ISBN 978-979-19256-0-0

PROCEEDING

THE 3RD INTERNATIONAL CONFERENCE ON MATHEMATICS AND STATISTICS

BOGOR, 5 – 6 AUGUST 2008

Mathematics and Statistics: bridge for academia, business, and government in the entrepreneurial era

3rd ICoMS 2008

Inventory Model With Gamma Distribution	Hadi Sumadibrata, Ismail Bin Mohd	642
Accuracy Analysis Of Naive Bayesian Anti-Spam Filter	Ruslam, Armin Lawi, And Sri Astuti Thamrin	649
A New Method For Generating Fuzzy Rules From Training Data And Its Application In Financial Problems	Agus Maman Abadi, Subanar, Widodo, Samsubar Saleh	655
The Application Of Laws Of Large Numbers In Convergence Concept In Probability And Distribution	Georgina M. Tinungki	662
An Empirical Bayes Approach for Binary Response Data in Small Area Estimation	Dian Handayani, Noor Akma Ibrahim, Khairil A. Notodiputro, MOhd. Bakri Adam	669
Statistical Models For Small Area Estimation	Khairil A Notodiputro, Anang Kurnia, and Kusman Sadik	677
Maximum Likelihood Estimation For The Non-Separable Spatial Unilateral Autoregressive Model	Norhashidah Awang, Mahendran Shitan	685
Small Area Estimation Using Natural Exponential Families With Quadratic Variance Function (Nef-Qvf) For Binary Data	Kismiantini	691
Using An Extended And Ensemble Kalman Filter Algorithm For The Training Of Feedforward Neural Network In Time Series Forecasting	Zaqiatud Darojah, M. Isa Irawan, And Erna Apriliani	696
Estimation Of Outstanding Claims Liability And Sensitivity Analysis: Probabilistic Trend Family (PTF) Model	Arif Herlambang, Dumaria R Tampubolon	704
Expected Value Of Shot Noise Processes	Suyono	711
Modelling Malaysian Wind Speed Data Via Two Paramaters Weibull	Nur Arina Basilah Kamisan, Yong Zulina Zubairi, Abdul Ghapor Hussin, Mohd. Sahar Yahya	718
Application Of Latin Hypercube Sampling And Monte Carlo Simulation Methods: Case Study The Reliability Of Stress Intensity Factor And Energy Release Rate Of Indonesian Hardwoods	Yosafat Aji Pranata And Pricillia Sofyan Tanuwijaya	726
The Development Of Markov Chain Monte Carlo (Mcmc) Algorithm For Autologistic Regression Parameters Estimation	Suci Astutik, Rahma Fitriani, Umu Sa'adah, And Agustin Iskandar	734
A Note About Dh-Fever Estimation With ARIMAX Models	Elly Ana, Dwi Atmono Agus W	741
Evaluation Of Additive-Innovational Outlier Identification Procedure For Some Bilinear Models	[I]smail, M.I., Mohamed, I.B., Yahya, M.S.	745

