



# FORUM STATISTIKA DAN KOMPUTASI

Vol. 15 No. 2, Oktober 2009

Pemilihan Model Regresi Linier Multilevel Terbaik ( <i>Choice the Best Linear Regression Multilevel Models</i> )	Bertho Tantular Aunuddin Hari Wijayanto	1 - 7
Metode Prediksi Tak-Bias Linear Terbaik dan bayes Berhirarki untuk Pendugaan Area Kecil Berdasarkan Model State Space ( <i>Best Linear Unbiased Prediction and Hierarchical Bayes Methods for Small Area Estimation Using State Space Models</i> )	Kusman Sadik	8 - 13
Pendugaan Regresi Semiparametrik dengan Pendekatan Model Campuran Linear ( <i>Estimation of Semiparametric Regression using Linear Mixed Model Approach</i> )	Anik Djuraidah	14 - 18
Pemodelan Kalibrasi Peubah Ganda dengan Pendekatan Regresi Sinyal P-spline ( <i>Multivariate Calibration Models using P-spline Signal Regression Approach</i> )	Tonah Ahmad Ansori Mattjik Khairil Anwar Notodiputro	19 - 25
Metode Pendugaan Matriks Ragam-Peragam Dalam Analisis Regresi Komponen Utama ( <i>Variance-Covariance Matrix Estimation Method for Principal Component Regression</i> )	Itasia Dina S. Dian Kusumaningrum Yani Suryani	26 - 34

Departemen Statistika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Institut Pertanian Bogor  
Bogor, Indonesia

## **METODE PREDIKSI TAK-BIAS LINEAR TERBAIK DAN BAYES BERHIRARKI UNTUK PENDUGAAN AREA KECIL BERDASARKAN MODEL STATE SPACE**

*(Best Linear Unbiased Prediction and Hierarchical Bayes Methods  
for Small Area Estimation Using State Space Models)*

Kusman Sadik

Departemen Statistika FMIPA Institut Pertanian Bogor

### **ABSTRACT**

*There have been two main topics developed by statisticians in a survey, i.e. sampling techniques and estimation methods. The current issues in estimation methods related to estimation of a particular domain having small size of samples or, in more extreme cases, there is no sample available for direct estimation. Sample survey data provide effective reliable estimators of totals and means for large area and domains. But it is recognized that the usual direct survey estimator performing statistics for a small area, have unacceptably large standard errors, due to the circumstance of small sample size in the area. The most commonly used models for this case, usually in small area estimation, are based on generalized linear mixed models. Some time happened that some surveys are carried out periodically so that the estimation could be improved by incorporating both the area and time random effects. In this paper we propose a state space model which accounts for the two random effects and is based on two equation, namely transition equation and measurement equation. Based on a evaluation criterion, the proposed hierarchical Bayes estimator turns out to be superior to both estimated best linear unbiased prediction (BLUP) and the direct survey estimator. The posterior variances which measure accuracy of the hierarchical Bayes estimates are always smaller than the corresponding variances of the BLUP and the direct survey estimates.*

**Key words:** *Generalized linear mixed model, hierarchical Bayes, best linear unbiased prediction, prior and posterior function, generalized variance function, block diagonal covariance, state space model.*

### **PENDAHULUAN**

Survei rutin yang dilakukan oleh pemerintah suatu negara, umumnya didesain untuk memperoleh statistik nasional. Artinya, survei semacam ini didesain untuk inferensia bagi daerah yang luas. Persoalan muncul ketika dari survei seperti ini ingin diperoleh informasi untuk area yang lebih kecil, misalnya informasi pada level propinsi, kabupaten, bahkan mungkin level kecamatan. Ukuran contoh pada level area tersebut biasanya sangat kecil sehingga statistik yang diperoleh akan memiliki ragam yang besar.

Bahkan bisa saja pendugaan tidak dapat dilakukan karena area tersebut tidak terpilih menjadi contoh dalam survei. Oleh karena itu dikembangkan metode pendugaan parameter yang dapat mengatasi hal ini. Metode tersebut dikenal dengan nama metode pendugaan area kecil (*small area estimation*, SAE).

Statistik area kecil semacam itu telah menjadi perhatian para statistisi dunia secara sangat serius sejak sepuluh tahun terakhir ini (misalnya Ghosh and Rao, 1994; Rao dan Yu, 1994; Rao, 2003; Russo *et.al.*, 2005; Chambers dan Chandra, 2006; Lahiri, 2008). Berbagai metode SAE telah dikembangkan khususnya menyangkut metode yang berbasis model. Area kecil tersebut didefinisikan sebagai himpunan bagian dari populasi dimana suatu peubah menjadi perhatian. Pendekatan klasik untuk menduga parameter area kecil didasarkan pada aplikasi model desain penarikan contoh (*design-based*) yang dikenal sebagai pendugaan langsung (*direct estimation*). Namun, metode pendugaan langsung pada sub-populasi tidak memiliki presisi yang memadai karena kecilnya jumlah contoh yang digunakan untuk memperoleh dugaan tersebut.

Menurut Lahiri (2008), metode pendugaan tidak langsung pada area kecil pada dasarnya memanfaatkan informasi area sekitarnya (*neighbouring areas*) dan sumber data di luar area yang statistiknya ingin diperoleh. Dalam hal ini, model dikembangkan dengan asumsi bahwa keragaman didalam area kecil peubah respon dapat diterangkan oleh hubungan keragaman yang bersesuaian pada informasi penyerta (*auxiliary*) yang berupa pengaruh tetap, sedangkan keragaman spesifik area kecil diasumsikan dapat diterangkan oleh informasi tambahan yang berupa pengaruh acak area.

