ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 216, 2008

*Dorota Bartosińska**

BAYES ESTIMATION IN AGRICULTURAL SAMPLE SURVEYS IN POLAND

ABSTRACT. Direct estimators used in sample surveys usually provide parameters' estimates for country and regions. They do not provide estimates for smaller cross-sections (age, gender etc.) or smaller geographical areas (subregions, counties, towns and communes). One of the possibilities to obtain such estimates is Bayes approach. It is based on known information beyond the sample. There were considered two Bayes estimators: empirical and hierarchical to obtain precise estimates for counties in agricultural sample surveys carried out by Central Statistical Office in Poland. Additional source of information was Census of Agriculture, whose data are correlated with data from agricultural sample surveys.

Key words: Bayes estimation, agricultural sample survey, small area estimation.

I. INTRODUCTION

In sample surveys carried out by official statistical services both in Poland and other countries, direct estimators are usually used, based only on results from a sample. They provide parameters' estimates for basic cross-sections of a country as a whole and for large areas like regions. However, they do not provide estimates for smaller cross-sections such as age, gender etc. or smaller geographical areas such as subregions, counties, towns and communes. One of the possibilities to obtain such estimates is Bayes approach, based on known information beyond the sample.

The aim of the paper is to estimate parameters for counties in agricultural sample surveys carried out by the Central Statistical Office in Poland. Two Bayes estimators were considered: the empirical one and the hierarchical one. Additional information was provided by Census of Agriculture (CA).

The author describes briefly the applications of Bayes approach in small area estimation in Poland up to now, sources of basic and auxiliary data and applied estimators There are also presented some results of Bayes estimation and conclusions from the analysis.

* Ph.D., Chair of Statistics and Econometrics, University of Maria Curie Skłodowska in Lublin.

II. PREVIOUS APPLICATIONS OF BAYES APPROACH IN SMALL AREA ESTIMATION IN POLAND

So far, Bayes approach to small area estimation was applied in household surveys to obtain precise estimates of structure by number of persons (Kordos, Paradysz, 2000). It was also applied in labour force surveys, where the number of the unemployed, employed and economically inactive was estimated (Bracha, Lednicki, Wieczorkowski, 2004; Kordos, 2006; Kubacki, 2006).

The attempts at application of Bayes estimation were also made in agricultural sample surveys (Kordos, Paradysz, 2000; Bartosińska, 2005). In both the later research Census of Agriculture was used as source of auxiliary data. In the first one livestock inventory in 1999 for regions, and livestock inventory and crop acreage in 1998 for counties were estimated. In the second one some agricultural characteristics for counties in 1998 and 2001 were estimated. Some results of this research are given below.

III. SOURCES OF BASIC AND AUXILIARY DATA

Surveys of land use, crop acreage and livestock inventory were the sources of basic data. These surveys are called as June Agricultural Surveys (JAS). They were carried out by the Central Statistical Office in Poland in 1998 and 2001. The totals for counties were estimated on example of the Lublin region. The sample selected for JAS 1998 involved about 10 thousand farms from the Lublin region. It was about 3.2% of the population. In JAS 2001 the sample for the Lublin region involved 5437 farms. It was about 1.7% of the population.

Census of Agriculture carried out by the Central Statistical Office in Poland in 1996 was the source of auxiliary. There were about 300 thousand farms surveyed during this census in the Lublin region. Census of agriculture uses the same concepts, definitions and classifications as agricultural sample surveys following a given census.

IV. APPLIED ESTIMATORS

As mentioned earlier, two Bayes estimators were considered to obtain precise estimates for counties: empirical and hierarchical one.

The empirical Bayes estimator of the total of the variable of interest Y for d th small area is given by (Kordos, Paradysz, 2000):

$$y_{d,EB} = \frac{d^2(y_d)}{d^2(y_d)+d^2(y_{d,SYN,R})} y_{d,SYN,R} + \frac{d^2(y_{d,R})}{d^2(y_d)+d^2(y_{d,SYN,R})} y_d, \qquad (1)$$

where:

y_d – direct estimator of the total of Y for d th small area,

$d^2(y_d)$ – variance of direct estimator of the total of Y for d th small area,

$y_{d,SYN,R}$ – regression estimator of the total of Y for d th small area,

$d^2(y_{d,SYN,R})$ – variance of regression estimator of the total of Y for d th small area.

The hierarchical Bayes estimator of the total of the variable of interest was calculated as a result of simulation by Monte Carlo Markov Chain method using software WinBugs.

