GENERALIZED MULTILEVEL LINEAR MODEL DENGAN PENDEKATAN BAYESIAN UNTUK PEMODELAN DATA PENGELUARAN PERKAPITA RUMAHTANGGA

Azka Ubaidillah[1], Anang Kurnia[2], Kusman Sadik[2]

[1]Politeknik Statistika STIS, Jakarta
[2]Departemen Statistika, Institut Pertanian Bogor, Bogor
e-mail: [1]azka@stis.ac.id, [2]akstk29@gmail.com, [2]kusmansadik@gmail.com

Abstrak

Data pengeluaran perkapita rumahtangga merupakan salah satu informasi penting sebagai pendekatan untuk mengukur tingkat kemakmuran dan kesejahteraan di suatu daerah. Data tersebut sangat diperlukan oleh pemerintah baik di pusat maupun daerah dalam merumuskan, melaksanakan dan mengevaluasi pelaksanaan pembangunan. Penelitian ini akan menganalisis model yang tepat untuk pemodelan data pengeluaran perkapita rumahtangga yang memperhitungkan kekhususan data BPS yang memiliki struktur hirarki dan pola distribusi data yang memiliki karakteristik skewed kanan. Pemodelan dilakukan dengan menggunakan distribusi Log-normal tiga parameter (LN3P) dan Log-logistik tiga parameter (LL3P) dengan struktur satu tingkat (unilevel) dan dua tingkat (multilevel). Proses pendugaan parameter dilakukan dengan metode Markov Chain Monte Carlo (MCMC) dan algoritma Gibbs Sampling. Hasil penelitian menunjukkan bahwa pada model unilevel, model LL3P lebih baik dari model LN3P. Sedangkan pada model multilevel, model LN3P lebih baik dari model LL3P. Hasil penelitian juga menunjukkan model terbaik untuk pemodelan data pengeluaran perkapita rumahtangga adalah model multilevel LN3P dengan intercept sebagai komponen berhirarki dengan nilai Deviance Information Criterion (DIC) terkecil.

Kata kunci : *Generalized Multilevel Linear Model*, LL3P, LN3P, MCMC, Pengeluaran perkapita rumahtangga.

Abstract

Household per capita expenditure data is one of the important information as an approach to measure the level of prosperity in an area. Such data is needed by the government, both at the central and regional levels in formulating, implementing and evaluating the implementation of development programs. This research is aimed at modeling the household per capita expenditure data which takes into account the specificity of BPS data which has a hierarchical structure, and data distribution pattern which has the right skewed characteristic. The modeling is done by using the three parameters of Log-normal distribution (LN3P) and the three parameters of Log-logistics (LL3P) with a single level (unilevel) and two levels (multilevel) structure. The parameter estimation process is done by Markov Chain Monte Carlo (MCMC) method and Gibbs Sampling algorithm. The results showed that on the unilevel model, the LL3P model is better than the LN3P model. While in multilevel model, LN3P model is better than LL3P model. The results also show that the best model for modeling household per capita expenditure data is the LN3P multilevel model with the smallest Deviance Information Criterion (DIC) value.

Keywords*: Generalized Multilevel Linear Model, LL3P, LN3P, MCMC, Household per capita expenditure.*

PENDAHULUAN

Salah satu tujuan Negara Republik Indonesia yang sekaligus menjadi amanat konstitusi sebagaimana termaktub dalam Pembukaan Undang-Undang Dasar 1945 adalah memajukan kesejahteraan umum. Untuk mewujudkan tujuan tersebut maka pemerintah melakukan perumusan perencanaan, melaksanakan pembangunan dan secara berkesinambungan melakukan evaluasi atas pelaksanaan pembangunan. Keberhasilan pembangunan secara umum diukur dari tingkat kemakmuran dan kesejateraan rakyat. Salah satu pendekatan pengukuran tingkat kemakmuran dan kesejahteraan rakyat adalah pengeluaran perkapita rumahtangga.

Pada dasarnya faktor yang berpengaruh dalam masalah kemakmuran dan kesejahteraan dapat dikategorikan dalam dua hal pokok yaitu paradigma perilaku dan paradigma kebijakan (Atika dan Pirmansyah, 2011). Paradigma perilaku terkait dengan upaya masing-masing individu dan rumahtangga dalam mencapai kesejahteraan. Sedangkan paradigma kebijakan terkait dengan kondisi ekonomi, politik dan kebijakan pemerintah. Hal ini menunjukkan bahwa kondisi di luar rumahtangga itu sendiri juga mempengaruhi perbedaan tingkat kesejahteraan. Dengan demikian tingkat kesejahteraan rumahtangga di suatu wilayah dipengaruhi oleh faktor internal dan eksternal rumahtangga tersebut.

Pada umumnya, data di bidang sosial seperti data pengeluaran perkapita rumahtangga memiliki struktur data yang berhirarki dimana unit-unit di tingkat yang lebih rendah, yaitu rumahtanga tersarang (*nested*) atau terkelompok dalam unit-unit di tingkat yang lebih tinggi yaitu wilayah (desa/kelurahan/dsb). Oleh karena itu pemodelan pengeluaran perkapita rumahtangga hendaknya mempertimbangkan kombinasi antara karakteristik rumahtangga dan karakteristik wilayahnya.

Keterbatasan analisis statistik klasik untuk data berstruktur hirarki adalah tidak diperhitungkannya struktur hirarki data. Teknik yang biasa digunakan adalah teknik agregasi, disagregasi dan pemodelan regresi menurut kelompok (Goldstein, 1995; Raudenbush dan Byrk, 2002; De Leeuw dan Meijer, 2008). Analisis dengan metode klasik hanya dilakukan pada salah satu tingkatan data yaitu tingkat individu dengan menggunakan teknik disagregasi, atau di tingkat kelompok dengan menggunakan teknik agregasi. Namun demikian, teknik klasik tersebut akan sangat berpengaruh dari sisi metodologi dan statistiknya (De Leeuw dan Meijer, 2008).

Menurut Goldstein (1995) dan Hox (1995), penggunaan model *multilevel* untuk data bestruktur hirarki memiliki beberapa kelebihan. Model *multilevel* dapat digunakan untuk menganalisis informasi dari beberapa tingkatan yang berbeda menjadi satu analisis statistik. Model *multilevel* memperhitungkan pengaruh variasi setiap tingkat data terhadap variasi respon. Hal ini memungkinkan peneliti untuk mengetahui variasi di setiap tingkatan data terhadap variasi respon.

