STATISTICS IN TRANSITION-new series, Summer 2012
Vol. 13, No. 2, pp. 261—278

ESTIMATION OF PARAMETERS FOR SMALL AREAS USING HIERARCHICAL BAYES METHOD IN THE CASE OF KNOWN MODEL HYPERPARAMETERS

Jan Kubacki[1]

ABSTRACT

In the paper the method of parameters estimation using hierarchical Bayes (HB) method in the case of known model hyperparameters for *a priori* conditionals was presented. This approach has some advantage in comparison with subjective model parameters selection because of more simulation stability and allows obtaining estimates that has more regular distribution. As an example the data about average per capita income from Polish Household Budget Survey for counties (NUTS4) and auxiliary variables from Polish Tax Register (POLTAX) were used. The computation was done using WinBUGS software and R-project environment with R2WinBUGS package, which control the simulations in WinBUGS, and coda package, which allows performing the analysis of simulation results. In the paper sample code in R-project that can be used as a pattern for further similar applications was also presented. The efficiency of hierarchical Bayes estimation with other small area methods was compared. Such comparison was done for HB and EBLUP techniques, for which some consistency related to the precision of estimates obtained using both techniques was achieved.

Key words: Small area estimation, hierarchical Bayes estimation, WinBUGS.

1. Introduction

Small area estimation methods are obviously used in the situations where there is a need to "borrow strength" to determine the estimation using sample survey, but the sample of considered subpopulation is not large enough, what causes too large estimation error. Here "small area" can be understood as smaller administrative units (for example counties – in Polish poviats) or specific groups extracted from the population (for example specific socio-economic groups). This problem can concern also mini-domains or rare features, which are observed with

[1] Centre for Mathematical Statistics, Statistical Office in Łódź, ul. Suwalska 29, 93-176 Łódź, Poland. E-mail: j.kubacki@stat.gov.pl.

smaller frequency, and because of this the estimates of such variables may cause difficulties even for larger administrative units (for example regions). The estimates for income from unemployment benefits for regions from Household Budget Survey may be a good example here. Relative estimation error here may be sometimes large and may exceed 20%. Application of the small area methods may be justified in such a case.

The small area estimation methodology has been systematically developed since 1980's. Here we can mention books from J.N.K Rao (2003) and N.T. Longford (2005) and Mukhopadhyay (1998). In Polish literature one can also find some examples of more comprehensive studies of this topic. Here we can point out works by Bracha, Lednicki and Wieczorkowski (2003, 2004), Domański and Pruska (2001), Gołata (2004), Dehnel (2003) and Żądło (2008). Small area issues were also the topic of many scientific conferences. Here we can recall one of the first small area estimation conference that was held in Warsaw in 1992 (see Kalton, G., Kordos, J., and Platek, R., 1993) and series of the conferences entitled "Small Area Estimation" that have been organized every two years since 2005. First was the conference organized in Jyväskylä, Finland (see http://www.stat.jyu.fi/sae2005/index.html), than conference that took place in 2007 in Pisa, Italy (see http://sae2007.dsm.unipi.it), next was the conference organized in 2009 in Elche, Spain (see http://icio.umh.es/congresos/sae2009) and the last conference took place in 2011 in Trier, Germany (see http://www.uni-trier.de/index.php?id=30789). Small area estimation topics were also presented at the conferences that were organized in Poland. Here we can mention the "Survey Sampling in Economic and Social Research" conference that is organized by the University of Economics in Katowice (see http://web2.ue.katowice.pl/metoda) and the conference "Multivariate Statistical Analysis" that is organized by University of Łódź (see http://www.msa.uni.lodz.pl). Thus, we can see that literature related to the small area estimation is relatively large and contains wide theoretical material, with application examples, what allows for implementation of small area methods in statistical practice.

Hierarchical Bayes estimation method is one of the most often applied small area estimation method. In the last years the growth of interest of this technique is observed. Here we can mention for example PhD thesis that was prepared by M. Vogt (2010) and B. Liu (2009). This method assumes that both *a priori* distributions $f(\lambda)$ of model parameters and conditional distributions $f(\mu,y|\lambda)$ of small area parameters μ (given the model parameter values) are known. Here also data from survey y should be included. Using Bayes theorem one can obtain *a posteriori* distribution $f(\mu|y)$. In simple cases such distribution can be obtained analytically, but more complex cases require special computational methods using MCMC (Markov Chain Monte Carlo) techniques, which are implemented numerically using Gibbs sampler methods.

2. Hierarchical Bayes (HB) method – application for small areas

Here the assumption for HB method will be presented more accurate. First, it is assumed, that we should obtain the following *a posteriori* distribution:

$$f(\mu \mid \mathbf{y}) = \int f(\mu, \lambda \mid \mathbf{y}) d\lambda \tag{2.1}$$

Using Bayes inference we can obtain the following dependence:

$$f(\mu,\lambda \mid \mathbf{y}) = \frac{f(\mathbf{y},\mu \mid \lambda) f(\lambda)}{f_1(\mathbf{y})} \tag{2.2}$$

where $f_1(\mathbf{y})$ is the marginal distribution and has the form:

$$f_1(\mathbf{y}) = \int f(\mathbf{y},\mu \mid \lambda) f(\lambda) d\mu d\lambda \tag{2.3}$$