In both Bayes estimators the following linear regression model was used:

$$\mathrm{y_d} = \mathrm{X_d^T}\beta + \mathrm{u} + \mathrm{e_d}; \qquad (2)$$

where:

y_d – estimate of the total of the variable of interest Y for d th small area,

$\mathbf{X}_d = [X_{dj}]$ – matrix of the totals of auxiliary variables for d th small area,

$\beta_{\mathrm{area}} = [\beta_j]$ – vector of k area-level regression parameters,

u – model-based random variable,

e_d – design-based random variable for d th small area.

The direct estimator is rather inefficient for small areas and it serves as a benchmark against which other estimators can be compared. The direct estimator of the total of the variable of interest Y for d th small area is given by:

$$\mathrm{y_d} = \sum_{\mathrm{i=1}}^{\mathrm{n_d}} \frac{\mathrm{y_{di}}}{\pi_{\mathrm{di}}}; \qquad (3)$$

where:

y_{di} – value of the variable of interest Y for ith unit in dth small area,

π_{di} – inclusion probability for ith unit in dth small area.

V. RESULTS OF ESTIMATION

The estimation precision for the counties, when direct estimator based only on the sample survey data is used, was low. Coefficients of variation (CV) of direct estimates fluctuated between 4.4 and 45.0% for two features of smaller variation: numbers of cows and pigs; and between 5.8 and 99.6% for two features of larger variation: crop acreage of sugar beet and rape.

There were used linear regression to obtain both Bayes estimates. The dependent variable was taken from JAS and independent variables were taken from CA. Independent variables were chosen by backward step by step regression method. As independent variables there were included the variables of interest from CA. Other potential independent variables were either weakly correlated with the variable of interest or strongly correlated with other independent variables. That's why they had to be removed from the regression models. Some results of regression are presented in Table 1.

Table 1

Some results of regression JAS data on CA data

Dependent variable	Independent variable	In	Intercept	Slope	R^2
Number of cows in 1998	Number of cows in 1996	thous.	0.020	1.046	0.974
Number of cows in 2001		thous.	0.648	0.826	0.875
Number of pigs in 1998	Number of pigs in 1996	thous.	–3.404	1.222	0.963
Number of pigs in 2001		thous.	11.632	0.879	0.794
Crop acreage of sugar beet in 1998	Crop acreage of sugar beet in 1996	ha	–87.148	0.919	0.947
Crop acreage of sugar beet in 2001		ha	358.520	0.603	0.896
Crop acreage of rape in 1998	Crop acreage of rape in 1996	ha	145.353	1.518	0.451
Crop acreage of rape in 2001		ha	85.661	1.140	0.766

Source: own calculations based on data from the GUS.

The Bayes estimation precision where the regression models were used was significantly better than direct estimation precision for all analysed features for all counties. In Figure 1 coefficients of variation of EB and HB estimates were compared to direct estimates on the example of one feature: the number of pigs in JAS 2001. The figure shows that all CVs are significantly smaller for EB and HB than for direct estimates.

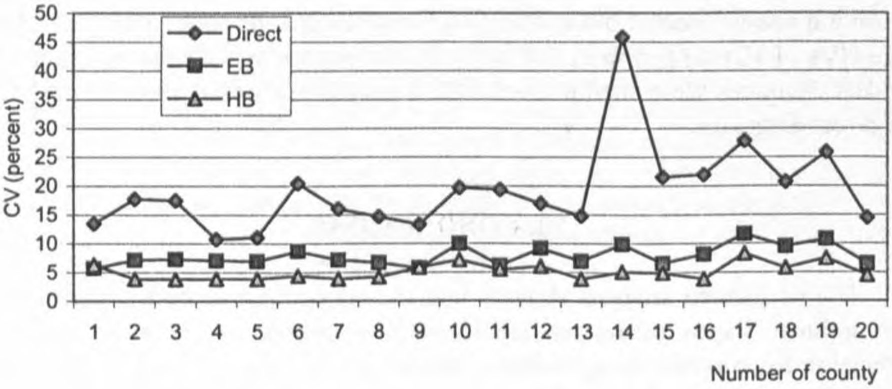

Figure 1. Coefficients of variation of number of pigs' estimates for counties in the Lublin region in 2001

Table 2 presents CVs of direct, EB and HB estimates of all studied features for all counties in the Lublin region.

CVs of EB fluctuated between 1.6 and 11.7% for two features of smaller variation: numbers of cows and pigs; and between 3.6 and 96.6% for two features of larger variation: crop acreage of sugar beet and rape. Average CVs of EB ranged from 2.8 to 40.4% for particular features. Average CVs of EB estimates were smaller by 6.2–21.8 percentage points than average CVs of direct estimates.

