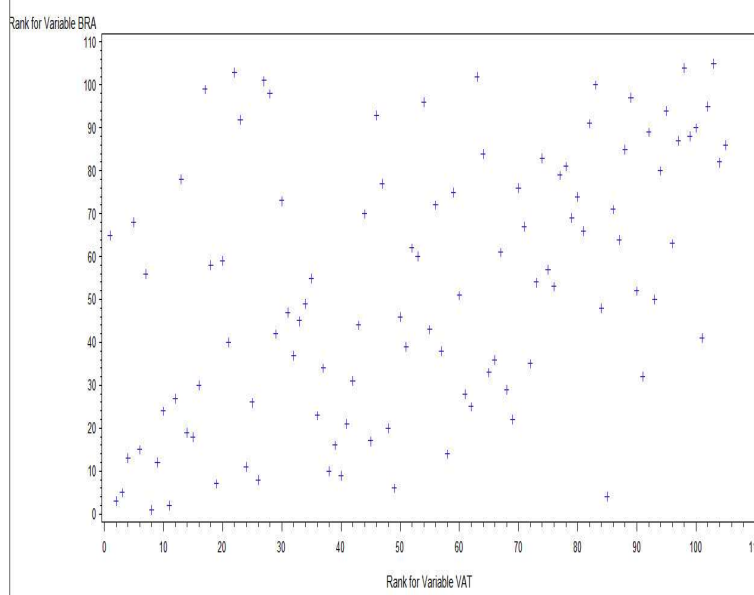
Figur 22: Inkrementella betalningar, dagsnivå, vatten mot brand



