

Republic of Bulgaria Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Bulgaria during calendar years 2008/2009 as part of the fourth round of the Business Environment and Enterprise Performance Survey (BEEPS IV), a joint initiative of the World Bank Group (“WB”) and the European Bank for Reconstruction and Development (“EBRD”). It is an enterprise survey whose objective is to gain an understanding of firms’ perception of the environment in which they operate. The survey was until now administered three times at three years interval. This has added an important element of dynamics in the study of business environment in transition countries.

The 2008 survey was restructured to improve cross-country comparability and to make it compatible with the Enterprise Surveys the Enterprise Analysis Unit of the World Bank has been implementing in the past two years in other regions of the world.

The objective of the survey is to obtain feedback from enterprises in client countries on the state of the private sector as well as to help in building a panel of enterprise data that will make it possible to track changes in the business environment over time, thus allowing, for example, impact assessments of reforms.

Through interviews with firms in the manufacturing and services sectors, the survey will assess the constraints to private sector growth and create statistically significant business environment indicators that are comparable across countries.

The report outlines and describes the sampling design of the data, the data set structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

2. Sampling Structure

2. The sample for the Bulgaria was selected using stratified random sampling, following the methodology explained in the Sampling Manual¹. Stratified random sampling² was preferred over simple random sampling for several reasons³:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following

¹ The complete text can be found at http://www.enterprisesurveys.org/documents/Implementation_note.pdf

² A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., “Elementary Survey Sampling”, Fifth Edition).

³ Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/regions.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and oblast (region). The original sample design with specific information of the industries and regions chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into 23 manufacturing industries, 2 services industries -retail and IT-, and one residual sector as defined in the sampling manual. Each sector had a target of 90 interviews.

5. Size stratification was defined following the standardized definition for the rollout: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees)⁴. For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6. Regional stratification was defined in six regions. These regions are Severozapaden, Severen Tsentralen, Severoiztochen, Yugozapaden, Yuzhen Tsentralen and Yugoiztochen.

3. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments for the selected regions were required. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. For most countries covered in BEEPS IV, two sample frames were used. The first was supplied by the World Bank and consisted of enterprises interviewed in BEEPS 2005. The World Bank required that attempts should be made to re-interview

⁴ The panel firms from BEEPS 2005 with less than 5 employees are included in the 5 to 19 strata.

establishments responding to the BEEPS 2005 survey where they were within the selected geographical regions and met eligibility criteria. That sample is referred to as the Panel. The second sample frame used for the survey in Bulgaria was purchased from the Bulgarian National Statistical Institute [BNSI], which is the governmental statistical institution. The frame contained a full list of establishments with more than five employees in the target sectors of the survey. This was from the latest available version published in 2007, although it related to updates at the end of 2005.

9. The quality of the frame was assessed at the onset of the project. The frame proved to be useful though it showed positive rates of non-eligibility, repetition, non-existent units, etc. These problems are typical of establishment surveys, but given the impact these inaccuracies may have on the results, adjustments were needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of contacts to complete the survey was 34% (249 out of 737 establishments).

Sample Frame Bulgaria

Source: BNSI (Bulgarian National Statistical Institute) 2007

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Severozapaden	5-19	913	887	1,450	3,250
	20-99	481	66	309	856
	100+	147	1	31	179
Severozapaden Total		1,541	954	1,790	4,285
Severen Tsentralen	5-19	1,079	969	1,983	4,031
	20-99	645	109	442	1,196
	100+	202	4	50	256
Severen Tsentralen Total		1,926	1,082	2,475	5,483
Severoiztochen	5-19	858	970	2,532	4,360
	20-99	452	114	694	1,260
	100+	137	8	99	244
Severoiztochen Total		1,447	1,092	3,325	5,864
Yugozapaden	5-19	2,259	2,168	6,538	10,965
	20-99	1,212	341	1,961	3,514
	100+	368	47	354	769
Yugozapaden Total		3,839	2,556	8,853	15,248
Yuzhen Tsentralen	5-19	2,174	1,407	3,402	6,983
	20-99	1,075	150	817	2,042
	100+	279	5	90	374
Yuzhen Tsentralen Total		3,528	1,562	4,309	9,399
Yugoiztochen	5-19	1,073	1,074	2,824	4,971
	20-99	521	127	694	1,342
	100+	153	7	86	246
Yugoiztochen Total		1,747	1,208	3,604	6,559
Grand Total		14,028	8,454	24,356	46,838

