

METODE BAYES EMPIRIK UNTUK MEMODELKAN DATA CACAHAN DENGAN PEUBAH PENYERTA PADA PENDUGAAN AREA KECIL

NADYA CINDY EKA PUTRI, FERRA YANUAR, HAZMIRA YOZZA

*Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : nadyacindy22@gmail.com*

Diterima 9 Maret 2019 Direvisi 7 April 2019 Dipublikasikan 7 Mei 2019

Abstrak. Metode Bayes Empirik merupakan suatu metode pada *Small Area Estimation* (SAE) yang menggunakan metode Bayes dalam pendugaan parameternya. *Small Area Estimation* (SAE) didefinisikan sebagai suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil, oleh karena itu diperlukan informasi tambahan agar diperoleh dugaan yang lebih akurat. Metode Bayes Empirik ini dapat digunakan pada data cacahan dengan menggunakan model Poisson-Gamma tanpa peubah penyerta dan dengan peubah penyerta. Penelitian ini mengkaji penggunaan metode Bayes Empirik berbasis model Poisson-Gamma pada area kecil baik secara teoritis maupun empiris melalui simulasi. Hasil penelitian ini menunjukkan penduga Bayes Empirik dengan peubah penyerta memberikan hasil yang lebih baik dengan standar error yang lebih kecil dibandingkan dengan penduga langsung dan penduga Bayes Empirik tanpa peubah penyerta.

Kata Kunci: Metode Bayes Empirik, *Small Area Estimation* (SAE), model Poisson-Gamma

1. Pendahuluan

Small Area Estimation (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil. Dalam metode SAE ini untuk meningkatkan keefektifan dari data yang akan digunakan serta menurunkan standar error penduga parameternya maka digunakan penduga tak langsung (*indirect estimation*). Untuk memperoleh dugaan area kecil dengan penduga tak langsung, terdapat beberapa metode pendugaan yang diperkenalkan oleh Rao dalam [3] seperti penduga prediksi tak bias linier terbaik empirik atau *empirical best linear unbiased prediction* (EBLUP) untuk data kontinu, penduga Bayes empirik atau *empirical Bayes* (EB), dan penduga Bayes hierarkhi atau *hierarchical Bayes* (HB) untuk data biner dan cacahan [5].

Metode Bayes Empirik adalah salah satu metode pada SAE menggunakan metode Bayes dalam pendugaannya. Pada metode Bayes Empirik ini, penduga

parameter lebih dipengaruhi oleh informasi dari data (empirik) karena kekurangan informasi untuk menentukan distribusi priornya.

Salah satu metode Bayes Empirik yang sering digunakan adalah model Poisson-Gamma. Model Poisson merupakan peluang baku untuk data cacahan namun dalam keadaan tertentu nilai harapan dan ragamnya tidak sama, yang dinamakan overdispersi. Hal itu dapat diatasi dengan mengambil informasi tambahan pada distribusi prior untuk mengatasi kelebihan ragam dari pengamatan. Distribusi prior yang dapat dipilih adalah distribusi Gamma, sehingga model Poisson yang akan diuji menjadi model dengan dua tahap yaitu model Poisson-Gamma. Wakefield dalam [6] membuat metode penduga Bayes Empirik pada SAE dengan model Poisson-Gamma dengan peubah penyerta. Untuk itu, maka pada penelitian ini, akan dilakukan pendugaan area kecil pada tiga kasus yaitu: penduga langsung, penduga Bayes Empirik (penduga tak langsung) tanpa peubah penyerta, dan penduga Bayes Empirik (penduga tak langsung) dengan peubah penyerta.

2. Pendugaan Area Kecil (*Small Area Estimation/SAE*)

Pendugaan area kecil merupakan konsep terpenting dalam pendugaan parameter secara tidak langsung di suatu area yang relatif kecil dalam percontohan survei (*survey sampling*). Metode pendugaannya digunakan untuk menduga karakteristik dari subpopulasi (domain yang lebih kecil). Metode pendugaan ini diperlukan karena pendugaan langsung (*direct estimation*) pada subpopulasi tidak memiliki presisi yang memadai karena kecilnya jumlah sampel yang digunakan untuk memperoleh dugaan tersebut.

