

ISBN : 978-979-17979-0-0

Prosiding

Seminar Nasional Mahasiswa S3 Matematika

REVITALISASI DAN SOSIALISASI DIRI UNTUK BERPERAN AKTIF DALAM PENINGKATAN KUALITAS PENELITIAN & PENDIDIKAN MATEMATIKA DI INDONESIA

Editor :

**Muslim Ansori
Ismail Djakaria
Dhoriva Urwatul Wutsqa
Agus Maman Abadi
M. Andy Rudhito
Umu Sa'adah
Karyati
Hasih Pratiwi**

Forum Komunikasi Mahasiswa S-3 Matematika se-Indonesia
Bekerja Sama Dengan Jurusan Matematika FMIPA UGM



Makalah Bidang Statistika

Estimating Conditional Intensity of Point Processes Models (<i>Nurtiti Sanusi, Sutawanir D., I Wayan M, Wahyu T</i>)	328
Generator Frame Wavelet Ketat dari Kelas Box Spline Multivariate (<i>Mahmud Yunus, Janny Lindiarni, Hendra Gunawan</i>)	334
Generalized Regression dalam Pendugaan Area Kecil (<i>Anang Kurnia, Khairil A. Notodiputro, Asep Saefuddin, dan I Wayan Mangku</i>)	346
Chain Ladder Method as a Gold Standard to Estimate Loss Reserves (<i>Aceng K. Mutaqin, Dumaria R. Tampubolon, Sutawanir Darwis</i>)	356
Optimalisasi Respon Ganda pada Metode Respon Permukaan (<i>Response Surface</i>) dengan Pendekatan Fungsi <i>Desirability</i> (<i>Hari Sakti Wibowo, I Made Sumertajaya, Hari Wijayanto</i>)	365
Sifat Penaksir Simulated MLE pada Respon Multinomial (<i>Jaka Nugraha</i>)	376
Estimasi Model Probit pada Respons Biner Multivariat Menggunakan MLE dan GEE (<i>Jaka Nugraha</i>)	386
Analisis Komponen Utama Temporar (<i>Ismail Djakaria</i>)	399
Pendekatan <i>Dynamic Linear Model</i> atau <i>State Space Model</i> pada Pendugaan Area Kecil (<i>Small Area Estimation</i>) (<i>Kusman Sadik dan Khairil Anwar Notodiputro</i>)	411
Pengembangan Algoritma EM untuk Data Tidak Lengkap (<i>Incomplete Data</i>) pada Model Log-Linear (<i>Kusman Sadik</i>)	422
CRUISE sebagai Metode Berstruktur Pohon (<i>Tree-Structured</i>) pada Data Non-Biner (<i>Kusman Sadik</i>)	433
Model <i>Proportional Hazards Cox</i> dengan <i>Missing Covariates</i> (<i>Nurkaromah Dwidayati, Sri Haryatmi, Subanar</i>)	447
Aplikasi Mixture Distribution dalam Pemodelan Resiko Aktuaria (<i>Adhitya Ronnie Effendie</i>)	459
Analisis Komponen Utama Probabilistik pada Data <i>Missing</i> (<i>Ismail Djakaria</i>)	464
Makalah Bidang Terapan	
Bilangan Ramsey untuk Graf Bintang S_6 dan Graf Bipartit Lengkap $K_{2,N}$, $N=2,3,4$ (<i>Isnaini Rosyida</i>)	476
H -supermagic labelings of c copies of some graphs (<i>Tita Khalis Maryati, A. N. M. Salman, Edy Tri Baskoro, Irawati</i>)	485

Pendekatan *Dynamic Linear Model* atau *State Space Model* pada Pendugaan Area Kecil (*Small Area Estimation*)

Kusman Sadik dan Khairil Anwar Notodiputro
Departemen Statistika, FMIPA IPB
Jl. Raya Dramaga, Kampus IPB Dramaga, Bogor
e-mail : kusmansadik@yahoo.com