Menurut Sadik *et.al.* (2008), model pendugaan parameter area kecil akan sangat membantu khususnya dalam menyediakan kebutuhan data dan informasi yang akurat untuk kebutuhan daerah di Indonesia seperti level propinsi, kabupaten/kota atau bahkan kecamatan dengan memanfaatkan keakuratan data BPS pada level nasional, tanpa harus mengeluarkan biaya besar untuk mengumpulkan data sendiri. Dengan demikian, secara nasional akan cukup banyak biaya yang bisa dihemat sehingga dapat dialokasikan untuk pembiayaan pembangunan lainnya.

Penelitian ini bertujuan untuk mengkaji konsep dan sifat-sifat statistik model SAE yang memasukkan pengaruh acak area dan waktu (*state-space*) melalui pendekatan prediksi tak-bias linear terbaik / PTLT (*best linear unbiased prediction* / BLUP) dan Bayes berhirarki.

## MODEL DASAR SAE

Model SAE yang didasarkan pada pendugaan tidak langsung dapat menggunakan pendekatan model secara umum. Misalkan diasumsikan bahwa  $\theta_i = g(\bar{Y}_i)$  untuk beberapa spesifikasi  $g(\cdot)$  dihubungkan dengan data penyerta spesifik pada area  $i$ ,  $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})^T$  melalui suatu model linear

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + b_i v_i, \quad i = 1, \dots, m \quad (1)$$

dimana  $b_i$  adalah konstanta positif yang diketahui dan  $\boldsymbol{\beta}$  adalah vektor berukuran  $p \times 1$ . Sedangkan  $v_i$  adalah pengaruh acak spesifikasi area yang diasumsikan bebas dan menyebar identik dengan  $E_m(v_i) = 0$  dan  $V_m(v_i) = \sigma_v^2 (\geq 0)$ , atau  $v_i \sim \text{iid} (0, \sigma_v^2)$ . Pendugaan tidak langsung untuk rata-rata populasi di area kecil  $i$ , diperlukan informasi mengenai penduga langsungnya yaitu  $\hat{\theta}_i$ . Dengan menggunakan metode James-Stein akan diperoleh:

$$\hat{\theta}_i = \theta_i + e_i \quad (2)$$



$$\hat{\theta}_i = \mathbf{x}_i^T \boldsymbol{\beta} + b_i v_i + e_i, \quad i = 1, \dots, m \quad (3)$$

dimana galat penarikan contoh (*sampling error*)  $e_i$  adalah bebas dengan  $E_p(e_i|\theta_i) = 0$  dan  $V_p(e_i|\theta_i) = \psi_i$ , atau  $v_i \sim iid N(0, \sigma_v^2)$ .

### PENGEMBANGAN MODEL SAE

Model SAE yang dipaparkan di atas hanya memasukkan pengaruh acak area untuk memperkecil kuadrat tengah galatnya. Namun, terkadang beberapa survei diulang pelaksanaannya secara berkala, misalnya Susenas. Untuk survei yang demikian ini pendugaan tidak langsung di suatu area kecil dapat ditingkatkan efisiensi pendugaannya dengan memasukkan pengaruh acak area dan waktu (Rao dan Yu, 1994; Rao, 2003). Model tersebut terdiri dari model galat penarikan contoh

$$\hat{\theta}_{it} = \theta_{it} + e_{it}, \quad t = 1, \dots, T; \quad i = 1, \dots, m \quad (4)$$

dan model penghubung (*linking model*)

$$\theta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + v_i + u_{it} \quad (5)$$

Pada model di atas,  $\hat{\theta}_{it}$  adalah penduga langsung dari survei untuk area kecil  $i$  pada waktu  $t$ ,  $\theta_{it}$  adalah fungsi untuk rata-rata pada area  $i$  pada waktu  $t$ , dan  $e_{it}$  adalah galat penarikan contoh yang menyebar normal dengan nilai harapan 0 dan mempunyai matriks koragam blok-diagonal  $\boldsymbol{\Psi}$ . Sedangkan  $\mathbf{x}_{it}$  adalah vektor spesifikasi area yang mungkin berubah menurut waktu  $t$ . Sebagaimana diasumsikan sebelumnya bahwa  $v_i \sim iid(0, \sigma_v^2)$  dan  $u_{it}$  diasumsikan mengikuti proses sebagai berikut:

$$\begin{aligned} u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \text{ dan} \\ \varepsilon_{it} &\sim iid N(0, \sigma^2) \end{aligned} \quad (6)$$

Berdasarkan model di atas,  $\theta_{it}$  tergantung pada pengaruh area spesifik  $v_i$  dan area oleh waktu spesifik  $u_{it}$  yang berkorelasi antar waktu. Model tersebut merupakan bentuk umum model *state space* sebagai berikut:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t; & E(\boldsymbol{\varepsilon}_t) &= 0, & E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T) &= \boldsymbol{\Sigma}_t \\ \boldsymbol{\alpha}_t &= \mathbf{H}_t \boldsymbol{\alpha}_{t-1} + \mathbf{A}_t \boldsymbol{\eta}_t; & E(\boldsymbol{\eta}_t) &= 0, & E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^T) &= \boldsymbol{\Gamma} \end{aligned} \quad (7)$$

dimana  $\boldsymbol{\varepsilon}_t$  dan  $\boldsymbol{\eta}_t$  tidak berkorelasi antar waktu. Persamaan pertama disebut sebagai persamaan transisi (*transition equation*), sedangkan persamaan kedua disebut persamaan pengukuran (*measurement equation*) yang terkait dengan persamaan pertama (Pfeffermann, 2002; Durbin, 2004; Zivot, 2006). Model ini pada dasarnya merupakan kasus khusus model campuran linear terampat / MCLT (*generalized linear mixed model* / GLMM) tetapi dengan melakukan *updating* dugaan pada tiap waktu, yaitu

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i^T + \mathbf{e}_i \quad (8)$$

dimana  $\mathbf{y}_i = \hat{\theta}_i$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})^T$ ,  $\mathbf{Z}_i = (\mathbf{1}_T, \mathbf{I}_T)$ ,  $\mathbf{v}_i^T = (v_i, \mathbf{u}_i^T)$ ,  $\mathbf{e}_i = (\mathbf{e}_{i1}, \dots, \mathbf{e}_{iT})^T$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ ,  $\mathbf{1}_T$  adalah  $T \times 1$  vektor 1 dan  $\mathbf{I}_T$  adalah matriks identitas berordo  $T$ . Sedangkan

$$\mathbf{G}_i = \begin{bmatrix} \sigma_v^2 & \mathbf{0}^T \\ \mathbf{0} & \sigma^2 \boldsymbol{\Lambda} \end{bmatrix}, \quad \mathbf{R}_i = \boldsymbol{\Psi}_i \quad (9)$$

dengan  $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}$  adalah  $T \times T$  matriks koragam  $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})^T$  dengan elemen ke-( $t,s$ )  $\rho^{|t-s|}/(1 - \rho^2)$ .

