Modeling dependence of claims in insurance using autoregressive conditional duration models

- Sawsse Araichi (University Lyon 1, Laboratory SAF - Institute of high commercial studies, Sousse, Tunisia, Laboratory research for economy, management and quantitative finance)
- Christian De Peretti (University Lyon 1, Laboratory SAF)
- Lotfi Belkacem (Institute of high commercial studies, Sousse, Tunisia, Laboratory research for economy, management and quantitative finance)

2013.9
Modeling Dependence of Claims In Insurance Using Autoregressive Conditional Duration Models

Sawssen Araichi\textsuperscript{a,b}, Christian de Peretti\textsuperscript{a}, and Lotfi Belkacem\textsuperscript{b}

\textsuperscript{a}Laboratory of Actuarial and Financial Sciences (SAF, EA2429), Institute of Financial and Insurance Sciences(ISFA School), University Claude Bernard Lyon 1, France.
\textsuperscript{b}Laboratory Research for Economy, Management and Quantitative Finance, Institute of High Commercial Studies of Sousse, Tunisia.

February 12, 2013

Abstract

This paper proposes to analyze the dynamic behavior of claims amounts in the insurance company using Autoregressive Conditional Duration Models (ACD). We highlight that Gamma ACD model and a new model called the Generalized Extreme Value ACD are more appropriate to describe the behavior of the process of claims, and to forecast the conditional expected loss of the lines Auto Damage and Auto Liability. Furthermore, we derive a parametric VaR ACD model to evaluate a coverage amount of these claims. Using backtesting techniques, the VaR ACD model provides a good estimation of risk.

Keywords: Claims Amounts, Temporal Dependence, ACD models, diagnostic test, Value at Risk ACD.

1 Introduction

Insurance is considered as a transfer of risks from the insured to the insurer. It must have a level of liability (equity and reserves), which enables it to absorb unfavorable risks and to be solvent in future years. A fundamental problem in insurance is to predicting the distribution of the total claims amount in a given time period, in order to evaluate premiums and reserves and to avoid ruin of the company. For that actuaries must adequately model the distribution of amounts. Among parametric distributional families, which have been extensively studied for modeling insurance loss process. The Lognormal model used by (Kremer, 1982) in his ANOVA approach, and the Negative binomial model of (Richard, 2000). Also (Cummins et al., 1990) introduced the Generalized Beta of the Second Kind (GB2) for modeling insurance loss payment, they have shown that the GB2 family includes many commonly used distributions, such as the Lognormal, Gamma, Weibull, the Burr and Generalized...
Gamma Distributions. Actuarial literature have shown that the exponential family distributions is the more adequate to model the amount of Loss (Mikosch, 2006). (Cooray and Ananda, 2005) model actuarial data with a composite Lognormal Pareto Distribution, and they have shown that this model can take into account the tail behavior for both small and large losses. For a comprehensive overview of modeling actuarial data we refer readers to (England and Verrall, 2002). The common shortcoming of these models, that they ignore the temporal dependence that may exist between the loss payment of the insurance company. In the actuarial science, these claims amounts are generally considered as a random sum and assumed to be a sequence of independent and identically distributed see(Bühlmann, 1996) and (Rolski et al., 1999).

Actuaries seem to believe that identifying the evolution of claim amounts in time improve their prediction in the future. One approach is the use of time series to forecast on the characteristics of the dynamic effect of the claims over time . for exemple; the Autoregressive Moving Average models (ARMA) and Autoregressive Conditional Heteroskedasticity models that can be found in (Box and Jenkins, 1976), (Nelson and Van Ness, 1973) and (Newbold, 1982). However, there are few empirical investigations of actuarial models analysing the temporal dependence between claims.

The main purpose of this paper is to specify the temporal dependence between the claims amounts, to introduce a new model called Generalized extreme value ACD for modeling extreme amounts, and to develop a methodology to obtain an expression and its estimate for the Value at Risk of the one day ahead predictive distribution of the claims, conditionally on the past and current losses. Specifically we introduce a new method for analysing insurance risks through the use of the time series models, especially the Autoregressive Conditional Duration Models. Indeed, we model two lines of business of the insurance company, the Auto Damage and the Auto Liability, we find that the temporal dependence of these two lines can be expressed by a GAMMA and GEV Autoregressive Conditional Duration model. This paper is organized as follows: section 2 is devoted to literature models, section 3 presents the standard Autoregressive Conditional Duration model, specifications and several of its extensions, existing checking tests, and methods to develop VaR ACD model. Section 4 reports the empirical results, followed by concluding remarks and some open questions in section 5.

2 Models of Literature

2.1 Models of Independence

A key objective of the actuarial risk theory is the assessment of a risk of the contracts portfolio, to estimate, premiums, claims amounts, reinsurance and other decision variables such as the probable maximum loss. In the risk theory the total cost of claims over a period is often calculated by taking into account the frequency and the size of individual claims. Consider a portfolio of insurance contracts for a period of time. The aggregation of claims is generally regarded as the sum of the losses generated by each contract:

\[ S = \sum_{i=1}^{n} X_i \]
In actuarial science two basic models were proposed to model these amounts of claims: the Individual and the Collective model. In the individual model of risk, we are interested in modeling the distribution of total claims $S$. The total costs of accidents caused by $N$ contracts are written as, $X_1, \ldots, X_n$, these variables are assumed to be independent but not identically distributed. This takes into account the characteristics of each claim and the heterogeneity of the portfolio. The collective model has an important role in the development of risk theory in actuarial science, it models the total claims as a sum of independent and identically distributed variables. This model does not distinguish the contracts in the portfolio but sees it as a series of shocks caused by the occurrence of disasters. For more details about these models, we refer readers to the book (Goovaerts et al., 2001). All these models are developed on a key assumption: independence between risks. Towards this context, dependencies between the non-life risks may be one of the causes of under pricing or insufficient funds to cover the cost of claims corresponding to a year. We present above, some models of dependence presented in the actuarial theory.

2.2 Models of Dependence

In actuarial science, individual risks are usually assumed to be independent. However independence assumption does not always reflect reality. In many lines of business, the introduction of dependence at the portfolio is needed to represent the effects of events hitting several policies simultaneously like car accident, earthquakes, epidemics and so on. Consequently, individual risks are certainly not independent but merely depend on each other. Several notions of dependence were introduced in the literature to model the fact that large amount of risk tend to be associated with large amount of the others. (Goovaerts and Dhaene, 1996) looked for the type of dependence between individuals that gives rise to the riskiest aggregate claims.