Abstrak

Pengembangan teknik penarikan sampel (*sampling technique*) dan pengembangan metodologi pendugaan parameter populasi (*estimation methods*) merupakan dua topik utama yang menjadi perhatian para statistisi dalam membahas persoalan survei. Adapun persoalan mutakhir dalam metodologi pendugaan adalah menyangkut pendugaan untuk domain survei yang memiliki sampel kecil atau bahkan tidak memiliki sampel satupun, (Rao, 2003; Chambers dan Tzavidis, 2006). Misalnya, beberapa survei ekonomi yang dilakukan di Indonesia berskala nasional. Pada survei yang demikian banyaknya sampel rumah tangga untuk tiap kabupaten dalam suatu propinsi sangat kecil (*small area*). Bahkan bisa terjadi kabupaten tertentu tidak terpilih sebagai daerah survei sehingga sampel rumah tangga dari kabupaten tersebut tidak ada. Metode umum yang dikembangkan untuk pendugaan area kecil (*small area estimation / SAE*) adalah model yang berbasis pada *generalized linear mixed model* (GLMM). Beberapa pengembangannya saat ini mulai didiskusikan oleh para statistisi di dunia. Salah satunya adalah memasukkan model *state space* pada komponen galat dalam GLMM.

Kata kunci : *direct estimation, indirect estimation, small area estimation (SAE), dynamic linear model general, linear mixed model (GLMM), empirical best linear unbiased prediction (EBLUP), block diagonal covariance, state space model.*

Pendahuluan

Survei menjadi salah satu bagian penting dari proses pengambilan keputusan yang berbasis pada data. Sehingga survei sudah dilakukan baik di lembaga penelitian swasta maupun negeri. Bahkan kebijakan publik suatu negara sangat dipengaruhi oleh data-data hasil survei.

Sangat beragam persoalan yang ditemui dalam survei. Namun demikian, ada dua topik utama yang menjadi perhatian para statistisi selama tahun-tahun terakhir ini. Topik tersebut menyangkut persoalan pengembangan teknik penarikan contoh (*sampling technique*) dan pengembangan metodologi pendugaan parameter populasi (*estimation methods*).

Biasanya statistik diperoleh dari suatu survei yang didisain untuk memperoleh statistik nasional. Artinya survei semacam ini didisain untuk inferensia bagi daerah (*domain*) 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.

Dalam survei ini area yang dimaksud mungkin saja direpresentasikan oleh objek survei yang jumlahnya sangat kecil sehingga analisis yang didasarkan hanya pada objek-objek tersebut menjadi sangat tidak dapat diandalkan. Untuk mengatasi hal ini diperlukan metode pendugaan yang menggabungkan antara informasi didalam area yang dimaksud dengan informasi diluar area tersebut.

Statistik area kecil (*small area statistic*) telah menjadi perhatian para statistisi dunia secara sangat serius sejak sepuluh tahun terakhir ini (misalnya Dol, 1991; Ghosh and Rao, 1994; Lahiri, 2003; dan Rao, 2003; Chambers dan Tzavidis, 2006). Berbagai metode pendugaan area kecil (*small area estimation*) telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based area estimation*).

Pendugaan Langsung

Survei digunakan untuk mendapatkan informasi mengenai parameter populasi dengan mengefektifkan biaya yang tersedia. Secara lebih luas survei secara praktis tidak hanya digunakan untuk menduga total polulasi tetapi juga untuk menduga keragaman subpopulasi atau domain. Domain dapat didefinisikan sebagai daerah geografik, sosio-demografi, dan sebagainya. Dalam konteks survei, penduga dikatakan langsung (direct estimator) apabila pendugaan terhadap parameter populasi di suatu domain hanya didasarkan pada data contoh yang diperoleh dari domain tersebut. Misalnya, pendugaan rata-rata tingkat pengeluaran rumah tangga perbulan di suatu kabupaten didasarkan hanya pada data survei yang diperoleh dari kabupaten tersebut. Informasi lain yang berada di luar domain kabupaten tersebut tidak diperhitungkan.