Figur 23: Inkrementella betalningar, månadsnivå, vatten mot brand

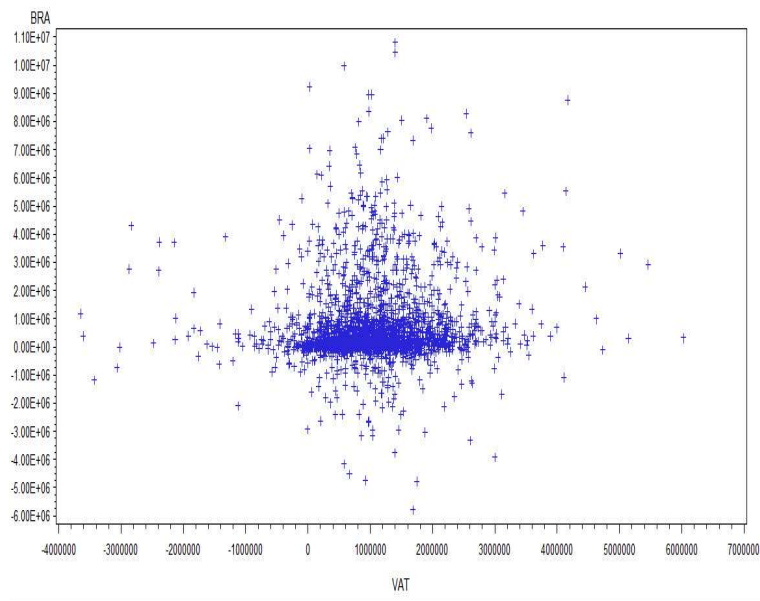


Figur 24: Inkrementella betalningar ranger, dagsnivå, vatten mot brand

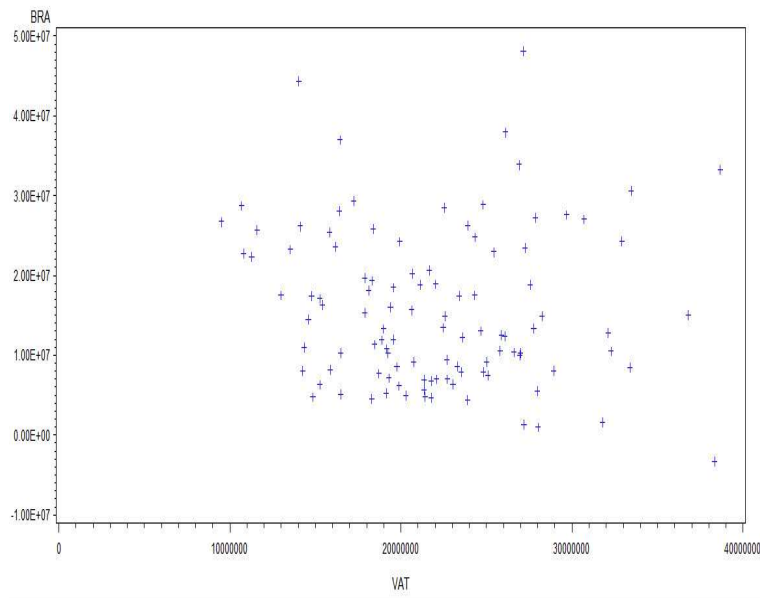


Figur 25: Inkrementella betalningar ranger, månadsnivå, vatten mot brand

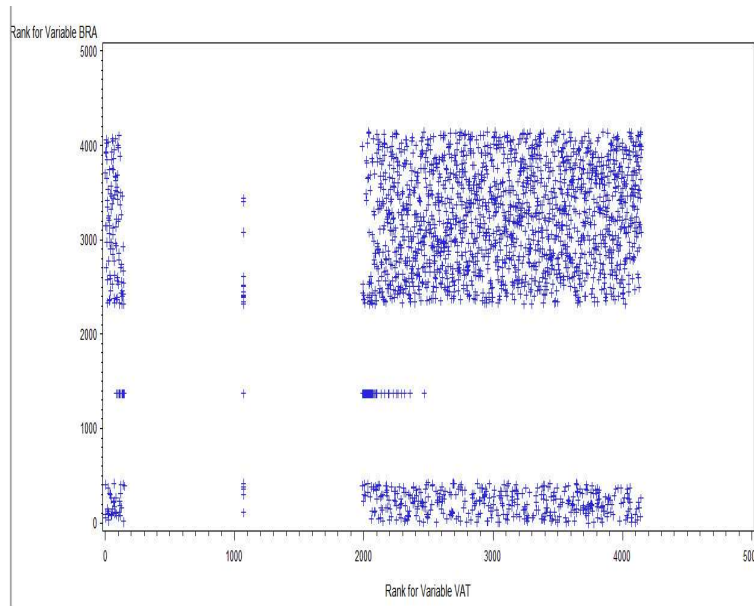




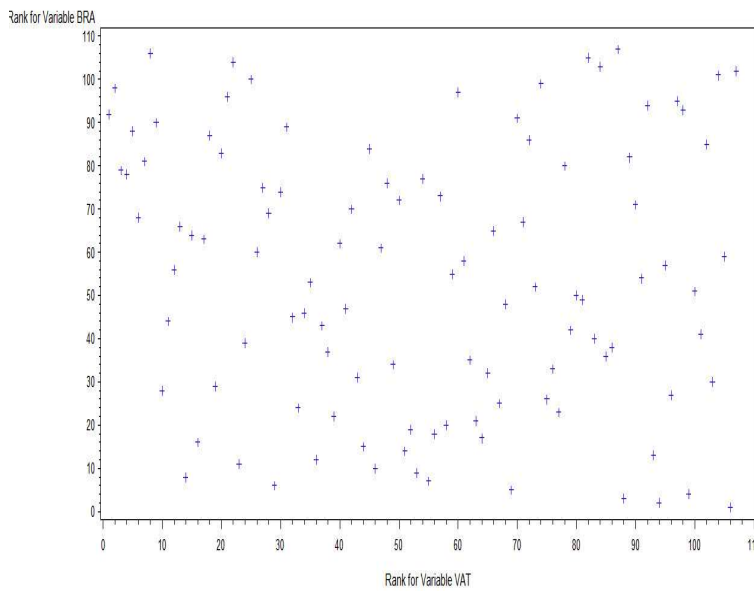
Figur 26: Inkrementella taxationer, dagsnivå, vatten mot brand