Sectors included in the Sample:

Original Sectors	Manufactures: 15 to 37 Services: 52 Residual: 45 to 51, 55 to 64 and 72
Added Sectors	No

4. Data Base Structure:

10. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors (manufacturing, services and IT). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

11. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “*ECA*” indicate questions used in the previous rollout (2005) and, therefore, they may not be found in the implementation of the rollout in other Countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

12. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

13. As noted above, there are 3 levels of stratification: industry, size and region. Different combinations of these variables generate the strata cells for each industry/region/size combination. A distinction should be made between the variable *a4a* and *d1a2* (*industry expressed as ISIC rev. 3.1 code*). The former gives the establishment’s classification into one of the chosen industry-strata, whereas the latter gives the actual establishment’s industry classification (*four digit code*) in the sample frame.

14. All of the following variables contain information from the sampling frame and were defined with the sampling design. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for

researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

-*a2* is the variable describing sampling regions (oblasts)

-*a6a*: coded using the same standard for small, medium, and large establishments as defined above.

-*a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 36), and retail, and IT for services (52, and 72 respectively).

-*id2005*: The variable contains the firm ids of the panel firms

-*id2007*: The variable contains the firm ids of the panel firms interviewed in 2007. (available only in Bulgaria, Albania, and Croatia)

15. The surveys were implemented following a 2 stage procedure. In the first stage a screener questionnaire was applied over the phone to determine eligibility and to make appointments; in the second stage, a face-to-face interview took place with the Manager/Owner/Director of each establishment. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

16. Note that there are additional variables for location (*a3x*), industry (*d1a2*), and size (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advance users are advised to use these variables for analytical purposes.

17. Variable *a3x* indicates the actual location of the establishment. There may be divergences between the location in the sampling frame and the actual location, as establishments may be listed in one place but the actual physical location is in another place.

18. Variable *d1a2* indicates the actual ISIC code of the main output of the establishment as answered by the interviewee. This is probably the most accurate variable to classify establishments by activity.

19. Variables *l1*, *l6* and *l8* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

20. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results.

5. Universe Estimates

21. Universe estimates for the number of establishments in each cell in Bulgaria were produced for the strict, weak and median eligibility definitions. The estimates were the multiple of the relative eligible proportions.

22. Appendix C shows the overall estimates of the numbers of establishments based on the strict, weak and median relative estimates.

6. Weights

23. Since the sampling design was stratified and employed differential sampling individual observations should be properly weighted when making inferences about the population. Under stratified random sampling unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or *pw* in Stata.)⁵

24. Special care was given to the correct computation of the weights. Considering the varying quality of the sample frames, it was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, out of order, no tone in the phone line, answering machine, fax line, wrong address or moved away and could not get the new references) The information required for the adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews. Please, note that panel firms with less than 5 employees were also included in the eligible sample and special coded zero was used in a6a and a6b (sample and screener size) to reflect those cases.

25. For some units it was impossible to determine eligibility because the contact was not successfully completed. Consequently, different assumptions as to their eligibility result in different universe cells' adjustments and in different sampling weights. Three sets of assumptions were considered:

a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *w_strict*.