Pendugaan langsung umumnya didasarkan pada teknik penarikan contohnya (sampling technique). Teknik semacam ini telah dikembangkan oleh Sewnson dan Wretman (1992), dan Thompson (1997). Metode yang didasrak pada pemodelan (model-based) juga telah dikembangkan, misalnya seperti yang dilakukan oleh Dorfman dan Royall (2001).

Pada pendugaan yang berbasis pada rancangan survei (design-based), pembobot rancangan $w_j(s)$ memiliki peranan penting dalam membentuk penduga berbasis rancangan \hat{Y} bagi Y . Pembobot ini tergantung pada s dan elemen j ($j \in s$). Salah satu bentuk pembobot yang penting adalah $w_j(s) = 1/\pi_j$, dimana $\pi_j = \sum_{\{s:j \in s\}} p(s)$, $j=1, 2, \dots, N$. Apabila tidak informasi penyerta (auxiliary information), maka penduga langsung dapat diekspresikan sebagai $\hat{Y} = \sum_{\{s:j \in s\}} w_j(s) y_j$. Dalam kasus ini, rancangan tidak berbias apabila terpenuhi $\sum_{\{s:j \in s\}} p(s) w_j(s) = 1$ untuk $j=1, 2, \dots, N$. Pembobot ini merupakan bentuk umum dari penduga Horvitz-Thompson

Rao (1979) telah menunjukkan bahwa penduga kuadratik tidak bias dan tidak negatif dari ragam \hat{Y} adalah

$$v(\hat{Y}) = -\sum_{j < k} \sum_{j, k \in s} w_{jk}(s) b_j b_k \left(\frac{y_j}{b_j} - \frac{y_k}{b_k} \right)^2$$

dimana $w_{jk}(s)$ merupakan pembobot dari suatu rancangan survei, b_j dan b_k adalah konstanta.

Pendugaan langsung ini dapat pula menggunakan informasi penyerta (auxiliary information) yang ada pada domain yang bersangkutan. Misalkan informasi penyerta dalam domain tertentu, $\mathbf{X} = (X_1, \dots, X_p)^T$, tersedia dan vektor x_j untuk $j \in s$ terobservasi, sehingga data (y_j, x_j) untuk setiap elemen $j \in s$ terobservasi. Suatu penduga yang mengefisienkan informasi penyerta ini adalah generalized regression (GREG) yang bisa ditulis sebagai berikut:

$$\hat{Y}_{GR} = \hat{Y} + (\mathbf{X} - \hat{\mathbf{X}})^T \hat{\mathbf{B}}$$

dimana $\hat{\mathbf{X}} = \sum_{\{s; j \in s\}} w_j(s) \mathbf{x}_j$ dan $\hat{\mathbf{B}} = (\hat{B}_1, \dots, \hat{B}_p)^T$ adalah solusi dari persamaan kuadrat terkecil terboboti dari contoh (*sample weighted least squares*):

$$(\sum_{\{s; j \in s\}} w_j(s) \mathbf{x}_j \mathbf{x}_j^T / c_j) \hat{\mathbf{B}} = \sum_{\{s; j \in s\}} w_j(s) \mathbf{x}_j y_j / c_j$$

dengan $c_j (> 0)$ sebagai konstanta spesifik.