Data pengeluaran rumahtangga memiliki karakteristik khusus dengan nilai yang selalu positif dan memiliki frekuensi tinggi untuk pengeluaran perkapita rumahtangga golongan menengah ke bawah, sedangkan untuk golongan rumahtangga menengah ke atas memiliki frekuensi yang relatif rendah. Distribusi yang sesuai untuk pola data tersebut dan banyak digunakan untuk analisis di bidang sosial ekonomi adalah distribusi Log-normal dan distribusi Log-logistik (Johnson, Kotz dan Balakrishnan, 1995b). Sesuai dengan karakteristik pengeluaran perkapita rumahtangga yang tidak pernah nol, maka digunakan distribusi Log-normal dan Log-logistik yang diperluas dengan penambahan satu parameter yang selanjutnya dikenal dengan distribusi Log-normal tiga parameter dan distribusi Log-logistik tiga parameter (Ismartini dkk, 2012).

Beberapa penelitian yang menggunakan konsep *multilevel* dalam analisis penelitiannya antara lain dilakukan oleh Ha dkk (2005) yang melakukan penelitian mengenai pemodelan dengan menggunakan metode *multilevel mixed* *linear* untuk data survival CGD (Chronic Granulomatous Deseas). Zimmer dkk (2010) menggunakan model hirarki linier untuk memperkirakan perbedaan pedesaan/perkotaan dalam status fungsional transisi pada masyarakat Cina yang berumur 55 tahun ke atas selama 2 tahun dan memperkirakan derajat dimana tingkat sosial ekonomi individu dan komunitas merupakan penentu dalam menjelaskan perbedaan status tersebut. Sementara Anderson dan Wells (2010) menggunakan pendekatan Bayesian hirarki regresi pada pemodelan hirarki untuk data kelompok longitudinal penelitian hukum empiris.

Keterbatasan model hirarki sederhana seperti hirarki linier klasik dengan pendekatan *Maximum Likelihood* adalah apabila jumlah sampel yang kecil dan tidak setimbang terdapat pada model dengan tingkat yang lebih tinggi, maka inferensia statistiknya ada kemungkinan menjadi tidak dapat dipercaya (Raudenbush dan Bryk, 2002). Sehingga untuk mengatasi permasalahan tersebut maka digunakan pendekatan Bayesian pada model hirarki atau *Hierarchical Bayesian* (HB) dimana menurut Raudenbush dan Bryk (2002), model HB memiliki keuntungan yaitu mampu mengatasi permasalahan pemodelan hirarki untuk jumlah data yang sedikit dan tidak seimbang baik pada tingkat bawah maupun pada tingkat yang lebih tinggi.

Pada tahun 2012, Ismartini dkk mengembangkan model linier hirarki dengan pendekatan Bayesian untuk pemodelan data pengeluaran perkapita rumahtangga berbasis distribusi Log-normal tiga parameter (LN3P) dan distribusi Log-logistik tiga parameter (LL3P). Namun penelitian yang dilakukan oleh Ismartini belum dilakukan kajian yang lebih detail dari sisi konsep *Generalized Linear Model* (GLM).

Tujuan makalah ini yaitu memodelkan pengeluaran perkapita rumahtangga di Kota Jambi dengan menggunakan model *unilevel* dan *multilevel* berbasis pada distribusi LN3P dan distribusi LL3P. Proses pemodelan dimulai dengan membentuk model paling sederhana, yaitu model *unilevel* dengan tanpa kovariat, sampai model yang kompleks yaitu model *multilevel*. Kemudian model-model tersebut dibandingkan untuk diperoleh model terbaik dengan menggunakan kriteria *Deviance Information Criterion* (DIC).

TINJAUAN PUSTAKA

1. Model *Multilevel* Linier

Model *multilevel* merupakan model regresi yang mengakomodasi adanya struktur data hirarki atau data bersarang. Dalam sturktur data hirarki ini, variabel respon diukur pada level mikro saja, sedangkan variabel prediktor diukur baik di level mikro maupun di level makro (Goldstein, 1995; Hox, 2010). Sesuai dengan konsep hirarki, maka model *multilevel* menghasilkan persamaan regresi bertingkat, yaitu koefisien regresi di tingkat lebih rendah diregresikan lagi di tingkat yang lebih tinggi. Ismartini dkk. (2012) pada penelitiannya tentang pemodelan pengeluaran per kapita rumahtangga, menjelaskan model *multilevel* pada level mikro menggambarkan hubungan antara variabel respon (pengeluaran per kapita rumahtangga) dengan beberapa variabel prediktor yang merupakan karakteristik rumahtangga. Sedangkan model pada level makro menjelaskan hubungan antara koefisien model pada level mikro dengan karakteristik wilayah.

Persamaan model mikro untuk setiap kelompok adalah sebagai berikut:

$$Y_{ik} = \beta_{0k} + \beta_{1k}X_{1ik} + \cdots + \beta_{Pk}X_{Pik} + e_{ik}, \qquad (1)$$

dimana i = 1,2, ...,n_k dan k = 1,2, ...,K, atau jika dinyatakan dalam bentuk matriks menjadi:

$$\mathbf{y}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \mathbf{e}_k \qquad (2)$$

dengan:

$\mathbf{y}_k = [y_{1k} \quad y_{2k} \quad \cdots \quad y_{n_k k}]^T$,

$$\mathbf{X}_k = \begin{bmatrix} 1 & x_{11k} & x_{21k} & \cdots & x_{P1k} \\ 1 & x_{12k} & x_{22k} & \cdots & x_{P2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_k k} & x_{2n_k k} & \cdots & x_{Pn_k k} \end{bmatrix},$$

$\boldsymbol{\beta}_k = [\beta_{0k} \quad \beta_{1k} \quad \cdots \quad \beta_{Pk}]^T$,