Metode yang dikembangkan dalam pendugaan area kecil adalah metode pendugaan tidak langsung dengan memanfaatkan kekuatan area di sekitarnya dan sumber data di luar area. Atau dengan kata lain pendugaan tidak langsung dapat dilakukan dengan memanfaatkan peubah-peubah tambahan dalam menduga parameter. Peubah-peubah pendukung tersebut dapat berupa informasi tambahan yang didapatkan dari area lain yang serupa, survei terdahulu yang dilakukan di daerah yang sama (seperti kasus) atau peubah lain yang berhubungan dengan peubah yang ingin diduga [3].

Terdapat dua ide utama diterapkan untuk mengembangkan metode pendugaan pada area kecil:

- (1) Asumsi bahwa keragaman di dalam area kecil bagi variabel responnya dapat diterangkan oleh hubungan keragaman yang bersesuaian pada informasi tambahan atau disebut dengan model pengaruh tetap (*fixed effect models*).
- (2) Asumsi keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan yang kemudian disebut sebagai model pengaruh acak (*random effect*).

Kedua asumsi tersebut kemudian bergabung membentuk model pengaruh campuran (*mixed models*) [3].

Pada SAE terdapat dua jenis model, yaitu model level area dan model level unit. Skop kajian pada makalah ini dibatasi untuk model level area dasar. Pada

model level area dasar, data pendukung yang tersedia hanya sampai level area tertentu. Model level area menghubungkan penduga langsung area kecil dengan data pendukung dari domain lain untuk setiap area. Pada model level area dasar diasumsikan bahwa parameter area kecil yang akan diduga, yang berkaitan dengan data penyerta area kecil $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$ dan mengikuti model linier sebagai berikut [3]:

$$\theta_i = x_i \underline{\beta} + v_i. \quad (2.1)$$

dimana $i = 1, 2, \dots, m$ dan $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ merupakan vektor koefisien regresi yang berukuran $p \times 1$. Sedangkan, v_i adalah pengaruh acak area kecil yang diasumsikan memiliki sebaran identik dan saling bebas, dengan $v_i \sim N(0, \sigma_v^2)$ sering kali diasumsikan sebagai pengaruh acak yang normal.

Untuk melakukan inferensi tentang rata-rata area kecil \bar{Y}_i pada persamaan (2.1), dapat diasumsikan bahwa penduga langsung \widehat{Y}_i ada, dan

$$\widehat{\theta}_i = g(\widehat{Y}_i). \quad (2.2)$$

dimana $i = 1, 2, \dots, m$, *sampling error* $e_i \sim N(0, \psi_i)$, dan ψ_i diketahui.

Dengan menggabungkan persamaan (2.1) dan (2.2) maka diperoleh model gabungan sebagai berikut:

$$\widehat{\theta}_i = x_i^T \underline{\beta} + v_i + e_i. \quad (2.3)$$

untuk $i = 1, 2, \dots, m$.

2.1. Penduga Area Kecil Untuk Kasus Tertentu

Misalkan suatu area dapat dibagi kedalam m area kecil yang tidak saling tumpang tindih. Misalkan Y_i adalah banyaknya kasus tertentu yang terjadi pada area ke- i . Y_i biasanya diasumsikan menyebar menurut sebaran Poisson ($e_i \mu_i \theta_i$) dimana e_i adalah nilai harapan banyaknya kasus tertentu terjadi di suatu area yang dihitung dari $e_i = n_i \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m n_i}$, e_i diasumsikan bukan merupakan peubah acak. dan dapat dipandang sebagai banyaknya objek yang beresiko mendapatkan atau mengalami kasus tertentu tersebut. θ_i adalah parameter yang menyatakan resiko relatif kejadian kasus tertentu tersebut terjadi di area ke- i .