Model By Spectral Regression Methods	Iriawan, Suhartono	
Application Of Cluster Analysis To Developing Core Collection In Plant Genetic Resources	Sutoro	875
Small Area Estimation With Time And Area Effects Using A Dynamic Linear Model	Kusman Sadik And Khairil Anwar Notodiputro	880
Statistical Analysis Of Wind Direction Data	Ahmad Mahir Razali, Arfah Ahmad, Azami Zaharim And Kamaruzzaman Sopian	886
Generalized Additive Mixed Models in Small Area Estimation	Anang Kurnia, Khairil A. Notodiputro, Asep Saefuddin, I Wayan Mangku	891
Kernel Principal Component Analysis In Data Visualization	Ismail Djakaria, Suryo Guritno, Sri Haryatmi	898
GARCH Models And The Simulations	Nelson Nainggolan, Budi Nurani Ruchjana And Sutawanir Darwis	906
Rainfall Prediction Using Bayesian Network	Hera Faizal Rachmat, Aji Hamim Wigena, and Erfiani	911
Identifying Item Bias Using The Simple Volume Indices And Multidimensional Item Response Theory Likelihood Ratio (Irt-Lr) Test	Heri Retnawati	916
Ordinary Kriging And Inverse Distance Weighting For Mapping Soil Phosphorus In Paddy Field	Mohammad Masjkur, Muhammad Nuraidi and Chichi Noviant	924
K-Means Clustering Visualization On Agriculture Potential Data For Villages In Bogor Using Mapserver	Imas S. Sitanggang, Henri Harianja, and Lailan Syaufina	932
Some Methods To Estimate The Number Of Components In A Mixture	M. A. Satyawan, A. H. Wigena, Erfiani	941
A Probabilistic Model For Finding A Repeat Triplet Region In DNA Sequence	Tigor Nauli	947
Application Of Spherical Harmonics In Determination Of Tec Using Gps Observable	Mardina Abdullah, Siti Aminah Bahari, Baharudin Yatim, Azami Zaharim, Ahmad Mahir Razali	954
Testing Structure Correlation Of Global Market By Statistic Vvsv	Erna Tri Herdiani, and Maman A. Djauhari	961
Exploring the MAUP from a spatial perspective	Gandhi Pawitan	967
Estimation of RCA(1) Model using EF: A new procedure and its robustness	1Norli Anida Abdullah, 2Ibrahim Mohamed, 3Shelton Peiris	996
Second Order Linear Elliptic Operators In The Unit Square	Abdul Rouf Alghofari	1008

POSTER

Study Of Fractional Factorial Split-Plot Experiment	Sri Winarni, Budi Susetyo, and Bagus Sartono	1012
Improving Model Performance For Predicting Poverty Village Category Using Neighborhood Information In Bogor	Bagus Sartono, Utami Dyah S, and Zulhelmi Thaib	1019
Ammi Models On Count Data: Log-Bilinear Models	Alfian Futuhul Hadi H. Ahmad Ansori Mattjik I Made Sumertajaya Halimatus Sa'diyah	1026
Prediction Of Oil Production Using Non Linear Regression By Sdpro Software (Special Program Package)*)	Budi Nurani R , and Kartlos J. Kachiashvili	1038
An Implementation Of Spatial Data Mining Using Spatial Autoregressive (Sar) Model For Education Quality Mapping At West Java*)	Atje Setiawan A. , Retantyo Wardoyo , Sri Hartati , and Agus Harjoko	1045
Validation Of Training Model For Robust Tests Of Spread	Teh Sin Yin, and Abdul Rahman Othman	1056
Spectral Approach For Time Series Analysis	Kusman Sadik	1063
The ACE Algorithm for Optimal Transformations in Multiple Regression	Kusman Sadik	1066
The Relation Between The Students' Interaction And The Construction Of Mathematical Knowledge	Rini Setianingsih	1069
Application of Auto Logistic Regression Spatial Model using Variogram Based Weighting Matrix to Predict Poverty Village Category	Utami Dyah Syafitri, Bagus Sartono, Vinda Pratama	1075

SMALL AREA ESTIMATION WITH TIME AND AREA EFFECTS USING A DYNAMIC LINEAR MODEL

Kusman Sadik and Khairil Anwar Notodiputro

Department of Statistics, Institut Pertanian Bogor
Jl. Meranti, Wing 22 Level 4, Kampus IPB Darmaga, Bogor 16680 – Indonesia

e-mail : kusmans@ipb.ac.id, khairiln@ipb.ac.id

Abstract. *There is a growing demand for reliable small area statistics in order to asses or to put into policies and programs. 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 circumtance of small sample size in the area. In fact, sample sizes in small areas are reduced, due to the circumtance that the overall sample size in a survey is usually determined to provide specific accuracy at a macro area level of aggregation, that is national territories, regions and so on. The most commonly used models for this case, usually in small area estimation, are based on generalized linear mixed models (GLMM). 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.*