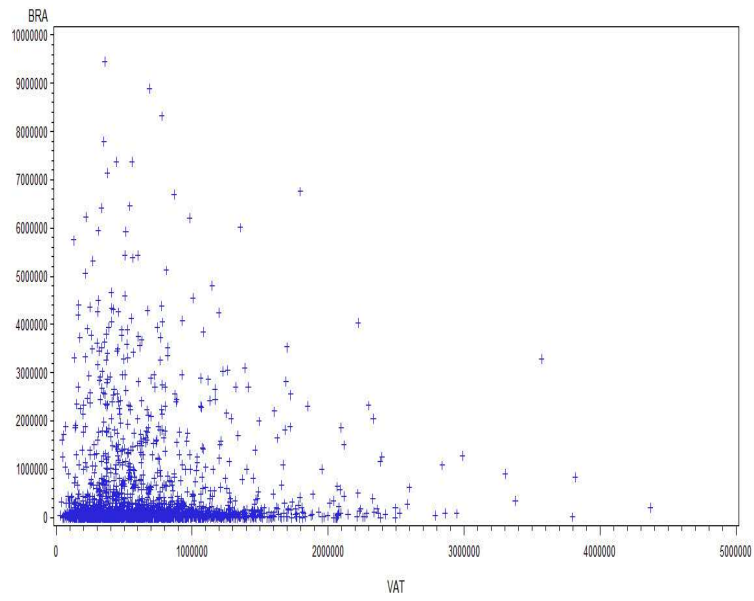
Figur 27: Inkrementella taxationer, månadsnivå, vatten mot brand



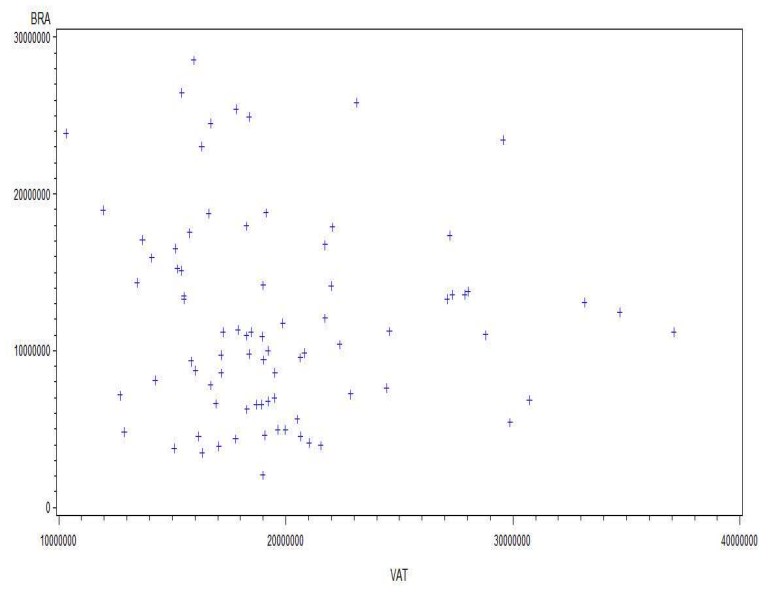
Figur 28: Inkrementella taxationer ranger, dagsnivå, vatten mot brand



