Jurnal UJMC, Jilid 1, No 1, Hal. 1-10
ISSN : 2460-3333

PEMODELAN *COPULA* CLAYTON UNTUK PREDIKSI KLAIM PADA DATA LONGITUDINAL DENGAN *EXCESS ZEROS*

Anaviroh[1] dan Adhitya Ronnie Effendie[2]

[1]Universitas Islam Darul 'Ulum Lamongan, ana.math07@gmail.com
[2]Universitas Gadjah Mada, adhityaronnie@yahoo.com

Abstract. This papper discuss about longitudinal data models of claim counts with excess-zeros, in which time-dependence of the claim counts is modeled by using a copula function. The copula approach extensively to model the serial dependence of the claim counts in car insurance, to model this serial dependence of the claim counts (between the history and future claims). The maximum likelihood is applied to estimate the parameters of the discrete copula model. A two-step procedure is proposed to estimate the parameters and predict the claim counts of the next period using the estimated parameters.

Keywords*: car insurance, longitudinal data, copula, excess zeros.*

Abstrak. Model longitudinal diaplikasikan untuk pemodelan banyak klaim dengan data yang memiliki nilai nol berlebihan dan dependensi antar waktu klaim dimodelkan menggunakan fungsi *copula*. Pendekatan *copula* digunakan untuk memodelkan serial dependensi banyak klaim asuransi kendaraan bermotor (antara klaim pada waktu yang lalu dan klaim pada waktu yang akan datang). *Maximum likelihood* diaplikasikan untuk mengestimasi parameter dari model *copula* diskrit. Dua langkah yang dilakukan yaitu mencari nilai estimasi parameter dan memprediksi banyak klaim pada periode selanjutnya menggunakan nilai parameter tersebut.

Kata Kunci: *asuransi kendaraan bermotor, data longitudinal, copula, excess zeros.*

1 Pendahuluan

Salah satu asuransi non jiwa yang populer dan banyak diminati masyarakat adalah asuransi kendaraan bermotor. Di berbagai negara suransi kendaraan bermotor merupakan peraih pendapatan total premi yang terbesar [3]. Pada asuransi kendaraan bermotor memungkinkan pemegang polis mengajukan klaim berulang kali dalam satu periode pertanggungan. Banyaknya klaim asuransi merupakan variabel respon diskrit yang diasumsikan berdistribusi Poisson. Pada konsep GLM, metode analisis yang dapat digunakan untuk mengetahui hubungan antara variabel respon yang berdistribusi Poisson dengan variabel penjelasnya dikenal sebagai model regresi Poisson [4]. Pada suatu observasi, banyaknya klaim asuransi yang bernilai nol mungkin muncul dengan jumlah yang banyak, karena pemegang polis bisa saja tidak mengajukan klaim pada beberapa periode pertanggungan. Oleh karena itu, jika terjadi overdispersi akibat adanya terlalu banyak nilai nol (*excess zero*) pada variabel respon maka digunakan distribusi *Zero-Inflated* Poisson (ZIP).

Pada data longitudinal, selain hubungan antara variabel respon dan kovariat yang diperhatikan, juga memperhatikan hubungan dependensi antara hasil observasi yang diperoleh secara berulang untuk individu yang sama [9]. Salah satu struktur dependensi yang tersedia dalam analisis data longitudinal adalah *autoregressive* order satu (AR1). Model ini digunakan untuk mengukur dependensi data yang memiliki ketergantungan terhadap waktu. Model *autoregressive* order satu (AR1) menggunakan informasi respon sebelumnya untuk memodelkan respon sekarang, yang secara umum merupakan prinsip dari model Markov.

Copula adalah suatu fungsi yang dapat menggabungkan beberapa distribusi marginal menjadi distribusi gabungan, dengan asumsi terdapat hubungan dependensi antara distribusi marginal-marginalnya [5]. Metode ini memiliki fleksibilitas, dimana distribusi marginal dari variabel-variabelnya dapat dibedakan atau bahkan dapat diperoleh distribusi gabungan dari distribusi marginal yang tidak diketahui. *Copula* Archimedean merupakan salah satu kelas *copula* yang sangat penting dan sering digunakan aplikasinya pada bidang keuangan dan asuransi. Sesuai model dalam [8], *copula* akan digunakan untuk mendapatkan distribusi kumulatif gabungan antara banyak klaim periode sekarang dengan periode sebelumnya, yang terdapat pada model Markov orde satu, selanjutnya diaplikasikan untuk prediksi.