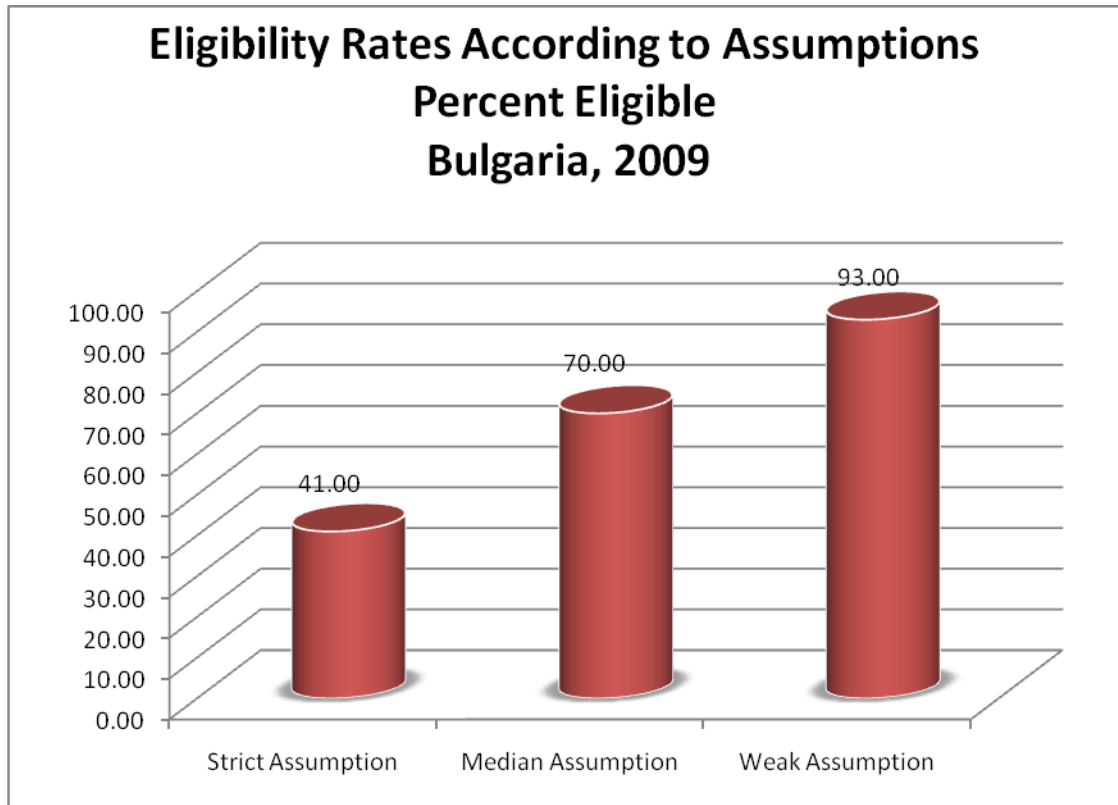
b- Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median*.

c- Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to finalize a contact are assumed eligible. This includes establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was

⁵ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

impossible to find a new address. The resulting weights are included in the variable *w_weak*. Note that under the weak assumption only observed non-eligible units are excluded from universe projections.

The following graph exhibits the different eligibility rates under each set of assumptions.



26. Within each of these assumptions regarding eligibility a pair of weight sets was calculated. The first set of estimates calculated proportions using the raw sample count for each cell. However, the achieved sample numbers in many cells were small. Hence, those eligibility rates, and the adjusted universe cells projections, are subject to relatively large sampling variations. Therefore a second set of more robust estimates (collapsed weights) was also produced. These estimates made use of the multiples of the relative eligibility rates for each industry, size, and region. Those relative rates were based on much larger samples than the individual cells and thus produced values with smaller sampling variations. The data sets include only these robust weights.

Please note that for the purpose of the weights computations all panel firms were considered to be part of the current universe, although technically they are not randomly selected.

7. Appropriate use of the weights

27. As discussed above, under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.

28. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)⁶

29. From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed⁷. If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