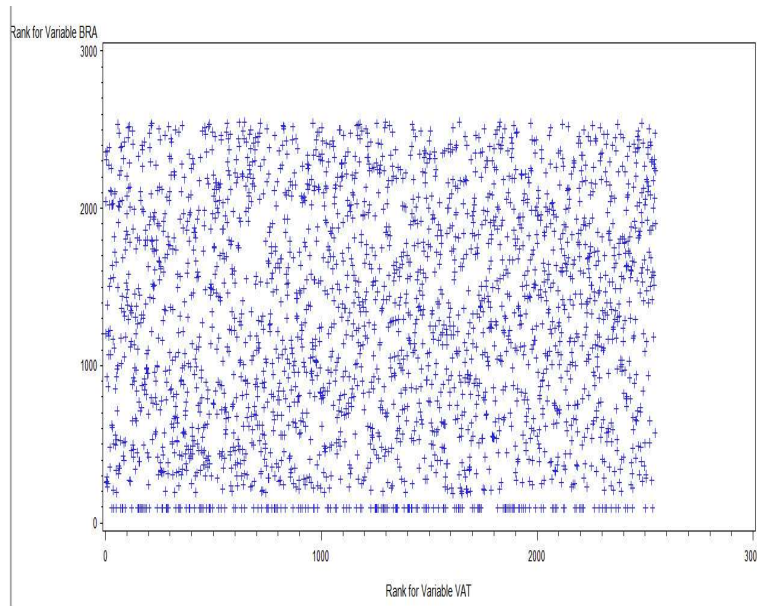
Figur 29: Inkrementella taxationer ranger, månadsnivå, vatten mot brand



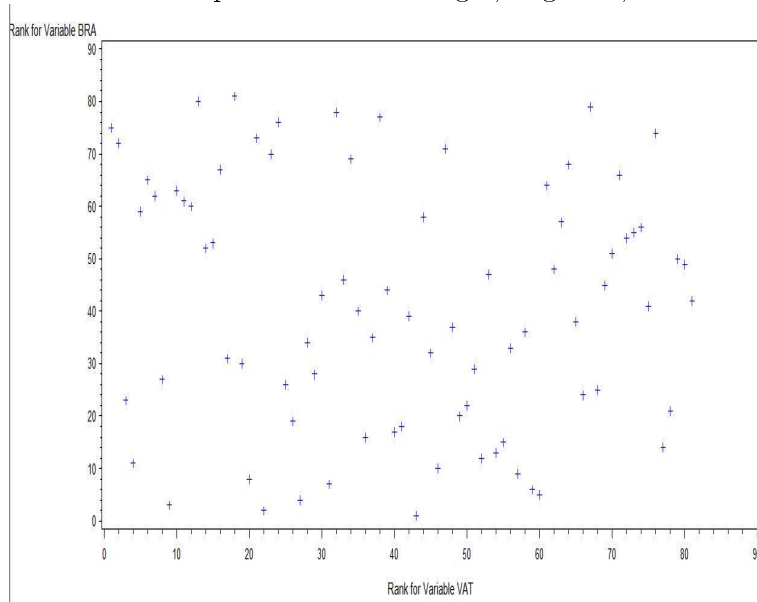
Figur 30: Data baserat på skadedatum, dagsnivå, vatten mot brand



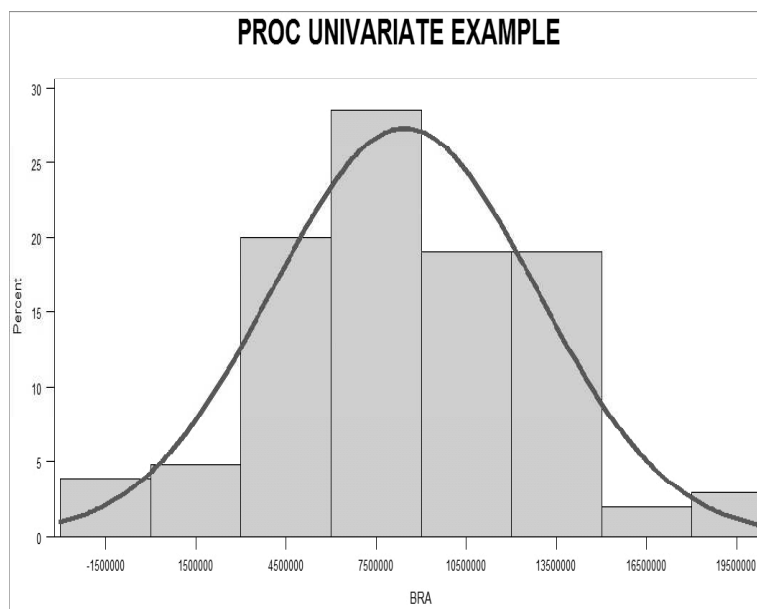
Figur 31: Data baserat på skadedatum, månadsnivå, vatten mot brand