2 Kajian Teori

2.1 *Generalized Linear Models* (GLM)

Terdapat tiga komponen utama yang membentuk GLM, yaitu asumsi distribusi, komponen sistematik, dan fungsi penghubung (*link function*) [2]. Secara umum, variabel random respon $Y_1,....,Y_m$ dengan $E(Y_i)=\mu_i$ diasumsikan mempunyai fungsi densitas dari keluarga Eksponensial. Komponen sistematik dalam GLM berbentuk prediktor *linear*. Prediktor *linear* menghubungkan dan memberi spesifikasi pengaruh variabel penjelas X_i ke mean dari respon Y_i dalam bentuk $\eta_i = X_i\beta$ yang merupakan kombinasi *linear* antara koefisien regresi dengan kovariat. Fungsi penghubung monoton g sedemikian sehingga $g(\mu_i)=X_i\beta$ merupakan fungsi yang menghubungkan mean respon $\mu_i = E(Y_i \mid X_i)$ dengan kovariat $X_i\beta$. Misalkan $Y_1,....,Y_m$ adalah variabel random independen, suatu fungsi penghubung disebut fungsi penghubung kanonik apabila $g(\mu_i)=\theta$ dengan θ adalah parameter kanonik dalam

$$f(y_i)=\exp\left(\frac{y_i\theta_i-\psi(\theta_i)}{\phi}+c(y_i,\phi)\right) \tag{1}$$

dengan $\psi(.)$ dan $c(.)$ merupakan fungsi yang diketahui, ϕ adalah parameter skala, dan $f(y)$ merupakan fungsi probabilitas variabel random Y yang termasuk dalam keluarga Eksponensial.

2.2 Rantai Markov

Proses stokastik $X = \{X_n, n = 0,1,2 \dots\}$ mempunyai ruang status berupa himpunan berhingga atau himpunan terbilang, yang secara umum dinotasikan sebagai himpunan $\{0,1, \dots\}$. Jika pada waktu n proses tersebut berada pada status i, maka dinotasikan dengan $X_n = i$ [6].

Definisi 1. *Suatu proses stokastik waktu diskrit* $\{X_n, n = 0,1,2 \dots\}$ *disebut memiliki sifat markov jika untuk semua* $i_0, \dots, i_{n-1}, i, j$ *dan semua* $n > 0$ *berlaku*

$$P\{X_{n+1} = j | X_n = i, \dots, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}.$$

Dapat diartikan bahwa probabilitas perilaku tertentu di masa yang akan datang, hanya bergantung kondisi pada satu langkah sebelumnya dan tidak bergantung pada informasi tambahan dimasa lalu.

2.3 *Copula*

Copula adalah suatu fungsi yang menggabungkan beberapa distribusi marginal menjadi distribusi bersama. *Copula* merupakan pendekatan yang berguna untuk memahami dan mendeteksi struktur dependensi variabel random. Konsep *copula* pertama kali diperkenalkan oleh A. Sklar pada tahun 1959. Kelebihan dari pendekatan *copula* adalah distribusi marginalnya tidak harus sama [5].

Definisi 2. *Copula berdimensi n yang dinotasikan dengan* C *adalah fungsi distribusi multivariat* F *dari variabel-variabel random* $X_1, X_2 \dots, X_n$ *dengan distribusi marginalnya* $F_1, F_2 \dots, F_n$ *berdistribusi uniform standar yaitu* $X_i \sim UNIF\ (0,1)$, $i = 1,2, \dots, n$. *Fungsi copula ini merupakan fungsi yang memiliki domain* $[0,1]^n$ *dan range* $[0,1]$, *yang dilambangkan dengan* $C: [0,1]^n \rightarrow [0,1]$.

2.4 *Copula* Archimedean

Copula Archimedean merupakan salah satu kelas dari copula yang spesial, karena beberapa alasan diantaranya adalah *copula* ini mudah dikonstruksikan, banyak variasi keluarga *copula* yang masuk ke dalam kelas ini, dan struktur dependensinya bervariasi. *Copula* Archimedean sering digunakan diberbagai bidang aplikasi, diantaranya pada bidang keuangan dan bidang asuransi. Secara umum, bentuk *copula* Archimedean adalah

$$C(u, v) = \varphi^{[-1]}\big(\varphi(u) + \varphi(v)\big),$$

dengan $0 \leq u, v \leq 1$. Dengan demikian, $C(u, v)$ adalah *copula* Archimedean dan φ merupakan fungsi pembangkit (generator) dari *copula* C dengan $\varphi(0) = \infty$ dan $\varphi(1) = 0$ sehingga $\varphi^{[-1]} = \varphi^{-1}$.