They consider a portfolio of dependent risks $(X_1, X_2, \ldots, X_n)$ having a given two point distribution in 0 and $\alpha_i > 0$, that:

$Pr(X_i+1 = 0 \mid X_i = 0) = 1,$ (i = 1, 2, ..., n − 1) and $Pr(X_i+1 = \alpha_i \mid X_i = \alpha_i = 1),$ (i = 2, ..., n).

They have proven that the dependency between the risks $X_i$ gives rise to the riskiest aggregate claims random variable in the sense that it has the largest stop loss premiums.

$\pi_S(d) = E(max(S − d, 0))$ for $d \geq 0$.

Another paper of (Goovaerts and Dhaene, 1996), where they have shown that the well known Compound Poisson approximation of the aggregate independent claims still perform well when claims are dependent. For each $X_i = I_i V_i$, with $V_i$ is the total claim amount and $I_i$ is an indicator of risk. The model of aggregate claims is presented as:

$F^{cp}(s) = \sum_{n=0}^{s} Pr(k = n) F^n(s)$ (s = 0, 1...).

where $k$ is distributed as a Poisson random variable with parameter $\lambda$ given by:

$\lambda = \sum_{i=1}^{n} q_i \gamma_i$; is the probability of having a risk in the portfolio, and $F(s)$ is the distribution given by:
\[ F(s) = 1/\lambda \sum_{i=1}^{n} q_i Pr(V_i \leq s \mid I_i = 1) \].

(Denuit et al., 2001) consider a weak form of positive dependence, known as positive cumulative dependence (PCD), presented as follows: \( X_1, X_2, \ldots, X_n \), with marginal distribution functions \( F_1, F_2, \ldots, F_n \), then:

\[
X_1^\perp + X_2^\perp + \ldots + X_n^\perp \leq X_1 + X_2 + \ldots + X_n \leq X_1^U + X_2^U + \ldots + X_n^U
\]

where \( X_1^\perp, X_2^\perp, \ldots, X_n^\perp \) represent independent version of \( X_1, X_2, \ldots, X_n \) and \( X_1^U, X_2^U, \ldots, X_n^U \) represent the comonotonic versions of \( X_1, X_2, \ldots, X_n \).

\( X_i^U = F_i^{-1}(U), X_2^U = F_2^{-1}(U), X_n^U = F_n^{-1}(U) \). Where \( U \) denotes a random variable uniformly distributed on the unit interval \([0,1] \) and \( F_i^{-1} \) is the quantile function associated to the distribution function \( F_i \) of \( X_i \).

Therefore, the authors show that, any risk averse decision maker will prefer \( X_1^\perp + X_2^\perp + \ldots + X_n^\perp \) over \( X_1 + X_2 + \ldots + X_n \) when the risks are PCD. Also they show that the premium of a sum of PCD risks is maximal if the risks are comonotonic and minimal if the risks are mutually independent. So, if we consider a premium amount \( H(x) \) to any risk \( X \) they found that:

\[
H[\sum_{i=1}^{n} X_i^\perp] \leq H[\sum_{i=1}^{n} X_i] \leq H[\sum_{i=1}^{n} X_i^U]
\]

(Genest et al., 2002) use a compound Poisson distribution to approximate the distribution of the total claim amount where dependence between the policies arises through mixtures. they provide a model to explain the effect of dependence on the total claim amount when the contracts are linked through an Archimedean copula model.

All these models, introduce the structure dependence between the number of occurrence of the risks, they ignore the temporal dependence that may exist, and they fail to specify the processes followed by claims that make possible the forecast of future amounts.

Forecasting the dynamic process of claims and setting up suitable reserves to meet claims is an important mission of the actuaries. In this context, (Gerber, 1982) examined the probability of ruin in a model where the annual gains of an insurance company are dependent random variables. He used a linear model which is often used in time series analysis including the autoregressive model and the moving average model as a special cases.

(Cummins, 1985), test the performance of the econometric ARIMA models in forecasting two paid claim cost series:

\[ \log(y_t) = \beta_0 + \beta_1 \log W_t + u_t \]

where: \( y_t \) are the average paid claim costs in quarter \( t \), \( W_t \) is the total non-farm private sector compensation per man-hour, and \( u_t \) is the disturbance. They model two lines of business and show that forecasting with ARIMA models can produce a clear gain in forecast accuracy for the line damage liability insurance that can lead to a substantial quantitative difference in premium collections. For the line bodily injury coverage the authors found no clear gain accuracy from adopting econometric models. However, no
accuracy would be lost. Harvey and Fernandes, (1989) developed a time series model for studying a number of claims and their payment amounts. Their model is based on the Gaussian distribution, they assume that claim amounts are lognormally distributed and they satisfy the model: \( y_{jt} = \mu_t + \varepsilon_{jt}, \ j = 1, ..., N_t \) and \( t = 1, ..., T. \) where: \( \varepsilon_{jt} \) are mutually and serially uncorrelated disturbances with variance \( \sigma^2_{\varepsilon}. \) They have shown that this model is not very appropriate to the claim amounts and they suggest to use the Gamma distribution.

Also El-Bassiouni and El-Habashi, (1991) forecast monthly compulsory motor insurance claims based on Box Jenkins methodology. The model is given by:

\[
(1 - B^{12})Y_t = \alpha + (1 - \beta B^{12})\alpha_t
\]

where:

- \( Y_t = \ln(X_t) \).
- \( X_t \): are the claims at month \( t \).
- \( B \): is the backward shift operator.
- \( \alpha_t \): is a white noise process with variance \( \sigma^2, \alpha, \beta \) and \( \sigma^2 \) are unknown parameters.

They show that their model forecast well the future claims and can cover all the actual claims of the year.

In addition, Promislow, (1991), (Zhang, 2005), determine the ruin probability expression of the claims under a linear time series model, they model claims using the moving average and the autoregressive models.

All models above are based on the normality assumption that may lead to unreliable finding. Cossette et al., (2011) have introduced the Poisson autoregressive model and the Poisson moving average model to analyse the dependence relationship in time between the claim frequencies. They have shown that the structural dependence of Poisson autoregressive model and the Poisson moving average model has in impact on the stop Loss and the risk measures. In fact, the stop loss premium increase and both the Value at Risk and the Tail Value at Risk increase with this dependence.

Most of the temporal dependence models proposed to date fail to forecast the dependence relationship in time between the claims amounts and to describe the variation of claims. they model the amount of claims assuming the normality assumption, which is not real in the context of claims amounts in insurance. Also, claims amounts involve some extreme amounts, requiring new specific statistical techniques. So, we introduce in this paper a new model called the Generalized extreme value ACD model.