$\mathbf{e}_k = [e_{1k} \quad e_{2k} \quad \cdots \quad e_{n_k k}]^T$,

Pembentukan model makro dilakukan dengan menjadikan koefisien regresi pada model mikro, β_{rk} , r = 0,1,2, ..., P dalam persamaan (2) sebagai variabel respon yang dijelaskan oleh model makro. Bentuk persamaan makro adalah sebagai berikut:

$$\beta_{rk} = \gamma_{0r} + \gamma_{1r} Z_{1k} + \cdots + \gamma_{Lr} Z_{Lk} + u_{rk} \ ,$$

atau jika dinyatakan dalam bentuk matriks adalah

$$\boldsymbol{\beta}_r = \mathbf{Z}\boldsymbol{\gamma}_r + \mathbf{u}_r \qquad (3)$$

dengan:

$\boldsymbol{\beta}_r = [\beta_{r1} \quad \beta_{r2} \quad \cdots \quad \beta_{rK}]^T$,

$$\mathbf{Z} = \begin{bmatrix} 1 & z_{11} & z_{21} & \cdots & z_{L1} \\ 1 & z_{12} & z_{22} & \cdots & z_{L2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_{1K} & z_{2K} & \cdots & z_{LK} \end{bmatrix},$$

$\boldsymbol{\gamma}_r = [\gamma_{0r} \quad \gamma_{1r} \quad \cdots \quad \gamma_{Lr}]^T$,

$\mathbf{u}_r = [u_{r1} \quad u_{r2} \quad \cdots \quad u_{rK}]^T$.

Asumsi model hirarki dua tingkat tersebut adalah: (Hox, 1995; De Leeuw dan Kreft, 2006)

a. Level mikro
   i. Residual bersifat independen atau $Cov(e_{ik}, e_{i^*k}) = 0 \ , i \neq i^*$
   ii. e_{ik} berdistribusi normal
   iii. $E(\mathbf{e}_k) = \mathbf{0}$
   iv. $Var(\mathbf{e}_k) = \sigma^2_{[e]k} \mathbf{I}_{n_k}$

   dengan $\sigma^2_{[e]k}$ adalah *varians* residual model mikro kelompok ke-k.

b. Level makro
   i. Residual bersifat independen atau $Cov(u_{rk}, u_{r^*k^*}) = 0 \ , \ r \neq r^*$ dan $k \neq k^*$
   ii. u_{rk} berdistribusi normal
   iii. $E(\mathbf{u}_r) = \mathbf{0}$
   iv. $Var(\mathbf{u}_r) = \mathbf{T}_r$, dengan:

$$\mathbf{T}_k = \begin{bmatrix} \sigma^2_{[u]r11} & \sigma^2_{[u]r21} & \cdots & \sigma^2_{[u]rK1} \\ \sigma^2_{[u]r12} & \sigma^2_{[u]r22} & \cdots & \sigma^2_{[u]rK2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{[u]r1K} & \sigma^2_{[u]r2K} & \cdots & \sigma^2_{[u]rKK} \end{bmatrix}$$

   dengan $\sigma^2_{[u]rkk}$ adalah *varians* residual model makro untuk koefisien regresi ke-r pada kelompok ke-k.

c. Antar level
   Residual level mikro bersifat independen terhadap residual level makro atau $Cov(e_{ik}, u_{rk}) = 0$.

Persamaan gabungan dari persamaan (2) dan (3) adalah: (Raudenbush dan Byrk, 2002; De Leeuw dan Kreft, 2006; Goldstein, 1995)

$$\mathbf{y}_k = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k + \mathbf{e}_k \qquad (4)$$

dengan:

$\mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma}$ adalah komponen tetap (deterministik) dan $\mathbf{X}_k \mathbf{u}_k + \mathbf{e}_k$ adalah komponen random (stokastik) dalam model *multilevel*,

$E(\mathbf{y}_k) = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma}$,

$Var(\mathbf{y}_k) = \mathbf{X}_k \mathbf{T} \, \mathbf{X}_k^T + \sigma_k^2 \mathbf{I}_{n_k}$.

2. *Generalized Linear Model*

Generalized Linear Model (GLM) merupakan perluasan dari *Linear Model*. Terdapat tiga komponen dalam GLM yaitu komponen random, komponen sistematik dan komponen fungsi penghubung. Komponen random mencakup variabel yang mempunyai distribusi yang termasuk dalam keluarga eksponensial, misalnya binomial, poisson, normal, gamma, dsb. Komponen sistematik adalah kombinasi linier dari fungsi nilai harapan komponen acak dengan kovariatnya. Sedangkan fungsi penghubung merupakan bentuk dari fungsi parameter natural. Formula distribusi keluarga eksponensial adalah (Dobson, 2002):

$$f(y;\theta) = a(y)b(\theta)\exp[c(\theta)d(y)] \qquad (5)$$

Pada penelitian ini digunakan asumsi pola distribusi LN3P dan distribusi LL3P. Jika *Y* adalah variabel random yang berdistribusi Log-normal 3 parameter (LN3P) yang dinotasikan $Y \sim LN3\left(\mu_{[y]}, \sigma^2_{[y]}, \lambda\right)$, maka fungsi kepadatan peluang (pdf) dari *Y* dapat ditulis sebagai berikut (Johnson dkk., 1995a dan Aitchison, 1957):

$$f(y|\mu, \sigma^2, \lambda) = \frac{1}{(y-\lambda)\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log(y-\lambda)-\mu)^2\right], \tag{6}$$

dimana $y > \lambda$, $\lambda > 0$, $\mu > 0$ dan $\sigma > 0$, dengan μ adalah parameter lokasi (*location*), σ adalah parameter skala (*scale*) dan λ adalah parameter batas (*threshold*).

Persamaan (6) di atas selanjutnya dapat ditulis kembali dalam bentuk persamaan:

$$f(y|\mu, \sigma^2, \lambda) = \exp\left[\frac{\mu}{\sigma^2}\log(y - \lambda) - \frac{\mu}{2\sigma^2} - \frac{1}{2\sigma^2}(\log(y-\lambda))^2 - \log(y-\lambda) - \frac{1}{2}\log\sqrt{2\pi\sigma^2}\right] \tag{7}$$

Dari persamaan (7) tersebut terlihat bahwa distribusi LN3P termasuk keluarga eksponensial. Dengan cara yang sama untuk distribusi LL3P juga dapat dituliskan persamaan distribusi keluarga eksponensial.