2.4.1 *Copula* Clayton

Copula Archimedean satu parameter dibentuk menggunakan generator $\varphi_\gamma(t)$, dengan *index* parameter γ. Dengan memilih satu fungsi generator, maka akan diperoleh subkelas bagian atau famili dari *copula* Archimedean, diantaranya

subfamili Gumbel, Clayton, Frank, dan lain sebagainya. *Copula* Clayton di definisikan sebagai berikut:
Untuk setiap $u, v \in [0,1]^2$

$$C_\gamma(u,v) = (u^{-\gamma} + v^{-\gamma} - 1)^{-1/\gamma}, \gamma > 0. \quad (2)$$

Dengan fungsi generatornya adalah

$$\varphi_\gamma(t) = (t^{-\gamma} - 1), \quad (3)$$

dan fungsi inversnya adalah

$$\varphi_\gamma{}^{-1}(t) = (t+1)^{-\frac{1}{\gamma}}. \quad (4)$$

2.5 *Generalized Logistic Models*

Regresi logistik merupakan salah satu model statistika yang dapat digunakan untuk menganalisis pola hubungan antara sekumpulan variabel independen dengan suatu variabel dependen bertipe katagoris atau kualitatif. Kategori dari variabel dependen dapat terdiri atas dua kemungkinan nilai (*dichotomus*), seperti ya/tidak, sukses/gagal, dan lainnya, atau lebih dari dua nilai (*polychotomous*), seperti sangat tidak setuju, tidak setuju, setuju dan sangat setuju. Jika variabel dependen terdiri dari dua kategori (bivariat dependen), maka disebut dengan *Generalized Logistic Models*. Dalam [1], bentuk umum dari *Generalized Logistic Models* adalah

$$P(Y_1 = k, Y_2 = l) = \frac{\exp(x_i^T \kappa_{kl})}{\exp(x_i^T \kappa_{00}) + \exp(x_i^T \kappa_{01}) + \exp(x_i^T \kappa_{10}) + \exp(x_i^T \kappa_{11})} \quad (5)$$

dengan $\kappa_{00} = 0, k, l = 0,1$.

3 Pemodelan Dependensi Klaim dengan *Copula*

3.1 Model Spesifikasi

Terdapat sebanyak n polis, yang masing-masing diamati selama Ti tahun. Lamanya interval waktu pada suatu observasi adalah $d_{i,t}$. Variabel random yang menyatakan banyaknya klaim yang terjadi dalam setiap tahun observasi, dinotasikan dengan $N_{i,t}$ dengan $i = 1,2 \dots, n$ dan $t = 1, \dots, Ti$. Informasi lain yang digunakan dalam model adalah variabel penjelas, yang dinotasikan dengan $\boldsymbol{z_{i,t}}$. Secara umum pemodelan banyak klaim $N_{i,t}$, untuk pemegang polis i, pada waktu t mengikuti distribusi Poisson dengan rata-rata frekuensi klaim $\lambda_{i,t}$:

$$\Pr(N_{i,t} = n_{i,t}) = \frac{\lambda_{i,t}{}^{n_{i,t}}}{n_{i,t}!} \exp(-\lambda_{i,t}) \quad (6)$$

dengan $\lambda_{i,t} = d_{i,t} exp(z_{i,t}^T \beta)$, menunjukkan hubungan antara banyak klaim dengan informasi kovariat.

3.2 Struktur Dependensi *Autoregressive* Order 1 *(*AR 1*)*

Sesuai [7] struktur dependensi *autoregressive* order 1 pada data longitudinal dimodelkan sebagai model markov orde pertama, dengan modelnya sesuai dengan Definisi 2.1, yaitu

$$Pr\left(N_{i,t} = n_{i,t} \middle| n_{i,t-1}, \dots, n_{i,1}\right) = Pr\left(N_{i,t} = n_{i,t} \middle| n_{i,t-1}\right),$$

yang dapat dituliskan sebagai:

$$Pr\left(N_{i,t} = n_{i,t} \middle| n_{i,t-1}\right) = Pr\left(N_{i,t} \leq n_{i,t} \middle| n_{i,t-1}\right) - Pr\left(N_{i,t} \leq n_{i,t} - 1 \middle| n_{i,t-1}\right) \tag{7}$$