Penduga langsung diperoleh dengan memaksimumkan fungsi kemungkinan (*likelihood*). Disini diasumsikan bahwa banyaknya pengamatan suatu kasus terjadi pada area ke- i , y_i memiliki distribusi Poisson sebagai berikut:

$$y_i \sim \text{Poisson}(e_i \mu_i \theta_i)$$

dengan $\mu_i = \mu(x_i, \underline{\beta})$, e_i adalah nilai harapan dari banyaknya kasus pada area ke- i dan θ_i adalah resiko relatif pada area ke- i . Fungsi kepekatan peluang y_i adalah sebagai berikut

$$f(y_i) = \frac{\exp(-e_i \mu_i \theta_i) (e_i \mu_i \theta_i)^{y_i}}{y_i!} \quad (2.4)$$

dengan fungsi kemungkinannya dan logaritma natural dari fungsi kemungkinannya adalah sebagai berikut:

$$\begin{aligned}
 L(e_i\mu_i\theta_i|y_i) &= f(y_i) \\
 &= \frac{\exp(-e_i\mu_i\theta_i)(e_i\mu_i\theta_i)^{y_i}}{y_i!} \\
 \ln(L(e_i\mu_i\theta_i|y_i)) &= -e_i\mu_i\theta_i + y_i \ln(-e_i\mu_i\theta_i) - \ln(y_i!) \tag{2.5}
 \end{aligned}$$

Nilai penduga langsung bagi θ_i diperoleh dari $y_i \sim Poisson(e_i\mu_i\theta_i)$ dengan cara sebagai berikut:

$$\begin{aligned}
 \frac{d}{d\theta_i} \ln(L(e_i\mu_i\theta_i|y_i)) &= -e_i\mu_i + \frac{y_i e_i\mu_i}{e_i\mu_i\theta_i} \\
 e_i\mu_i &= \frac{y_i}{\theta_i} \\
 \hat{\theta}_i &= \frac{y_i}{e_i\mu_i} \tag{2.6}
 \end{aligned}$$

dimana θ_i adalah penduga kemungkinan maksimum bagi θ_i .

2.2. Metode Bayes Empirik atau Empirical Bayes (EB)

Metode **Bayes Empirik** yaitu metode yang menggunakan inferensia dari estimasi posterior untuk menduga parameter dimana informasi parameter pada sebaran awal tidak diketahui, kemudian data dimanfaatkan untuk memperoleh nilai dugaan parameter [4].

Model Poisson merupakan model peluang baku untuk data cacahan. Pada model ini terdapat keterbatasan dalam nilai harapan ragam ketika digunakan untuk pendugaan parameter tunggal. Umumnya, data cacahan sering mengalami overdispersi. Oleh karena itu, dikembangkanlah suatu formula yang dapat mengatasi permasalahan tersebut. Yakni dengan menggunakan model dua tahap untuk data cacahan yang dikenal sebagai model Poisson-Gamma.

Model Poisson-Gamma dapat dijelaskan sebagai berikut: Y_i adalah banyaknya pengamatan suatu kasus pada area ke- i ($i = 1, 2, \dots, m$) yang belum diketahui, e_i adalah nilai harapan banyaknya suatu kasus pada area ke- i , θ_i adalah resiko relatif area ke- i yang tidak diketahui dan m menyatakan jumlah area, sedangkan v dan α adalah parameter prior yang belum diketahui.

Pada tahap pertama diasumsikan bahwa $Y_i \sim Poisson(e_i\mu_i\theta_i)$ untuk $i = 1, 2, \dots$ dengan fungsi kepekatan peluang sebagai berikut:

$$f(y_i|\theta_i) = \frac{\exp(-e_i\mu_i\theta_i)(e_i\mu_i\theta_i)^{y_i}}{y_i!} \tag{2.7}$$

dimana $y_i = 1, 2, \dots$, $\mu_i = \mu(\underline{x}_i, \underline{\beta})$ merupakan model regresi, sehingga $\underline{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})^T$ merupakan vektor peubah penyerta tetap dan $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ merupakan vektor koefisien regresi.