Selanjutnya, nilai harapan dan varians model dapat dinyatakan sebagai berikut:

$$E(\ln(y-\lambda)) = \mu \tag{8}$$

$$Var(\ln(y-\lambda)) = \sigma^2 \tag{9}$$

Fungsi nilai harapan model merupakan fungsi identitas dengan persamaan:

$$f(\mu) = \mu = \boldsymbol{X\beta} \tag{10}$$

Kemudian, untuk pemodelan GLM dengan struktur data berhirarki, maka GLM diperluas menjadi *Generalized Multilevel Linear Model* (GMLM) dengan formula fungsi nilai harapan sebagai berikut:

$$f(\mu) = \mu = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k \tag{11}$$

3. Pengujian Distribusi Data Respon

Pada penelitian ini digunakan uji Anderson-Darling untuk pengujian distribusi data variabel respon. Salah satu alasan digunakannya uji Anderson-Darling adalah bahwa uji Anderson-Darling lebih fleksibel daripada uji Kolmogorov-Smirnov (Anderson dan Darling, 1952). Hal ini karena uji Anderson-Darling merupakan modifikasi dari uji Kolmogorov-Smirnov dimana dilakukan penggabungan fungsi bobot sehingga uji Anderson-Darling menjadi lebih fleksibel.

Formula hipotesis uji Anderson-Darling adalah sebagai berikut:

H_0 : Data mengikuti suatu pola fungsi distribusi tertentu

H_1 : Data tidak mengikuti suatu pola fungsi distribusi tertentu

Menurut Anderson-Darling (1954), misal $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ dimana *n* adalah banyaknya pengamatan, maka statistik ujinya adalah sebagai berikut:

$$w_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log u_i + \log(1-u_{n-i+1})], \tag{12}$$

dimana $u_i = F(x_i)$ adalah fungsi distribusi kumulatif. Nilai kritis dari uji Anderson-Darling dirumuskan sebagai berikut:

$$CV = \frac{0,752}{1+\frac{0,75}{n}+\frac{2,25}{n^2}}, \tag{13}$$

dimana *CV* (*Critical Value*) adalah nilai kritis. H_0 ditolak jika $w_n^2 > CV$.

4. Analisis Bayesian

Metode Bayesian mengacu nama ilmuwan Thomas Bayes (1702-1761) yang menemukan perlakuan matematika untuk masalah *non trivial* dari inferensi Bayesian. Thomas Bayes menemukan suatu penyelesaian untuk kasus khusus yang kemudian dikenal dengan nama Teorema Bayesian. Selanjutnya Teorema Bayesian dipopulerkan oleh Matematikawan asal Prancis, Peirre-Simon Laplace dengan istilah peluang Bayesian.

Berbeda dengan teori statistika klasik (*frequentist*), analisis bayesian memperlakukan semua parameter yang tidak diketahui sebagai variabel random dan memiliki distribusi (Boldstad, 2007). Teorema bayesian didasarkan pada distribusi *posterior* yang merupakan perpaduan antara distribusi *prior* (informasi masa lalu sebelum dilakukan observasi) dan data observasi yang digunakan untuk menyusun fungsi *Likelihood* (Box dan Tiao, 1973). Hubungan distribusi *posterior* dengan distribusi *prior* dan *Likelihood* dapat ditulis sebagai berikut.

Distribusi *posterior* ∝ *likelihood* × Distribusi *prior*

Pada teorema Bayes, apabila terdapat parameter $\boldsymbol{\theta}$ yang diberikan oleh data observasi *y*, maka distribusi probabilitas untuk *posterior* $\boldsymbol{\theta}$ pada data *y* akan proporsional dengan perkalian antara distribusi *prior* $\boldsymbol{\theta}$ dan fungsi *Likelihood* $\boldsymbol{\theta}$ yang diberikan oleh data *y*. Secara matematis dapat ditulis sebagai berikut:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y})}$$

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{14}$$

dimana $f(\boldsymbol{\theta}|\mathbf{y})$ merupakan distribusi *posterior* yang proporsional dengan perkalian antara fungsi *Likelihood* $f(\mathbf{y}|\boldsymbol{\theta})$ dan distribusi *prior* $f(\boldsymbol{\theta})$.

4.1 *Markov Chain Monte Carlo*

Untuk mendapatkan pendugaan parameter dari distribusi *posterior* melalui proses integrasi seringkali sulit dilakukan apabila melibatkan persamaan integral yang sangat kompleks. Oleh karena itu penyelesaian perhitungan pendugaan parameter seringkali dilakukan secara numerik, salah satunya adalah teknik *Markov Chain Monte Carlo* (MCMC). Menurut Carlin (1992) pendekatan MCMC sangat efektif untuk mengurangi beban komputasi dalam menyelesaikan persamaan integrasi yang kompleks. Scollnik (2011) mengemukakan bahwa metode ini memungkinkan proses simulasi dengan mengambil sampel acak dari model stokastik yang sangat rumit.

Ide dasar dari MCMC adalah membangkitkan data sampel dari distribusi *posterior* sesuai proses *Markov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif sehingga diperoleh kondisi yang konvergen terhadap *posterior* (Ntzoufras, 2009). Kondisi tersebut harus memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* (Boldstad, 2010; Tailor dan Carlin, 1998) yaitu:

a. *Irreducible*, artinya sampel parameter yang dibangkitkan melalui proses MCMC adalah bersifat random.
b. *Aperiodic*, artinya sampel parameter yang dibangkitkan tersebut tidak memiliki pola yang periodik dalam domain nilai tertentu.
c. *Recurrent*, artinya perubahan sampel parameter terjadi secara stabil dalam domain nilai tertentu.

5. *Hierarchical Bayesian (HB)*

Raudenbush dan Byrk (2002) menyatakan bahwa secara umum proses pembentukan model HB dua tingkat diawali dengan membentuk model mikro sesuai persamaan (2) sebagai *Likelihood* dari data observasi yang memiliki parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\Omega}$, dengan $\boldsymbol{\Omega} = Var(\mathbf{y})$ sehingga fungsi *Likelihood* adalah $f(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\Omega})$. Selanjutnya ditentukan prior dari parameter-parameter yang tidak diketahui dan dilakukan secara bertingkat yaitu *two stage prior* (untuk model hirarki dua tingkat). *Stage-1 prior* berdasarkan model makro sesuai persamaan (2) yang dinyatakan dalam notasi $P_1(\boldsymbol{\beta}|\boldsymbol{\gamma},\mathbf{T})$ dengan $\boldsymbol{\gamma}$ adalah matriks koefisien regresi model makro dan $\mathbf{T}$ adalah matriks varians kovarians dari u_{rk}.