Basically, the economic motivations behind the Autoregressive Conditional Duration Models are: First, due to the clustering of risks, claims insurance occur in clusters. This implies that amounts of losses exhibit a significant serial correlation. So the claim amounts are described as self exciting when the past evolution impacts the probability of future events. Second, like the GARCH models, the Autoregressive Conditional Duration Models and their alternatives have proven their success in capturing the clustering effects. For this reason, it seems interesting to model the cluster behavior of risks in insurance.
3 The Autoregressive Conditional Duration Models

(Engle and Russell, 1998) introduced the Autoregressive Conditional Duration Models (ACD), which present a new way of modeling irregularly spaced financial transaction data. They used the similar idea of the Generalised Autoregressive Conditional Heteroscedastic (GARCH) models to develop the ACD models based on exponential (EACD) and weibull (WACD) distributions. Their explicit objective is to model the intertemporally correlated duration. At present, this model is considered as the most powerful tools, as provided by several applications to real financial data ((Engle and Russell, 1998), (Engle, 2000) and (Engle, 2002). (Bauwens and Giot, 2000) introduced the Log-ACD model, which implies a non linear relation between the duration and its lags. (Grammig and Maurer, 2000) developed an ACD model based on the Burr distribution. They showed that their model provides a greater flexibility than the EACD and WACD models. (Zhang et al., 2001) proposed the Threshold Autoregressive Conditional Duration (TACD) model to allow the expected duration to depend non linearly on past information. (Lunde, 1999) extend the ACD model to Generalized Gamma duration that allows a non monotonic hazard function taking for instance both tub shaped or inverted U-shaped forms. (Ghysels et al., 2004) presented the stochastic volatility duration (SVD) model for processes that involve time varying uncertainty. In the following, we present three specifications of ACD models: EACD, WACD and GACD models, and we introduce a new one called GEVACD model.

3.1 The Standard ACD model

In the following, we present the ACD model adapted to the context of insurance.

Let \( X_i \) represent the claims amounts (loss) paid at the day \( i \).

The ACD model assumes that \( X_i \) follow a mutiplicative structure:

\[
X_i = \psi_i \epsilon_i \text{ that } \epsilon_i \sim \text{i.i.d with } E(\epsilon_i) = 1 \tag{1}
\]

That \( \psi_i = E(X_i \mid X_{i-1}) \) is the conditional expected loss. Engle and Russell (1998) specify an autoregressive structure for this conditional mean:

\[
\psi_i = \omega + \sum_{j=1}^{p} \alpha_j X_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{2}
\]

So \( \psi_i \) depends on \( p \) past observations and \( q \) past expected observations. The autoregressive structure implies that small observations (respectively low), are more likely to be followed by small observations. Thus the model accounts for the clustering effect on the time elapsed between events.

With the constraints on the coefficient \( \omega > 0 \) and \( \sum (\alpha_i + \beta_j) < 1 \).

\[A \text{ variable } T \text{ follows a Burr distribution has a probability density function } f(t) = \frac{\mu x^{\mu-1}}{(1+\sigma^2\mu x)^{\frac{\mu}{\sigma^2}+1}}.\]
3.2 Estimation of ACD model

Estimating parameters of the ACD model is based on maximizing the log likelihood function. Let \( f(\epsilon, \theta) \) be the density function for \( \epsilon \) with parameters \( \theta_\epsilon \), knowing that the standardized claims amounts \( \epsilon_i \) are independent and identically distributed with a mean of unity.

Let \( g(x_i | \Omega_{i-1}; \theta_x) \) be the probability density function for \( X_i \) knowing the observation of the data set available at time \( t_{i-1} \).

This implies that \( g(x_i | \Omega_{i-1}, \theta_x) = g(\frac{x_i}{\psi_i}, \theta_x) \), then \( g(x_i | \Omega_{i-1}) = \frac{1}{\psi_i} f(\frac{x_i}{\psi_i}, \theta_\epsilon) \).

where: \( \theta = (\theta_x, \theta_\epsilon) \) is the vector of all the unknown parameters.

The log likelihood function is given by:

\[
L(\theta) = \sum_{i=1}^{N} \log g(x_i | \Omega_{i-1}, \theta) = \sum_{i=1}^{N} [\log f(\frac{x_i}{\psi_i}, \theta_\epsilon) - \log(\psi_i)]
\]  

(3)

In the following, we present the ACD model with Exponential, Weibull, Gamma and Generalized extreme value distributions.

Exponential ACD model

Engle and Russell, 1998, introduced the EACD model in their seminal paper, the authors assume that the scale parameter \( = 1 \). Let \( X_i = \psi_i \epsilon_i \) with \( \epsilon_i \sim \exp(1) \).

So \( P(X_i < x_i) = P(\psi_i \epsilon_i < x_i) = P(\epsilon_i < \frac{x_i}{\psi_i}) = 1 - \exp(-\frac{x_i}{\psi_i}) \). Thus the density function of the conditional claims amounts is:

\[
f(x_i | \Omega_{i-1}) = \frac{1}{\psi_i} \exp(-\frac{x_i}{\psi_i})
\]  

(4)

The log likelihood of the model:

\[
L(x_i | \Omega_{i-1}) = -\sum_{i=1}^{N} [\log(\psi_i) + \frac{x_i}{\psi_i}]
\]  

(5)

Weibull ACD model

Engle and Russell assume a model that the residuals follow a Weibull distribution with a shape parameter \( k \) and a scale parameter \( \lambda = 1 \) : \( \epsilon_i \sim W(k, 1) \). We define the density function of WACD as :

\[
f(x_i | \Omega_{i-1}) = \left(\frac{x_i \Gamma(1 + \frac{1}{k})}{\psi_i}\right)^k \frac{k}{x_i} \exp\left(-\frac{x_i \Gamma(1 + \frac{1}{k})}{\psi_i}\right)^k
\]  

(6)

The log likelihood function is estimated as:

\[
L(x_i | \Omega_{i-1}) = \sum_{i=1}^{N} [k \log\left(\frac{\Gamma(1 + \frac{1}{k})x_i}{\psi_i}\right) + \log\left(\frac{k}{x_i}\right) - \left(\frac{x_i \Gamma(1 + \frac{1}{k})}{\psi_i}\right)^k]
\]  

(7)
Gamma ACD model

The GACD model is a particular case of the Generalized Gamma Model introduced by [Lunde, 1999], so \( \epsilon_i \sim \text{Gamma}(k, 1) \). The density function and the log likelihood are calculated as:

\[
f(x_i \mid \Omega_{i-1}) = \frac{x_{i}^{k-1}}{\psi_i^k \Gamma(k)} \exp(-\frac{x_i}{\psi_i}) \quad (8)
\]

\[
L(x_i \mid \Omega_{i-1}) = \sum_{i=1}^{N} ((k - 1)\log(x_i) - k\log(\psi_i) - \log(\Gamma(k)) - \frac{x_i}{\psi_i}) \quad (9)
\]