As it was mentioned in the introduction, in particular cases to perform such calculations the knowledge about *a priori* distributions is needed. This knowledge can be used in construction of particular models for small areas. In the case considered here we take into account the type A model, and, speaking more precisely, basic area level model, which has the following form:

$$\hat{\theta}_i = \mathbf{z}_i^T \beta + b_i v_i + e_i \tag{2.4}$$

where $\hat{\theta}_i$ is small area estimator of particular variable for small area i, $\mathbf{z_i}$ is vector of explanatory variable, β is vector of regression coefficients, b_i is known positive constants, v_i represents the model error, and e_i represents the sample design error. It is often assumed, that the values of component v_i constitutes variables that are independent and identically distributed (iid) having the following properties:

$$E_m(v_i) = 0, V_m(v_i) = \sigma_v^2 \tag{2.5}$$

where E_m is the expected value for the component v for model, and V_m is the model variance. It is assumed for design error, that (for direct estimates)

$$E_p(e_i \mid \theta_i) = 0, V_p(e_i \mid \theta_i) = \psi_i \tag{2.6}$$

It is also assumed that estimation error for direct estimates ψ_i is also known. Taking into consideration the (2.4-2.6) and assuming that the distribution of model error σ_v^2 is also known and has the inverse Gamma distribution $G^{-1}(a,b)$ having parameters a and b (where a is the shape parameter and b is the scale parameter) the hierarchical model can be written in the following form:

(i) $\hat{\theta}_i \mid \theta_i, \beta, \sigma_v^2 \overset{ind}{\sim} N(\theta_i, \psi_i)$ $i=1,\ldots m$

(ii) $\theta_i \mid \beta, \sigma_v^2 \overset{ind}{\sim} N(\mathbf{z}_i^T \beta, b_i^2 \sigma_v^2)$ $i=1,\ldots m$

(iii) $f(\beta) \sim 1$

(iv) $\sigma_v^2 \mid \beta, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}} \sim \mathrm{G}^{-1}(a,b)$ (2.7)

and here the case of known distribution of σ_v^2 and "flat" prior for β, given by $f(\beta)\sim 1$ is considered. It is also assumed that (in contrast to model (10.3.1) from Rao book), values of the parameters a and b in Gamma distribution for σ_v^2 are known, what is a good approximation for the model from paragraph 10.3.3 in Rao. These values can be obtained from empirical distribution of model estimates that can be determined from linear regression models. Because models that have identical explanatory variables and similar variability of the estimates for both direct estimates and regression coefficients are considered, such approximation may lead to correct estimates of *a posteriori* for hierarchical model. According to Rao suggestion (p. 237) "when σ_v^2 is assumed to be known and $f(\beta)\sim 1$, the HB and BLUP approaches under normality lead to identical point estimates and measures of variability". However, it should be noted that model (10.3.1) in our opinion reflects the variability of σ_v^2 slightly less, what leads to consistency but with more simplified variance measure (see for example equation (7.1.6) in Rao)

$$MSE(\tilde{\theta}_i^H) = E(\tilde{\theta}_i^H - \theta_i)^2 = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) \qquad (2.8)$$

Thus, taking into consideration such variability, obtained estimates are more consistent with EBLUP estimates (and incorporating full model variability). More details about this issue will be presented in experimental section.

3. Markov chain Monte Carlo (MCMC) methods

Assuming that $\boldsymbol{\eta} = (\boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T$ is the vector of small area parameters $\boldsymbol{\mu}$ and model parameters $\boldsymbol{\lambda}$, it should be noted that for more complex models, which model (2.7) is a good example of, obtaining a sample from *a posteriori* distribution that has the form like (2.2) may be difficult because of complex nature of the denominator $f_1(y)$. Application of MCMC method in such a case may allow avoiding such difficulties. Here Markov chain $\{\boldsymbol{\eta}^{(k)}, k=0,1,2,\ldots\}$ is constructed, that the distribution of $\boldsymbol{\eta}^{(k)}$ is converged to unique stationary distribution given by $f(\boldsymbol{\eta}|\mathbf{y})$ denoted by as $\pi(\boldsymbol{\eta})$. Thus, neglecting the first d samples (drawing in the burn-in phase), we can obtain D dependent samples $\boldsymbol{\eta}^{(d)},\ldots,\boldsymbol{\eta}^{(d+D)}$, drawing from the target distribution $f(\boldsymbol{\eta}|\mathbf{y})$. Such sample is independent from starting point $\boldsymbol{\eta}^{(0)}$.

Such Markov chain construction requires that one-step transition probability $P(\boldsymbol{\eta}^{(k+1)}, \boldsymbol{\eta}^{(k)})$ be dependent only on the current state $\boldsymbol{\eta}^{(k)}$. As a consequence it leads to the conclusion, that conditional distribution of $\boldsymbol{\eta}^{(k+1)}$ given $\boldsymbol{\eta}^{(0)},\ldots, \boldsymbol{\eta}^{(k)}$ is independent on the chain history $\{\boldsymbol{\eta}^{(0)},\ldots,\boldsymbol{\eta}^{(k-1)}\}$. In such case the stationary condition for the transition kernel should be satisfied:

$$\int \pi(\boldsymbol{\eta}^{(k)}) P(\boldsymbol{\eta}^{(k+1)} \mid \boldsymbol{\eta}^{(k)}) d\boldsymbol{\eta}^{(k)} = \pi(\boldsymbol{\eta}^{(k+1)}) \qquad (3.1)$$

The equation (3.1) shows, that if $\boldsymbol{\eta}^{(k)}$ can be obtained from $\pi(\cdot)$, then also $\boldsymbol{\eta}^{(k+1)}$ can be obtained from $\pi(\cdot)$. It is also necessary to ensure that the distribution of $\boldsymbol{\eta}^{(k)}$ given $\boldsymbol{\eta}^{(0)}$, denoted as $P^{(k)}(\boldsymbol{\eta}^{(k)}| \boldsymbol{\eta}^{(0)})$ converge to $\pi(\boldsymbol{\eta}^{(k)})$ regardless of that how the $\boldsymbol{\eta}^{(0)}$ is chosen. Thus, the chain considered here should be irreducible and aperiodic. Irreducible means that for all starting points η (0) the chain reach some not empty set in the state space with positive likelihood. Aperiodicity means, that the chain should not oscillate between different set of states in a periodical manner.