Selanjutnya, diasumsikan bahwa parameter θ_i berdistribusi Gamma atau $\theta_i \sim Gamma(\alpha, \alpha)$ dengan rata-rata 1 dan ragam $\frac{1}{\alpha}$ dengan fungsi kepekatan peluang

sebagai berikut:

$$\pi(\theta_i) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\alpha\theta_i), \theta_i > 0 \quad (2.8)$$

Fungsi kepekatan peluang bersama $f(y_i, \theta_i)$ adalah perkalian antara $f(y_i|\theta_i)$ dan $\pi(\theta_i)$ sebagai berikut:

$$f(y_i, \theta_i) = \frac{\exp(-e_i\mu_i\theta_i)(e_i\mu_i\theta_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\alpha\theta_i), \theta_i > 0 \quad (2.9)$$

Dari fungsi kepekatan peluang bersama pada Persamaan (2.9) akan diperoleh fungsi kepekatan peluang marginalnya dengan cara sebagai berikut:

$$\begin{aligned} m(y_i) &= \int_0^\infty f(y_i, \theta_i) d\theta_i \\ &= \frac{\exp(-e_i\mu_i\theta_i)(e_i\mu_i\theta_i)^{y_i}}{y_i!} \frac{\alpha^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\alpha\theta_i) d\theta_i \\ &= C(y_i + \alpha - 1, \alpha - 1) \left(\frac{\alpha}{\mu_i e_i + \alpha} \right)^\alpha \left(1 - \frac{\alpha}{\mu_i e_i + \alpha} \right)^{y_i} \end{aligned} \quad (2.10)$$

Dapat dilihat bahwa fungsi marjinal dari fungsi kepekatan peluang bersama merupakan fungsi dari sebaran Binomial Negatif dengan rata-rata dan ragam sebagai berikut:

$$E(Y_i) = e_i\mu_i, \quad (2.11)$$

$$Var(Y_i) = e_i\mu_i \left(1 + \frac{e_i\mu_i}{\alpha} \right). \quad (2.12)$$

Diperoleh rata-rata dan ragam dari fungsi marjinal dengan ragam meningkat sebagai fungsi kuadrat dari rata-rata. Selanjutnya, akan ditentukan fungsi posterior yaitu dengan pembagian antara Persamaan (2.9) dengan Persamaan (2.10):

$$\begin{aligned} \pi(\theta_i|y_i, \beta, \alpha) &= \frac{f(y_i, \theta_i)}{m(y_i)} \\ &= \frac{(e_i\mu_i + \alpha)^{y_i + \alpha}}{\Gamma(y_i + \alpha)} \theta_i^{y_i + \alpha - 1} \exp(-(e_i\mu_i + \alpha)\theta_i), \theta_i > 0 \end{aligned} \quad (2.13)$$

Dari fungsi posterior di atas maka dapat diidentifikasi bahwa $\theta_i|y_i, \beta, \alpha \sim \text{Gamma}(y_i + \alpha, e_i\mu_i + \alpha)$. penduga Bayes bagi θ_i yang diperoleh dari nilai harapan fungsi posterior di atas adalah sebagai berikut:

$$E(\theta_i|y_i, \beta, \alpha) = \widehat{\theta}_i^B(\beta, \alpha) = \frac{y_i + \alpha}{e_i\mu_i + \alpha} \quad (2.14)$$

dan dengan ragam posterior sebagai berikut:

$$Var(\theta_i, y_i, \beta, \alpha) = g_i(y_i, \hat{\beta}, \alpha) = \frac{y_i + \alpha}{(e_i\mu_i + \alpha)^2} \quad (2.15)$$

Maka penduga Bayes Empirik atau *Empirical Bayes* (EB) bagi θ_i adalah sebagai berikut:

$$\begin{aligned} \widehat{\theta}_i^{EB} &= E(\theta_i|y_i, \hat{\beta}, \alpha) = \frac{y_i + \hat{\alpha}}{e_i\mu_i + \hat{\alpha}} \\ &= \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i)(\hat{\mu}_i) \end{aligned} \quad (2.16)$$