## Metode Prediksi Tak-bias Linear Terbaik (PTLT)

Model linear campuran terampat yang digunakan sebagai model dasar area kecil dengan pengaruh acak area dan waktu (*state space*) di atas dapat pula dinyatakan sebagai berikut  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}$ , dimana  $\mathbf{y} = \hat{\theta}_i$ ,  $\mathbf{X} = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})^T$ ,  $\mathbf{Z} = (\mathbf{1}_T, \mathbf{I}_T)$ ,  $\mathbf{v}^T = (v_i, \mathbf{u}_i^T)$ ,  $\mathbf{e} = (\mathbf{e}_{i1}, \dots, \mathbf{e}_{iT})^T$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ ,  $\mathbf{1}_T$  adalah  $T \times 1$  vektor 1 dan  $\mathbf{I}_T$  adalah matriks identitas berordo  $T$ . Pada persamaan di atas,  $\mathbf{y}$  adalah vektor dari pengamatan contoh,  $\mathbf{X}$  dan  $\mathbf{Z}$  adalah matriks yang diketahui, dan  $\mathbf{v}$  dan  $\mathbf{e}$  adalah saling bebas dengan nilai tengah  $\mathbf{0}$  dan matriks koragam  $\mathbf{G}$  dan  $\mathbf{R}$ , tergantung pada beberapa parameter  $\boldsymbol{\phi}$  yang disebut sebagai komponen ragam.

Ragam bagi  $\mathbf{y}$  yaitu  $\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}^T$ . Penduga linear yang diinginkan adalah bersifat tidak bias dan memiliki ragam minimum. Suatu penduga linear  $\hat{\mu} = \mathbf{a}^T\mathbf{y} + b$  adalah tidak bias bagi  $\mu = \mathbf{1}^T\boldsymbol{\beta} + \mathbf{m}^T\mathbf{v}$  berdasarkan model campuran di atas, yaitu  $E(\hat{\mu}) = E(\mu)$  jika dan hanya jika  $\mathbf{a}^T\mathbf{X} = \mathbf{1}^T$  dan  $b = 0$ . Nilai kuadrat tengah galat / KTG (*mean square error* / MSE) bagi penduga tidak bias linear  $\hat{\mu}$  dapat dinyatakan sebagai  $\text{KTG}(\hat{\mu}) = \text{KTG}(\hat{\mu}) = \mathbf{a}^T\mathbf{V}\mathbf{a} - 2\mathbf{a}^T\mathbf{Z}\mathbf{G}\mathbf{m} + \mathbf{m}^T\mathbf{G}\mathbf{m}$

Selanjutnya  $\text{KTG}(\hat{\mu})$  ini diminimumkan dengan kendala  $\mathbf{a}^T\mathbf{X} = \mathbf{1}^T$ . Penyelesaiannya menggunakan metode pengganda Lagrange, yaitu

$$L(\mathbf{a}, \boldsymbol{\lambda}) = \mathbf{a}^T\mathbf{V}\mathbf{a} - 2\mathbf{a}^T\mathbf{Z}\mathbf{G}\mathbf{m} + \mathbf{m}^T\mathbf{G}\mathbf{m} + 2(\mathbf{a}^T\mathbf{X} - \mathbf{1}^T)\boldsymbol{\lambda} \quad (10)$$

Apabila persamaan di atas disubstitusikan ke dalam kendala  $\mathbf{a}^T\mathbf{X} = \mathbf{1}^T$ , maka akan diperoleh penyelesaian bagi  $\boldsymbol{\lambda}$  yaitu  $\boldsymbol{\lambda} = -(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{1} + (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}\mathbf{m}$

Penduga  $h(\boldsymbol{\phi}, \mathbf{y})$  tergantung pada komponen ragam  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^T$ , tetapi di dalam praktik nilai komponen  $\boldsymbol{\phi}$  tidak akan diketahui, sehingga PTLT harus dilakukan secara empirik (PTLTE). Misalkan  $h(\boldsymbol{\phi}, \mathbf{y})$  dituliskan sebagai  $h(\boldsymbol{\phi})$  dan  $h(\hat{\boldsymbol{\phi}}, \mathbf{y})$  dituliskan sebagai  $h(\hat{\boldsymbol{\phi}})$ , maka MSE bagi  $h(\hat{\boldsymbol{\phi}})$  dapat dinyatakan sebagai berikut:

$$\text{KTG}[h(\hat{\boldsymbol{\phi}})] = \text{KTG}[h(\boldsymbol{\phi})] + E[h(\hat{\boldsymbol{\phi}}) - h(\boldsymbol{\phi})]^2, \quad (11)$$

Melalui pendekatan deret Taylor diperoleh persamaan bahwa  $E[h(\hat{\boldsymbol{\phi}}) - h(\boldsymbol{\phi})]^2 = E[\mathbf{d}(\boldsymbol{\phi})^T(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})]^2$  dengan  $\mathbf{d}(\boldsymbol{\phi}) = \partial h(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$ , selanjutnya dapat dinyatakan

$$E[\mathbf{d}(\boldsymbol{\phi})^T(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})]^2 = \text{tr}[\mathbf{A}(\boldsymbol{\phi})E(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})^T], \quad (12)$$

dimana  $\mathbf{A}(\boldsymbol{\phi})$  adalah matriks koragam dari  $\mathbf{d}(\boldsymbol{\phi})$ .