Tahap selanjutnya adalah menentukan *stage-2 prior* yang dinyatakan dalam notasi $P_2(\boldsymbol{\gamma},\boldsymbol{\Omega},\mathbf{T})$. Dengan demikian distribusi *posterior* model HB adalah fungsi yang proporsional terhadap perkalian *Likelihood*, *stage-1 prior* dan *stage-2 prior* yang dinyatakan dalam notasi sebagai berikut:

$$P(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \mathbf{T}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Omega})P_1(\boldsymbol{\beta}|\boldsymbol{\gamma}, \mathbf{T})P_2(\boldsymbol{\gamma}, \boldsymbol{\Omega}, \mathbf{T}) \quad (15)$$

Distribusi *prior* yang digunakan untuk masing-masing elemen vektor parameter model HB berdasarkan distribusi LN3 adalah sebagai berikut:

$$\begin{aligned} &\lambda_k \sim N\left(\mu_{[\lambda]k}, \sigma^2_{[\lambda]k}\right), \\ &\beta_{rk} \sim N\left(\mu_{[\beta]rk}, \tau_{[\beta]rk}\right), \\ &\tau_{[y]k} \sim \text{Gamma}\left(a_{\tau_{[y]k}}, \ b_{\tau_{[y]k}}\right) \\ &\gamma_{lr} \sim N\left(\mu_{\gamma_{lr}}, \sigma^2_{\gamma_{lr}}\right). \end{aligned} \quad (16)$$

Penentuan nilai parameter distribusi *prior* dilakukan dengan cara gabungan antara *conjugate prior* dan *informative prior* (berdasarkan data). Proses penentuan nilai parameter tersebut dilakukan secara berulang-ulang dimana hasil (*posterior*) pada setiap percobaan dijadikan informasi untuk memperbaiki *prior* model sehingga diperoleh hasil estimasi yang konvergen dan memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* (Boldstad, 2010; Tailor dan Carlin, 1998), proses perbaikan prior berulang tersebut dikenal dengan *two-step* MCMC (Iriawan, 2012).

6. Pemodelan

Pemodelan data pengeluaran perkapita rumahtangga dengan distribusi LN3P dan distribusi LL3P dalam penelitian ini meliputi:

a. Model *unilevel*, yang meliputi:

a.1 GLM_min, yaitu model GLM minimal (parameter hanya berupa intersep). Persamaan modelnya adalah:

$$f(\mu_i) = \beta_o \quad (17)$$

a.2 GLM_mod, yaitu model GLM dengan kovariat (X_1, D_2, X_3, D_4). Persamaan modelnya adalah:

$$f(\mu_i) = \beta_o + \beta_1 X_{1i} + \cdots + \beta_4 D_{4i} \quad (18)$$

b. Model *multilevel*, yang meliputi:

b.1 GLMM_Int, yaitu model GLM dengan parameter intersep yang bervariasi menurut wilayah dan tanpa kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\mu_{ik}) = \beta_{ok} + \beta_1 X_{1ik} + \cdots + \beta_p X_{pik} + \mathrm{u}_k \quad (19)$$

b.2 GLMM_X, yaitu model GLM dengan parameter intersep dan koefisien kovariat di model level pertama yang bervariasi menurut wilayah dan tanpa kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\boldsymbol{\mu}_k) = \mathbf{X}_k \boldsymbol{\beta}_k + \mathbf{u}_k \quad (20)$$

b.3 GMLM_Int, yaitu model GMLM dengan struktur hirarki pada parameter intersep (terdapat kovariat di model level kedua). Persamaan modelnya adalah:

$$f(\mu_{ik}) = \gamma_{00} + \gamma_{01} Z_{1k} + \cdots + \gamma_{0l} Z_{lk} + \beta_1 X_{1ik} + \cdots + \beta_p X_{pik} + u_k \quad (21)$$

b.4 Model GMLM_X, yaitu model GMLM dengan struktur hirarki pada intersep dan kovariat di model level pertama dan terdapat kovariat di model level kedua. Persamaan modelnya adalah:

$$f(\boldsymbol{\mu}_k) = \mathbf{X}_k \mathbf{Z}_k \boldsymbol{\gamma} + \mathbf{X}_k \mathbf{u}_k \quad (22)$$

7. Pemilihan Model Terbaik

Strategi pemilihan model terbaik dalam penelitian ini menggunakan *Deviance Information Criterion* (DIC) yang dirumuskan sebagai berikut:

$$DIC = D(\bar{\boldsymbol{\theta}}_m, m) + 2p_m \quad (23)$$

dimana p_m adalah jumlah parameter efektif dari model. Menurut Ntzoufras (2009), DIC merupakan kriteria seleksi model terbaik yang hampir sama dengan *Akaike's Information Criterion* (AIC).

METODOLOGI

1. Sumber Data

Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS), yaitu:
Data karakteristik individu dan rumahtangga di Kota Jambi yang berasal dari Susenas 2011 dan data karakteristik desa/kelurahan di Kota Jambi yang berasal dari Podes 2011.

2. Variabel Penelitian

Variabel respon yang digunakan yaitu data pengeluaran perkapita perbulan yang diperoleh dengan cara menghitung pengeluaran rumahtangga per bulan dibagi dengan jumlah anggota rumahtangganya. Sedangkan variabel prediktor dapat dilihat pada Tabel 1 dan Tabel 2 berikut ini.