dengan $\hat{\gamma}_i = \frac{e_i \hat{\mu}_i}{e_i \hat{\mu}_i + \hat{\alpha}}$, $\theta_i = \frac{y_i}{e_i \hat{\mu}_i}$. θ_i yaitu penduga langsung dari θ_i yang diperoleh dari persamaan (2.6), y_i adalah banyaknya pengamatan suatu kasus dan e_i adalah nilai harapan banyaknya suatu kasus. μ_i diperoleh dari dengan memodelkan hubungan antara variabel y_i dengan variabel-variabel parameter lain yang diasumsikan mempengaruhi y_i tersebut melalui suatu model regresi tertentu, model yang dipi-lih pada kasus itu adalah model regresi Binomial Negatif yang dinyatakan dengan $\mu_i = \exp(\underline{x}_i^T, \underline{\beta})$.

3. Pembahasan

Pemetaan laju mortalitas atau terjadinya suatu penyakit pada area kecil seperti kanker atau malaria merupakan alat yang digunakan secara luas pada penelitian-penelitian dalam bidang kesehatan masyarakat. Model pemetaan penyakit, misalkan suatu daerah yang digunakan untuk pemetaan penyakit dibagi kedalam m area kecil yang tidak saling tumpang tindih.

Penduga langsung (kasar) bagi θ_i diberikan oleh SMR (*standardize mortality ratio*) yang merupakan penduga sederhana pada pemetaan penyakit, $\hat{\theta}_i = \frac{y_i}{e_i}$ dimana y_i menyatakan banyaknya kasus yang teramati dan e_i menyatakan nilai harapan banyaknya suatu kasus, $i = 1, 2, \dots, m$, dimana $e_i = n_i \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m n_i}$ adalah banyaknya orang pada periode tersebut beresiko pada area ke- i tersebut dan e_i dianggap bukan merupakan peubah acak.

Wakefield memperkenalkan model Poisson-Gamma dengan peubah penyerta, yang mana pada asumsi pertama diasumsikan bahwa $y_i \sim Poisson(e_i \mu_i \theta_i)$ dan asumsi kedua bahwa $\theta_i \sim Gamma(\alpha, \alpha)$ yang dianggap sebagai prior [6].

Untuk melihat kelayakan model yang akan diujikan diperoleh persamaan regresi didapatkan dari model regresi binomial negatif yang ditulis sebagai berikut:

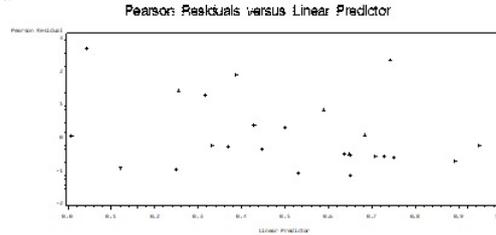
$$\ln\left(\frac{\hat{y}_i}{e_i}\right) = -0,2090 + 0,0139x_i$$

dengan kriteria kelayakan modelnya dapat dilihat pada Tabel (1) berikut:

Tabel 1. Kriteria kelayakan model

Kriteria	DF	Value	Value/DF
Devian	23	26,553	1,155
Skala Devian	23	26,553	1,155
<i>Pearson Chi-Square</i>	23	27,779	1,208
Skala <i>Pearson X2</i>	23	27,779	1,208

Dari Tabel (1) terlihat bahwa model regresi binomial negatif ini layak digunakan, karena nilai Value/DF untuk masing-masing kriteria nilainya kurang dari 2 ($Value/DF < 2$), serta untuk menguji kelayakan model regresi binomial dapat dilihat grafis hubungan antara Prediktor Linier dengan sisaan Pearson yang disajikan dalam Gambar (1).



Gambar 1. Grafis Prediktor linier dengan sisaan Pearson.

Dari gambar grafis tersebut dapat dilihat bahwa persamaan regresi mencerminkan sebaran dari data dan sisaan yang cenderung tidak berpola. Sehingga dapat dikatakan bahwa terpenuhinya kelayakan model tersebut.