$$\text{tr}[\mathbf{A}(\hat{\boldsymbol{\phi}})E(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})^T] = \text{tr}[(\nabla\mathbf{b})\mathbf{V}(\nabla\mathbf{b}^T)E(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})^T], \quad (13)$$

dimana  $\nabla\mathbf{b}^T = \text{col}_{1 \leq j \leq p}(\partial\mathbf{b}^T/\partial\phi_j)$  dan  $\mathbf{b}^T = \mathbf{m}^T\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}$ . Selanjutnya dapat diperoleh bahwa

$$\text{KTG}[h(\hat{\boldsymbol{\phi}})] = \text{KTG}[h(\boldsymbol{\phi})] + \text{tr}[(\nabla\mathbf{b}^T)\mathbf{V}(\nabla\mathbf{b}^T)^TE(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})^T], \quad (14)$$

## Metode Bayes Berhirarki

Metode PTLT yang dipaparkan di atas memerlukan beberapa kondisi tertentu yang harus dipenuhi, diantaranya adalah pengasumsian bahwa parameter peubah tetap ( $\boldsymbol{\beta}$ ) dan ragam penarikan contohnya adalah konstanta atau tidak memiliki fungsi sebaran tertentu (Prasad dan Rao, 1990; Schall, 1991; Solomon dan Cox, 1992). Padahal kenyataannya, sangat dimungkinkan parameter tersebut bukan suatu konstanta, melainkan memiliki suatu fungsi sebaran tertentu. Metode Bayes dapat digunakan untuk

mengatasi persoalan tersebut. Sehingga metode Bayes lebih fleksibel daripada metode PTLT yang memerlukan berbagai kondisi tertentu.

Pada model SAE yang memasukkan pengaruh acak area dan waktu, yaitu  $y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + v_i + u_{it} + e_{it}$  dan  $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ , dimana galat penarikan contoh  $e_{it}$  diasumsikan menyebar normal dengan nilai tengah nol dan matriks koragam blok-diagonal  $\boldsymbol{\Sigma}$  dengan blok  $\boldsymbol{\Sigma}_i$  (matriks berukuran  $T \times T$ ) diketahui. Matriks koragam  $\boldsymbol{\Sigma}_i$  untuk setiap area ke- $i$  dan setiap waktu  $t$  dapat diperoleh melalui metode *generalized variance function* (GVF). Model tersebut dapat pula dinyatakan dalam bentuk matriks sebagai berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{u} + \mathbf{e} \quad (15)$$

dengan

$$\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_m^T)^T, \quad \mathbf{X}_i^T = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}), \quad \mathbf{Z} = \mathbf{I}_m \otimes \mathbf{1}_T,$$

$$\mathbf{v} = (v_1, \dots, v_m)^T, \quad \mathbf{u} = (\mathbf{u}_1^T, \dots, \mathbf{u}_m^T)^T, \quad \mathbf{e} = (\mathbf{e}_1^T, \dots, \mathbf{e}_m^T)^T$$

dimana  $\mathbf{u}_i^T = (u_{i1}, \dots, u_{iT})$ ,  $\mathbf{e}_i^T = (e_{i1}, \dots, e_{iT})$ ,  $\mathbf{1}_T$  adalah vektor satuan, dan  $\mathbf{I}_m$  adalah matriks identitas berordo  $m$ , dan  $\otimes$  adalah perkalian Kronecker.  $E(\mathbf{v}) = \mathbf{0}$ ,  $\text{Cov}(\mathbf{v}) = \sigma_v^2 \mathbf{I}_m$ ,  $E(\mathbf{u}) = \mathbf{0}$ ,  $\text{Cov}(\mathbf{u}) = \sigma^2 \mathbf{I}_m \otimes \boldsymbol{\Gamma} = \sigma^2 \mathbf{R}$ ,  $E(\mathbf{e}) = \mathbf{0}$ ,  $\text{Cov}(\mathbf{e}) = \boldsymbol{\Sigma} = \text{blok-diagonal}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m)$ , dan  $v$ ,  $u$ , dan  $e$  adalah bebas,  $\boldsymbol{\Gamma}$  adalah matriks  $T \times T$  dengan elemen  $\rho^{|i-j|}/(1-\rho)$ . Berdasarkan persamaan tersebut maka  $\text{Cov}(\mathbf{y}) = \mathbf{V} = \boldsymbol{\Sigma} + \sigma^2 \mathbf{R} + \sigma_v^2 \mathbf{Z}\mathbf{Z}^T = \text{blok-diagonal}(\boldsymbol{\Sigma}_i + \sigma^2 \boldsymbol{\Gamma} + \sigma_v^2 \mathbf{J}_T) = \text{blok-diagonal}(\mathbf{V}_i)$  dengan  $\mathbf{J}_T = \mathbf{1}_T \mathbf{1}_T^T$ .