8. Non-response

30. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

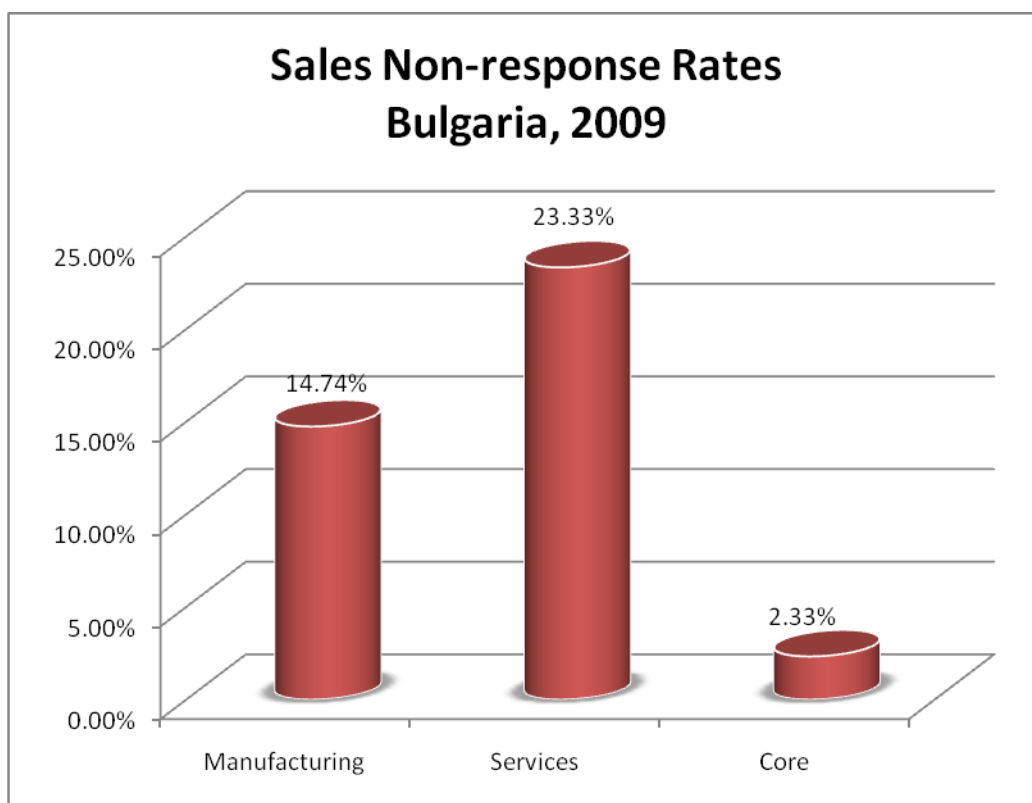
31. Item non-response was addressed by two strategies:

a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as (-8).

b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by type of questionnaire. Please, note that the coding utilized in this dataset does not allow us to differentiated between “Don’t know” and “refuse to answer”, thus the non-response in the table below reflects both categories (DKs and NAs).

⁶ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

⁷ The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.



32. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Up to 4 attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

33. As the following graph shows, the number of contacted establishments per realized interview was 2.48. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units.



34. Details on rejections rates, eligibility rates, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to the Bulgaria. All enterprise surveys suffer from these shortcomings but in very few cases they have been made explicit.

References

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996

Appendix A

Cell Weights – Bulgaria

Strict Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	913	145	98
	20-99	96		196
	100+	41		
Severen Tsentralen	5-19	159	138	389
	20-99	124	8	68
	100+	21		24
Severoiztochen	5-19	402	63	158
	20-99	83	20	34
	100+	13		5
Yugozapaden	5-19	112	59	109
	20-99	34	17	49
	100+	11	2	8
Yuzhen Tsentralen	5-19	106	77	199
	20-99	39	17	34
	100+	85		22
Yogoiztochen	5-19	88	31	370
	20-99		6	143
	100+			

Collapsed Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	161	67	98
	20-99	96		59
	100+	41		
Severen Tsentralen	5-19	186	90	230
	20-99	124	8	68
	100+	21		24
Severoiztochen	5-19	186	90	230
	20-99	83	20	34
	100+	13		5
Yugozapaden	5-19	161	67	109
	20-99	34	17	59
	100+	11	2	8
Yuzhen Tsentralen	5-19	106	77	236
	20-99	39	17	45
	100+	85		22
Yogoiztochen	5-19	88	31	236
	20-99		6	45
	100+			

Cell Weights – Bulgaria

Weak Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	838	131	113
	20-99	109		285
	100+	45		
Severen Tsentralen	5-19	211	182	657
	20-99	207	13	144
	100+	33		49
Severoiztochen	5-19	761	118	381
	20-99	198	48	103
	100+	30		15
Yugozapaden	5-19	276	146	343
	20-99	106	53	193
	100+	32	7	32
Yuzhen Tsentralen	5-19	137	99	328
	20-99	63	27	71
	100+	131		43
Yogoiztochen	5-19	240	84	1,286
	20-99		22	623
	100+			

Collapsed Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	343	143	113
	20-99	109		201
	100+	45		
Severen Tsentralen	5-19	298	143	471
	20-99	207	13	144
	100+	33		49
Severoiztochen	5-19	298	143	471
	20-99	198	48	103
	100+	30		15
Yugozapaden	5-19	343	143	343
	20-99	106	53	201
	100+	32	7	32
Yuzhen Tsentralen	5-19	137	99	489
	20-99	63	27	117
	100+	131		43
Yogoiztochen	5-19	240	84	489
	20-99		22	117
	100+			

