

# Analysis of repeated business surveys in Ukraine

Oksana Honchar<sup>1</sup> and Tetiana Ianevych<sup>2</sup>

<sup>1</sup>Scientific & Technical Complex for Statistical Research, National Academy of Statistics,  
Accounting & Audit, e-mail: ohonchar@list.ru

<sup>2</sup>Taras Shevchenko National University of Kyiv, e-mail: yakovenkot@gmail.com

## Abstract

In the paper methodology of repeated surveys is considered on example Ukrainian investment survey. Short characteristic and sample design of quarter investment survey are presented. The allocation depending on the fraction  $\alpha$  of the sample size is considered. Horvitz-Thompson and regression estimator using information from previous survey are compared using simulation study.

*Keywords:* Investment survey, repeated surveys, Horvitz-Thompson estimator, regression estimator, bias, mean square error.

## 1 Introduction

Most surveys in short-term business statistics are repeated with some periodicity. In most cases they are conducted monthly or quarterly but sometimes also have weekly or two-month periodicity. For such surveys special methodology should be used.

As an example we consider the Ukrainian investment survey, which plays an important role as a source of data regarding capital investment, along with its components, and provides the data for constructing the capital investment index, which is one from the main indicators of economical statistics. The capital investment survey is conducted with annual and quarter periodicity. The main aim of annual survey of capital investment is estimation of the structure whereas quarter survey has estimation of changes as an objective.

All enterprises with capital investment in a reporting year should be observed. In the quarter survey only enterprises that have significant size of capital investment are observed. Actually only enterprises which took part in other quarter surveys (mostly in structural business survey) can be observed quarterly.

The first problem in the quarter survey is in the future frame for this survey will be formed on the central level (in SSSU). This population will be formed from the business register. In 2010 the frame included 689 042 enterprises. For the quarter survey some threshold is used to take part in the survey. Until now it was defined by capital investment itself since before to obtain questionnaire two questions were asked to a potential respondent: 1) if enterprise had capital investment in this quarter and if yes 2) what was size of capital investment. If size exceeded some threshold (different by activities) an enterprise was observed. However forming of population on the central level does not presume knowledge of capital investment size so using it for threshold definition is impossible.

The second problem concerns with quality of the quarter data from the investment survey. Quarter capital

investment was not estimated for enterprises that were not observed. So we have a problem of underestimating capital investment size in quarter surveys. It can be no problem for capital investment index calculation because size of underestimation is proportional to proper size of capital investment by domains. But it becomes problem using absolute values of capital investment.

## 2 Sampling design

The main characteristics of Ukrainian investment survey are presented in Table 1.

Table 1: Characteristics of Ukrainian investment survey

Kind of economic activity	• All kinds
Geographical coverage	• All regions in Ukraine (27)
Unit	• Enterprise
Periodicity	• Quarter; • Annual
Main variable	• Total capital investment
Kind of survey	• quarterly: – now: cut-off with census; – plan: cut-off with sample: census for large and middle enterprises; sample for small enterprises; • annually: census
Levels of publication	• National level; • Regions; • Kinds of economic activities (2digit by NACE); • Institutional sector of economy; • Organization and legal form of management

Sample frame for quarter investment survey is formed from the business statistical register. Then some data is added from the previous year's annual structural business survey and investment survey for sampling design construction. Sample frame consists of three subpopulations: financial sector of economy, non-financial sector and new enterprises. For sample survey small enterprises are separated in terms of financial and nonfinancial sectors with all new enterprises covered regardless of size.

For small enterprises stratification by activity (2-digit of NACE) and numbers of employees (0, 1-2, 3-5, 6-9, 10-19, and 20-50) is used. Then outliers are detected. Enterprise is atypical (outlier) if the condition  $|y_{hi} - \bar{y}_h| < 3 \cdot \sigma_h$  does not hold. Here  $y_{hi}$  – capital investment for unit  $i$  in stratum  $h$ ,  $\bar{y}_h$  – mean of capital investment in stratum  $h$ ,  $\sigma_h$  – standard deviation of capital investment in stratum  $h$ .