Untuk mengetahui nilai probabilitas bersyarat pada Persamaan 7, maka harus diketahui nilai parameter-parameter yang terdapat dalam model. Parameter diperoleh dari estimasi dengan metode *Maximum Likelihood*. Karena setiap polis diobservasi selama T_i tahun, maka fungsi likelihood untuk setiap polis i adalah

$$\Pr\left(N_{i,1} = n_{i,1}, \dots, N_{i,t} = n_{i,t}\right)$$
$$= Pr\left(N_{i,1} = n_{i,1}\right) \prod_{j=2}^{T_i} Pr\left(N_{i,j} = n_{i,j} | N_{i,j-1} = n_{i,j-1}\right). \tag{8}$$

Karena pemegang polis memungkinkan tidak mengajukan klaim pada dua tahun berturut-turut, maka Persamaan 8 mengakomodasi nilai nol yang berlebihan. Akibatnya, hubungan antara banyak klaim dengan kovariat dari $Pr\left(N_{i,1} = n_{i,1}\right)$ pada Persamaan 8, dapat dimodelkan dengan *Zero-Inflated* Poisson (ZIP). Fungsi densitas dari distribusi ZIP adalah

$$P\left(N_{i,1} = n_{i,1}\right) = \begin{cases} \omega_{i,1} + \left(1 - \omega_{i,1}\right) e^{-\lambda_{i,1}}, & n_{i,1} = 0 \\ \dfrac{\left(1 - \omega_{i,1}\right) e^{-\lambda_{i,1}} \lambda_{i,1}{}^{n_{i,1}}}{n_{i,1}!}, & n_{i,1} > 0, 0 \leq \omega_{i,1} \leq 1 \end{cases} \tag{9}$$

dengan $\lambda_{i,1} = d_{i,1} e^{\left(\boldsymbol{z}_{i,1}^{\boldsymbol{T}} \boldsymbol{\beta}\right)}$ dan $ln\left[\frac{\omega_{i,1}}{1-\omega_{i,1}}\right] = \left(\boldsymbol{z}_{i,1}^{\boldsymbol{T}} \boldsymbol{\alpha}\right)$.

Terdapat empat bentuk, dari definisi likelihood pada Persamaan 8, yang masing-masing serupa satu sama lain. Sebagai contoh, CDF gabungan dari $\left(N_{i,t-1} \leq n_{i,t-1}, N_{i,t} \leq n_{i,t}\right)$ dapat dinyatakan sebagai

$$Pr\left(N_{i,t-1} \leq n_{t-1}, N_{i,t} \leq n_t\right) = p_{00}^{i,t} + p_{01}^{i,t} H_{01}^{i,t}(n_{i,t}) + p_{10}^{i,t} H_{10}^{i,t}(n_{i,t-1}) + p_{11}^{i,t} H_{11}^{i,t}(n_{i,t-1}, n_{i,t}) \tag{10}$$

dengan

$$H_{01}^{i,t}(n_{i,t}) = \Pr\left(N_{i,t} \leq n_t \middle| N_{i,t-1} = 0, N_{i,t} > 0\right) \tag{11}$$
$$H_{10}^{i,t}(n_{i,t}) = \Pr\left(N_{i,t-1} \leq n_{t-1} \middle| N_{i,t} = 0, N_{i,t-1} > 0\right) \tag{12}$$
$$H_{11}^{i,t}(n_{i,t}) = \Pr\left(N_{i,t-1} \leq n_{t-1}, N_{i,t} \leq n_t \middle| N_{i,t-1} > 0, N_{i,t} > 0\right) \tag{13}$$
$$p_{00}^{i,t} = \Pr\left(N_{i,t-1} = 0, N_{i,t} = 0\right) \tag{14}$$
$$p_{01}^{i,t} = \Pr\left(N_{i,t-1} = 0, N_{i,t} > 0\right) \tag{15}$$
$$p_{10}^{i,t} = \Pr\left(N_{i,t-1} > 0, N_{i,t} = 0\right) \tag{16}$$
$$p_{11}^{i,t} = \Pr\left(N_{i,t-1} > 0, N_{i,t} > 0\right) \tag{17}$$