Cell Weights – Bulgaria

Median Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	865	121	106
	20-99	105		251
	100+	49		
Severen Tsentralen	5-19	197	153	561
	20-99	182	10	116
	100+	34		47
Severoiztochen	5-19	699	97	319
	20-99	170	37	81
	100+	31		14
Yugozapaden	5-19	203	96	230
	20-99	73	33	121
	100+	27	5	24
Yuzhen Tsentralen	5-19	118	76	257
	20-99	51	20	52
	100+	127		38
Yogoiztochen	5-19	173	54	845
	20-99		13	384
	100+			

Collapsed Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Severozapaden	5-19	261	98	106
	20-99	105		131
	100+	49		
Severen Tsentralen	5-19	276	119	398
	20-99	182	10	116
	100+	34		47
Severoiztochen	5-19	276	119	398
	20-99	170	37	81
	100+	31		14
Yugozapaden	5-19	261	98	230
	20-99	73	33	131
	100+	27	5	24
Yuzhen Tsentralen	5-19	118	76	359
	20-99	51	20	81
	100+	127		38
Yogoiztochen	5-19	173	54	359
	20-99		13	81
	100+			

Appendix B

Status Codes - Total

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	257
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	5
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	4
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	28
Ineligible	5. The establishment has less than 5 permanent full time employees	1
	6. The firm discontinued businesses	16
	7. Not a business: private household	2
	8. Ineligible activity: education, agriculture, finances, governments...	25
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	50
	92. Line out of order	45
	93. No tone	0
	10. Answering machine	1
	11. Fax line - data line	10
	12. Wrong address/ moved away and could not get the new references	66
	13. Refuses to answer the screener	199
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	22
	151. Out of target - outside the covered regions, firm moved abroad	6
	152. Out of target - firm moved abroad	0
	Total	737

Response Outcomes - Total

Complete interviews (<i>Total</i>)	288
Incomplete interviews	1
Eligible in process	2
Refusals	3
Out of target	44
Impossible to contact	172
Ineligible - coop.	6
Refusal to the Screener	199
Total	715

PANEL

Complete interviews (<i>Total</i>)	118
Incomplete interviews	0
Eligible in process	0
Refusals	1
Out of target	20
Impossible to contact	23
Ineligible - coop.	0
Refusal to the Screener	52
Total	214

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	83
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	4
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	4
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	28
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	11
	7. Not a business: private household	1
	8. Ineligible activity: education, agriculture, finances, governments...	8
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	7
	92. Line out of order	5
	93. No tone	0
	10. Answering machine	0
	11. Fax line - data line	1
	12. Wrong address/ moved away and could not get the new references	10
	13. Refuses to answer the screener	52
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	Total	214

FRESH SAMPLE

Complete interviews (<i>Total</i>)	163
Incomplete interviews	1
Eligible in process	2
Refusals	1
Out of target	23
Impossible to contact	149
Ineligible - coop.	6
Refusal to the Screener	139
Total	484

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	166
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	1
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	0
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	1
	6. The firm discontinued businesses	4
	7. Not a business: private household	1
	8. Ineligible activity: education, agriculture, finances, governments...	17
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	43
	92. Line out of order	40
	93. No tone	0
	10. Answering machine	1
	11. Fax line - data line	9
	12. Wrong address/ moved away and could not get the new references	56
	13. Refuses to answer the screener	139
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	22
	151. Out of target - outside the covered regions, firm moved abroad	6
	152. Out of target - firm moved abroad	0
	Total	506

ICS

Complete interviews (<i>Total</i>)	7
Incomplete interviews	0
Eligible in process	0
Refusals	1
Out of target	1
Impossible to contact	0
Ineligible - coop.	0
Refusal to the Screener	8
Total	17

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	8
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	0
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	0
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	1
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	0
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	0
	92. Line out of order	0
	93. No tone	0
	10. Answering machine	0
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	0
	13. Refuses to answer the screener	8
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	Total	17