Pada metode Bayes berhirarki, untuk  $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}$  maka  $\text{Var}(\boldsymbol{\theta}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}) = \text{Var}(\mathbf{Z}\mathbf{v}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T$ , maka model di atas dapat dituliskan menjadi

$$\text{Level 1 : } \mathbf{Y} \mid \boldsymbol{\theta} \sim \mathbf{N}(\boldsymbol{\theta}, \mathbf{R})$$

$$\text{Level 2 : } \boldsymbol{\theta} \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}^T) \quad (16)$$

$$\text{Level 3 : } f(\mathbf{R}) \propto 1, f(\mathbf{G}) \propto 1, \text{ dan } f(\boldsymbol{\beta}) \propto 1$$

Nilai tengah area kecil pada satu waktu  $T$ ,  $\theta_{iT} = (\mathbf{x}_{iT}^T \boldsymbol{\beta} + v_i + u_{iT})$ , merupakan kasus khusus dari kombinasi linear  $\tau = \mathbf{l}^T \boldsymbol{\beta} + \mathbf{l}_1^T \mathbf{v} + \mathbf{l}_2^T \mathbf{u}$  yang dapat diduga berdasarkan sebaran posteriornya pada pendekatan Bayes berhirarki dimana  $\mathbf{l} = \mathbf{x}_{iT}$ ,  $\mathbf{l}_1$  adalah vektor  $m$  dengan nilai 1 pada posisi ke- $i$  dan 0 pada selainnya, serta  $\mathbf{l}_2$  adalah vektor  $mT$  dengan nilai 1 pada posisi ke- $iT$  dan 0 pada selainnya (Rao dan You, 2002; Sadik, 2008). Berdasarkan asumsi kenormalan galat  $v_i$ ,  $u_{it}$ , dan  $e_{it}$ , maka

$$\text{KTG}[\hat{\theta}_{iT}(\rho)] = \text{KTG}(\tilde{\theta}_{iT}) + E\{\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)\}^2,$$

$$= g_{1iT}(\sigma^2, \sigma_v^2, \rho) + g_{2iT}(\sigma^2, \sigma_v^2, \rho) + g_{3iT}(\sigma^2, \sigma_v^2, \rho) \quad (17)$$

dimana

$$g_{1iT}(\sigma^2, \sigma_v^2, \rho) = \sigma_v^2 + \frac{\sigma^2}{1-\rho^2} - (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)^T \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T) \quad (18)$$

$$g_{2iT}(\sigma^2, \sigma_v^2, \rho) = \{\mathbf{x}_{iT} - \mathbf{X}_i^T \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\}^T (\mathbf{X}^T \mathbf{V}_i^{-1} \mathbf{X})^{-1}$$

$$\times \{\mathbf{x}_{iT} - \mathbf{X}_i^T \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\} \quad (19)$$

$$E\{\tilde{\theta}_{iT} - \hat{\theta}_{iT}(\rho)\}^2 \approx \text{tr}(\boldsymbol{\Omega}^T \mathbf{V} \boldsymbol{\Omega} \boldsymbol{\Psi}) = g_{3iT}(\sigma^2, \sigma_v^2, \rho), \quad (20)$$

$\boldsymbol{\Psi}$  adalah matriks koragam  $2 \times 2$  dari penduga tidak bias  $\tilde{\sigma}^2(\rho)$  dan  $\tilde{\sigma}_v^2(\rho)$ , dan  $\boldsymbol{\Omega} = (\partial \mathbf{b} / \partial \sigma^2, \partial \mathbf{b} / \partial \sigma_v^2)$  dengan  $\mathbf{b}^T = (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)^T \mathbf{V}_i^{-1}$ . Penghitungan turunan  $\partial \mathbf{b} / \partial \sigma^2$  dan  $\partial \mathbf{b} / \partial \sigma_v^2$  dapat diperoleh dengan menyederhanakan  $\boldsymbol{\Omega}^T \mathbf{V} \boldsymbol{\Omega} = \mathbf{A}$ , dimana  $\mathbf{A}$  adalah matriks simetrik  $2 \times 2$  dengan elemen

$$a_{11} = \{\boldsymbol{\gamma}_T - \boldsymbol{\Gamma} \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\}^T \mathbf{V}_i^{-1} \{\boldsymbol{\gamma}_T - \boldsymbol{\Gamma} \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\}$$

$$a_{22} = \{\mathbf{1}_T - \mathbf{J}_T \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\}^T \mathbf{V}_i^{-1} \{\mathbf{1}_T - \mathbf{J}_T \mathbf{V}_i^{-1} (\sigma_v^2 \mathbf{1}_T + \sigma^2 \boldsymbol{\gamma}_T)\}$$

$$a_{12} = a_{21} = \{\gamma_T - \Gamma \mathbf{V}_i^{-1}(\sigma_v^2 \mathbf{1}_T + \sigma^2 \gamma_T)\}^T \mathbf{V}_i^{-1} \{\mathbf{1}_T - \mathbf{J}_T \mathbf{V}_i^{-1}(\sigma_v^2 \mathbf{1}_T + \sigma^2 \gamma_T)\} \quad (21)$$

Pada model SAE dengan pengaruh acak area dan waktu berdasarkan model *state space* di atas,  $\hat{\theta}_{it} = \theta_{it} + e_{it}$ ,  $\theta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + v_i + u_{it}$ , dan  $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ , dan  $\varepsilon_{it} \sim iid N(0, \sigma^2)$ ,  $t = 1, \dots, T$ ;  $i = 1, \dots, m$  serta  $v_i \sim iid N(0, \sigma_v^2)$ , ada dua pendekatan Bayes berhirarki yang dikaji dalam penelitian ini (Sadik dan Notodiputro, 2009), yaitu

Model BB-1:

- Level 1 :  $\hat{\theta}_{it} | \theta_{it} \sim N(\theta_{it}, \psi_{it})$ ,  $\psi_{it}$  diketahui
- Level 2 :  $\theta_{it} | \boldsymbol{\beta}, u_{it} \sim N(\mathbf{x}_{it}^T \boldsymbol{\beta} + u_{it}, \sigma_v^2)$
- Level 3 :  $u_{it} | u_{i,t-1} \sim N(u_{i,t-1}, \sigma^2)$
- Level 4 :  $\boldsymbol{\beta}$  menyebar *flat*( $-\infty, \infty$ ),  $\sigma_v^2$  dan  $\sigma^2$  menyebar Gamma

Model BB-2:

- Level 1 :  $\hat{\theta}_{it} | \theta_{it}, \sigma_{it}^2 \sim N(\theta_{it}, [1-f]\sigma_{it}^2/n_{it})$
- Level 2 :  $([n_{it}-1]s_{it}^2)/\sigma_{it}^2 \sim \chi^2(n_{it}-1)$
- Level 3 :  $\theta_{it} | \boldsymbol{\beta}, u_{it} \sim N(\mathbf{x}_{it}^T \boldsymbol{\beta} + u_{it}, \sigma_v^2)$
- Level 4 :  $u_{it} | u_{i,t-1} \sim N(u_{i,t-1}, \sigma^2)$
- Level 5 :  $\boldsymbol{\beta}$  menyebar *flat*( $-\infty, \infty$ ),  $\sigma_v^2$  dan  $\sigma^2$  menyebar Gamma

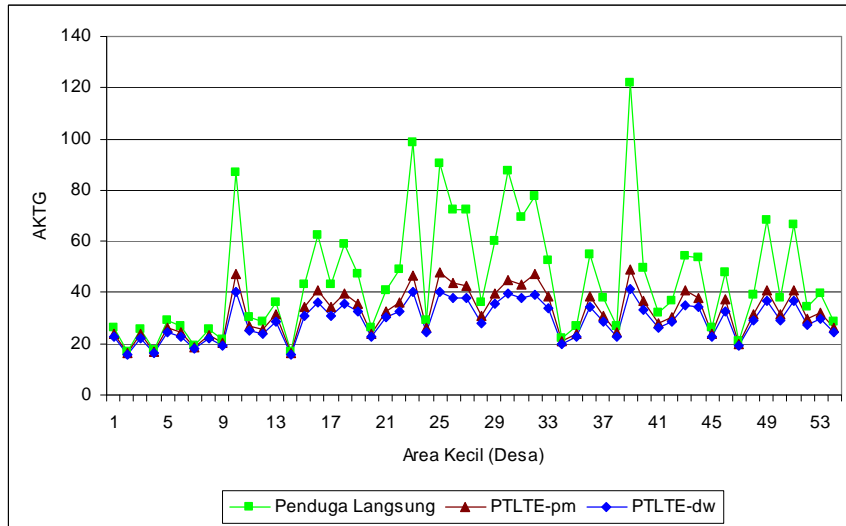
## PENERAPAN PADA DATA

Sebagai kajian empirik, data yang digunakan adalah data aktual dan data simulasi. Sumber data aktual adalah data Survei Sosial-Ekonomi Nasional (Susenas) 2001-2005 khususnya yang menyangkut rata-rata pengeluaran perkapita perbulan di wilayah kota Bogor, serta data Potensi Desa (Podes) yang digunakan sebagai salah satu sumber data bagi pengaruh tetap pada tiap desa yang diamati. Data simulasi digunakan untuk mengkaji sifat-sifat statistik dari model yang digunakan pada beberapa kondisi tertentu.

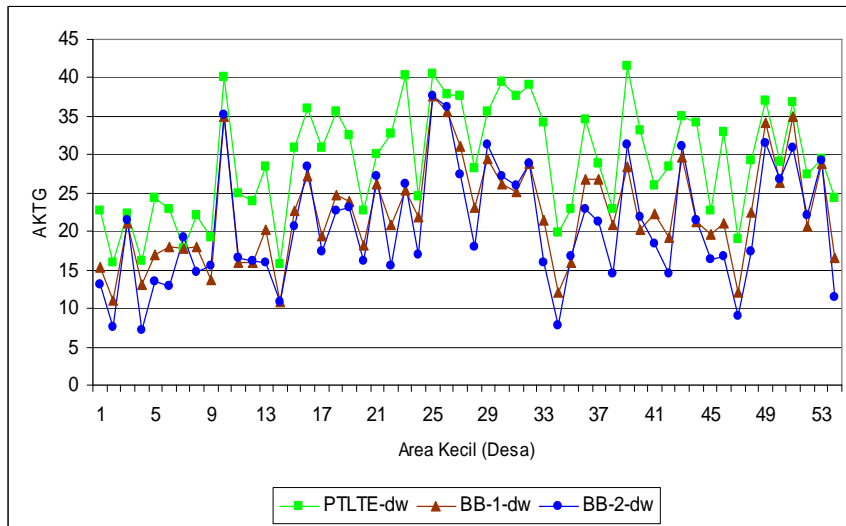
Pada Gambar 1, terlihat bahwa untuk rata-rata pengeluaran perkapita pada data Susenas wilayah Kota Bogor, nilai akar kuadrat tengah galat / AKTG (*root mean square error / RMSE*) lebih kecil pada metode PTLT empirik / PTLTE (*empirical best linear unbiased prediction / EBLUP*) data penampang melintang (*cross sectional data*) dibandingkan dengan AKTG pada metode pendugaan langsung. Sementara AKTG pada metode PTLTE data deret waktu cenderung lebih kecil daripada metode PTLTE data penampang melintang. Ini menunjukkan bahwa pengaruh acak area maupun pengaruh sintetik vektor kovariat berfungsi mengkalibrasi hasil pendugaan langsung yang hanya didasarkan pada data survei semata. Selanjutnya, pengaruh acak area dan waktu maupun pengaruh sintetik vektor kovariat berfungsi memperbaiki hasil pendugaan metode PTLTE yang hanya didasarkan pada data survei pada satu tahun saja.

Perbandingan antara metode PTLTE dengan Bayes berhirarki untuk data deret waktu rata-rata pengeluaran perkapita pada data Susenas wilayah Kota Bogor disajikan pada Gambar 2. Berdasarkan hasil tersebut terlihat bahwa AKTG cenderung lebih kecil pada metode Bayes berhirarki model 2 dibandingkan dengan AKTG pada metode Bayes berhirarki model 1 maupun metode PTLTE. Hal ini mengindikasikan bahwa metode Bayes berhirarki model 2 lebih baik daripada model 1 dalam menurunkan AKTG. Namun demikian, penurunan AKTG tersebut tidak sebesar penurunan yang terjadi ketika metode PTLTE dibandingkan dengan Bayes berhirarki.