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

1. Introduction

The problem of small area estimation is how to produce reliable estimates of area (domain) characteristics, when the sample sizes within the areas are too small to warrant the use of traditional direct survey estimates. The term of small area usually denote a small geographical area, such as a county, a province, an administrative area or a census division. From a statistical point of view the small area is a small domain, that is a small sub-population constituted by specific demographic and socioeconomic group of people, within a larger geographical areas. Sample survey data provide effective reliable estimators of totals and means for large areas and domains. But it is recognized that the usual direct survey estimators performing statistics for a small area, have unacceptably large standard errors, due to the circumstance of small sample size in the area. In fact, sample sizes in small areas are reduced, due to the circumstance that the overall sample size in a survey is usually determined to provide specific accuracy at a macro area level of aggregation, that is national territories, regions ad so on (Datta and Lahiri, 2000).

Demand for reliable small area statistics has steadily increased in recent years which prompted considerable research on efficient small area estimation. Direct small area estimators from survey data fail to borrow strength from related small areas since they are based solely on the sample data associated with the corresponding areas. As a result, they are likely to yield unacceptably large standard errors unless the sample size for the small area is reasonably large(Rao, 2003). Small area efficient statistics provide, in addition of this, excellent statistics for local estimation of population, farms, and other characteristics of interest in post-censual years.

Small area statistics are important tools for planning policies in specific regional and administrative areas. But important is also the information demand from other sectors, such as private, especially for questions related to local social and economics conditions, in local area marketing research, and so on. The small area statistics are based on a collection of statistical methods that “borrow strength” form related or similar small areas through statistics models that connect variables of interest in small areas with vectors of supplementary data, such as demographic, behavioral, economic notices, coming from administrative, census and specific sample surveys records. Small area efficient statistics provide, in addition of this, excellent statistics for local estimation of population, farms, and other characteristics of interest in post-censual years.

2. Direct estimation

Sample surveys have long been recognized as cost-effective means of obtaining information on wide-ranging topics of interest at frequent intervals over time. They are widely used in practice to provide estimates not only for the total population of interest but also for a variety of sub-populations (domains). Domain may be defined by geographic areas or socio-demographic groups or other sub-population. Examples of geographic domain (area) include a state/province, county, municipality, school district, unemployment insurance region, metropolitan area, and health service area. On other hand, a socio-demographic domain may refer to a specific age-sex-race group within a large geographic area. An example of other domains is the set of business firms belonging to a census division by industry group (Rao, 2003).

Direct estimation usually is based on under the design-based or repeated sampling framework. The technique have been developed by Sewnson dan Wretman (1992). Model-based methods have also been used to develop direct estimators and associated inferences. Such methods provide valid conditional inferences refering to the particular sample that has been drawn, regardless of the sampling design (see Pffermann, 2002).

A sampling design is used to select a sample s from U with probability $p(s)$. The sample selection probability $p(s)$ can depend on know design variables such as stratum indicator variables and size measures of clusters. Design weights $w_j(s)$ play an important role in constructing design-based estimator $\hat{Y}$ of Y. These basic weights may depend both on s and the element j ($j \in s$). An important choice is $w_j(s)=1/p_j$, where $p_j = \Sigma_{\{s:j\in s\}}p(s)$, j=1, 2, .., N. In the absence of auxiliary population information, we use expansion estimator $\hat{Y} = \Sigma_{\{s:j\in s\}}w_j(s)y_j$. In this case, the design-unbiased condition reduces to $\Sigma_{\{s:j\in s\}}p(s)w_j(s) = 1$, j=1, 2, …, N. The choice $w_j(s)=1/p_j$ satisfies the unbiased condition and leads to the well-known Horvits-Thompson estimator.

Rao (2003) has shown that a non-negative unbiased quadratic estimator of variance of $\hat{Y}$ is necessarily of the form

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

where the weights $w_{jk}(s)$ satisfy the unbiased condition and the non-zero constants b_j and b_k.