Generalized extreme value ACD model

The existence of high claims amount in the insurance company encourage actuary to model with an extreme value distribution. In our investigate, the novel form, is to model with a temporal extreme value distribution. We develop a new methodology based on ACD model, assuming that the standardized claims amounts follow a Generalized extreme value (GEV) distribution, with \( E(\epsilon_i) = 1 \). To the best of our knowledge, this is the first research to take into account the temporal dependence between extreme value of claim amounts in the insurance. The distribution and the density function of the GEVACD model are presented as follows:

\[
F(x_i) = \exp\left\{-\left[1 + \frac{(\frac{x_i}{\psi_i} - \mu)(1 - \Gamma(1 + \xi))}{(\mu - 1)}\right]^{\frac{1}{\xi}}\right\} \quad (10)
\]

\[
f(x_i \mid \Omega_{i-1}) = \frac{1 - \Gamma(1 + \xi)}{\xi(\mu - 1)\psi_i} \left[1 + \frac{(\frac{x_i}{\psi_i} - \mu)(1 - \Gamma(1 + \xi))}{(\mu - 1)}\right]^{-\frac{1}{\xi}} \exp\left\{-\left[1 + \frac{(\frac{x_i}{\psi_i} - \mu)(1 - \Gamma(1 + \xi))}{(\mu - 1)}\right]^{\frac{1}{\xi}}\right\} \quad (11)
\]

And the log likelihood of the GEVACD model are then formalized as:

\[
L(x_i \mid \Omega_{i-1}) = \log(1 - \Gamma(1 + \xi)) - \log((\mu - 1)\psi_i\xi) + (-\frac{1}{\xi} - 1)\log(1 + \frac{(\frac{x_i}{\psi_i} - \mu)(1 - \Gamma(1 + \xi))}{(\mu - 1)})
\]

\[-\left[1 + \frac{1}{(\mu - 1)}(\frac{x_i}{\psi_i} - \mu)(1 - \Gamma(1 + \xi))\right]^{\frac{1}{\xi}}\]

With shape parameter \( \xi \) and location parameter \( \mu \).
3.3 Test of Residuals

Ljung Box Test

A model of dependence can be tested by using the Ljung Box test for autocorrelation. It can be defined as follows:

\( H_0 \): The data are independently distributed.
\( H_1 \): The data are correlated.

The test statistic is:
\[
Q_{LB} = N(N + 2) \sum_{j=1}^{p} \frac{\rho_j^2}{N-j},
\]
where \( N \) is the number of observations, \( \rho_j \) is the sample autocorrelation at lag \( j \), and \( p \) the number of autocorrelation. \( Q \sim \chi_{1-\alpha}^2(p) \), where \( \chi_{1-\alpha}^2 \) the \( \alpha \) quantile of the chi-squared distribution with \( p \) degrees of freedom.

The ACD process assumes that residuals are i.i.d with no serial correlations. The approach used by (Engle and Russell, 1998) consists of applying the Ljung Box statistic to check for remaining serial dependence. Large value of the Ljung Box statistic indicate model inadequacy.

Lagrange Multiplier Test LM

The LM test assesses the null hypothesis that a series of residuals exhibit no conditional heteroscedasticity (ARCH effect), against the alternative the existence of ARCH effect. It can be reduced to compute the Lagrange multiplier statistic:
\[
LM = N \star R^2,
\]
where \( N \) is the sample and \( R^2 \), is the R-squared obtained from regression.

After estimating ACD models and choosing the best one to perform the data, insurance company should calculate the amount of coverage allowing it to meet its obligations to its policyholders. For that the new solvency 2 framework suggest to evaluate capital using the Value at risk measure at 99.5% confidence level. (Quantitatives Impact studies 5 (2010)). In the following, we derive the Value at Risk ACD at confidence level \( \alpha \) and we proceed to implement a backtesting procedure.

3.4 Evaluation of the coverage amount: VaR ACD model and backtesting

Value at Risk constitutes the most popular measure of risk, it refers to the maximum expected loss that will not be exceeded under normal market conditions over a predetermined period (H), at a given confidence level \( \alpha \) (Jorion, 1997). It corresponds to:
\[
Pr(Loss(H) \leq VaR_\alpha) = \alpha
\]

Primarily, the VaR is a banking model integrated via Basel 2, and used to estimate capital requirement to support the market risk.

In the insurance context, the new framework solvency 2 requires that insurers must have
a solvency capital allowing them to be solvent with a higher probability (99.5%). So, they adopt the VaR model providing a measure of prospective risk of the insurance. In our study, our contribution is to develop a VaR ACD model adapting to forecast the coverage amount of claims in the non-life insurance. We suggest various ACD distributions for calculating the VaR.

### 3.4.1 VaR ACD models

Similarly to VaR GARCH models, we derive the VaR ACD model as follows:

- We suppose that the claims amounts follow a parametric ACD distribution $F(x)$.

- We divide the sample of claims in two samples 70% and 30%, then we estimate the parameters of ACD model of the first sample from the period 1 to $T$.

- We forecast the conditional expected claims for the period $T+1$:

\[
\hat{\psi}_i = \hat{\omega} + \sum_{j=1}^{p} \hat{\alpha}_j x_{i-j} + \sum_{j=1}^{q} \hat{\beta}_j \psi_{i-j} \tag{13}
\]

$\hat{\psi}_i$ is the one step ahead conditional expected forecast.

We note $VaR_{T+1}(\alpha)$ the Value at Risk forecast at the confidence level $\alpha$, at time $T+1$:

\[
Pr(X_{T+1} < VaR_{T+1}|\Omega_T) = \alpha \tag{14}
\]

$X_{T+1}$: The amount of claims forecast at time $T+1$.

$\Omega_T$: Available information at time $T$.

We have:

\[
Pr(\psi_{T+1}\epsilon_{T+1} < VaR_{T+1}|\Omega_T) = \alpha
\]

\[
Pr(\epsilon_{T+1} < \frac{VaR_{T+1}}{\psi_{T+1}}|\Omega_T) = \alpha
\]

Then it is straightforward to derive the: $VaR_{T+1|T}(\alpha) = \psi_{T+1}F^{-1}(\alpha)$

From this expression, we can deduce VaR forecast with various ACD distributions.

The VaR EACD model is estimated as follows:

\[
VaR_{T+1|T}(\alpha) = -Log(1 + \alpha)\psi_{T+1} \tag{15}
\]

and the VaR WACD model is

\[
VaR_{T+1|T}(\alpha) = \frac{-Log(1 - \alpha)\psi_{T+1}}{\Gamma(1 + \frac{1}{\kappa})} \tag{16}
\]
The estimation of VaR GACD model is
\[ \text{VaR}_{T+1|T}(\alpha) = \hat{\psi}_{T+1} G^{-1}(\alpha, \hat{k}) \] (17)

\( G^{-1} \): is the inverse of the Gamma function with shape parameter \( k \).