Metode Bayes berhirarki model 2 pada dasarnya merupakan kalibrasi terhadap model 1, dimana pada model 1 ragam penarikan contoh dianggap diketahui sedangkan pada model 2 dianggap tidak diketahui. Sebagaimana dijabarkan di atas, penurunan AKTG ini sebagai akibat adanya penguraian komponen ragam yang terdapat di dalam model, termasuk komponen ragam yang diakibatkan oleh fluktuasi tingkat pengeluaran perkapita antar tahun.



Gambar 1. Plot AKTG untuk Penduga Langsung, PTLTE Data Penampang Melintang, dan PTLTE Data Deret Waktu – Susenas Wilayah Kota Bogor



Gambar 2. Plot AKTG untuk PTLTE, Bayes Berhirarki Model-1 dan Model-2 Data Deret Waktu - Susenas Wilayah Kota Bogor

Data simulasi digunakan untuk mengetahui pengaruh waktu ( $T$ ) dan korelasi diri ( $\rho$ ) pada model SAE berdasarkan model *state space* yang diterapkan pada data Susenas. Proses simulasi berdasarkan model di atas, yaitu  $\hat{\theta}_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + v_i + u_{it} = \theta_{it} + v_i + u_{it}$  dan  $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ . Simulasi dilakukan untuk banyaknya waktu ( $T$ ) yang berbeda yaitu  $T = 3, 5, \text{ dan } 7$ , dengan lima nilai korelasi diri yaitu  $\rho = 0.01, 0.25, 0.45, 0.60, \text{ dan } 0.85$ . Selanjutnya dibandingkan relatif bias dan nilai AKTG antara metode PTLTE dengan metode Bayes berhirarki untuk model 1 dengan model 2, yang hasilnya disajikan pada Tabel 1 dan Tabel 2.

Pada Tabel 1, BB-1 adalah metode Bayes berhirarki model 1 dan BB-2 adalah metode Bayes berhirarki model 2. Berdasarkan tabel tersebut terlihat bahwa secara umum bias relatif mutlak tidak berbeda antara metode PTLTE dengan Bayes berhirarki model 1 dan 2. Meskipun pada waktu  $T$  yang lebih besar dan korelasi diri yang juga



besar, bias relatif mutlak Bayes berhirarki model 2 cenderung lebih kecil daripada dua pendekatan lainnya.

Tabel 1. Nilai Rata-rata Bias Relatif Mutlak (ARB) pada Metode PTLTE dan Bayes Berhirarki untuk Data Simulasi

Korelasi Diri ( $\rho$ )	$T = 3$		
	PTLTE	BB-1	BB-2
0.01	0.0000104	0.0000097	0.0000089
0.25	0.0000111	0.0000112	0.0000098
0.45	0.0000097	0.0000107	0.0000117
0.60	0.0000122	0.0000101	0.0000106
0.85	0.0000109	0.0000121	0.0000111
$\rho$	$T = 5$		
	PTLTE	BB-1	BB-2
0.01	0.0000103	0.0000108	0.0000101
0.25	0.0000112	0.0000113	0.0000099
0.45	0.0000104	0.0000108	0.0000118
0.60	0.0000121	0.0000100	0.0000105
0.85	0.0000108	0.0000120	0.0000110
$\rho$	$T = 7$		
	PTLTE	BB-1	BB-2
0.01	0.0000102	0.0000098	0.0000105
0.25	0.0000116	0.0000114	0.0000100
0.45	0.0000099	0.0000109	0.0000119
0.60	0.0000119	0.0000099	0.0000104
0.85	0.0000107	0.0000118	0.0000109

Tabel 2. Nilai Rata-rata AKTGR pada Metode PTLTE dan Bayes Berhirarki untuk Data Simulasi

Korelasi Diri ( $\rho$ )	$T = 3$		
	PTLTE	BB-1	BB-2
0.01	0.2034	0.1823	0.1141
0.25	0.2029	0.1824	0.0991
0.45	0.1954	0.1613	0.0825
0.60	0.1957	0.1327	0.0626
0.85	0.1903	0.1178	0.0311
$\rho$	$T = 5$		
	PTLTE	BB-1	BB-2
0.01	0.2034	0.1825	0.1075
0.25	0.2021	0.1734	0.1002
0.45	0.1913	0.1507	0.0813
0.60	0.1825	0.1328	0.0527
0.85	0.1677	0.1121	0.0312
$\rho$	$T = 7$		
	PTLTE	BB-1	BB-2
0.01	0.1989	0.1751	0.0987
0.25	0.1988	0.1593	0.0965
0.45	0.1823	0.1345	0.0701
0.60	0.1768	0.1278	0.0364
0.85	0.1609	0.1093	0.0297

Berdasarkan hasil pada Tabel 2, nilai akar kuadrat tengah galat relatif / AKTGR (*relative of root mean square error / RRMSE*) metode Bayes berhirarki model 2 cenderung lebih kecil dibandingkan dengan nilai AKTGR metode PTLTE dan Bayes

berhirarki model 1 pada waktu  $T$  dan korelasi diri  $\rho$  yang sama-sama besar. Artinya, pengaruh pengamatan antar waktu yang diakibatkan oleh banyaknya waktu  $T$  dan korelasi diri  $\rho$  dapat memperbaiki pendugaan parameter pada area kecil yang diindikasikan dengan menurunnya nilai AKTGR tersebut.

## SIMPULAN DAN SARAN

### Simpulan

Metode PTLT pada SAE memerlukan banyak kondisi tertentu yang harus dipenuhi, diantaranya adalah pengasumsian bahwa parameter peubah tetap dan ragam penarikan contohnya adalah konstanta atau tidak memiliki fungsi sebaran tertentu. Padahal kenyataannya, sangat dimungkinkan parameter tersebut bukan suatu konstanta, melainkan memiliki suatu fungsi sebaran tertentu. Metode Bayes dapat digunakan untuk mengatasi persoalan tersebut. Sehingga metode Bayes lebih fleksibel daripada metode PTLT yang memerlukan berbagai kondisi tertentu.