Tabel 1. Variabel Prediktor pada Model Mikro (Karakteristik rumahtangga)

Var	Keterangan
X_1	Jumlah anggota rumahtangga (ART)
D_2	=1, jenis lantai terluas dari keramik/ubin/tegel/teraso =0, jenis lantai terluas dari lainnya
X_3	Luas lantai perkapita
D_4	=1, sumber energi untuk memasak dari LPG/gas =0, sumber energi untuk memasak dari lainnya

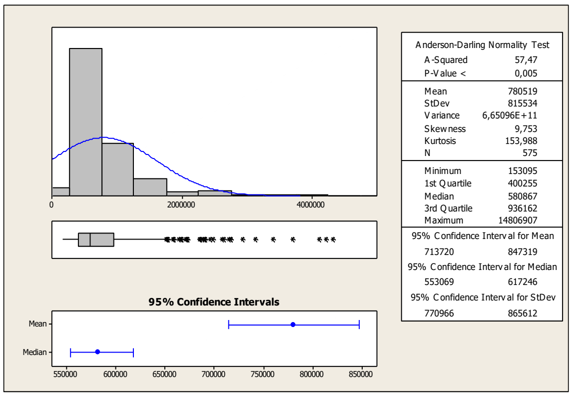
Tabel 2. Variabel Prediktor pada Model Makro (Karakteristik wilayah Kelurahan)

Var	Keterangan
Z_1	Rasio SLTA per 10.000 penduduk
Z_2	Rasio Puskesmas per 10.000 penduduk

HASIL DAN PEMBAHASAN

1. Karakteristik Variabel Respon

Karakteristik variabel respon perlu diketahui untuk menentukan bentuk distribusi *likelihood* yang nantinya akan diterapkan pada pemodelan dengan pendekatan Bayesian. Gambar 1 berikut dapat menjelaskan karakteristik variabel respon (pengeluaran perkapita rumahtangga perbulan).



Gambar 1. Karakteristik data pengeluaran perkapita rumahtangga perbulan di Kota Jambi

Dari Gambar 1 di atas terlihat bahwa terdapat dua ciri pola data pengeluaran perkapita rumah tangga perbulan yaitu nilai datanya positif dan memiliki ekor yang menceng ke kanan. Beberapa jenis distribusi yang memiliki kedua ciri tersebut diantaranya adalah distribusi Log-normal 2 parameter, Log-normal 3 Parameter, Log-logistik 3 parameter dan Weibull 3 parameter.

2. Distribusi Variabel Respon

Merujuk pada penelitian yang dilakukan oleh Ismartini dkk. (2012), maka pada penelitian ini dilakukan pengujian data pengeluaran perkapita rumahtangga perbulan di Kota Jambi dengan asumsi awal bahwa data mengikuti pola distribusi Log-normal tiga parameter (LN3P) dan distribusi Log-logistik tiga parameter (LL3P). Dari Tabel 3 terlihat bahwa distribusi data pengeluaran perkapita rumahtangga dapat dimodelkan dengan distribusi LN3P dan distribusi LL3P.

Tabel 3. Hasil Uji Anderson Darling

Distribusi	Banyaknya amatan	Nilai AD	Critical value
LN3P	575	0,44956	2,5018
LL3P	575	0,63205	2,5018

Selanjutnya, distribusi LN3P dan distribusi LL3P dimodelkan dengan metode *Generalized Liniear Model* (GLM) karena kedua distribusi tersebut dapat dikategorikan sebagai keluarga eksponensial.

3. Pemodelan Pengeluaran Perkapita Rumahtangga

Proses pendugaan parameter model LN3P dan LL3P dengan pendekatan Bayesian dilakukan dengan menggunakan MCMC dan algoritma *Gibbs Sampling* dengan iterasi sebanyak 10.000, *thin* 10 dan *burn-in* sebanyak 1000 iterasi.

➢ **Pemodelan *Unilevel***

Hasil pemodelan *unilevel* ditampilkan pada Tabel 4 sebagai berikut:

Tabel 4. Hasil Pendugaan Parameter Model *Unilevel* Distribusi LN3P dan LL3P

Parameter	Model Berdistribusi LN3P		Model Berdistribusi LL3P	
	GLM_min	GLM_mod	GLM_min	GLM_mod
λ (*Threshold*)	125000 (100.2)	124900 (100.7)	124900 (100.3)	124900.0 (102.1)
β_0	13.06 (0.03159)	12.99 (0.01987)	13.05 (0.03188)	12.99 (0.01989)
β_1	-	-0.09684 (0.004649)	-	0.3091 (0.01707)
β_2	-	0.3109 (0.01699)	-	0.01077 (0.00058)
β_3	-	0.01075 (0.000589)	-	0.2862 (0.0191)
β_4	-	0.2879 (0.01987)	-	0.3091 (0.01707)
DIC	16369.3	16081.0	16372.2	16061.0

Dari Tabel 4 terlihat bahwa model *unilevel* yang terbaik adalah model GLM_mod berdistribusi LL3P, yaitu model dengan variabel X_1, D_2, X_3, dan D_4 sebagai kovariat.

Model GLM_mod dengan empat kovariat (X_1, D_2, X_3, dan D_4) selanjutnya digunakan sebagai basis pemodelan *multilevel*.

➢ **Pemodelan *Multilevel***

Kondisi berbeda terjadi pada model *multilevel*, dimana model LN3P secara umum lebih baik daripada model LL3P jika dilihat dari nilai DIC sebagaimana ditampilkan pada Tabel 5 berikut ini.

Tabel 5. Nilai *Deviance* dan DIC untuk Pemodelan *Multilevel*

Distribusi	Model	*Deviance*	DIC
LN3P	GLMM_Int	15950	15975,5
	GLMM_X	15940	15966,4
	GMLM_Int*	**15870**	**15922.5**
	GMLM_X*	16630	16654,4
LL3P	GLMM_Int	15950	15992,6
	GLMM_X	15940	15979,9
	GMLM_Int*	**15870**	**15935.9**
	GMLM_X*	16660	16694,4

Keterangan :

*) Model GMLM menggunakan empat kovariat di model level pertama (X_1, D_2, X_3, dan D_4) dan dua kovariat di model level kedua (Z_1 dan Z_2). Penggunaan kovariat tersebut didasarkan pada signifikansi parameter dan kebaikan model

Tabel 5 menunjukkan bahwa model GMLM_Int berdistribusi LN3P mempunyai nilai DIC yang paling kecil di antara model *multilevel* lainnya. Selain itu, jika dibandingkan dengan model *unilevel* terbaik (GLM_mod2 berdistribusi LL3P), model GMLM_Int berdistribusi LN3P mempunyai nilai DIC yang lebih kecil. Dengan demikian, model terbaik di antara semua model yang dihasilkan pada penelitian ini adalah model *multilevel* berdistribusi LN3P dengan struktur hirarki pada parameter intersep (GMLM_Int berdistribusi LN3P).

4. Model Terbaik

Uraian berikut ini menjelaskan lebih detail model terbaik (GMLM_Int berdistribusi LN3P) yang dihasilkan, meliputi pemilihan distribusi prior, hasil pendugaan parameter dan diagnosa model.