4. Gibbs sampler

The computational implementation of MCMC can be performed using the method called Gibbs sampler. We briefly present this method here. The Gibbs sampler assumes that we obtain the series of the samples $\boldsymbol{\eta}^{(k)}$ with partitioning $\boldsymbol{\eta}$ vector into blocks $\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_r$. These blocks can contain one or more elements. For example, for basic area level model we have $\boldsymbol{\mu}=(\theta_1,\ldots,\theta_m)^T=\boldsymbol{\theta}$ and $\lambda=(\beta^T,\sigma_v^2)^T$. In such case $\boldsymbol{\eta}$ can be constituted with the following blocks $\eta_1=\beta$, $\eta_2=\theta_1,\ldots$, $\eta_{m+1}=\theta_m$, $\eta_{m+1}=\sigma_v^2$, assuming that $r=m+2$. It is also required that the following Gibbs conditional should be considered: $f(\boldsymbol{\eta}_1 \big| \boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_r,\mathbf{y})$, $f(\boldsymbol{\eta}_2 \big| \boldsymbol{\eta}_1,\boldsymbol{\eta}_3,\ldots,\boldsymbol{\eta}_r,\mathbf{y}),\ldots,f(\boldsymbol{\eta}_r \mid \boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_{r-1},\mathbf{y})$. The Gibbs sampler uses the conditionals mentioned above in construction of the transition kernel $P(\cdot|\cdot)$, for which stationary distribution of the Markov chain is equal to $\pi(\boldsymbol{\eta})=f(\boldsymbol{\eta}|\mathbf{y})$. This result is the consequence of the fact that $f(\boldsymbol{\eta}|\mathbf{y})$ is uniquely determined by the Gibbs conditionals.

Gibbs sampler algorithm can be described as follows:

Step 0. Choose the starting point $\boldsymbol{\eta}^{(0)}$ for components $\eta_1^{(0)},\ldots,\eta_r^{(0)}$, assuming, that k is equal 0. We can for example choose as the starting points the REML estimates for model parameters λ and EB estimates for μ parameters. But it can be an arbitrary set of points.

Step 1. Generate $\eta^{(k+1)}=(\eta_1^{(k+1)},\ldots,\eta_r^{(k+1)})$ in the following way. Draw $\eta_1^{(k+1)}$ using $f(\eta_1 \mid \eta_2^{(k)},\ldots,\eta_r^{(k)},y)$, than $\eta_2^{(k+1)}$ using $f(\eta_2 \mid \eta_1^{(k+1)},\eta_3^{(k)},\ldots,\eta_r^{(k)},y),\ldots$, and finally draw $\eta_r^{(k+1)}$ from $f(\eta_r \mid \eta_1^{(k+1)},\ldots,\eta_r^{(k+1)},y)$

Step 2. Set the $k=k+1$ and go to step 1.

The steps 1-2 constitute one cycle for each k. The sequence $\{\boldsymbol{\eta}^{(k)}\}$ generated by Gibbs sampler is the Markov chain with stationary distribution $\pi(\boldsymbol{\eta})=f(\boldsymbol{\eta}|y)$.

5. Assumptions for hierarchical model and model hyperparameters

As it was shown earlier (see (2.7)), the hierarchical model should contain several assumptions connected with *a priori* distributions that include the sampling scheme, the model that explains the observations and the model variability. Because in the paper estimates for counties (poviats) are considered, some difficulties here that arise mainly from too small sample size should be overcome. Direct estimates and their standard error were determined using a specific technique that assumes using balanced repeated replication technique (BRR) in situations where application of BRR is possible and bootstrap method, where using the BRR is impossible. This method was analyzed earlier (see Kubacki, Jędrzejczak and Piasecki (2011) or Kubacki, Jędrzejczak (2011)) and reveals effectiveness of such approach. The comparison of bootstrap precision estimates with other techniques, including Taylor linearization methods, indicates that both these techniques are nearly consistent. It should be noted that BRR method is applied now in Polish Household Budget Survey.

In the work considered here the following variables describing some income related categories were investigated:

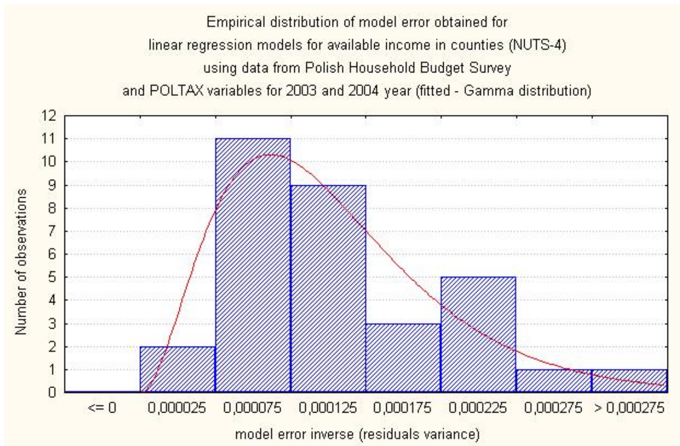
- available income
- income from hired work
- income from self-employment
- income from social security benefits
- retirement pays
- pensions resulting from inability to work
- family pensions
- income from other social benefits
- unemployment benefits.