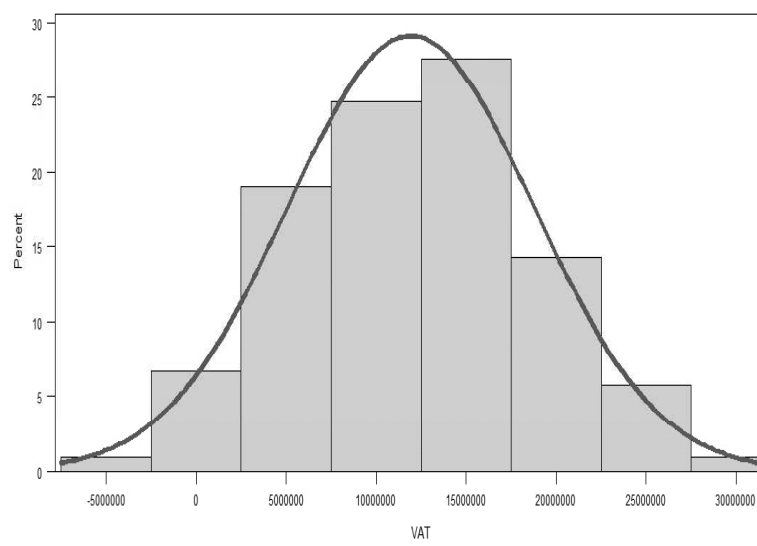
Figur 32: Data baserat på skadedatum ranger, dagsnivå, vatten mot brand



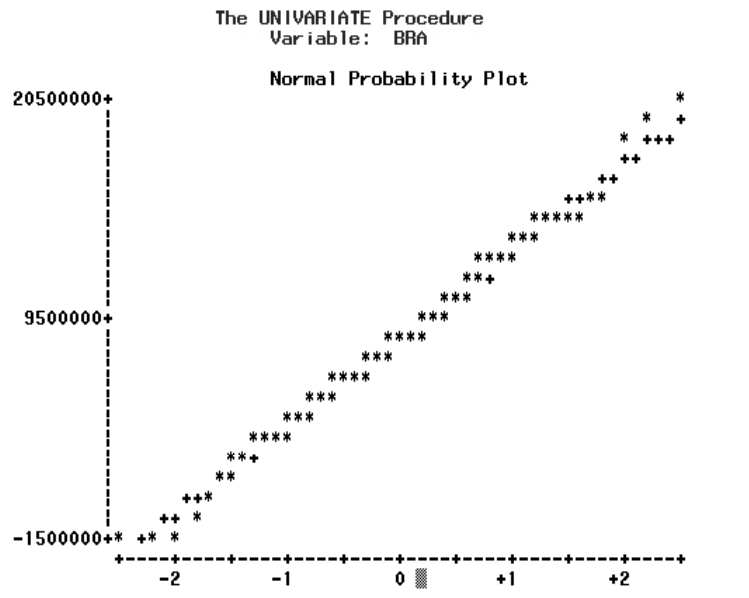
Figur 33: Data baserat på skadedatum ranger, månadsnivå, vatten mot brand