Distribusi prior yang digunakan untuk model GMLM_Int LN3P adalah:

$$\lambda \sim N(124950{,}1.0\text{E}-07),$$
$$\beta_{0k} \sim N\left(\mu_{[\beta]0k}, \tau_{[\beta]0k}\right),$$
$$\tau_{[y]k} \sim \text{Gamma}(6.296{,}0.687)$$
$$, \qquad (24)$$
$$\tau_{[\beta]0k} \sim \text{Gamma}(10{,}10),$$
$$\gamma_{00} \sim N(13.18{,}167.3126021)$$
$$,$$
$$\gamma_{10} \sim N(0.065936{,}1721321.967)$$

Adapun hasil pendugaan parameter model terbaik ditampilkan pada Tabel 6 berikut ini:

Tabel 6. Hasil Pendugaan Parameter Model GMLM_Int Berdistribusi LN3P

Parameter	Mean	sd	2.5%	97.5%	Sample
λ (*Threshold*)	121100	3008.0	115100	1.27E+5	9001
γ_{00}	13.13	0.07217	12.99	13.27	9001
γ_{01}	0.06592	7.6E-4	0.06441	0.06741	9001
γ_{02}	0.06593	7.599E-4	0.06445	0.0674	9001
β_1	-0.0921	0.005058	-0.1022	-0.0824	9001
β_2	-	-	-	-	-
β_3	0.3136	0.01699	0.281	0.3469	9001
β_4	0.01114	6.07E-4	0.00992	0.0123	9001
β_5	0.279	0.01974	0.2399	0.3175	9001

Tabel 6 menunjukkan bahwa semua parameter model dapat diduga secara signifikan. Hal ini terlihat dari nilai *Credible Interval* (antara 2.5% sampai dengan 97.5%) yang tidak mengandung nilai nol. Secara matematis dapat ditulis dalam persamaan sebagai berikut:

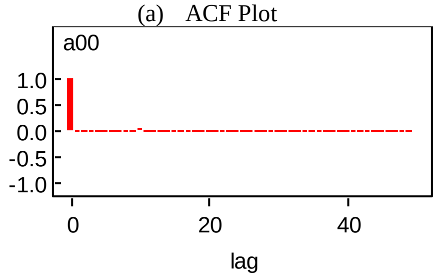
$$f(\mu_{ik}) = 13.13 + 0.066Z_{4k} + 0.066Z_{5k}$$
$$-0.0921\,X_{1ik} + 0.3136D_{3ik} + 0.0111X_{4ik}$$
$$+0.279\,D_{5ik} + u_k$$

Dari persamaan model terlihat bahwa kovariat model di level kedua (proporsi banyaknya SMU terhadap jumlah penduduk dan proporsi puskesmas terhadap jumlah penduduk) berpengaruh positif terhadap pengeluaran perkapita rumah tangga masing-masing sebesar exp(0.066)=1.068 kali. Adapun pengaruh kovariat model di level pertama terlihat bahwa variabel jumlah anggota rumahtangga berpengaruh negatif terhadap pengeluaran perkapita rumahtangga sebesar exp(-0.0921)=0.91 kali. Selanjutnya, variabel jenis lantai terluas dari keramik/ keramik/ ubin/ tegel/ teraso berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar sebesar exp(0.3136)=1.368 kali, kemudian variabel luas lantai perkapita berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar exp(0.0111)=1.011 kali, dan variabel sumber energi untuk memasak berpengaruh terhadap pengeluaran perkapita rumahtangga sebesar exp(0.279)=1.322 kali.

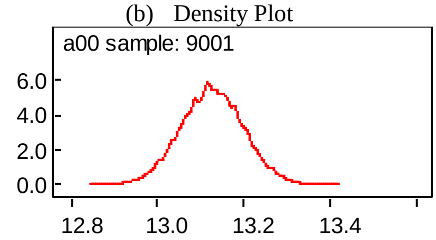
- **Diagnosa Model**

Diagnosa model terbaik ditampilkan dalam Gambar 2. Dari Gambar 2 terlihat bahwa sampel parameter yang dibangkitkan dari distribusi posterior untuk menduga parameter sudah memenuhi persyaratan pemodelan Bayesian. Hal ini terlihat dari plot ACF yang menunjukkan kondisi *Irreducible* (sampel parameter yang dibangkitkan melalui proses MCMC bersifat random) dan *Aperiodic* (tidak memiliki pola periodik). Kemudian, dari *Density* plot terlihat bahwa sampel parameter yang dibangkitkan melalui proses MCMC sudah sesuai dengan distribusi normal dan dari *Serial* plot terlihat bahwa pergerakan sampel parameter tercapai kondisi *recurrent* (stabil dalam domain nilai tertentu). Dengan terpenuhinya kondisi *Irreducible*, *Aperiodic* dan *Recurrent* maka dapat disimpulkan hasil pendugaan parameter model telah memenuhi sifat-sifat *Markov Chain* yang *strongly ergodic* sehingga model layak digunakan untuk keperluan analisis statistik.

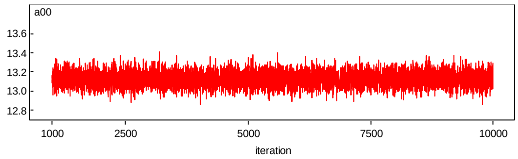
(a) ACF Plot



(b) Density Plot



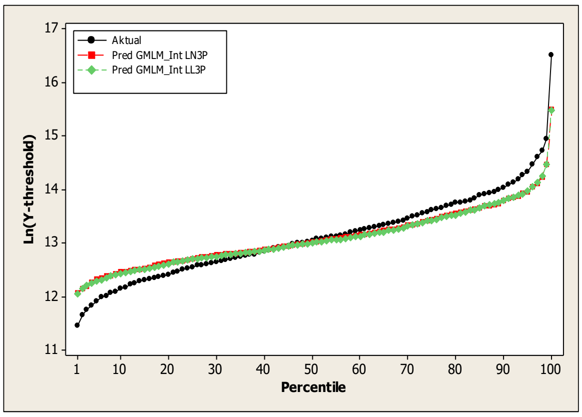
(c) Serial Plot



Gambar 2. ***Diagnostic* plot model GMLM_Int distribusi LN3P untuk parameter γ_{00}**

- **Prediksi Model**

  Prediksi model GMLM_Int berdistribusi LN3P dan distribusi LL3P ditampilkan dalam Gambar 3 sebagai berikut:



Gambar 3 Plot persentil data aktual dan hasil prediksi model GMLM berdistribusi LN3P dan LL3P

Dari Gambar 4.3 terlihat bahwa hasil prediksi antara model GMLM berdistribusi LN3P dan model GMLM berdistribusi LL3P mempunyai kemiripan dimana plot prediksi kedua model tersebut berhimpitan. Namun masih terlihat bahwa kemampuan prediksi kedua model masih perlu ditingkatkan, khususnya untuk persentil bawah dan persentil atas yang terdapat gap antara data aktual dengan hasil prediksi model.