The explanatory variables for the regression models come from POLTAX register and describe the following categories of income:

1. income from salary, related to employment
2. income from pension, rent (domestics)
3. income from economic activity carried out personally
4. income from property rights
5. income from tenancy or lease
6. income from other sources
7. income from special kind of agriculture production
8. discount from income (revenue) of universal insurance premium contribution
9. discount from tax (lump sum) of universal health insurance premium contribution,

and variables 5,6,7 were linked in one value (as a sum). These data was aggregated at the county - NUTS-4 - level (the anonymous POLTAX file contains the information about administrative unit down to NUTS-5 level) and then the indicator about average income from the mentioned above sources was determined by dividing the sums of this variable for NUTS-4 by the facto population (number of persons) for particular NUTS-4 unit. Such kind of explanatory variables was used for all target variables mainly because of time limit in the considered project. However, it seems that other sources of explanatory data could be used here. Here we can mention data from Polish Social Insurance Company (ZUS) and Labour Offices. This can be treated as an interesting investigation proposition due to the fact that the definitions of the described POLTAX variables only partially corresponds with Household Budget Survey income variables that can weaken the models for small areas.

Figure.1. Empirical distribution of model error obtained for linear regression for available income in counties (NUTS-4) using data from Polish Household Budget Survey and POLTAX variable for 2003 and 2004 year (fitted with Gamma distribution)



Source: Own calculations.

Parameters of distributions for model (2.7) were determined using shape parameters and scale parameters for Gamma distribution estimated from empirical distribution, achieved from the NUTS-4 level models (constructed separately for each region-voivodship NUTS-2). An example of such distribution is shown above.

6. Implementation of the hierarchical model in WinBUGS

In computation the WinBUGS and R-project software was used (also modules R2WinBUGS, coda and MASS). Special macro for R-project was prepared (its simplified example will be shown later), which was used as a connector with data input, performing necessary computations (including simulations in WinBUGS) and automatic visualization (here coda module was used).

In simulations the following computational schema was used. Similar schema was also used in earlier works that was done for hierarchical Bayes applications for small areas. Here we can mention two works: “Small Area Estimation with R Unit 5: Bayesian Small Area Estimation" (see Gomez-Rubio, V., 2008) and „Bayesian Spatial Modeling: Propriety and Applications to Small Area Estimation with Focus on the German Census 2011" (see Vogt, M., 2010). This scheme was as follows.

In the situation presented here Y[p] is related to the direct estimates, their estimation error tau[p], values from A[p] to G[p] are determined by values of explanatory variables for the model, parameters a0 and b0 come from empirical distribution of model error for linear regression and alphas are related to the linear regression coefficients.

```
model
{
  for(p in 1 : N) {
    Y[p] ~ dnorm(mu[p], tau[p])
    mu[p] <- alpha[1] + alpha[2] * A[p] + alpha[3] * B[p] + alpha[4] * C[p] + alpha[5] *
D[p] + alpha[6] * E[p] + alpha[7] * F[p] + alpha[8] * G[p] + u[p]
    u[p] ~ dnorm(0, precu)
  }
  precu ~ dgamma (a0,b0)
  alpha[1] ~ dflat()
  alpha[2] ~ dflat()
  alpha[3] ~ dflat()
  alpha[4] ~ dflat()
  alpha[5] ~ dflat()
  alpha[6] ~ dflat()
  alpha[7] ~ dflat()
  alpha[8] ~ dflat()
  sigmau<-1/precu
}
```

The macro in R-project environment has a (simplified) form like the code presented below. The code includes (for clarity of expression) only sections that present how the model parameters are determined and where simulations are done - with WinBUGS call. The rest of the code has more orderliness character and includes loading the necessary packages (here RODBC, R2WinBUGS and MASS is needed), setting the gamma parameters for σ_v^2 (here fitdistr function is called), reading the input data for particular region (here functions from RODBC package is used), and – after completing the simulations in WinBUGS – arranging the results and estimating the mean and variance (previously using read.coda function) as well as saving the results to the file (here standard cat and format function is used).

```
# determining the model parameters
model_HB<-paste("C:/Documents and Settings/PTS/Moje
dokumenty/model_kongres_demo.txt", sep = "")
infile  <- "coda1.txt"
indfile <- "codaindex.txt"
burn_in <- 3000
a0 <- dochg_shape
b0 <- dochg_rate
data <- list(N=N, Y=Y, tau=tau, A=A, B=B, C=C, D=D, E=E, F=F, G=G, a0=a0, b0=b0)
model <- lm( Y ~ 1 + A + B + C + D + E + F + G)
mod_smry <- summary(model)
alpha <- as.vector(mod_smry$coefficients[,1])
sigma_2 <- (mod_smry$sigma)*(mod_smry$sigma)
precu <- 1/sigma_2
u <- vector(mode = "numeric", length = N)
inits <- list(list(alpha=alpha, precu=precu, u=u))
parameters <- c("mu", "alpha", "precu", "u")

# simulations -  WinBUGS call
sim_HB <- bugs(data, inits, parameters, model_HB,n.chains=1, n.burnin = 1,
n.iter=10000, n.thin = 1)
```

7. Results and discussion

As it was mentioned earlier estimates from model for HB method (including assumptions for model (2.7)) have similar values as for EBLUP estimator, both for point estimates and for estimation error. The method applied here allows also for obtaining relatively stable simulation history, and the distributions for linear model μ have normal distribution. Normality is achieved also for model error components, and the distribution of σ_v^2 reveals consistency with Gamma distribution. The simulation history also does not have autocorrelation and achieve stability already from the beginning of the simulation. Below, the results of computations for Wielkopolskie voivodship were presented.

Some specific attribute for the computations here is the presence of autocorrelation for model error component in the case of Oborniki county (u[13] denotation). It is connected with relatively low direct estimation error, compared with simulation history for other counties. Such behaviour in MCMC simulation is observed also for other explanatory variables. But existence of such autocorrelation does not change much the normality of their distribution.