One main object of this paper is to develop a methodology to obtain an expression for the Value at Risk of the GEVACD model of the one day ahead predictive distribution of the claims. So the VaR GEVACD estimation is
\[ \text{VaR}_{T+1|T}(\alpha) = \hat{\psi}_{T+1} \left[ \frac{(\mu - 1)}{1 - \Gamma(1 - \xi)} (\log(\alpha) - \xi - 1) + \mu \right] \] (18)

After evaluating the VaR, we now turn our attention to accuracy of the generated out of sample forecasts in calculating reliable VaR measures. For that, we use a backtesting procedure. The VaR forecast must neither overestimate nor underestimate the true VaR. More specifically, the insurance company should allocate the adequate amount of capital to cover the risks. Consequently, we need to set up adequate techniques validating or not the amount of risk.

As defined by (Jorion, 1997), backtesting consists in verifying if actual losses are in line with projected losses. This involves a systemic comparison of the history of the model generated VaR forecasts with actual losses and relies on testing over VaR violations. In the following, we present the backtesting procedure applied to the VaR ACD model.

Let us denote \( x_{T+1} \), the amount of claims at time \( T+1 \). \( \text{VaR}_{T+1|T}(\alpha) \), is the coverage amount anticipated conditionally to an information set \( \Omega_T \) available at time \( T \) for the \( \alpha \) confidence level. Let \( I_{T+1}(\alpha) \) be the violation or hit variable associated to the ex-post observation of a \( \alpha \) VaR violation at time \( T+1 \):
\[ I_{T+1}(\alpha) = \begin{cases} 
1 & \text{if } x_{T+1} > \text{VaR}_{T+1|T}(\alpha) \\
0 & \text{otherwise} 
\end{cases} \] (19)

It means that at time \( T+1 \), the observed loss exceeds the VaR forecasts. As suggested by (Christoffersen, 1998), VaR forecasts are valid if and only if the violation sequence \( \{I_T\} \) satisfies the following two assumptions.
- The unconditional coverage (UC) hypothesis: the probability of a violation must be equal to the \( 1 - \alpha \) confidence level: the probability that the loss exceeds the VaR must be equal to \( 1 - \alpha \).
\[ \Pr(I_T(\alpha) = 1) = E(I_T(\alpha)) = 1 - \alpha \] (20)
- The independence (IND) hypothesis: VaR violations observed at two different dates for the same coverage rate must be independently distributed. In other words, past VaR violations do not hold information about current and future violations.
4 Empirical Analysis

4.1 Data description

The data set is composed of two lines of business from an insurance company, Auto Damage and Auto liability over the period January 2002, to December 2007. The first guarantee covers all damaged caused to the car, even if the insured is responsible of the accident. The second, is the guarantee of damage caused to others, belongs to the Civil Responsibility lines of business.

Focusing on the daily claims amounts of these two lines, we have 2061 and 2038 observations respectively for the Auto Damage and Auto liability. Indeed, for each day \( j \), we have aggregated all the claims amounts occurring at this day.

In the following we present, some summary statistics, the distribution plots of the two lines of business and the corresponding autocorrelations.
Examining the claims amounts distributions reveals that both lines of business exhibit high positive skewness and excess kurtosis indicating that distributions are shifted to the left. The Jarque Bera test of normality rejects the normality of claims amounts distributions. Thus, our data sets are very heavy-tailed.
Figure 3: plot of the lines Auto Damage and Auto Liability

Figure 4: Autocorrelations of the lines Auto Damage and Auto Liability
From Figure 3, we can observe the phenomenon of clustering of the claims amounts; high (respectively low) amounts tend to be followed by high (respectively) low amounts, that may be due to the arrival process of claims or the covered system adopted by the company. Indeed, due to weather events, claims can arrive simultaneously. Also, insurance company can pay the large losses in the same period. Figure 4 and 5 depict the plots of the Autocorrelations and the Partial autocorrelations of the claims amounts: we observe that the correlations are not large in magnitude, but they clearly indicate serial dependence in the data. Also, we observe that the line Auto Damage and Auto liability exhibit a positive correlations. Therefore, positive dependence properties expressing the notion that 'large' or 'small' value of the claims amounts tend to occur together, this called clustering phenomenon. These features of claims amounts challenges us to apply econometric techniques. Moreover, recent framework solvency 2, argue that claims amount may convey information about the solvency capital in the future and should, therefore be modeled as well. Motivated by these considerations, we specify the stationarity of the data and we will specify the dynamic process of it. We used an Augmented Dickey Fuller Test for a unit root in the data sample. It is carried out under the null hypothesis: a unit root in the data.
Table 1: Augmented Dickey Fuller Test

<table>
<thead>
<tr>
<th>Lines of Business</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto Damage</td>
<td>−22.33</td>
<td>0.000</td>
</tr>
<tr>
<td>Auto Liability</td>
<td>−27.41</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2: Ljung Box Test

<table>
<thead>
<tr>
<th>Lines of Business</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto Damage</td>
<td>308.7</td>
<td>0.00</td>
</tr>
<tr>
<td>Auto Liability</td>
<td>42.2</td>
<td>6.910^{-9}</td>
</tr>
</tbody>
</table>

Examining the stationary of the series, table 1 shows that both series are stationary, the test of Augmented Dickey Fuller exhibit a p-value < 5%, so we have no unit root in the amounts of claims of the lines of business. From table 2 we remark that for the lines of business, the Ljung Box statistics are very large. Auto Damage has a Ljung Box Statistic of 308.7 and 42.2 for Auto Liability. The null hypothesis of white noise is easily rejected for both lines of business based on the critical value 18.31. Due to the clustering and correlation phenomenon between claims amounts, we are interesting to specify the dynamic movement of claims. So we use an autoregressive process, that Loss payment amounts depend on past observations. Indeed Autoregressive Conditional Duration Models are applied to allow us to highlight the dynamic behavior of the claims. We estimate different models: EACD, WACD, GACD and GEVACD. Table 3 and 4 report estimation results.