KESIMPULAN DAN SARAN

1. Kesimpulan

Dari hasil analisa yang sudah diuraikan pada bab sebelumnya, maka dapat diperoleh kesimpulan sebagai berikut:

1. Model LL3P memberikan hasil pemodelan yang lebih baik daripada model LN3P untuk pemodelan *unilevel*. Sedangkan untuk pemodelan *multilevel*, model LN3P memberikan hasil pemodelan yang lebih baik dari model LL3P.
2. Model GMLM_Int LN3P (model *multilevel* berdistribusi LN3P dengan struktur hirarki pada parameter intersep) adalah model terbaik untuk pemodelan data pengeluaran perkapita rumahtangga dengan nilai DIC terkecil.

2. Saran

Dari plot antara data aktual dan hasil prediksi model ternyata masih didapatkan gap yang semakin besar untuk persentil menengah ke bawah dan persentil menengah ke atas. Selain itu, nilai Deviance dan DIC yang diperoleh dari pemodelan GMLM masih cukup tinggi. Oleh karena itu disarankan untuk penelitian selanjutnya agar menambah level model menjadi tiga level. Harapannya dengan menambah level model akan memperkecil nilai Deviance dan mengurangi besarnya gap antara data aktual dan prediksi model.

DAFTAR PUSTAKA

Anderson, T.W. dan Darling, D.A. 1954. A test Goodness of Fit. *Journal of American Statistical Association*, Volume 49, Issue 268, hal. 765-769.

Anderson, T.W. dan Darling, D.A .1952. Asymptotic Theory of Certain “Goodness of Fit” Criteria Based on stochastic Process. *The Annals of Mathematical Statistics*, Vol. 23, No. 2, hal. 193-212.

Anderson, W., dan Wells, M.T. 2010. A Bayesian Hierarchical Regression Approach to Clustered and Longitudinal Data in Empirical Legal Studies. *Journal of Empirical Legal Studies*, Volume 7, Issue 4, hal. 634-663.

Boldstad, W.M. 2007. *Introduction to Bayesian Statistics. 2nd Edition*. Wiley, New Jersey.

Box, G.E.P. dan Tiao, G.C. 1973. *Bayesian Inference in Statistical Analysis*. John Willey And Sons, Inc : New York.

BPS Kota Jambi. 2011. *Kota Jambi Dalam Angka 2011*. Badan Pusat Statistik Kota Jambi, Provinsi Jambi.

Carlin, B.P. 1992. A SimpleMonte Carlo Approach to Bayesian Graduation. *Transactions of the Society of Actuaries* XLIV, hal. 55–76.

Carlin, B. P. dan Chib, S. 1995. Bayesian model choice via Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, Ser. B.* Vol.57 No.3, hal.473–484.

De Leeuw, J. dan Kreft, I. 2006. Random Coefficient Models for Multilevel Analysis. *Departement of Statistics Paper*. Departement of Statistics, UCLA, Los Angeles. http://preprints.stat.ucla.edu/496/dLK_jes.pdf (19 Juli 2010).

Dobson, A.J. 2002. *An Introduction to Generalized Linear Models, Second Edition*. Chapman & Hall, London

Gelman, A., dan Hill, J. 2006. *Data Analysis Using Regression and Multilevel / Hierarchical Models*. Cambridge University Press.

Goldstein, H. 1995. *Multilevel Statistical Models*. 2nd *edition*. Arnold, London. ISBN-10: 111995682X, hal. 382.

Ha, I.D., dan Lee. Y. 2005. Multilevel Mixed Linear Models for Survival Data. *Lifetime Data Analysis*, 11, hal. 131-142.

Hox, J.J. 1995. *Applied Multilevel Analysis*, 1st *edition*, TT-Publikaties, Amsterdam, hal. 119.

Iriawan, N. 2012. *Pemodelan dan Analisis Data Driven*. ITS Press : Surabaya.

Ismartini, P., Iriawan, N., Setiawan, dan Ulama, B.S.S. 2012. Toward a Hierarchical Bayesian Framework for Modelling the Effect of Regional Diversity on Household Expenditure. *Journal of Mathematics and Statistics*, Vol.8, No.2, hal. 283-291.

Kaashoek, J.F. dan Van Dijk, H.K. 2002. Neural Network Pruning Applied to Real Exchange Rate Analysis. *Journal of Forecasting*, 21, pp. 559-577.

McCullagh, P., dan Nelder, J.A. 1989. *Generalized Linear Models*, *Second Edition*. Chapman & Hall, London

Raudenbush, S.W. dan A.S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd *edition*. Sage Publications, Thousand Oaks. ISBN-10: 076191904X, hal. 485.

Scollnik, D. P. M. 2011. *An Introduction To Markov Chain Monte Carlo Methods And Their Actuarial Applications*. Handbook of Markov Chain Monte Carlo Chapter 1, hal.114-165 : Chapman & Hall/CRC Handbooks of Modern Statistical Method.

Steenbergen, M. R., dan Jones, B. S. 2002. Modelling Multilevel Data Structure. *American Journal of Political Science*, Vol. 46, No. 1, hal. 218-237.

Tailor, H.M., dan Carlin, S. 1998. *An Introduction to Stochastic Modelling 3rd edition*. Academic Press, San Diego.

Zimmer, Z., Wen, M., dan Kaneda, T. 2010. A Multi-Level Analysis of Urban/Rural and Socioeconomic Differences in Functional Health Status Transition among Older Chinese. *Social Science & Medicine* 71, hal. 559-567.