Total sample size is about 100000 units (Table 2).

Table 2: Sample size in Ukrainian investment survey in 2010

Subpopulation	Size	%
Large & middle enterprises	35 139	100
Outliers and small strata	2 088	100
Small enterprises observed with $p < 1$	59 292	9.5
New enterprises	3 305	10
Total	99 824	14.5

Sample with such design gives quite good results for total investment but the problem of small domain estimation still exists. For estimation Horvitz-Thompson and regression estimates are calculated. As auxiliary information, the number of employees in current period is used.

### 3 Using information from previous survey

As long as this is repeated survey, for improvement of estimates accuracy, as auxiliary information, we decided to use capital investment from previous year's survey.

Resources of investigation are:

1. Quarter data on capital investment in 2010;
2. Annual data in 2009, 2010;
3. Statistical register data in 2010.

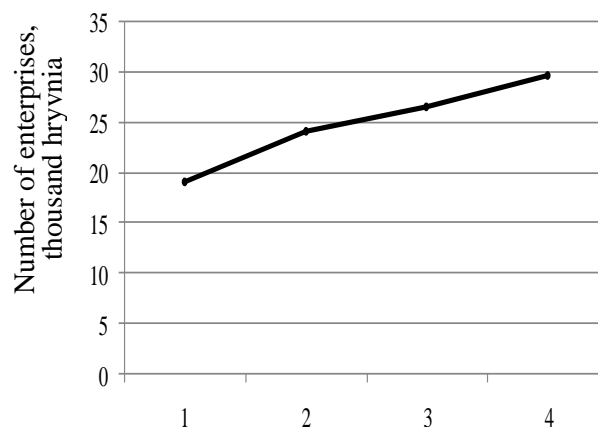
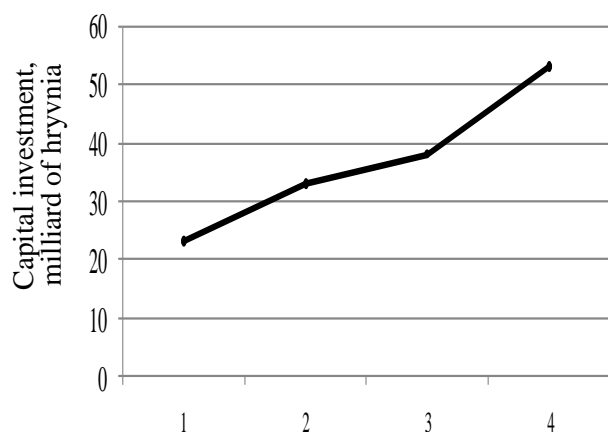
Table 3: Distribution of enterprises by capital investment size by quarters in 2010

Period	Capital investment, thousand hryvnia	Rate of quarter capital investment in annual investment, %	Number of enterprises
quarter 1	23 220 146	13,7	19 017
quarter 2	33 432 883	19,7	23 658
quarter 3	38 397 367	22,7	26 416
quarter 4	52 946 719	31,2	29 590
Sum of quarters	147 997 115	<b>87,3</b>	×
Annual survey	169 434 303	100,0	71 352

Table 4: Distribution of enterprises by quarters in 2010

Criterion	Number of enterprises	Rate of enterprises, %
Enterprises which were observed:		
in one quarter	14 317	19,9
in two quarters	8 563	11,9
in three quarters	8 138	11,3
in four quarters	10 706	14,9
only in annual survey	30 354	42,1
Number of enterprises reported in annual survey	72 078	100,0

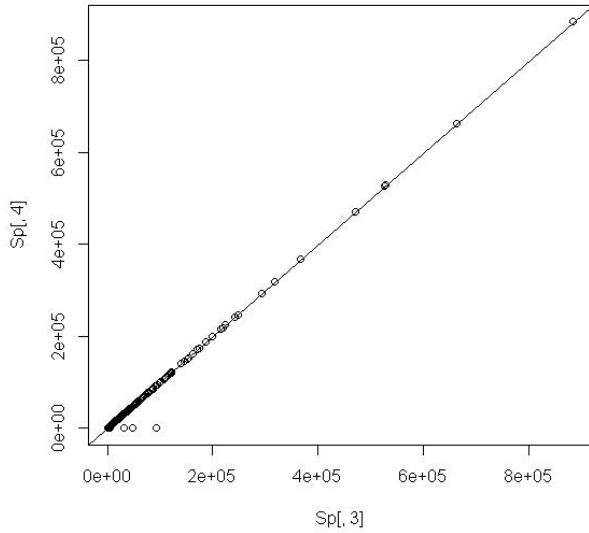
Capital investment and number of enterprises by quarters in 2010