The dependencies above for MCMC simulations are observed also for other variables, but fitting the data is sometimes weaker. Achieving normality in such situations may indicate that the assumptions about normality for distributions about estimates and model errors may be in such situation satisfied. However, it is difficult to say whether this fact can be confirmed empirically, because in real situations the change of socio-economic conditions often can be observed what may change the level of the phenomenon (for example because of prize changes and GDP changes), so observed regularities may be characteristic for hypothetical populations often know as superpopulations.

The computations performed for Wielkopolskie voivodship reveal differences between estimation error for EBLUP and HB method, but for majority of similar models the estimation error estimates obtained using these two methods are relatively close. The comparison of REE distribution is presented in Figure 6. However, some differences are observed, and are shown in Figure 7. It is evident from that distribution, that for most cases the HB method has higher REE reduction, then EBLUP estimator. However, REE reduction for EBLUP has more flat patterns that REE reduction for HB method.

Table 1. Values of available income estimate obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 and Wielkopolskie voivodship with their precision estimate and relative estimation

error reduction obtained using direct estimation method and EBLUP method using REML technique

County (NUTS-4 unit)	Available income						
	Direct estimates			Estimates for EBLUP method (REML variant - SAE package)			REE reduction
	Para-meter estimate	Estima-tion error	REE (%)	Para-meter estimate	Estima-tion error	REE (%)	
Chodzieski	599.35	63.27	10.56	560.36	33.69	6.01	1.756
Czarnkowsko-Trzcianecki	503.02	80.88	16.08	565.86	28.15	4.97	3.233
Gnieźnieński	506.33	47.71	9.42	586.35	34.20	5.83	1.616
Gostyński	556.11	76.08	13.68	575.33	29.10	5.06	2.705
Grodziski	530.14	51.71	9.75	534.09	36.75	6.88	1.417
Jarociński	731.52	129.69	17.73	581.59	28.62	4.92	3.603
Kępiński	552.20	16.41	2.97	555.65	21.06	3.79	0.784
Kolski	634.46	54.89	8.65	545.68	33.38	6.12	1.414
Koniński	530.42	78.88	14.87	537.14	36.91	6.87	2.164
Kościański	547.35	43.21	7.89	563.69	31.80	5.64	1.399
Krotoszyński	580.99	52.75	9.08	560.27	30.59	5.46	1.663
Nowotomyski	759.51	196.83	25.92	561.16	42.17	7.51	3.449
Obornicki	667.71	4.06	0.61	667.25	4.36	0.65	0.932
Ostrowski	619.02	37.61	6.08	615.20	31.63	5.14	1.182
Ostrzeszowski	579.69	43.57	7.52	569.91	33.26	5.84	1.288
Pilski	728.53	94.61	12.99	625.12	38.75	6.20	2.095
Pleszewski	598.08	86.58	14.48	571.59	34.60	6.05	2.392
Poznański	683.95	85.94	12.57	754.01	43.44	5.76	2.181
Rawicki	694.54	63.63	9.16	571.44	42.61	7.46	1.229
Słupecki	526.62	52.33	9.94	555.81	33.18	5.97	1.665
Szamotulski	588.32	45.80	7.78	586.45	34.01	5.80	1.342
Średzki	594.31	54.73	9.21	610.19	30.31	4.97	1.854
Śremski	670.13	57.39	8.56	583.47	35.56	6.09	1.405
Turecki	457.04	48.58	10.63	513.10	42.97	8.37	1.269
Wągrowiecki	505.59	51.85	10.26	573.06	29.37	5.12	2.001
Wolsztyński	567.58	44.26	7.80	575.68	33.81	5.87	1.328
Wrzesiński	568.85	39.09	6.87	580.39	30.90	5.32	1.291
Złotowski	558.94	45.12	8.07	567.62	31.85	5.61	1.438
m. Kalisz	635.61	13.24	2.08	638.30	16.99	2.66	0.783
m. Konin	699.53	119.79	17.12	622.06	52.57	8.45	2.026
m. Leszno	664.60	74.80	11.26	690.10	53.55	7.76	1.450
m. Poznań	931.31	44.42	4.77	915.60	46.62	5.09	0.937

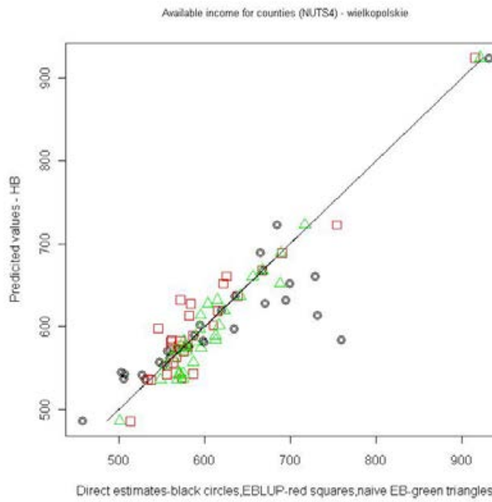
Source: Own calculations.

Table 2. Values of available income estimate obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 and Wielkopolskie voivodship with their precision estimate and relative estimation error reduction obtained using direct estimation method and hierarchical Bayes estimation