Banyak klaim pada waktu *t* mengikuti distribusi Poisson dengan rata-rata $\lambda_{i,t}$, dan pada waktu (*t-1*) berdistribusi Poisson dengan rata-rata $\lambda_{i,t-1}$, maka Persamaan 11 dan Persamaan 12 dapat dinyatakan sebagai

$$\mathrm{H}_{01}^{\mathrm{it}}(y) = \frac{F_{i,t}(y) - \exp\left(-d_{i,t}e^{(\boldsymbol{z}_{i,t}^{T}\boldsymbol{\beta})}\right)}{1 - \exp\left(-d_{i,t}e^{(\boldsymbol{z}_{i,t}^{T}\boldsymbol{\beta})}\right)}, y \geq 0 \tag{18}$$

dan

$$\mathrm{H}_{10}^{\mathrm{it}}(x) = \frac{G_{i,t-1}(x) - \exp\left(-d_{i,t-1}e^{\left(z_{i,t-1}^{T}\beta\right)}\right)}{1 - \exp\left(-d_{i,t-1}e^{\left(z_{i,t-1}^{T}\beta\right)}\right)}, x \geq 0. \tag{19}$$

CDF gabungan $\mathrm{H}_{11}^{\mathrm{i,t}}(x,y)$ dengan marginalnya $\tilde{G}_{i,t-1}(x)$ dan $\tilde{F}_{i,t}(y)$, dapat diperoleh dengan fungsi copula $C_{\gamma}(.,.)$, yaitu:

$$\mathrm{H}_{11}^{\mathrm{i,t}}(x,y) = C_{\gamma}\left(\tilde{G}_{i,t-1}(x),\ \tilde{F}_{i,t}(y)\right), \mathrm{x} \geq 0, \mathrm{y} \geq 0 \tag{20}$$

dengan $\tilde{G}_{i,t-1}(x)$ adalah distribusi *modified* Poisson yang didefinisikan pada Persamaan 19, dan $\tilde{F}_{i,t}(y)$ didefinisikan pada Persamaan 20. Sementara $G_{i,t-1}(x)$ dan $F_{i,t}(y)$, masing-masing adalah CDF Poisson dengan rata-rata yang dihubungkan dengan kovariat, yaitu

$$\lambda_{i,t} = d_{i,t} \exp\,(z_{i,t}^{T}\beta) \text{ dan } \lambda_{i,t-1} = d_{i,t-1} \exp\,(z_{i,t-1}^{T}\beta).$$

Namun, $p_{00}^{i,t}, p_{01}^{i,t}, p_{10}^{i,t}, p_{11}^{i,t}$ dapat direparameter menggunakan model *Generalized Logistic Model* [1], yaitu

$$p_{kl}^{i,t} = \frac{\exp(x_i^T\kappa_{kl})}{\exp(x_i^T\kappa_{00}) + \exp(x_i^T\kappa_{01}) + \exp(x_i^T\kappa_{10}) + \exp(x_i^T\kappa_{11})} \tag{21}$$

dengan $\kappa_{00} = 0, k, l = 0{,}1$ merupakan koefisien kovariat yang harus diestimasi.

3.3 Estimasi Parameter dan Prediksi

Misalkan $N_{i,t}$ menyatakan banyaknya klaim yang dilaporkan pemegang polis ke-i pada tahun-*t* dan $z_{i,t}$ adalah vektor kovariat dari pemegang polis ke-i pada tahun t, dengan; $i = 1,2, \ldots n$ dan $t = 1,2, \ldots T_i$. Selanjutnya, didefinisikan status indikator kejadian sebagai berikut:

$$c_{00}^{i,t} = I(N_{i,t-1} = 0, N_{i,t} = 0),\ c_{01}^{i,t} = I(N_{i,t-1} = 0, N_{i,t} > 0)$$

$$c_{01}^{i,t} = I(N_{i,t-1} > 0, N_{i,t} = 0),\ \ c_{11}^{i,t} = I(N_{i,t-1} > 0, N_{i,t} = 0)$$

dengan *I(A)* menotasikan indikator dari kejadian A, sehingga $c_{00}^{i,t} + c_{01}^{i,t} + c_{10}^{i,t} + c_{11}^{i,t} = 1$.