Direct estimation can also use auxiliary information. Suppose now that auxiliary information in the form of known population totals $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ is available and that the auxiliary vector xj for $j\in s$ is also observed, that is, the data (y_j, x_j) for each element $j\in s$ are observed. An estimator that makes efficient use of this auxiliary information is generalized regression (GREG) which may be written as:

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

where $\hat{\boldsymbol{X}} = \Sigma_{\{s:j\in s\}}w_j(s)\boldsymbol{x}_j$ and $\hat{\boldsymbol{B}} = (\hat{B}_1,\ldots,\hat{B}_p)^T$ is the solution of sample weighted least squares equations:

$$(\Sigma_{\{s:j\in s\}}\mathrm{w_j(s)}\mathbf{x}_j\mathbf{x}_j^T/\mathrm{c_j})\,\hat{\boldsymbol{B}} = \Sigma_{\{s:j\in s\}}\mathrm{w_j(s)}\mathbf{x}_j\mathrm{y_j}/\mathrm{c_j}$$

with specified constants cj(>0).

3. Indirect estimation in small area

A domain (area) is regarded as large (or major) if domain-specific sample is large enough to yield direct estimates of adequate precision. A domain is regarded as small if the domain-specific sample is not large enough to support direct estimates of adequate precision. Some other terms used to denote a domain with small sample size include local area, sub-domain, small subgroup, sub-province, and minor domain. In some applications, many domains of interest (such as counties) may have zero sample size.

Three types of indirect estimators can be identified : domain indirect, time indirect, and domain and time indirect. A domain indirect estimator makes use of y-values from another domain but not from another time period. A time indirect estimator uses y-values from another time period for the domain interest but not from another domain. On the other hand, a domain and time indirect estimator uses y-values from another domain as well as another time period (Rao, 2003).

In making estimates for small area with adequate level of precision, it is often necessary to use indirect estimators that borrow strength by using thus values of the ***variable*** of interest, y, from related areas and/or time periods and thus increase the effective sample size. These values are brought into the estimation process through a model (either implicit or explicit) that provides a link to related areas and/or time periods through the use of supplementary information related to y, such as recent census counts and current administrative records (Pfeffermann 2002; Rao 2003).

Methods of indirect estimation are based on explicit small area models that make specific allowance for between area variation. In particular, we introduce mixed ***models*** involving random area specific effects that account for between area variation beyond that explained by auxiliary variables included in the model. We assume that $\theta_i = g(\overline{Y}_i)$ for some specified $g(.)$ is related to area specific auxiliary data $\boldsymbol{z}_i = (z_{1i}, \dots, z_{pi})^T$ through a linear model

$$\theta_i = \boldsymbol{z}_i^T b + b_i v_i, \quad i = 1, \dots, m$$

where the b_i are known positive constants and b is the px1 vector of regression ***coefficients***. Further, the v_i are area specific random effects assumed to be independent and identically distributed (iid) with

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

4. Time series and cross-sectional models

Many sample surveys are repeated in time with partial replacement of the sample elements. For such repeated surveys considerable gain in efficiency can be achieved by borrowing strength across both small areas and time. Rao and Yu (1992, 1994) proposed an extension of the basic Fay-Herriot model to handle time series and cross-sectional data. Their model consist of a sampling error model

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

and a linking model

$$\theta_{it} = \boldsymbol{z}_{it}^T b + v_i + u_{it}$$

Here $\hat{\theta}_{it}$ is the direct survey estimator for small area i at time t, θ_{it} is a function of the small area mean, the e_{it} are sampling error normally distributed with zeros means and a known block diagonal covariance matrix y, and $\boldsymbol{z}_{it}$ is a vector of area specific covariates some of which may change with t, for example, administrative data. Further, $v_i \sim$ iid $(0, \sigma_v^2)$ and u_{it} are assumed to follow a random walk process for each i, that is,

$$u_{it} = u_{i,t-1} + \varepsilon_{it}, \text{ dan } \varepsilon_{it} \sim iid\ \text{N}(0, \sigma^2)$$

The errors $\{e_{it}\}$, $\{v_i\}$, and $\{\varepsilon_{it}\}$ are also assumed to be independent. The model on θ_{it} depends on both area specific effects v_i and the area-by-time specific effects u_{it} which are correlated across time for each i.