Table 2

Coefficients of variation of three different estimates for counties in the Lublin region (%)

Variable of interest	Year	Direct			EB			HB		
		Min	Average	Max	Min	Average	Max	Min	Average	Max
Number of cows	1998	4.4	11.8	17.9	1.6	2.8	7.1	1.8	2.4	5.0
	2001	7.7	14.7	25.5	3.8	5.4	9.1	2.8	3.7	7.8
Number of pigs	1998	7.1	14.1	28.0	2.4	4.4	7.3	2.3	3.1	5.5
	2001	10.8	19.2	45.0	5.7	7.9	11.7	3.8	5.2	8.3
Crop acreage of sugar beet	1998	5.8	29.8	72.3	3.6	25.5	73.0	0.5	1.1	3.3
	2001	12.3	37.4	99.0	5.6	21.8	77.0	4.7	16.7	83.7
Crop acreage of rape	1998	10.2	33.8	96.4	10.2	27.6	83.8	5.5	7.9	9.9
	2001	9.7	40.4	99.6	9.4	40.4	96.6	6.8	15.5	68.9

Source: own calculations based on data from the GUS.

CVs of HB fluctuated between 1.8 and 8.3% for two features of smaller variation and between 0.5 and 83.7% for two features of larger variation. Average CVs of HB ranged from 2.4 to 16.7% for particular features. Average CVs of HB estimates were smaller by 9.4–28.7 percentage points than average CVs of direct estimates.

VI. CONCLUSIONS

The performed analysis showed that the application of both empirical and hierarchical Bayes estimators significantly improves parameters' estimation precision for counties in agricultural sample surveys. Bayes estimation requires further research to test its usefulness to estimate other agricultural characteristics in other sample surveys, using other sources of auxiliary data, such as Integrated Administration and Control System (IACS).

REFERENCES

Bartosinska D. (2005), *Small Area Estimation Methods in Agricultural Sample Surveys (Using Data of Census of Agriculture)*, Warsaw School of Economics, (mimeo of doctoral dissertation in Polish).

Bracha Cz., Lednicki B., Wieczorkowski R. (2004), *Utilization of Composite Estimation Methods from the Labour Force Surveys by Region and Poviat in 2003*, Research Centre of the Central Statistical Office and the Polish Academy of Sciences, No. 299, Warsaw 2004 (in Polish).

Kordos J. (2006), Impact of Different Factors on Research in Small Area Estimation in Poland, *Statistics in Transition*, Vol. 7, No. 4, pp. 863-879.

Kordos J., Paradysz J. (2000), Some Experiments in Small Area Estimation in Poland, *Statistics in Transition*, Vol. 4, No. 4, pp. 679-697.

Kubacki J. (2006), Remarks on Using the Polish LFS Data and SAE Methods for Unemployment Estimation by County, *Statistics in Transition*, Vol. 7, No. 4, pp. 901-916.

Rao J.N.K. (2003), *Small Area Estimation*, John Wiley & Sons, New Jersey.

Dorota Bartosińska

ESTYMACJA BAYESOWSKA W REPREZENTACYJNYCH BADANIACH ROLNICZYCH W POLSCE

W badaniach reprezentacyjnych, prowadzonych przez statystykę publiczną w Polsce i innych krajach, są stosowane estymatory bezpośrednie, oparte wyłącznie na wynikach z próby. Dostarczają one ocen parametrów dla podstawowych przekrojów kraju jako całości i dla większych obszarów, jak województwa. Natomiast nie dają ocen dla mniejszych przekrojów, jak: wiek, płeć itp. oraz dla mniejszych obszarów, jak: podregiony, powiaty, miasta, gminy. Jedną z możliwości uzyskania takich ocen jest podejście bayesowskie, oparte na znanej informacji spoza próby. W artykule rozważa się dwa estymatory bayesowskie: empiryczny i hierarchiczny, aby uzyskać precyzyjne oceny parametrów dla powiatów w reprezentacyjnych badaniach rolniczych prowadzonych przez GUS w Polsce. Źródłem informacji dodatkowych jest pełny spis rolny. Zastosowanie tych estymatorów daje oceny parametrów dla powiatów o dużej precyzji, w przypadku istnienia znacznej korelacji między wynikami z pełnego spisu rolnego i z reprezentacyjnych badań rolniczych prowadzonych po danym spisie.