Misalkan $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ dan $\mathcal{L}_4$ adalah kontribusi likelihood untuk probabilitas kumulatif gabungan masing-masing dari $\Pr(N_{i,t-1} \leq n_{i,t-1}, N_{i,t} \leq n_{i,t})$, $\Pr(N_{i,t-1} \leq n_{i,t-1} - 1, N_{i,t} \leq n_{i,t})$, $\Pr(N_{i,t-1} \leq n_{i,t-1}, N_{i,t} \leq n_{i,t} - 1)$, dan $\Pr(N_{i,t-1} \leq n_{i,t-1} - 1, N_{i,t} \leq n_{i,t} - 1)$. Berdasarkan indikator kejadian maka $\mathcal{L}_1$dapat dituliskan sebagai

$$\mathcal{L}_1 = \left[p_{00}^{i,t}\right]^{c_{00}^{i,t}} \left[p_{01}^{i,t} H_{01}^{i,t}(n_{i,t})\right]^{c_{01}^{i,t}} \left[p_{10}^{i,t} H_{10}^{i,t}(n_{i,t-1})\right]^{c_{10}^{i,t}} \left[p_{11}^{i,t} H_{11}^{i,t}(n_{i,t-1}, n_{i,t})\right]^{c_{11}^{i,t}}. \tag{22}$$

Sama halnya untuk $\mathcal{L}_2, \mathcal{L}_3$ dan $\mathcal{L}_4$, diperoleh dengan mengganti $\left(n_{i,t-1}, n_{i,t}\right)$ dengan $\left(n_{i,t-1}-1, n_{i,t}\right)$ untuk $\mathcal{L}_2$, $\left(n_{i,t-1}, n_{i,t}-1\right)$ untuk $\mathcal{L}_3$, dan $\left(n_{i,t-1}-1, n_{i,t}-1\right)$ untuk $\mathcal{L}_4$.

Kontribusi likelihood secara keseluruhan dari *(i,t)* (pemegang polis –*i*, pada periode t), dinotasikan dengan $(\mathcal{L}_1-\mathcal{L}_2-\mathcal{L}_3+\mathcal{L}_4)_{i,t}$. Berdasarkan Persamaan 8, maka likelihood untuk pemegang polis *i* adalah

$$\Pr\left(N_{i,1}=n_{i,1},\dots,N_{i,T_i}=n_{i,T_i}\right)$$
$$=\Pr\left(N_{i,1}=n_{i,1}\right)\prod_{t=2}^{T_i}\frac{(\mathcal{L}_1-\mathcal{L}_2-\mathcal{L}_3+\mathcal{L}_4)_{i,t}}{\Pr\left(N_{i,t-1}=n_{i,t-1}\right)} \tag{23}$$

dengan nilai $Pr\left(N_{i,1}=n_{i,1}\right)$ mengikuti *Zero-Inflated Poisson* (ZIP), yang didefinisikan pada Persamaan 9.

Nilai prediksi frekuensi klaim dapat diperoleh setelah nilai parameter-parameter yang terdapat dalam model diestimasi. Model prediksi dengan copula yaitu

$$\Pr\left(n_{i,T+1}\mid n_{i1},\dots,n_{iT}\right)=\frac{\Pr_i\left(n_{i1},\dots,n_{i,T},n_{i,T+1}\right)}{\Pr_i\left(n_{i1},\dots,n_{i,T}\right)}$$

Berdasarkan sifat Markov diperoleh

$$\Pr\left(n_{i,T+1}\mid n_{i1},\dots,n_{iT}\right)=\frac{\Pr_i(n_{i,1})\prod_{t=2}^{T+1}\Pr_i\left(n_{it}\mid n_{it-1}\right)}{\Pr_i(n_{i,1})\prod_{t=2}^{T}\Pr_i\left(n_{it}\mid n_{it-1}\right)}$$
$$=\frac{\Pr_i(n_{i,T+1},n_{i,T})}{\Pr_i(n_{i,T})} \tag{24}$$

dengan

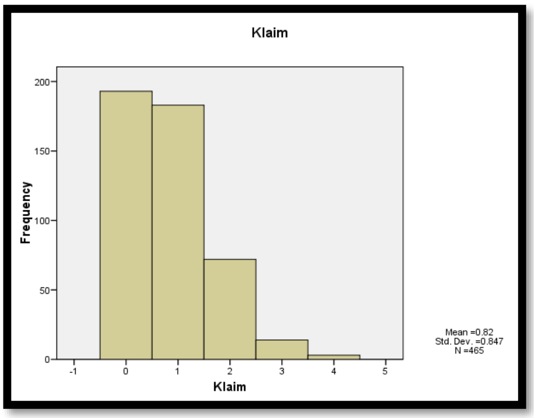
$$\Pr_i(n_{i,T+1},n_{i,T})$$
$$=\Pr_i(N_{i,T+1}\le n_{i,T+1},N_{i,T}\le n_{i,T})-\Pr_i(N_{i,T+1}\le n_{i,T+1}-1,N_{i,T}\le n_{i,T})$$
$$-\Pr_i(N_{i,T+1}\le n_{i,T+1},N_{i,T}\le n_{i,T}-1)+\Pr_i(N_{i,T+1}\le n_{i,T+1}-1,N_{i,T}\le n_{i,T}-1).$$