5. Generalized linear mixed model

Datta and Lahiri (2000), and Rao(2003) considered a general linear mixed model (GLMM) which covers the univariate unit level model as special cases:

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

Hence $\mathbf{v}$ and $\mathbf{e}^P$ are independent with $\mathbf{e}^P \sim \mathbf{N}(\mathbf{0}, \sigma^2 \text{y}^P)$ and $\mathbf{v} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{D}(1))$, where y^P is a known positive definite matrix and $\mathbf{D}(1)$ is a positive definite matrix which is structurally known except for some parameters l typically involving ratios of variance components of the form σ_i^2/σ^2. Further, $\mathbf{X}^P$ and $\mathbf{Z}^P$ are known design matrices and $\mathbf{y}^P$ is the N x 1 vector of population y-values. The GLMM form :

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

where the asterisk (*) denotes non-sampled units. The vector of small area totals (Y_i) is of the form $\mathbf{Ay}$ + $\mathbf{Cy}^*$ with $\mathbf{A} = \oplus_{i=1}^{m} \mathbf{1}_{n_i}^T$ and $\mathbf{C} = \oplus_{i=1}^{m} \mathbf{1}_{N_i - n_i}^T$ where $\oplus_{i=1}^{m} \mathbf{A}_u$ = blockdiag($\mathbf{A}_1, \ldots, \mathbf{A}_m$).

We are interested in estimating a linear combination, $\mu = \mathbf{1}^T b + \mathbf{m}^T \mathbf{v}$, of the regression parameters b and the realization of $\mathbf{v}$, for specified vectors, $\mathbf{l}$ and $\mathbf{m}$, of constants. For known d, the BLUP (*best linear unbiased prediction*) estimator of μ is given by (Rao, 2003)

$$\tilde{\mu}^H = t(d, \mathbf{y}) = \mathbf{1}^T \tilde{b} + \mathbf{m}^T \tilde{\mathbf{v}} = \mathbf{1}^T \tilde{b} + \mathbf{m}^T \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \tilde{b})$$

Model of indirect estimation, $\hat{\theta}_i = \mathbf{z}_i^T b + b_i v_i + e_i$, $i = 1, \ldots, m$, is a special case of GLMM with block diagonal covariance structure. Making the above substitutions in the general form for the BLUP estimator of m_i, we get the BLUP estimator of θ_i as:

$$\tilde{\theta}_i^H = \mathbf{z}_i^T \tilde{b} + \gamma_i (\hat{\theta}_i - \mathbf{z}_i^T \tilde{b}), \text{ where } \gamma_i = \sigma_v^2 b_i^2 / (y_i + \sigma_v^2 b_i^2), \text{ and}$$

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

6. Dynamic linear model

Many sample surveys are repeated in time with partial replacement of the sample elements. For such repeated surveys considerable gain in efficiency can be achieved by borrowing strength across both small areas and time. Their model consist of a sampling error model

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

$$\theta_{it} = \mathbf{z}_{it}^T b_{it}$$

where the coefficients $b_{it} = (\beta_{it0}, \beta_{it1}, \ldots, \beta_{itp})^T$ are allowed to vary cross-sectionally and over time, and the sampling errors e_{it} for each area i are assumed to be serially uncorrelated with mean 0 and variance y_{it}. The variation of b_{it} over time is specified by the following model:

$$\begin{bmatrix} \beta_{itj} \\ \beta_{ij} \end{bmatrix} = \mathsf{T}_j \begin{bmatrix} \beta_{i,t-1,j} \\ \beta_{ij} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_{itj}, \quad j = 0, 1, \ldots, p$$

It is a special case of the general dynamic linear model (state-space model) which may be expressed in the form

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t a_t + e_t; & E(e_t) = 0, \quad & E(e_t e_t^T) = St \\ a_t &= \mathbf{H}_t a_{t-1} + \mathbf{A} h_t; & E(h_t) = 0, \quad & E(h_t h_t^T) = G \end{aligned}$$

where e_t and h_t are uncorrelated contemporaneously and over time. The first equation is known as the *measurement equation*, and the the second equation is known as the *transition equation*. This model is a special case of the general linear mixed model but the state-space form permits updating of the estimates over time, using the Kalman filter equations, and smoothing past estimates as new data becomes available, using an appropriate smoothing algoritm.