County (NUTS-4 unit)	Available income						
	Direct estimates			Estimates using hierarchical Bayes method			REE reduction
	Parameter estimate	Estimation error	REE (%)	Parameter estimate	Estimation error	REE (%)	
Chodzieski	599.35	63.27	10.56	581.41	48.17	8.28	1.274
Czarnkowsko-Trzcianecki	503.02	80.88	16.08	544.23	52.31	9.61	1.673
Gnieźnieński	506.33	47.71	9.42	542.99	41.29	7.60	1.239
Gostyński	556.11	76.08	13.68	570.27	51.95	9.11	1.502
Grodziski	530.14	51.71	9.75	536.06	43.51	8.12	1.202
Jarociński	731.52	129.69	17.73	613.32	61.94	10.1	1.756
Kępiński	552.20	16.41	2.97	552.89	15.88	2.87	1.035
Kolski	634.46	54.89	8.65	597.46	45.14	7.56	1.145
Koniński	530.42	78.88	14.87	535.81	55.43	10.4	1.437
Kościański	547.35	43.21	7.89	556.84	36.49	6.55	1.205
Krotoszyński	580.99	52.75	9.08	575.07	41.91	7.29	1.246
Nowotomyski	759.51	196.83	25.92	583.59	78.11	13.4	1.936
Obornicki	667.71	4.06	0.61	667.47	4.09	0.61	0.993
Ostrowski	619.02	37.61	6.08	618.99	33.83	5.47	1.112
Ostrzeszowski	579.69	43.57	7.52	576.71	38.49	6.67	1.126
Pilski	728.53	94.61	12.99	660.38	63.46	9.61	1.351
Pleszewski	598.08	86.58	14.48	582.44	56.15	9.64	1.502
Poznański	683.95	85.94	12.57	722.69	64.37	8.91	1.411
Rawicki	694.54	63.63	9.16	631.92	53.71	8.50	1.078
Słupecki	526.62	52.33	9.94	541.89	42.12	7.77	1.278
Szamotulski	588.32	45.80	7.78	589.18	39.06	6.63	1.174
Średzki	594.31	54.73	9.21	601.61	43.88	7.29	1.263
Śremski	670.13	57.39	8.56	627.53	46.43	7.40	1.157
Turecki	457.04	48.58	10.63	485.88	45.10	9.28	1.145
Wągrowiecki	505.59	51.85	10.26	536.97	41.73	7.77	1.320
Wolsztyński	567.58	44.26	7.80	571.44	38.10	6.67	1.170
Wrzesiński	568.85	39.09	6.87	574.71	33.81	5.88	1.168
Złotowski	558.94	45.12	8.07	562.99	38.27	6.80	1.187
m. Kalisz	635.61	13.24	2.08	636.71	12.91	2.03	1.028
m. Konin	699.53	119.79	17.12	651.80	77.76	11.9	1.436
m. Leszno	664.60	74.80	11.26	689.04	63.24	9.18	1.226
m. Poznań	931.31	44.42	4.77	924.24	43.95	4.76	1.003

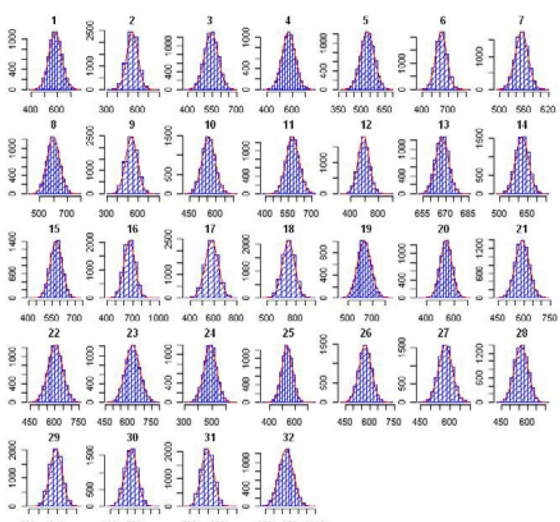
Source: Own calculations.

Figure 2. Observed vs. predicted plot for available income per capita estimates obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 and counties in Wielkopolskie voivodship estimated by direct estimator (black circles), EBLUP estimator (red squares) naïve EB estimator (green triangles) and hierarchical Bayes estimator



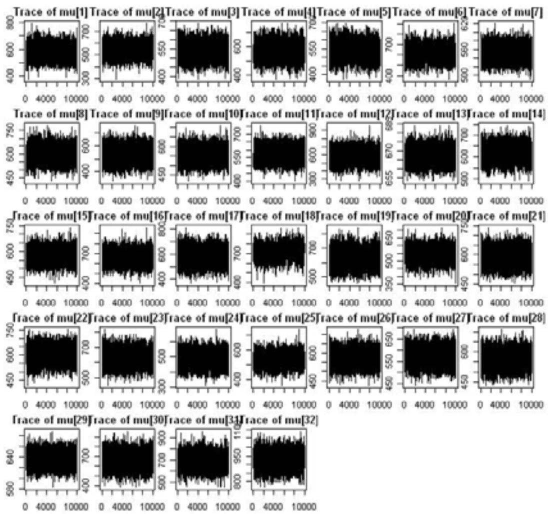
Source: Own calculations.

Figure 3. Plots of distributions of model estimates for available income per capita obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 year and counties in Wielkopolskie voivodship obtained by MCMC simulation using Gibbs sampler



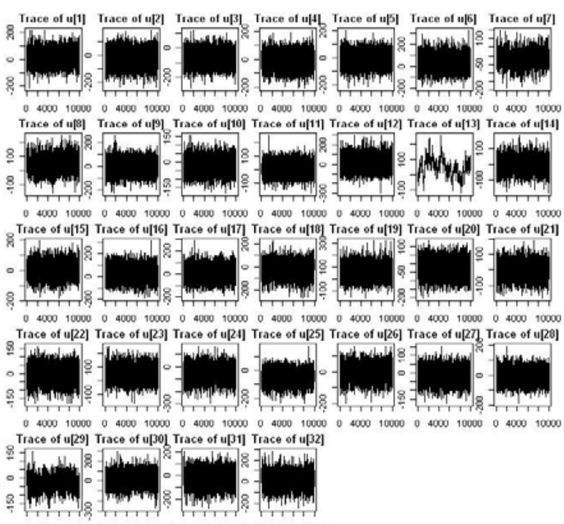
Source: Own calculations.