Appendix C

Eligibility Rules

Status Code	Eligibility Criteria		
	Strict	Weak	Median
1. Eligible establishment (Correct name and address)	1	1	1
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1	1	1
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	1	1	1
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	1	1	1
16. Panel firm - now less than five employees	1	1	1
5. The establishment has less than 5 permanent full time employees	0	0	0
6. The firm discontinued businesses	0	0	0
7. Not a business: Private household	0	0	0
8. Ineligible activity: education, agriculture, finances, governments...	0	0	0
91. No reply (after having called in different days of the week and in different business hours)	0	1	0
92. Line out of order	0	1	0
93. No tone	0	1	0
10. Answering machine	0	1	1
11. Fax line – data line	0	1	1
12. Wrong address/ moved away and could not get the new references	0	1	0
13. Refuses to answer the screener	0	1	1
14. In process (the establishment is being called/ is being contacted – previous to ask the screener)	0	0	0
151. Out of target – outside the covered regions, firm moved abroad	0	0	0
152. Out of target – firm moved abroad	0	0	0

Strict eligibility

= (Sum of the numbers with codes 1,2,3,4,&16) / Total

Weak eligibility

= (Sum of the numbers with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

Median eligibility

= (Sum of the numbers with codes 1,2,3,4,16,10,11, & 13) / Total

Bulgaria Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	21,849	43,359	33,468
Collapsed Cells	21,031	43,392	33,186

Appendix D

Questionnaires:

Problems for the understanding of questions (write question number)	<p>E11 - The respondents stated often that their company has "illegal" competitors, but this, in many cases, actually means that there are "illegal" companies in Bulgaria overall</p> <p>ECAj5a - This question needed clarification each time it was asked. Many respondents stated that they cannot make such a calculation or estimation.</p> <p>K2 - The different types seemed unclear and undistinguishable for some of the respondents</p>
Problems found in the navigability of – questionnaires (for example, skip patterns).	No special problems encountered
Comments on questionnaires length:	Too long - this is actually the biggest problem encountered. In many cases we had to arrange several appointments with the same respondent in order to be able to complete the questionnaire. It was very difficult to keep the respondent's attention till the end of the questionnaire.
Suggestions or other comments on the questionnaire:	<p>We would suggest shorter and fewer questions. Focus on fewer areas with target of 20 minute average duration.</p> <p>The financial part - mainly the N section - produces a lot of refusals. Maybe it could be shorter and asked in different way – ranges as opposed to precise/exact figures.</p> <p>The wording/style of the questionnaires is too heavy. We would suggest some revisions in order to make it easier for managing and following both by respondents and enumerators.</p>

Database

Comments on the data entry program	<p>Data entry program chosen: PERTS</p> <p>Comments: Big difficulties encountered during the data entry process.</p>
Comments on the data cleaning	<p>Concerning the process organized by TNS BBSS - each completed questionnaire was checked on three levels - by the enumerator itself, by the regional supervisor and by head-office team</p> <p>Concerning the data validation checks prepared by TNS Opinion – no special comments, maybe just few of the notifications seemed not applicable for the Bulgarian situation</p>

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	The beginning of year 2007 was very optimistic and enthusiastic in Bulgaria - joining EU was celebrated by almost all and was, actually, one of the only few points that were basis for public consensus in Bulgaria. Somewhere at the middle or even at the end of year 2007, the Bulgarians realized that the general economic and social situation will not change either dramatically or fast. Year 2008 started with more sober evaluations and expectations and continued with increase in pessimistic attitudes, especially after the first news and comments about the global crisis. Thus, at the end of year 2008, at the time of our survey, the most widespread feeling within the Bulgarians, employees and employers, was the uncertainty.
Relevant country events occurred during fieldwork:	Another source of pessimism was series of political scandals, mainly connected with the way of spending the money from EU funds. This is actually a process, started maybe before a year, which is still not completed. In the same time, there is still no effective sentence on any of the bigger scandals.
Other aspects:	Significant part of the EU funding targeted to Bulgaria was stopped after revealed malpractices in Bulgaria. This was one other reason for increased pessimism. Other public "burden" that add to the pessimistic attitudes is the widespread believe that the corruption in the country is strong and presented within all levels of government.