4.2 Estimation Results with ACD models

The following tables contain the results of estimation of parameters and their corresponding Student statistics test. We present only the model that the estimation results are valid, that are consistent with the constraints of the model. For the Ljung and LM test we present in the tables their statistics and their p-value.
Table 3: Estimation Results of the line Auto Damage

<table>
<thead>
<tr>
<th></th>
<th>EACD(1,1)</th>
<th>WACD(1,1)</th>
<th>GACD(1,1)</th>
<th>EACD(1,2)</th>
<th>WACD(1,2)</th>
<th>GACD(1,2)</th>
<th>GEVACD(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.042 (2.52) ***</td>
<td>0.061 (6.16) ***</td>
<td>0.061 (4.3) ***</td>
<td>0.046 (4.61) ***</td>
<td>0.05 (4.8) ***</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.083 (6.78) ***</td>
<td>0.093 (10.15) ***</td>
<td>0.039 (7.2) ***</td>
<td>0.073 (4.7) ***</td>
<td>0.079 (6.33) ***</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.91 (79.1) ***</td>
<td>0.9 (105.7) ***</td>
<td>0.96 (60.13) ***</td>
<td>0.635 (2.97) ***</td>
<td>0.678 (3.75) ***</td>
<td>0.75</td>
<td>0.97</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.29 (1.42) ***</td>
<td>0.24 (1.98) ***</td>
<td>0.2</td>
<td>-</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>-</td>
<td>1.3 (64.05) ***</td>
<td>1.7 (60.2) ***</td>
<td>-</td>
<td>1.316 (63.9) ***</td>
<td>1.7</td>
<td>-</td>
</tr>
<tr>
<td>$\xi$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.24</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.596</td>
</tr>
<tr>
<td>Mean</td>
<td>0.99</td>
<td>0.993</td>
<td>1.03</td>
<td>1.006</td>
<td>0.999</td>
<td>1.3</td>
<td>0.99</td>
</tr>
<tr>
<td>Variance</td>
<td>0.81</td>
<td>0.82</td>
<td>1.14</td>
<td>0.85</td>
<td>0.83</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>LJB Q(10)</td>
<td>21.17 (0.02)</td>
<td>24.5 (0.05)</td>
<td>10.32 (0.41)</td>
<td>25.45 (0.005)</td>
<td>21.7 (0.016)</td>
<td>39.4 (0.000)</td>
<td>65 (0.000)</td>
</tr>
<tr>
<td>LM (10)</td>
<td>22.7 (0.01)</td>
<td>26.13 (0.004)</td>
<td>10.02 (0.43)</td>
<td>15.33 (0.12)</td>
<td>17.12 (0.1)</td>
<td>11.3 (0.34)</td>
<td>5.6 (0.85)</td>
</tr>
</tbody>
</table>
Table 4: Estimation Results of the line Auto Liability

<table>
<thead>
<tr>
<th></th>
<th>EACD(1,1)</th>
<th>WACD(1,1)</th>
<th>GACD(1,1)</th>
<th>EACD(1,2)</th>
<th>WACD(1,2)</th>
<th>GACD(1,2)</th>
<th>GEVACD(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\omega)</td>
<td>0.054 (54.9) ***</td>
<td>0.052 (52.1) ***</td>
<td>0.055 (50.3) ***</td>
<td>0.059 (59.4) ***</td>
<td>0.052 (2.29) **</td>
<td>0.52</td>
<td>0.053</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.057 (7.5) ***</td>
<td>0.063 (9.27) ***</td>
<td>0.058 (5.13) ***</td>
<td>0.053 (5.11) ***</td>
<td>0.059 (6.41) ***</td>
<td>0.04</td>
<td>0.049</td>
</tr>
<tr>
<td>(\alpha_2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.94 (131.18) ***</td>
<td>0.939 (149.07) ***</td>
<td>0.93 (96.7) ***</td>
<td>0.48 (2.3) **</td>
<td>0.45 (2.59) ***</td>
<td>0.45</td>
<td>0.95</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.46 (2.3) **</td>
<td>0.49 (2.86) ***</td>
<td>0.48</td>
<td>-</td>
</tr>
<tr>
<td>(\kappa)</td>
<td>1.16 (68.6) ***</td>
<td>1.6</td>
<td>1.18 (68.8) ***</td>
<td>1.7</td>
<td>-</td>
<td>-</td>
<td>0.3</td>
</tr>
<tr>
<td>(\xi)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6</td>
</tr>
<tr>
<td>(\mu)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td>0.999</td>
<td>0.99</td>
<td>1.007</td>
<td>1.008</td>
<td>0.998</td>
<td>1.55</td>
<td>1.03</td>
</tr>
<tr>
<td>Variance</td>
<td>1.52</td>
<td>1.47</td>
<td>1.64</td>
<td>1.57</td>
<td>1.51</td>
<td>3.56</td>
<td></td>
</tr>
<tr>
<td>LJB Q(10)</td>
<td>13.01 (0.22)</td>
<td>12.51 (0.25)</td>
<td>12.22 (0.27)</td>
<td>18.06 (0.04)</td>
<td>14.84 (0.13)</td>
<td>9.59 (0.47)</td>
<td>8.34 (0.6)</td>
</tr>
<tr>
<td>LM (10)</td>
<td>12.19 (0.27)</td>
<td>12.84 (0.98)</td>
<td>13.62 (0.19)</td>
<td>7.22 (0.7)</td>
<td>7.98 (0.63)</td>
<td>9.15 (0.51)</td>
<td>9.4 (0.49)</td>
</tr>
</tbody>
</table>
For the two lines of business, we remark that all coefficients of the different models are significant, this observation is supported by the large $t$-statistics observed for all models on the parameters $\alpha_i$ and $\beta_i$, and $\sum(\alpha_i + \beta_i) < 1$, which allow the existence of conditional expected claims amounts. Also results reveals that residuals have a mean of approximately unity. The performance of the ACD models in detecting autocorrelations structure of loss payment, can be evaluated by examining the residuals (called the standardized claims amounts). For the line Auto Damage the autocorrelations are reduced from 308.7 to 23.2 an average, using the Standard models of (Engle and Russell, 1998) (Exponential and Weibull models). These models fail to pass the Ljung Box test at confidence level 5%, because (Ljung Box statistic> Critical value= 18.31), but we can not ignore that the autocorrelations are significantly reduced. For the GACD model, we observe that GACD(1,1) present a Ljung Box statistic lower than 18.31, so it exhibit no autocorrelations in the residuals of the Auto Damage. Examining the results of GEVACD(1,1) model, this model fail to capture the autocorrelations between residuals (p-value=0). This model is not adequate for the Auto Damage, that means; this line does not contain an extreme claims amounts. It is reasonable; this line is related to the material damage of the Auto. Regarding the LM test, the statistics indicate no ARCH effect for all models except the EACD(1,1) and the WACD(1,1) models.

For the Auto Liability, all models success to capture autocorrelations, we observe no autocorrelations and no ARCH effect in the standardized claims amounts. But, the GEVACD(1,1) present the lowest Ljung Box and LM statistics. This model is more adequate in modeling Auto Liability, that means; GEVCACD success to capture the temporal dependence between the extreme claims amounts. These high amounts is related to the corporal damage and all the consequences in terms of health of a person.