Jadi model prediksi dari banyak klaim tersebut merupakan penurunan langsung dari perumusan model Persamaan 7.

4 Studi Kasus

4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data klaim asuransi kendaraan bermotor yang diperoleh dari perusahaan asuransi "XYZ". Data diperoleh dari hasil pengamatan pada tahun 2009 sampai 2011, dari pengamatan terhadap 155 polis sehingga terdapat 465 data yang diamati. Informasi yang digunakan adalah banyaknya klaim yang dilaporkan oleh setiap pemegang polis yang digunakan sebagai variabel respon dan usia kendaraan yang dipandang sebagai variabel penjelas. Data klaim yang diperoleh disajikan pada histogram berikut ini.



Gambar 1. Histogram banyak klaim asuransi kendaraan bermotor

Dari histogram, terlihat bahwa data dengan nilai nol lebih banyak dibandingkan nilai yang lain, atau data dengan *excess zeros*. Statistik deskriptif frekuensi klaim dan usia kendaraan ditampilkan pada tabel berikut ini.

Tabel 1. Statistik deskriptif frekuensi klaim dan usia kendaraan

	Frekuensi klaim				**Usia kendaraan**			
Tahun	2009	2010	2011	Total	2009	2010	2011	Total
Mean	0,70	0,74	1,02	0,82	2,22	3,21	4,21	3,22
Minimum	0	0	0	0	1	2	3	2
Maksimum	4	3	4	4	9	10	11	11
Modus	0	1	1	0	1	2	3	3
Std. Deviasi	0,809	0,737	0,950	0,847	1,814	1,820	1,820	1,988

4.2 Estimasi Parameter Model *Copula* dan Prediksi

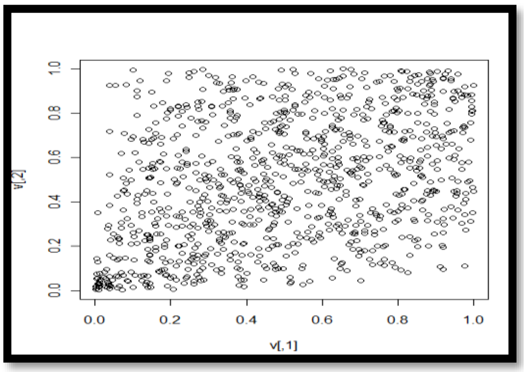
Analisis dilakukan dengan memperhatikan hubungan antara banyak klaim dengan usia kendaraan yang diasuransikan, serta hubungan banyak klaim yang diajukan pemegang polis antar waktu. Hubungan antara banyak klaim dengan usia kendaraan dicari dengan menggunakan konsep GLM, sedangkan hubungan antara banyak klaim pada tahun sekarang dengan tahun sebelumnya dimodelkan dengan *copula*.

Berdasarkan likelihoood yang diperoleh pada Persamaan 23, model bergantung pada parameter-parameter, yaitu

$$v' = [\beta, \kappa_{01}, \kappa_{10}, \kappa_{11}, \gamma].$$

Hasil estimasi parameter model *copula* dengan *software* R adalah $\beta = -0{,}019013, \kappa_{01} = 0{,}154846$ $\kappa_{10} = 0{,}225618$, $\kappa_{11} = 0{,}376208$, dan γ =0,746566. Parameter γ merupakan parameter dari *copula* Clayton, yang digunakan untuk mendapatkan CDF gabungan dari banyak klaim pada tahun 2009 dengan banyak klaim pada tahun 2010, dan banyak klaim pada tahun 2010 dengan banyak klaim pada tahun 2011. Dari parameter tersebut, akan disimulasikan 1000 data,untuk mengetahui karakteristik dari *copula* Clayton dengan parameter