Nilai hasil dugaan dengan penduga Bayes Empirik tanpa peubah penyerta dan dengan peubah penyerta hampir sama. Tetapi kedua penduga tersebut menghasilkan nilai dugaan yang sedikit berbeda jika dibandingkan dengan penduga langsung. Selanjutnya akan ditentukan metode penduga mana yang menghasilkan model terbaik, yaitu dengan membandingkan nilai standar errornya. Tabel (2) berikut menyajikan rata-rata nilai standar error untuk ketiga penduga tersebut.

Tabel 2. Perbandingan rata-rata nilai standar error antara penduga langsung dengan penduga tak langsung (EB).

Rata-rata	Resiko Relatif	Standar error
Penduga Langsung	1,071	0,723
Penduga EB tanpa peubah penyerta	1,006	0,041
Penduga EB dengan peubah penyerta	1,181	0,009

Berdasarkan Tabel (2) hasil estimasi menggunakan SAE dengan penduga Bayes Empirik berbasis model Poisson-Gamma dengan peubah penyerta mempunyai nilai standar error yang lebih kecil dibandingkan dengan pendugaan langsung dan penduga Bayes Empirik tanpa peubah penyerta. Sehingga dapat dikatakan bahwa penduga area kecil (*Small Area Estimation/SAE*) dengan penduga Bayes Empirik berbasis model Poisson-Gamma dengan peubah penyerta merupakan penduga yang lebih baik daripada dengan pendugaan langsung dan penduga Bayes Empirik berbasis model Poisson-Gamma tanpa peubah penyerta, karena memiliki ketelitian yang lebih baik. Hal ini disebabkan oleh peubah penyerta yang dapat dimodelkan secara optimum dengan peubah yang menjadi perhatian.

4. Kesimpulan

Dari pembahasan di atas, diperoleh kesimpulan sebagai berikut:

- (1) Penduga Bayes Empirik berbasis model Poisson-Gamma dengan peubah penyerta.

$$\widehat{\theta}_i^{EB} = E(\theta_i | y_i, \hat{\beta}, \alpha) = \frac{y_i + \hat{\alpha}}{e_i \mu_i + \hat{\alpha}} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i)(\hat{\mu}_i)$$

dengan $\hat{\gamma}_i = \frac{e_i \hat{\mu}_i}{e_i \hat{\mu}_i + \hat{\alpha}}$, $\theta_i = \frac{y_i}{e_i \hat{\mu}_i}$, dan $\mu_i = \exp(\underline{x}_i^T, \underline{\beta})$.

- (2) Penduga Bayes Empirik berbasis Poisson-Gamma dengan peubah penyerta dapat memberikan hasil dugaan yang lebih teliti dibandingkan dengan penduga langsung dan penduga Penduga Bayes Empirik berbasis Poisson-Gamma tanpa peubah penyerta.

Daftar Pustaka

- [1] Bain, L. J., Engelhardt, M., 1992, *Introduction to Probability and Mathematical Statistic*, Edisi Kedua, Duxbuy Press, California
- [2] Hogg, R. V., Craig, A. T., 1905, *Introduction to Mathematical Statistics*, Edisi Ketujuh, Pearson Education, Inc: United States of America
- [3] Rao, J. N. K., 2003, *Small Area Estimation*, New York, Willey
- [4] Kismiantini, 2007, *Penduga Statistik Area Kecil Berbasis Poisson-Gamma*, Tesis, tidak diterbitkan, Institut Pertanian Bogor
- [5] Marshall, R.J., 1991, Mapping disease and mortality rate Using Empirical Bayes Estimators, *Applied Statistics* **40**: 283 – 294
- [6] Wakefield, J., 2006, *Disease Mapping and Spatial Regression with Count Data*, *Biostatistics* **8**(2): 158 – 183
- [7] Walpole, E. R., 1986, *Pengantar Statistika Edisi Ketiga*, PT Gramedia Pustaka Umum, Jakarta
- [8] Walpole, E. R., Myers, R. H., 1995, *Ilmu Peluang Statistika Untuk Insinyur dan Ilmuwan*, Edisi Keempat, ITB, Bandung
- [9] Yanuar, F., 2015, The Use of Uniformative and Informative Prior Distribution in Bayesian SEM, *Global Journal of Pure and Applied Mathematics*, **11**(5) : 3259 – 3256