Pendugaan Tidak Langsung

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

Alternatif metode lain adalah dengan cara menghubungkan area tersebut dengan area lain melalui model yang tepat. Dengan demikian dugaan tersebut

merupakan dugaan tidak langsung (*indirect estimation*), dalam arti bahwa dugaan tersebut mencakup data dari domain yang lain. Rao (2003) menyebutkan bahwa prosedur pendugaan area kecil pada dasarnya memanfaatkan kekuatan area sekitarnya (*neighbouring areas*) dan sumber data diluar area yang statistiknya ingin diperoleh. Metode ini memiliki sejarah yang panjang tetapi baru mendapat perhatian dalam beberapa dekade terakhir untuk digunakan sebagai pendekatan pada pendugaan parameter area kecil. Dalam hal ini, dua ide utama digunakan untuk mengembangkan model untuk pendugaan parameter area kecil yaitu (1) asumsi bahwa keragaman didalam area kecil peubah respon dapat diterangkan seluruhnya oleh hubungan keragaman yang bersesuaian pada informasi tambahan, kemudian disebut model pengaruh tetap (*fixed effect models*), dan (2) asumsi keragaman spesifik area kecil tidak dapat diterangkan oleh informasi tambahan dan merupakan pengaruh acak area kecil (*random effect*). Gabungan dari dua asumsi tersebut membentuk model pengaruh campuran (*mixed models*)

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 , $z_i = (z_{1i}, \dots, z_{pi})^T$ melalui suatu model linear

$$\theta_i = z_i^T \boldsymbol{\beta} + b_i v_i, \quad i = 1, \dots, m$$

dimana b_i adalah konstanta positif yang diketahui dan $\boldsymbol{\beta}$ adalah vektor berukuran $p \times 1$. Sedangkan v_i adalah pengaruh acak (*random effects*) spesifikasi area yang diasumsikan bebas dan menyebar identik (*independent and identically distributed, iid*) dengan

$$E_m(v_i) = 0 \text{ dan } V_m(v_i) = \sigma_v^2 (\geq 0), \text{ atau } v_i \sim \text{iid} (0, \sigma_v^2)$$

Pendugaan tidak langsung untuk rata-rata populasi di area kecil i , (\bar{Y}_i) , diperlukan informasi mengenai penduga langsungnya yaitu \hat{Y} . Dengan menggunakan metode James-Stein akan diperoleh:

$$\hat{\theta}_i = g(\hat{Y}) = \theta_i + e_i$$

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

dimana galat penarikan contoh (*sampling error*) e_i adalah bebas dengan

$$E_p(e_i|\theta_i) = 0 \text{ dan } V_p(e_i|\theta_i) = \psi_i, \text{ atau } v_i \sim \text{iid}(0, \sigma_v^2)$$

GLMM dalam Pendugaan Area Kecil

Rao(2003) mengaitkan model-model di atas sebagai bagian dari *general linear mixed model* (GLMM) yang menggabungkan antara pengaruh tetap (*fixed effects*) dan pengaruh acak (*random effects*) dalam suatu model umum. Datta dan Ghosh (1991) mengemukakan formulasi model GLMM sebagai berikut :

$$\mathbf{y}^P = \mathbf{X}^P \boldsymbol{\beta} + \mathbf{Z}^P \mathbf{v} + \mathbf{e}^P$$

Pada model ini \mathbf{v} dan \mathbf{e}^P bebas dengan $\mathbf{e}^P \sim \mathbf{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Psi}^P)$ dan $\mathbf{v} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{D}(\boldsymbol{\lambda}))$, dimana $\boldsymbol{\Psi}^P$ adalah matrik definit positif yang diketahui dan $\mathbf{D}(\boldsymbol{\lambda})$ adalah matrik definit positif yang strukturnya diketahui. Sedangkan \mathbf{X}^P dan \mathbf{Z}^P adalah matrik rancangan dan \mathbf{Y}^P adalah vektor $N \times 1$ dari nilai y populasi. Matrik koragam bagi \mathbf{v} dan \mathbf{e} masing-masing adalah \mathbf{G} dan \mathbf{R} . Persamaan di atas dapat pula ditulis sebagai berikut:

$$\mathbf{y}^P = \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^* \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z} \\ \mathbf{Z}^* \end{bmatrix} \mathbf{v} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}^* \end{bmatrix}$$

dimana bagian yang ditandai *asterisk* (*) menunjukkan unit yang tidak tercakup dalam contoh (*nonsampled*). Vektor untuk total (Y_i) pada area kecil adalah berbentuk