Figure 4. Plots of simulation history for model estimates of available income per capita obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 year and counties in Wielkopolskie voivodship obtained using Gibbs sampler



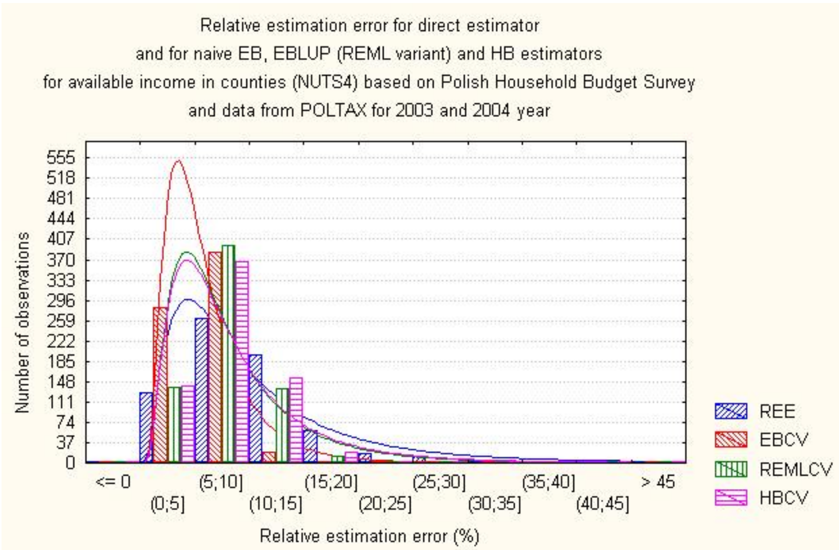
Source: Own calculations.

Figure 5. Plots of simulation history for model error of available income per capita obtained from Polish Household Budget Survey and selected variables from POLTAX register for 2003 and counties in Wielkopolskie voivodship obtained using Gibbs sampler



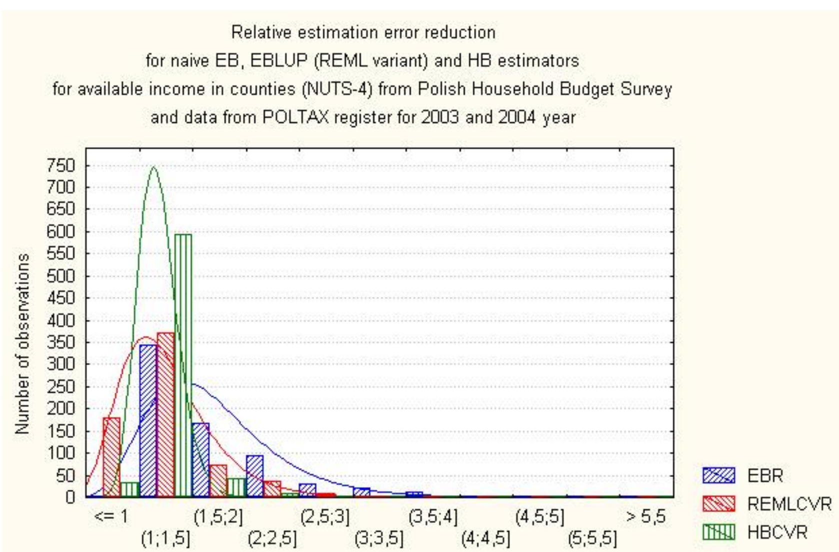
Source: Own calculations.

Figure 6. Distribution of relative estimation error for direct estimator and for naïve EB, EBLUP (REML variant) and HB estimators for available income in counties (NUTS4) based on Polish Household Budget Survey and data from POLTAX register for 2003 and 2004 year



Source: Own calculations.

Figure 7. Distribution of relative estimation error reduction for naïve EB, EBLUP (REML variant) and HB estimators for available income in counties (NUTS4) based on Polish Household Budget Survey and data from POLTAX register for 2003 and 2004 year



Source: Own calculations.

The differences observed for Wielkopolskie voivodship can be explained by weaker fit of the model. Such behaviour for Wielkopolskie region is visible also for ordinary regression models and that in fact can be a limitation on using the HB methods. However, it should be mentioned, that for more specific variables (for example for family pensions or unemployment benefits) the hierarchical models considered here have such an advantage that they rapidly achieve convergence, in contrast to loss of convergence, as it can be observed for some EBLUP models. It is, however, not the property of the hierarchical model itself, but the selection of the parameters of the model. As it was confirmed empirically, other parameters set for Gamma distribution using for σ_v^2 (as it was used for example in Vogt (2010) work, equal a=0.5, b=0.0005) do not behave properly for more specific variables. For that parameters of σ_v^2 the autocorrelation and sometimes the lack of stability (for example the oscillations for longer runs) are observed. Thus, application of such more general approach is not always efficient.

It should be noted here that such selection of parameters is only possible when more cases of similar models are available (as it was characteristic for counties models considered here). In more individual cases (for example when model for available income for voivodship is considered), the availability of more model cases is reduced. In such situation application of other strategy can be more suitable. One of such approaches (when σ_v^2 is not known) was shown in Rao book in part 10.3.3. The comparison of two of these methods may allow for more comprehensive assessment of methods used in this work.

8. Conclusions

In the paper the usefulness of estimates conducted by the hierarchical Bayes estimation in the case of known values of hyperparameters was demonstrated. Some consistency between hierarchical Bayes and other types of small area methods, for example EBLUP method, was shown. For this technique slightly better efficiency than EBLUP estimators was observed, but for less fitted model it could not be the rule. Because of good properties of computations shown in the paper (lack of autocorrelation and practically neglect of burn-in), it can be judged that such approach may be applied in practice. Unfortunately, in the situation described here some preliminary knowledge about the distribution of σ_v^2 is required, what may be sometimes difficult to obtain. In the case of counties it is, however, possible, and may be beneficial for practical reasons.