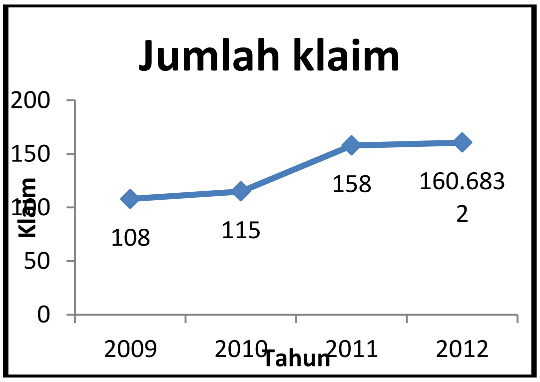
0,753429, yang diperoleh dari data. Hasil simulasi *copula* Clayton untuk 1000 data dengan parameter 0,753429 sebagai berikut



Gambar 2. *Scatter plot copula* Clayton

Pada Gambar 2, terlihat plot cenderung menyebar, tetapi ada sedikit plot yang terkonsentrasi pada ujung kiri bawah *scatter*. Jadi, karakteristik *copula* Clayton dengan parameter sebesar 0,753429 sesuai jika digunakan untuk memodelkan data dengan plot yang menyebar, namun terkonsentrasi pada ujung kiri bawah *scatter*.

Prediksi dapat diperoleh dengan mensubtitusikan nilai parameter yang diperoleh ke model prediksi pada Persamaan 24. Ekspektasi jumlah klaim yang akan diajukan pada tahun 2012, berdasarkan nilai parameter dari data adalah 160, 6832 klaim. Grafik prediksi jumlah klaim dapat dilihat sebagai berikut



Gambar 3. Grafik prediksi jumlah klaim

5 Kesimpulan

Berdasarkan uraian dapat disimpulkan beberapa hal sebagai berikut:

1. Salah satu struktur dependensi yang tersedia dalam analisis data longitudinal adalah AR 1. Model *AR 1* memasukan informasi respon sebelumnya untuk memodelkan respon sekarang yang secara umum merupakan prinsip dari model Markov orde satu.
2. Data pada setiap periode observasi merupakan data yang memiliki nilai nol berlebihan (*excess zeros*), sehingga sebagai asumsi awal data dimodelkan dengan distribusi *Zero Inflated* Poisson (ZIP), yaitu suatu distribusi yang digunakan untuk memodelkan variabel respon yang berupa data diskret dan mengandung nilai nol yang berlebihan.

3. Pada analisis data longitudinal dengan *excess zeros*, *copula* digunakan untuk mendapatkan fungsi distribusi kumulatif gabungan antara hasil observasi periode sekarang dengan periode sebelumnya.
4. Model yang dikonstruksikan merupakan suatu fungsi yang mengandung parameter, sehingga perlu diestimasi untuk memperoleh nilai parameternya. Metode estimasi parameter yang digunakan adalah pendekatan maksimum *likelihood*.
5. Nilai estimasi parameter yang diperoleh digunakan untuk mencari prediksi satu periode ke depan berdasarkan data observasi yang ada.

Daftar Pustaka

[1] Cox, D.R., dan Snell, E.J. 1989. *Analysis of Binary Data Second Ed*. Chapman and Hall. London.

[2] Jong, P.D., dan Heller, G. Z. 2008. *Generalized Linear Models for Insurance Data.* Cambridge University Press. Cambridge.

[3] Kaas, R., Goovaerts, M., Dhaene, J. , dan Denuit, M. 2001. *Modern Actuarial Risk Theory*. Kluwer Academic Publishers. Boston.

[4] Lambert, D. 1992. Zero Inflated Poisson Regression an Application to Defects in Manufacturing. *Technometri*. 34.

[5] Nelsen, B. 2006. *An Introduction to Copula (2nd ed.)*. Springer. New York.

[6] Ross, S.M. 1996. *Stochastic Processes second edition*. John Wiley & Sons. New York

[7] Vandenhende, F., dan Lambert, P. 2000. *Modeling Repeated Ordered Categorial Data Using Copula.* Discussion Paper 00-25. Institute of Statistics, University Chatolique de Louvain. Belgium.

[8] Zhao, X., dan Zhou, X. 2012. Copula Models for Insurance Claim Numbers with Excess Zeros and Time Dependence. *Insurance: Mathematics and Economic*. 50:191-199.

[9] Zeger, K.Y, Liang P.S., dan Albert, S. 1988. Models for Longitudinal Data: A Generalized Estimating Equation Approuach. *Biometric*. 44: 1049-1060.