We pass now to select the best model that well describe the dynamic effect of the data. For that, we are based on the Loglikelihood value, the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion).
Table 5: Loglikelihood, information criterion of ACD models for the line Auto Damage and Auto Liability

<table>
<thead>
<tr>
<th>Model</th>
<th>Auto Damage</th>
<th>Auto Liability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LogLikelihood</td>
<td>AIC</td>
</tr>
<tr>
<td>EACD(1,1)</td>
<td>20942</td>
<td>41878</td>
</tr>
<tr>
<td>EACD(1,2)</td>
<td>20905.95</td>
<td>41804</td>
</tr>
<tr>
<td>WACD(1,1)</td>
<td>20810.6</td>
<td>41613</td>
</tr>
<tr>
<td>WACD(1,2)</td>
<td>20777.09</td>
<td>41544</td>
</tr>
<tr>
<td>GACD(1,1)</td>
<td>21138</td>
<td>42268</td>
</tr>
<tr>
<td>GACD(1,2)</td>
<td>20888</td>
<td>41766</td>
</tr>
<tr>
<td>GEVACD(1,1)</td>
<td>20727</td>
<td>41442</td>
</tr>
</tbody>
</table>

From table 5, we can see that the GACD(1,1) model is the best one for modeling claims amounts for the two lines of business. In fact, the loglikelihood value of the Exponential and Weibull ACD are quite close, whereas the Gamma ACD produces considerably higher loglikelihood value. Using the information criterion: GACD(1,1) model maximizes the AIC and BIC criterion. So, the GACD(1,1) success to capture the temporal dependence between claims for the line Auto Damage. For the line Auto liability, we show that the GEVACD(1,1) model maximizes the Log-likelihood and the information criterion. It exhibit a better performance than the others models, because it takes into account the extremes claims of the Auto Liability.

To validate our selected model, we proceed to implement the cross validation procedure, which is a technique for assessing how accurately a predictive model will perform the data. For that:

- We partition the data of claims amounts into two complementary samples 70%, \((x_1, ..., x_{T^*})\) called the training sample and 30%, \((x_{T^*+1}, ..., x_N)\), called validation sample.

- We estimate parameters \((\hat{w}, \hat{\alpha}, \hat{\beta})\) of the ACD model on the training sample.

- We predict the conditional expected Loss of the period \((T^*+1, ..., N)\), and we deduce the claims amounts forecast \((\hat{x}_{T^*+1}, ..., \hat{x}_N)\).

- We calculate the Mean Squared Error and the Root Mean Squared Errors of the model on the validation sample:

\[
MSE = E[(\hat{x}_{T^*+1} - x_{T^*+1})^2]
\]

\[
RMSE = \sqrt{MSE}
\]

We present below the results of calculating the \(RMSE\) of different ACD models:
Table 6: Results of the RMSE of ACD models for the line Auto Damage and Auto Liability

<table>
<thead>
<tr>
<th>Model</th>
<th>Auto Damage</th>
<th>Auto Liability</th>
</tr>
</thead>
<tbody>
<tr>
<td>EACD(1,1)</td>
<td>6272.8</td>
<td>6415.2</td>
</tr>
<tr>
<td>EACD(1,2)</td>
<td>6256.3</td>
<td>6417</td>
</tr>
<tr>
<td>WACD(1,1)</td>
<td>6286.7</td>
<td>6428.3</td>
</tr>
<tr>
<td>WACD(1,2)</td>
<td>6261.3</td>
<td>6422.3</td>
</tr>
<tr>
<td>GACD(1,1)</td>
<td>6226.6</td>
<td>6370.9</td>
</tr>
<tr>
<td>GACD(1,2)</td>
<td>6318</td>
<td>6775.8</td>
</tr>
<tr>
<td>GEVACD(1,1)</td>
<td>6440</td>
<td>5705.1</td>
</tr>
</tbody>
</table>

From table 6, we remark that GACD(1,1) and GEVACD(1,1) present the smallest value of RMSE respectively for the Auto Damage and Auto Liability. So these two models constitute the adequate model for each line of business.

4.3 Comparison with Literature model

In the Actuarial Literature, the model used to detect the temporal dependence between claims amounts is particularly the Autoregressive Moving Average models (ARMA). These models generalize the autoregressive models and has the advantage of being more flexible to use, and provide a good approximations of data. These models are based on a formulation that the present value of the series is presented as a linear combination of its past values and the present value of a noise. A stationary process \((X_t)_{t \in \mathbb{Z}}\) has a representation ARMA\((p,q)\) if it satisfies the relation:

\[
X_t + \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \Leftrightarrow \phi(L) X_t = \Theta(L) \varepsilon_t
\]

To show the importance of our ACD models in modeling claims amounts. We estimate the two lines of business Auto Damage and Auto Liability with ARMA\((1,1)\) and ARMA\((2,2)\). Then, we calculate the amount of coverage with the ACD and the ARMA models and we highlight the difference between them.
Table 7: Estimation Results of the Auto Damage and the Auto Liability with ARMA(1,1)

<table>
<thead>
<tr>
<th></th>
<th>Auto Damage</th>
<th>Auto Liability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARMA(1,1)</td>
<td>ARMA(1,1)</td>
</tr>
<tr>
<td>$c$</td>
<td>1563.13</td>
<td>2640.59</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-0.79</td>
<td>-0.74</td>
</tr>
<tr>
<td>LJB Q(10)</td>
<td>11.56 (0.32)</td>
<td>LJB Q(10)</td>
</tr>
<tr>
<td></td>
<td>3.67 (0.96)</td>
<td></td>
</tr>
<tr>
<td>LM Q(10)</td>
<td>8.39 (0.59)</td>
<td>LM Q(10)</td>
</tr>
<tr>
<td></td>
<td>2 (0.99)</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>8021</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7128</td>
</tr>
</tbody>
</table>

Table 8: Estimation Results of the Auto Damage and the Auto Liability with ARMA(2,2)

<table>
<thead>
<tr>
<th></th>
<th>Auto Damage</th>
<th>Auto Liability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARMA(2,2)</td>
<td>ARMA(2,2)</td>
</tr>
<tr>
<td>$c$</td>
<td>1671.19</td>
<td>2987.74</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.18</td>
<td>-0.28</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-0.96</td>
<td>-0.99</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.17</td>
<td>0.26</td>
</tr>
<tr>
<td>LJB Q(10)</td>
<td>7.79 (0.64)</td>
<td>LJB Q(10)</td>
</tr>
<tr>
<td></td>
<td>2.1 (0.99)</td>
<td></td>
</tr>
<tr>
<td>LM Q(10)</td>
<td>8.96 (0.54)</td>
<td>LM Q(10)</td>
</tr>
<tr>
<td></td>
<td>2 (0.99)</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>8773</td>
<td>RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7968.7</td>
</tr>
</tbody>
</table>

From table 7 and 8, we observe a high constant, so all the weight is given to the constant, and the sum of the estimated parameters is very close to zero, consequently ARMA model can not explain the relation between the claims. Also, we remark that ARMA(1,1) and ARMA(2,2) capture the autocorrelations between claims: Ljung Box statistics < 18.3, and there is no ARCH effect in the residuals. Comparing the RMSE of the selected GACD(1,1) table 5 and 6 to ARMA models, we remark that GACD(1,1) and GEVACD(1,1) present the smallest value of RMSE both for the line Auto Damage and Auto Liability. So, we can conclude that our GACD and GEVACD models are more adequate than the ARMA models presented in the literature. Focusing on the residuals, we present in the ANNEX the QQplot of residuals from the estimating of ACD and ARMA models versus respectively to the Gamma, GEV and the normal distribution.