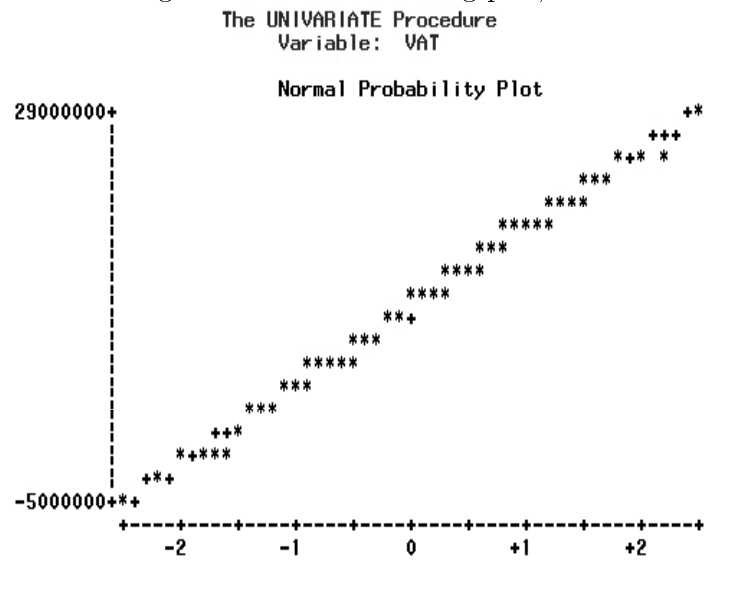
Figur 34: Histogram med normalfördelningskurva för Brand



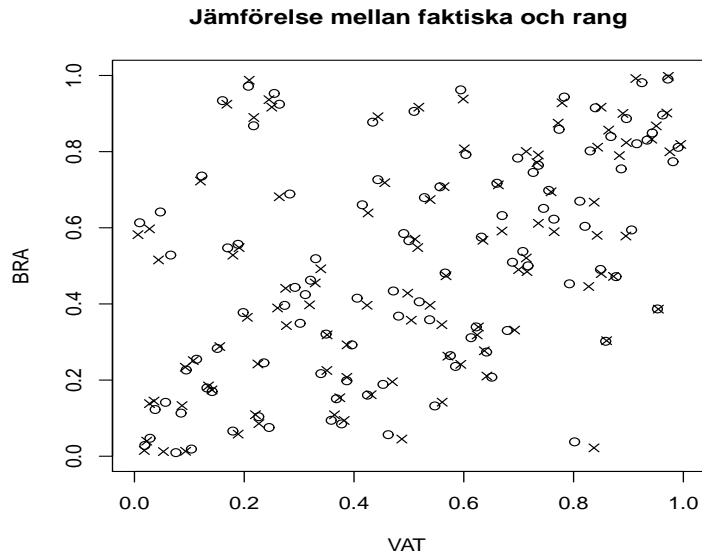
Figur 35: Histogram med normalfördelningskurva för Vatten