Penerapan pada data deret waktu Susenas, diketahui bahwa nilai AKTG lebih kecil pada metode pemodelan area kecil Bayes berhirarki dibandingkan dengan AKTG pada metode PTLTE. Hal ini mengindikasikan bahwa metode Bayes berhirarki lebih baik daripada PTLTE dalam menurunkan AKTG. Penurunan AKTG ini sebagai akibat adanya penguraian komponen ragam yang terdapat di dalam model, termasuk komponen ragam yang diakibatkan oleh fluktuasi tingkat pengeluaran perkapita antar tahun.

Berdasarkan kajian analitik pada model SAE yang memasukkan pengaruh acak area dan waktu, penduga PTLT yang berbasis pada metode kemungkinan maksimum terbatas dan penduga Bayes berhirarki sama-sama bersifat tidak bias. Namun, resiko atau KTG yang dihasilkan penduga Bayes lebih kecil daripada penduga PTLT, artinya dalam hal ini penduga Bayes lebih baik daripada penduga PTLT.

### Saran

Perlu penelitian lanjutan untuk mengembangkan metode ini di Indonesia khususnya beberapa keterbatasan yang terdapat dalam penelitian ini, yaitu (a) Model SAE yang dikaji dalam penelitian ini hanya mencakup peubah respon kontinu belum mencakup peubah respon diskret; (b) Model yang digunakan memasukkan pengaruh acak area dan waktu, masih memungkinkan untuk mengembangkan model yang memasukkan pengaruh acak area, waktu, dan pengaruh spasial; (c) Data deret waktu Susenas yang digunakan adalah 5 tahun, masih diperlukan eksplorasi lebih lanjut untuk mengetahui berbagai karakteristiknya dengan data deret waktu yang lebih panjang.

## DAFTAR PUSTAKA

- Chambers, R., dan Chandra, H. (2006). *Multipurpose Small Area Estimation*. Southampton Statistical Sciences Research Institute, Methodology Working Paper M06/06.
- Durbin, J. (2004). *State Space and Unobserved Component Models*. Cambridge University Press ([www.cambridge.org](http://www.cambridge.org)).
- Ghosh, M. and Rao, J.N.K. (1994). *Small area estimation: an appraisal*. *Statistical Sciences* **9**, 55-93.

- Lahiri, P. (2008). *Advanced Topics in Survey Methodology: Small Area Estimation*. Catatan Kuliah di *Joint Program in Survey Methodology*, University of Maryland, USA.
- Pfeffermann, D. (2002). *Small Area Estimation – New Developments and Directions*. *International Statistical Review*, **70**, 125-143.
- Prasad, N.G.N and Rao, J.N.K. (1990). *The estimation of the mean squared error of small-area estimators*. *Journal of the American Statistical Association*, **85**, 163-171.
- Rao, J.N.K. (2003). *Small Area Estimation*. New York : John Wiley and Sons.
- Rao, J.N.K. and Yu, M. (1994). *Small area estimation combining time series and cross-sectional data*. *Proceeding Survey Research Methods Section*, American Statistical Association, 1-9.
- Russo, C., M. Sabbatini dan R. Salvatore, (2005), "General Linear Models in Small Area Estimation : an assessment in agricultural surveys", Paper presented in The Mexsai Conference <[www.siap.sagarpa.gob.mx/mexsai/trabajos/t44.pdf](http://www.siap.sagarpa.gob.mx/mexsai/trabajos/t44.pdf)>, [29 April 2005]
- Sadik, K. and Notodiputro, K.A. 2009. *Hierarchical Bayes Estimation Using Time Series and Cross-sectional Data : A Case of Per-capita Expenditure in Indonesia*. Conference of Small Area Estimation, 29 Juni – 01 Juli 2009, Elche, Spanyol.
- Sadik, K. 2008. *Hierarchical Bayes Approach in Small Area Estimation*. Makalah Seminar Reguler di *Joint Program in Survey Methodology*, University of Maryland, USA. Oktober, 2008.
- Sadik, K., Notodiputro, K.A., Susetyo, B., Mangku, I.W. 2008. *Small Area Estimation With Time and Area Effects Using A Dynamic Linear Model*. The 3<sup>rd</sup> International Conference on Mathematics and Statistics (ICoMS-3). Institut Pertanian Bogor, Indonesia, 5-6 August 2008.
- Sadik, K. and Notodiputro, K.A. 2007. *A State Space Model in Small Area Estimation*. The 9<sup>th</sup> Islamic Countries Conference on Statistical Sciences 2007 (ICCS-IX), Kuala Lumpur: 12-14 December 2007.
- Sadik, K. dan Notodiputro, K.A. 2006. *Small Area Estimation with Time and Area Effects Using Two Stage Estimation*. *Proceeding at the First International Conference on Mathematics and Statistics, MSMSSEA*, 19-21 June 2006, Bandung.
- Sadik, K., Notodiputro, K.A., Susetyo, B., Mangku, I.W. 2006. *Pendugaan Area Kecil (Small Area Estimation) Berdasarkan Model yang Mengandung Langkah Acak (Random Walk)*. *Prosiding Seminar Nasional Matematika dan Konferda IndoMS wilayah Jabar & Banten*, 22 April 2006, Jurusan Matematika UNPAD, Bandung.
- Schall, R. (1991). *Estimation in generalized linear models with random effects*. *Biometrika* **78**, 719-727.
- Solomon, P.J. dan Cox, D.R. (1992). *Nonlinear components of variance models*. *Biometrika* **79**, 1-11.
- Zivot, E. (2006). *State Space Models and the Kalman Filter*. Cambridge University Press ([www.cambridge.org](http://www.cambridge.org)).