The vector a_t is known as the *state vector*. Let $\tilde{\boldsymbol{\alpha}}_{\mathbf{t-1}}$ be the BLUP estimator of a_{t-1} based on all observed up to time (t-1), so that $\tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}} = \mathbf{H} \tilde{\boldsymbol{\alpha}}_{\mathbf{t-1}}$ is the BLUP of a_t at time (t-1). Further, $\mathbf{P}_{t|t-1} = \mathbf{H}\mathbf{P}_{t-1}\mathbf{H}^T + \mathbf{A}G\mathbf{A}^T$ is the covariance matrix of the prediction errors $\tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}} - a_t$, where

$$\mathbf{P}_{t-1} = E(\tilde{\boldsymbol{\alpha}}_{\mathbf{t-1}} - a_{t-1})(\tilde{\boldsymbol{\alpha}}_{\mathbf{t-1}} - a_{t-1})^T$$

is the covariance matrix of the prediction errors at time (t-1). At time t, the predictor of a_t and its covariance matrix are updated using the new data ($\mathbf{y}_t$, $\mathbf{Z}_t$). We have

$$\mathbf{y}_t - \mathbf{Z}_t \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}} = \mathbf{Z}_t(a_t - \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}}) + e_t$$

which has the linear mixed model form with $\mathbf{y} = \mathbf{y}_t - \mathbf{Z}_t \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}}$, $\mathbf{Z} = \mathbf{Z}_t$, $\mathbf{v} = a_t - \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}}$, $\mathbf{G} = \mathbf{P}_{t|t-1}$ and $\mathbf{V} = \mathbf{F}_t$,

where $\mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t^T + S_t$. Therefore, the BLUP estimator $\tilde{\mathbf{v}} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{y}$ reduces to

$$\tilde{\boldsymbol{\alpha}}_{\mathbf{t-1}} = \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}} + \mathbf{P}_{t|t-1}\mathbf{Z}_t^T \mathbf{F}_t^{-1}(\mathbf{y}_t - \mathbf{Z}_t \tilde{\boldsymbol{\alpha}}_{\mathbf{t|t-1}})$$

7. Case study

Model of small area estimation can be applied to estimate average of households expenditure per month for each of m = 37 counties in East Java, Indonesia. Data that be used in this case study are data of Susenas (National Economic and Social Survey, BPS) 2003 to 2005.

Table 1. Design Based and Model Based Estimates of County Means and Estimated Standard Error