$\mathbf{A}\mathbf{y} + \mathbf{C}\mathbf{y}^*$ dengan $\mathbf{A} = \bigoplus_{i=1}^m \mathbf{1}_{n_i}^T$ dan $\mathbf{C} = \bigoplus_{i=1}^m \mathbf{1}_{N_i-n_i}^T$ dimana $\bigoplus_{i=1}^m \mathbf{A}_u = \text{blockdiag}(\mathbf{A}_1, \dots, \mathbf{A}_m)$.

Pada GLMM ini dilakukan pendugaan terhadap kombinasi linear dari parameter yaitu $\mu = \mathbf{1}^T\boldsymbol{\beta} + \mathbf{m}^T\mathbf{v}$. Rao (2003) mengemukakan bahwa untuk $\boldsymbol{\delta}$ tertentu yang diketahui maka penduga BLUP (*best linear unbiased prediction*) bagi μ adalah

$$\tilde{\mu}^H = t(\boldsymbol{\delta}, \mathbf{y}) = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \tilde{\mathbf{v}} = \mathbf{1}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \mathbf{GZ}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}})$$

dimana

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$$

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\boldsymbol{\delta}) = \mathbf{GZ}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}})$$

Model untuk pendugaan tidak langsung, yaitu $\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i + e_i$, $i = 1, \dots, m$, sebenarnya merupakan kasus khusus dari model GLMM, yaitu

$$\mathbf{y}_i = \hat{\theta}_i, \quad \mathbf{X}_i = \mathbf{z}_i^T, \quad \mathbf{Z}_i = b_i$$

dan

$$\mathbf{v}_i = v_i, \quad \mathbf{e}_i = e_i, \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$$

sedangkan

$$\mathbf{G}_i = \sigma_v^2, \quad \mathbf{R}_i = \psi_i$$

sehingga

$$\mathbf{V}_i = \psi_i + \sigma_v^2 b_i^2 \quad \text{dan} \quad \mu_i = \theta_i = \mathbf{z}_i^T \boldsymbol{\beta} + b_i v_i$$

Apabila persamaan pendugaan tidak langsung disubstitusikan ke dalam pendugaan GLMM akan diperoleh penduga BLUP bagi μ_i atau θ_i yaitu:

$$\tilde{\theta}_i^H = \mathbf{z}_i^T \tilde{\boldsymbol{\beta}} + \gamma_i (\hat{\theta}_i - \mathbf{z}_i^T \tilde{\boldsymbol{\beta}}), \quad \text{dimana} \quad \gamma_i = \sigma_v^2 b_i^2 / (\psi_i + \sigma_v^2 b_i^2), \quad \text{dan}$$

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_v^2) = \left[\sum_{i=1}^m \frac{\mathbf{z}_i \mathbf{z}_i^T}{\psi_i + \sigma_v^2 b_i^2} \right]^{-1} \left[\sum_{i=1}^m \frac{\mathbf{z}_i \hat{\theta}_i}{\psi_i + \sigma_v^2 b_i^2} \right]$$

Model Linear Dinamik atau *State Space*

Terkadang beberapa survei diulang pelaksanaannya secara berkala, misalnya Susenas. Untuk survei yang demikian ini pendugaan tidak langsung, seperti menduga tingkat kemiskinan di suatu kabupaten, dapat ditingkatkan efisiensi pendugaannya dengan memasukkan pengaruh acak area (*area random effects*) dan waktu (*time random effects*). Model *state space* merupakan salah satu alternatif metode pemodelan dalam pendugaan area kecil yang berbasis pada *general linear mixed model* (GLMM) dengan memperhatikan pengaruh acak waktu. Secara umum model *state space* dapat dinyatakan sebagai berikut:

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

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*). Model ini pada dasarnya merupakan kasus khusus GLMM tetapi dengan melakukan updating dugaan pada tiap waktu, penggunaan persamaan penyaring Kalman, dan pemulusan dugaan terakhir sebagai data baru.