First we compare the behavior of the H-T and regression estimator on the population of all enterprises. It was divided into two parts: those that had capital investment in 2009 and those that not.

These subpopulations are essentially different. So, in 2009 93% enterprises had not capital investments and it can be concluded from the data we have, that it is most likely that the enterprises that had not investment during one year would not have it during the next. The subpopulation of such enterprises is very homogeneous (it consists mostly of zeros) therefore, the use of the H-T estimator is appropriate there.

The subpopulation of enterprises that have positive capital investment in 2009 is not so homogeneous. So, for the beginning we compared the behavior of the H-T and regression estimator on this subpopulation



within simple random sampling. Besides this, we investigated what is the best allocation of the sample between this subpopulation. Let us consider 4 cases of the allocation depending on the fraction  $\alpha$  of the sample size  $n$ .

Table 5: Sample allocations

Sample sizes	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha \approx 0.77$
$n_p$	29646	35575	41504	45598
$n_0$	29646	23717	17788	13694

In the Table 5 the value  $n_p$  is the sample size for the subpopulation of enterprises that had capital investment in 2009, and the value  $n_0$  is the sample size for the subpopulation with zero investment. The last case ( $\alpha \approx 0.77$ ) corresponds to the situation when the enterprises with nonzero investments in 2009 is observed with probability one.

The parameter  $t$  that we are interesting in is the total capital investment for small enterprises in 2010. The estimates of this parameter has two terms  $\hat{t} = \hat{t}_p + \hat{t}_0$ , where  $\hat{t}_0$  is the estimate for the total investment for subpopulation of enterprises with zero investment in 2009. For  $\hat{t}_0$  we only considered the H-T estimator.  $\hat{t}_p$  is the estimate for subpopulation with nonzero investment and for this term we considered two alternatives – H-T and regression estimators. As an auxiliary variable for the regression estimator we used capital investment for the enterprise in the previous 2009 year. For annual surveys this data is strongly correlated, so we may wait for the essential improvement using the regression estimator. The picture shows the almost ideal linear dependence between 2009 and 2010 data.

Estimates for quarter survey will be presented at the workshop.

## 4 Simulation study

For comparison accuracy of estimators Monte-Carlo method with  $K = 10000$  simulations was used. As

accuracy indicators absolute relative bias  $ARB = \left| \frac{1}{K} \sum_{i=1}^K \hat{y}_d(s_i) - Y_d \right| / Y_d$  and relative root mean squared error

$RRMSE = \sqrt{\frac{1}{K} \sum_{i=1}^K (\hat{y}_d(s_i) - Y_d)^2} / Y_d$  are calculated. Simulation results are presented in Table 6.

Table 6: Comparison of the estimators

Estimator	Accuracy indicator	$\alpha=0.5$	$\alpha=0.6$	$\alpha=0.7$	$\alpha \approx 0.77$
$HT_p + HT_0$	ARB,%	0.09	0.03	0.02	
	RRMSE,%	4.03	2.97	1.87	
$GREG_p + HT_0$					0.007
	ARB,%				0.69
	RRMSE,%	0.01	0.009	0.01	
		0.52	0.53	0.62	

Remark. In last case ( $\alpha \approx 0.77$ ) the bias and the variation are caused only by estimation of enterprises with zero capital investment in 2009.

Estimation by domains is also very important for capital investment survey. Results of studying by activities and regions will be presented at the workshop.

## References

Särndal, C., Swensson, B. & Wretman, J. (2003). *Model assisted survey sampling*. Springer Verlag.