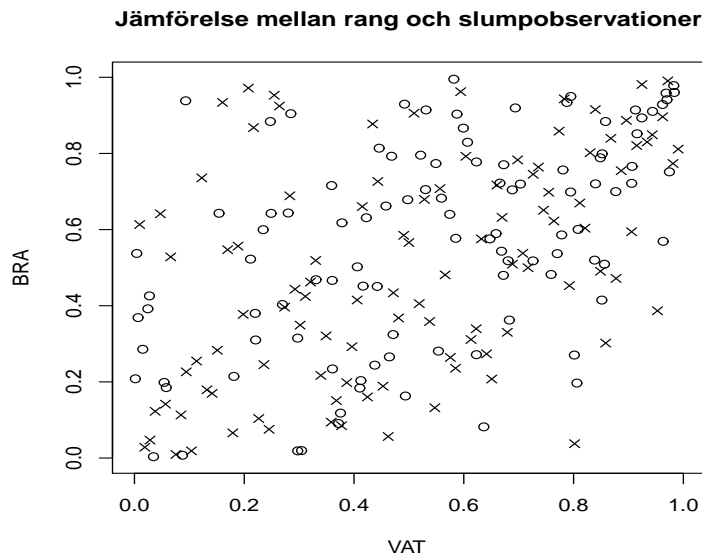
Figur 36: Normalfördelningsplot, Brand



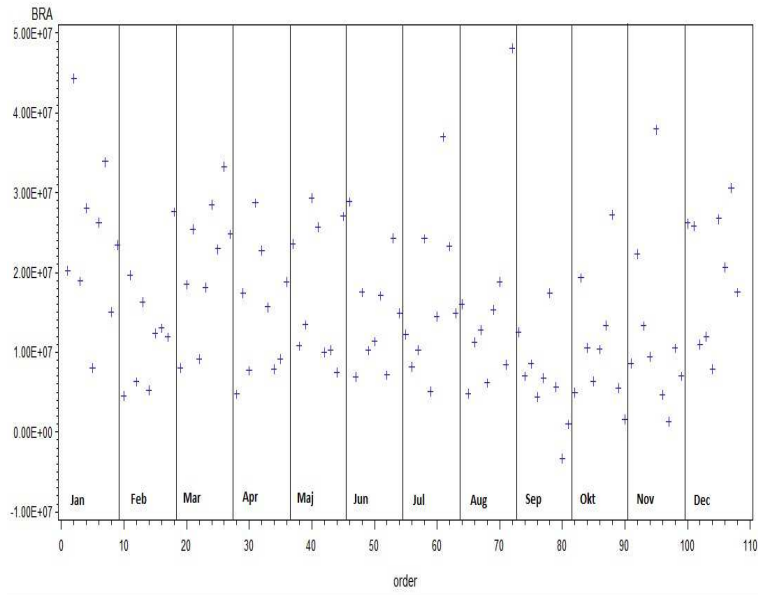
Figur 37: Normalfördelningsplot, Vatten



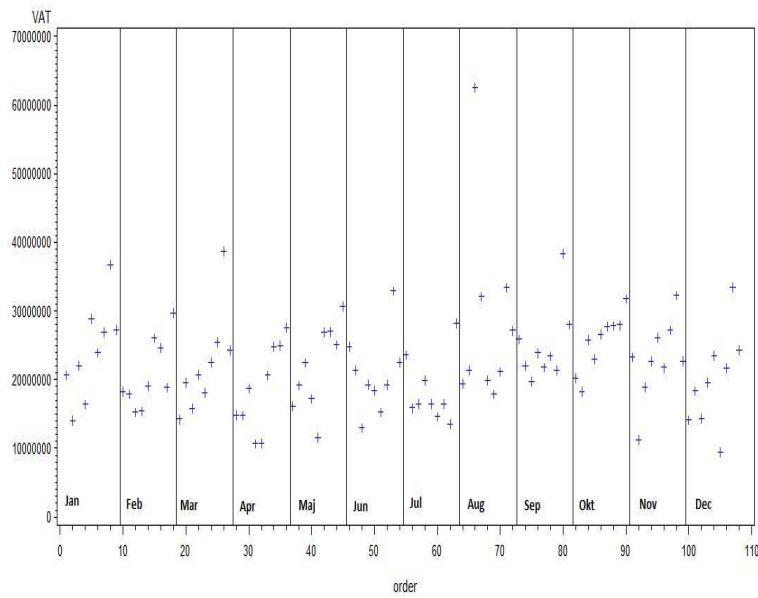
Figur 38: Scatterplot av rangobservationer,(o), mot dataobservationernas antagna värden vid  $U(0, 1)$ -fördelning, (x)



Figur 39: Scatterplot av slumpantal genererade från anpassad Gumbel copula till data,(o), mot rangobservationer,(x)



Figur 40: Månadstaxationer enligt inkrementellt data för brand, sorterat efter månad och år



Figur 41: Månadstaxationer enligt inkrementellt data för vatten, sorterat efter månad och år