Vektor $\boldsymbol{\alpha}_t$ disebut sebagai vektor *state*. Misalkan $\tilde{\boldsymbol{\alpha}}_{t|t-1}$ merupakan penduga BLUP bagi $\boldsymbol{\alpha}_{t-1}$ didasarkan pada semua pengamatan hingga waktu (t-1), sehingga

$$\tilde{\boldsymbol{\alpha}}_{t|t-1} = \mathbf{H} \tilde{\boldsymbol{\alpha}}_{t-1}$$

merupakan penduga BLUP bagi $\boldsymbol{\alpha}_t$ pada waktu (t-1). Selanjutnya dapat dinyatakan pula persamaan

$$\mathbf{P}_{t|t-1} = \mathbf{H} \mathbf{P}_{t-1} \mathbf{H}^T + \mathbf{A} \boldsymbol{\Gamma} \mathbf{A}^T$$

yaitu matrik koragam bagi penduga galat $\tilde{\boldsymbol{\alpha}}_{t|t-1} - \boldsymbol{\alpha}_t$, dimana $\mathbf{P}_{t-1} = E(\tilde{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{t-1}) E(\tilde{\boldsymbol{\alpha}}_{t-1} - \boldsymbol{\alpha}_{t-1})^T$ adalah matrik koragam bagi penduga galat pada waktu (t-1). Pada waktu t,

penduga bagi α_t dan matrik koragamnya diperbarui (*updated*) menggunakan data baru ($\mathbf{y}_t, \mathbf{Z}_t$), yaitu

$$\mathbf{y}_t - \mathbf{Z}_t \tilde{\alpha}_{t|t-1} = \mathbf{Z}_t(\alpha_t - \tilde{\alpha}_{t|t-1}) + \varepsilon_t$$

yang mempunyai bentuk model linear campuran dengan $\mathbf{y} = \mathbf{y}_t - \mathbf{Z}_t \tilde{\alpha}_{t|t-1}$, $\mathbf{Z} = \mathbf{Z}_t$, $\mathbf{v} = \alpha_t - \tilde{\alpha}_{t|t-1}$, $\mathbf{G} = \mathbf{P}_{t|t-1}$ dan $\mathbf{V} = \mathbf{F}_t$, dimana

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t^T + \Sigma_t$$

Contoh Kasus

Sebagai salah satu contoh kasus digunakan data simulasi untuk melihat beberapa karakteristik dari hasil pendugaannya. Pada simulasi ini digunakan 5 area ($i = 1, 2, \dots, 5$), waktu 5 periode ($t = 1, 2, \dots, 5$), dan masing-masing diulang sebanyak 100 kali. Beberapa yang disimulasi adalah $v_i \sim N(0, 3)$, $e_{it} \sim N(0, 2)$, dan $\varepsilon_{it} \sim N(0, 1)$ dengan 3 peubah untuk pengaruh tetap.

Tabel 1. Hasil Pendugaan EBLUB dan *State Space*

Area	Waktu	Nilai Sebenarnya (θ_{it})	<i>State Space</i>			EBLUB		
			$\hat{\theta}_{it}$	ARD	StdErr	$\hat{\theta}_{it}$	ARD	StdErr
1	1	49.69	49.44	0.51	0.9808	41.99	15.49	2.0865
1	2	37.27	37.22	0.12	0.9803	39.96	7.23	1.9958
1	3	41.61	41.49	0.30	0.9804	40.43	2.83	1.9992
1	4	40.77	40.63	0.35	0.9802	38.87	4.66	2.0297
1	5	37.70	37.67	0.07	0.9806	41.21	9.32	2.0286
2	1	45.12	44.77	0.77	0.9806	33.80	25.10	2.1830
2	2	32.77	32.71	0.18	0.9803	35.51	8.37	2.0254
2	3	32.02	32.00	0.06	0.9802	36.61	14.32	1.9956
2	4	48.19	48.01	0.37	0.9814	44.10	8.48	3.0813
2	5	28.77	28.71	0.20	0.9811	32.08	11.49	2.4513