Appendix E

Original Sample Design

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Severozapaden	5-19	3	3	2	8
	20-99	3	2	2	7
	100+	4		1	5
Severozapaden Total		10	5	5	20
Severen Tsentralen	5-19	4	4	3	11
	20-99	5	4	3	12
	100+	5	2	2	9
Severen Tsentralen Total		14	10	8	32
Severoiztochen	5-19	3	4	4	11
	20-99	3	4	4	11
	100+	3	3	4	10
Severoiztochen Total		9	11	12	32
Yugozapaden	5-19	8	9	11	28
	20-99	8	11	12	31
	100+	9	20	15	44
Yugozapaden Total		25	40	38	103
Yuzhen Tsentralen	5-19	8	6	5	19
	20-99	7	5	5	17
	100+	6	2	4	12
Yuzhen Tsentralen Total		21	13	14	48
Yugoiztochen	5-19	4	4	5	13
	20-99	4	4	4	12
	100+	3	3	4	10
Yugoiztochen Total		11	11	13	35
Grand Total		90	90	90	270

Appendix F

Local Agency team involved in the study:

Local Agency	Name: TNS BBSS Country: Bulgaria Membership of international organisation: ESOMAR, Gallup International Association Activities since: 1991
Name of Project Manager	Mrs Marchella Abrasheva - Regional Director and CEO of TNS BBSS
Name and position of other key persons of the project:	Mrs Lilia Stankova - Deputy Regional Director Mr Anton Valkovski - Head of Research Department Ms Martina Poriazova - Fieldwork Supervisor Mrs Eleonora Peeva - Deputy Fieldwork Manager Ms Snezhana Stefanova - IT Specialist
Enumerators involved:	Enumerators / recruiters: 40. The interviewers were in charge of setting the appointments fro the survey.
Other staff involved:	Fieldwork Coordinators: 18 Editing: 2 Data Entry: 1 Data Processing: 3

Sample Frame:

Characteristic of sample frame used:	Three sample frames used: Panel sample frame and ES 2007 sample frame, provided by the World Bank and TNS Opinion. TNS BBSS purchased samples for Blagoevgrad city, Haskovo city and Russe city, regions not included in the previous ES 2007 survey. These contained the list of all the enterprises in these regions and the respective economic sectors, with at least 5 employees. Only the panel included establishments with less than five employees.
Source:	The panel sample was provided by the World Bank. The Bulgarian National Statistical Institute provided the frame for the ES 2007 sample and the three additionally purchased samples as mentioned above.
Year of publication:	2007 for the ES 2007 sample with revisions implemented during the ES survey (2007) and 2008 for the three additionally purchased samples (Blagoevgrad city, Haskovo city and Russe city)
Comments on the quality of sample frame:	Overall good quality with comparatively higher amount of issues (outdated or insufficient information) within the Panel sample
Year and organism who conducted the last economic census	Still no such census is conducted in Bulgaria.
Other sources for companies statistics	Bulgarian National Statistical Institute

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: No specific issues noticed On regions: Higher level of refusals in South-Western region.
Comments on the response rate:	Very good response rate, especially taking into account the target respondent - top level manager or owner - and the interview length.
Comments on the sample design:	No special comments

Fieldwork:

Date of Fieldwork	September - December 2008
Country	Bulgaria
Interview number	Manufactures: 95 Services: 150 Core: 43
Problems found during fieldwork:	Difficulties in reaching top level managers within the bigger companies. There was an increased refusal rate, especially in comparison with several years ago. Nevertheless, the refusal rate enumerated is still very good for such type of survey.
Other observations:	No

Appendix H.**Survey Universe, Sample Population and Sampling Frames**

The following provides description of the general methodology used in BEEPS 2009.

The survey universe was defined as commercial, service or industrial business establishments with at least five full-time employees. Government departments including military, police, education, health and similar activities were excluded, as were those in primary industries including agriculture, mining, etc.

There are no up to date and reliable statistics relating to this universe in the countries being

surveyed in BEEPS IV. Consequently the universe size and characteristics have to be directly

estimated from the survey results themselves. This requirement increases the emphasis that has to be placed on the quality of the sample frame, because the validity of the results is predominantly a function of coverage and age of the sampling frame.

The criteria used to evaluate the available sampling frame in descending priority were those of:

- Coverage

- ✓ Up to datedness
- ✓ Availability of detailed stratification variables
- ✓ Location identifiers- address, phone number, email
- ✓ Electronic format availability
- ✓ Contact name(s)

The sample frames used for the surveys must consist of the lists of enterprises in each country that most optimally meet these requirements. The final selection was made by the TNS in collaboration with the World Bank and EBRD. For most countries covered in BEEPS IV two sample frames were used. The first frame was often an official frame of establishments supplied by the national statistical office of the country. The Enterprise Survey conducted for the World Bank in Albania in 2007/8 showed that a suitable frame did not exist for the country. Instead, the design returned to first principles, using a blocks enumeration methodology.