REFERENCES

BRACHA, CZ., LEDNICKI, B, WIECZORKOWSKI, R. (2003). Data Estimation for Polish Labour Force Survey for counties in 1995-2002 (in Polish - Estymacja danych z Badania Aktywności Ekonomicznej Ludności na poziomie powiatów dla lat 1995-2002), GUS, Warszawa.

BRACHA, CZ., LEDNICKI, B., WIECZORKOWSKI, R. (2004). Application of Complex Estimation Methods to the Disaggregation of data from Polish Labour Force Survey in 2003 (in Polish - Wykorzystanie złożonych metod estymacji do dezagregacji danych z Badania Aktywności Ekonomicznej Ludności w roku 2003), GUS, Warszawa, seria „Z prac Zakładu Badań Statystyczno-Ekonomicznych", z.299 .

CENTRAL STATISTICAL OFFICE (2000-2011). Household Budget Surveys (years 1999-2010) Statistical Information and Elaborations (in Polish Budżety gospodarstw domowych, (lata 1999-2010) Informacje i opracowania statystyczne), Warszawa, http://www.stat.gov.pl/gus/5840_3467_PLK_HTML.htm.

CENTRAL STATISTICAL OFFICE (2010). Polish Household Budget Survey Methodology (in Polish Metodologia Badania Budżetów Gospodarstw Domowych, Zeszyt metodologiczny zaopiniowany przez Komisję Metodologiczną GUS). Warszawa. http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL_WZ_meto_badania_bud__gospod__dom.pdf.

DEHNEL, G., (2003). Small Area Statistics as a Tool for Assessment of Regions Economic Development (In Polish: Statystyka małych obszarów, jako narzędzie oceny rozwoju ekonomicznego regionów), Wydawnictwo Akademii Ekonomicznej, Poznań.

DOMAŃSKI, CZ., PRUSKA, K. (2001). Methods of Small Area Statistics (in Polish - Metody statystyki małych obszarów), Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

GOŁATA, E. (2004). Indirect Estimation of Unemployment for the Local Labor Market (In Polish: Estymacja pośrednia bezrobocia na lokalnym rynku pracy), Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.

GOMEZ-RUBIO, V. (2008). "Small Area Estimation with R Unit 5: Bayesian Small Area Estimation", useR! 2008 11 August 2008, Dortmund (Germany), http://www.bias-project.org.uk/SAE_tutorial/useR08-tutorial.tgz.

KALTON, G., KORDOS, J., and PLATEK, R. (1993). Small Area Statistics and Survey Designs Vol. I: Invited Papers: Vol. 11: Contributed Papers and Panel Discussion, Warszawa, Główny Urząd Statystyczny.

KUBACKI, J. (2004). Application of the Hierarchical Bayes Estimation to the Polish Labour Force Survey, Statistics in Transition, Vol. 6, No. 5, 785-796. http://www.stat.gov.pl/cps/rde/xbcr/gus/PTS_sit_6_5.pdf.

KUBACKI, J. (2006). The Problems of Small Area Parameters Estimation in Polish Labor Force Survey (In Polish: Problematyka szacowania parametrów dla małych obszarów w badaniu aktywności ekonomicznej ludności), unpublished PhD thesis prepared in connection of PhD Studies in the College of Economic Analysis, Warsaw School of Economics.

KUBACKI, J., JĘDRZEJCZAK, A., PIASECKI, T. (2011). Application of Small Area Statistics Methods in Elaboration of Sample Surveys Results, Report from methodological study 3.065, Statistical Office in Łódź (in Polish Wykorzystanie metod statystyki małych obszarów do opracowania wyników badań statystycznych, Raport z pracy metodologicznej 3.065), Ośrodek Statystyki Matematycznej, Urząd Statystyczny w Łodzi.

KUBACKI, J., JĘDRZEJCZAK, A. (2011). The Comparison of Generalized Variance Function with Other Methods of Precision Estimation for Polish Household Budget Survey, Studia Ekonomiczne, Uniwersytet Ekonomiczny w Katowicach (in preparation).

LIU, B. (2009). Hierarchical Bayes Estimation and Empirical Best Prediction of Small Area Proportions, Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, http://drum.lib.umd.edu/bitstream/1903/9149/1/Liu_umd_0117E_10245.pdf.

LONGFORD, N.T. (2005). Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician, Springer-Verlag, New York.

MUKHOPADHYAY, P. (1998). Small Area Estimation in Survey Sampling, Narosa Pub House.

PLUMMER, M., BEST, N., COWLES, K. and VINES, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, vol. 6, 7-11.

RAO, J.N.K. (2003). Small Area Estimation, Wiley Interscience, Hoboken, New Jersey.

SALVATI, N., GÓMEZ-RUBIO, V. (2006). SAE: Small Area Estimation with R. R package version 0.07, http://www.bias-project.org.uk/software/SAE_0.07.zip.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N., and LUNN, D. (2003). WinBUGS User Manual, Version 1.4.

STURTZ, S., LIGGES, U., and GELMAN, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R., Journal of Statistical Software, 12(3), 1-16.

VENABLES, W.N. RIPLEY, B.D. (2002). Modern Applied Statistics with S, Fourth Edition. Springer, New York.

VOGT, M. (2010). Bayesian Spatial Modeling: Propriety and Applications to Small Area Estimation with Focus on the German Census 2011, PhD Thesis, University of Trier, http://ubt.opus.hbz-nrw.de/volltexte/2010/578/pdf/Dissertation_Martin_Vogt.pdf.

ŻĄDŁO, T. (2008). Elements of small area statistics with R software (in Polish - Elementy statystyki małych obszarów z programem R), Akademia Ekonomiczna Katowice.