Area	Waktu	Nilai Sebenarnya (θ_{it})	State Space			EBLUB			
			$\hat{\theta}_{it}$	ARD	StdErr	$\hat{\theta}_{it}$	ARD	StdErr	
3	1	28.55	28.56	0.05	0.9804	28.29	0.91	2.1730	
3	2	25.10	25.25	0.60	0.9801	31.26	24.54	1.9969	
3	3	33.64	33.58	0.19	0.9802	29.38	12.65	2.0615	
3	4	36.64	36.63	0.03	0.9805	34.38	6.16	2.2551	
3	5	30.15	30.21	0.21	0.9802	31.88	5.75	2.0134	
4	1	29.35	29.45	0.34	0.9802	31.24	6.45	2.4587	
4	2	20.74	20.91	0.80	0.9804	26.40	27.30	1.9973	
4	3	26.82	26.77	0.18	0.9817	22.34	16.70	2.4125	
4	4	26.04	26.06	0.09	0.9808	24.84	4.61	2.0760	
4	5	25.47	25.53	0.25	0.9805	26.09	2.43	2.0038	
5	1	29.58	29.74	0.55	0.9808	30.50	3.13	2.2742	
5	2	28.94	29.14	0.68	0.9811	31.44	8.65	2.4289	
5	3	18.06	18.27	1.16	0.9805	22.23	23.07	2.4996	
5	4	22.25	22.47	0.99	0.9802	26.44	18.85	2.0040	
5	5	31.72	31.73	0.03	0.9801	25.66	19.10	2.0410	
				0.36				11.50	

ARD : Percent Average Absolute Relative Deviance

Berdasarkan hasil pada Tabel 1, model *state space* pada GLMM dapat menurunkan nilai ARD dibandingkan dengan pendugaan EBLUP. Hasil ini juga menunjukkan bahwa apabila dalam pendugaan area kecil ada pengaruh waktu karena adanya survei yang diulang maka pengaruh waktu dapat dimasukkan dalam GLMM pada bentuk *state space*. Metode ini dapat diaplikasikan pada data-data survei yang dilakukan secara berulang, misalnya data Susenas BPS. Untuk meningkatkan efisiensi contoh (*sample*) untuk area kecil dapat digunakan data survei secara periodik tersebut.

Kesimpulan

Metode pendugaan area kecil (*small area estimation*) dapat digunakan untuk meningkatkan keakuratan pendugaan dengan cara meningkatkan efisiensi penggunaan contoh melalui fungsi hubung (*link function*) antara penduga langsung dengan pengaruh tetap dan pengaruh acak pada suatu area tertentu.

Apabila survei dilakukan berulang secara periodik maka peningkatan jumlah contoh efektif dapat dilakukan dengan cara memasukkan faktor waktu dalam model. Keefektifan akan optimal apabila pola pengaruh waktu dalam model dapat diidentifikasi secara baik.

Pustaka

- Chambers, R., Tzavidis, N. (2006). *M-quantile Models for Small Area Estimation*, forthcoming in *Biometrika*.
- Lahiri, P. (2003). *On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation*. *Statistical Science*, **18**, No. 2, 199–210
- Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley & Sons, Inc. New Jersey.
- Rao, J.N.K., dan Yu, M. (1994). *Small Area Estimation by Combining Time Series and Cross-Sectional Data*. Proceedings of the Section on Survey Research Method. American Statistical Association.
- Swenson, B., dan Wretman., J.H. (1989). *The Weighted Regression Technique for Estimating the Variance of Generalized Regression Estimator*. *Biometrika*, **76**, 527-537.
- Thompson, M.E. (1997). *Theory of Sample Surveys*. London: Chapman and Hall.