The quantile quantile plot of the standardized claims amounts against the assumed distribution can be used to check the validity of the model. We observe that the QQplot of the standardized claims of GACD(1,1) of the Auto Damage against the assumed distribution Gamma with shape parameter 1.7 and scale equal to 1, is more close than the residuals of ARMA model against the normal distribution (same interpretation for the Auto Liability.

We can conclude, that due the normality assumption of the ARMA models, these models fail to well describe the dynamic behavior of the claims, and ACD models seems to be more appropriate for forecasting the claims amounts. Next, we Calculate the coverage
amount using the VaR model and perform the backtesting procedure for showing the validity of the VAR (VaR are calculated with the unconditional mean).

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR&lt;sub&gt;99.5%&lt;/sub&gt;</th>
<th>number of hits</th>
<th>Probability of hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>GACD(1,1)</td>
<td>38621</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>GACD(1,2)</td>
<td>38531</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>GACD(2,2)</td>
<td>38598</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>EACD(1,1)</td>
<td>46069</td>
<td>3</td>
<td>0.0058</td>
</tr>
<tr>
<td>WACD(1,1)</td>
<td>39141</td>
<td>3</td>
<td>0.0058</td>
</tr>
<tr>
<td>EACD(1,2)</td>
<td>36205</td>
<td>5</td>
<td>0.00968</td>
</tr>
<tr>
<td>WACD(1,2)</td>
<td>30930</td>
<td>6</td>
<td>0.0116</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>26264</td>
<td>14</td>
<td>0.027</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>26748</td>
<td>19</td>
<td>0.0368</td>
</tr>
<tr>
<td>GEVACD(1,1)</td>
<td>50102</td>
<td>2</td>
<td>0.0033</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR&lt;sub&gt;99.5%&lt;/sub&gt;</th>
<th>number of hits</th>
<th>Probability of hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>EACD(1,1)</td>
<td>58367</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>WACD(1,1)</td>
<td>53352</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>GACD(1,1)</td>
<td>54379</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>EACD(1,2)</td>
<td>58597</td>
<td>3</td>
<td>0.005</td>
</tr>
<tr>
<td>WACD(1,2)</td>
<td>48233</td>
<td>5</td>
<td>0.008</td>
</tr>
<tr>
<td>GACD(1,2)</td>
<td>42118</td>
<td>7</td>
<td>0.011</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>35914</td>
<td>8</td>
<td>0.013</td>
</tr>
<tr>
<td>ARMA(2,2)</td>
<td>35382</td>
<td>11</td>
<td>0.018</td>
</tr>
<tr>
<td>GEVACD(1,1)</td>
<td>59552</td>
<td>3</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Estimating with parametric Value at Risk, reveals that VaR ACD model gives a coverage amount which is adequate to the risks supported by the insurance company. For all ACD models except EACD(1,2) and WACD(1,2), the probability of hits is significantly equal to the 0.5%, that reflects the validity of the VaR<sub>99.5%</sub>, therefore a good estimation of the risks.

For ARMA(1,1) and ARMA(2,2), the probability of violation is higher than 0.5%, indicating that VaR ARMA model underestimate the risk. It gives a coverage amount lower than risks supported by the company. So, we can conclude that predicting an amount of claims using VaR ACD model at a confidence level 99.5%, can ensure the solvency in the insurance company.
5 Conclusion

This paper propose to analyze the dynamic behavior of the claims amounts in the insurance company. Indeed claims is viewed as intertemporally correlated events. We treat dependence between the claims amounts as a stochastic process. For that we applied different ACD models to specify the best one who capture the autocorrelations for two lines of business Auto Damage and Auto Liability. The standardized claims amounts are tested with the Ljung Box statistic with 10 lags. The statistic suggest that ACD models does a good job of accounting for the temporal dependence in claim amounts. Unfortunately the ACD model may be too restrictive when claims amounts involve some extreme amounts. To solve this problem, we introduced a new specification of a Generalized extreme value version of ACD model, which is denoted the GEVACD model. This specification looks very promising in modeling the line Auto liability, and allowed us to highlight the dynamic of extreme claims amounts.

Within this framework, our model can be thought of as a model explaining the variation of the loss payment process. The fluctuations in payment are explained by the past behavior of the claims (autoregressive part of the model). We selected the GACD(1,1) and GEVACD model respectively for the lines Auto Damage and Auto Liability because they fit the data better and they easily pass the Ljung Box and the LM diagnostic test. Also, these two models maximize, Log likelihood and the information criterion, and depict the lowest errors (RMSE). Comparing our GACD and GEVACD models to ARMA models used in the literature, we remark that our selected models are more adequate than the ARMA models in fitting data. After that, we derive a parametric VaR ACD model to evaluate the coverage amount of claims, and we observe that VaR ACD model provides a good estimation of risks.

Acknowledgement

The authors would like to thank Mr. Frederic Planchet professor at Institute of Financial and Insurance Sciences for useful comments to revise the present work.

References


Annex A

Figure 6: QQplot of standardized claims amounts of the line Auto Damage versus Gamma(1,1.7)

Figure 7: QQplot of standardized claims amounts of the line Auto Liability versus Gamma(1,0.6)
Figure 8: QQplot of standardized claims amounts of the line Auto Damage versus GEV(0.24,1,0.596)

Figure 9: QQplot of standardized claims amounts of the line Auto Damage versus GEV(0.24,1,0.596)
Figure 10: QQplot of Residuals from ARMA(1,1) of the Auto Damage

Figure 11: QQplot of Residuals from ARMA(1,1) of the Auto Liability
Figure 12: QQplot of Residuals from ARMA(2,2) of the Auto Damage

Figure 13: QQplot of Residuals from ARMA(2,2) of the Auto Liability