County	Design Based (Direct Estimator)		Model Based (Indirect Estimator)			
			EBLUP		$\text{EBLUP}_{(ss)}$	
	$\hat{\mu}_i$	$s(\hat{\mu}_i)$	$\hat{\mu}_i^H$	$s(\hat{\mu}_i^H)$	$\hat{m}_i^{ss}$	$s(\hat{m}_i^{ss})$
Pacitan	4.89	0.086	3.89	0.062	5.23	0.038
Ponorogo	5.5	0.148	5.83	0.149	5.73	0.132
Trenggalek	5.3	0.135	6.89	0.155	5.65	0.161
Tulungagung	6.78	0.229	7.06	0.215	7.05	0.172
Blitar	5.71	0.132	5.74	0.198	6.12	0.141
Kediri	5.62	0.105	7.09	0.110	6.45	0.091
Malang	5.94	0.128	6.58	0.112	5.19	0.109
Lumajang	5.07	0.119	4.75	0.118	5.74	0.081
Jember	4.65	0.090	4.96	0.126	5.28	0.113
Banyuwangi	5.98	0.142	5.55	0.124	6.15	0.131
Bondowoso	4.53	0.127	4.64	0.092	5.43	0.105
Situbondo	4.67	0.104	5.89	0.085	4.44	0.074
Probolinggo	5.54	0.154	6.07	0.184	7.34	0.186
Pasuruan	6.31	0.151	4.95	0.121	6.39	0.109
Sidoarjo	9.33	0.169	9.46	0.177	8.32	0.123
Mojokerto	6.91	0.160	6.55	0.135	8.25	0.107
Jombang	6.09	0.131	5.06	0.130	5.96	0.091
Nganjuk	5.56	0.125	4.40	0.041	4.87	0.029
Madiun	5.5	0.139	5.16	0.116	5.46	0.121
Magetan	5.52	0.161	4.84	0.145	4.16	0.132
Ngawi	4.89	0.102	4.61	0.097	4.15	0.086
Bojonegoro	5.06	0.093	5.25	0.067	4.50	0.047
Tuban	6.02	0.114	5.75	0.061	6.47	0.046
Lamongan	6.29	0.106	6.47	0.123	5.69	0.065
Gresik	8.49	0.186	9.07	0.167	9.01	0.198
Bangkalan	6.61	0.140	5.69	0.091	7.00	0.076
Sampang	6.32	0.158	7.20	0.150	6.85	0.182
Pamekasan	5.78	0.107	6.10	0.126	5.93	0.109
Sumenep	5.48	0.108	5.76	0.077	5.09	0.032
Kota Kediri	8.01	0.159	7.60	0.157	7.11	0.144
Kota Blitar	7.98	0.191	7.63	0.159	8.51	0.182
Kota Malang	11.14	0.298	12.63	0.273	11.61	0.225
Kota Probolinggo	9.1	0.183	7.68	0.140	10.50	0.153
Kota Pasuruan	7.75	0.149	8.09	0.085	8.41	0.072
Kota Mojokerto	9.45	0.204	9.51	0.235	9.01	0.211

County	Design Based (Direct Estimator)		Model Based (Indirect Estimator)			
			EBLUP		$EBLUP_{(ss)}$	
	$\hat{\mu}_i$	$s(\hat{\mu}_i)$	$\hat{\mu}_i^H$	$s(\hat{\mu}_i^H)$	$\hat{m}_i^{ss}$	$s(\hat{m}_i^{ss})$
Kota Madiun	8.4	0.162	8.33	0.150	7.62	0.196
Kota Surabaya	11.45	0.328	11.81	0.353	11.16	0.321
Mean		0.149		0.138		0.124

Table 1 reports the design based and model based estimates. The design based estimates is direct estimator based on design sampling. EBLUP estimates, $\hat{\mu}_i^H$, use small are model with area effects (data of Susenas 2005). Whereas, EBLUP(ss) estimates, $\hat{m}_i^{ss}$, use small are model with area and time effects (data of Susenas 2003 to 2005). Estimated standard error denoted by s($\hat{\mu}_i$), s($\hat{\mu}_i^H$), and s($\hat{m}_i^{ss}$). It is clear from Table 1 that estimated standard error mean of model based is less than design based. The estimated standard error mean of EBLUP(ss) is less than EBLUP.

8. Conclusion

Small area estimation can be used to increase the effective sample size and thus decrease the standard error. For such repeated surveys considerable gain in efficiency can be achieved by borrowing strength across both small area and time. Availability of good auxiliary data and determination of suitable linking models are crucial to the formation of indirect estimators.

9. Reference

Datta, G.S, and Lahiri, P. (2000). A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors (BLUP) in Small Area Estimation Problems, *Statistica Sinica*, **10**, 613-627.

Pfeffermann, D. (2002). Small Area Estimation – New Developments and Directions. *International Statistical Review*, **70**, 125-143.

Pfeffermann, D. and Tiller, R. (2006). State Space Modelling with Correlated Measurements with Aplication to Small Area Estimation Under Benchmark Constraints. S3RI Methodology Working Paper M03/11, University of Southampton. Available from: http://www.s3ri.soton.ac.uk/ publications.

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. (1992). 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.