

Poland Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Poland 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 Poland 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 180 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 Central, Southern, Eastern, North-Western, South-Western, and Northern.

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 frame for Poland was the database of *Polskie Firmy*. The second frame, supplied by the World Bank/EBRD, consisted of enterprises interviewed in BEEPS 2005. The implementing contractor was asked to make their best efforts 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.

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 16% (559 out of 3,523 establishments).

Sample Frame Poland (Fresh)

Source: Polskie Firmy database

Region	Employees	Sector			
		Manufacturing	52	Residual	Grand Total
Central	5-19	2,657	3,762	6,016	12,435
	20-99	1,345	318	1,936	3,599
	100+	814	65	714	1,593
Central Total		4,816	4,145	8,666	17,627
Southern	5-19	2,102	4,258	6,442	12,802
	20-99	1,217	228	1,826	3,271
	100+	766	45	522	1,333
Southern Total		4,085	4,531	8,790	17,406
Eastern	5-19	1,298	2,725	4,464	8,487
	20-99	948	288	1,194	2,430
	100+	670	33	380	1,083
Eastern Total		2,916	3,046	6,038	12,000
North-western	5-19	1,568	2,504	4,280	8,352
	20-99	1,196	228	1,401	2,825
	100+	824	34	416	1,274
North-western Total		3,588	2,766	6,097	12,451
South-western	5-19	768	1,471	2,406	4,645
	20-99	606	115	782	1,503
	100+	464	25	214	703
South-western Total		1,838	1,611	3,402	6,851
Northern	5-19	1,992	3,043	5,538	10,573
	20-99	1,338	235	1,584	3,157
	100+	767	29	386	1,182
Northern Total		4,097	3,307	7,508	14,912
Grand Total		21,340	19,406	40,501	81,247

Sample Frame Poland (Panel)

Source: BEEPS 2005

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Central	< 5	20	9	12	41
	5-19	25	4	12	41
	20-99	20	1	6	27
	100+	1	5	6	12
Central Total		66	19	36	121
Southern	< 5	17	5	5	27
	5-19	16	2	3	21
	20-99	12		5	17
	100+	9	1	4	14
Southern Total		54	8	17	79
Eastern	< 5	1	4	5	10
	5-19	1	1	1	3
	20-99	2		3	5
	100+	3	1	3	7
Eastern Total		7	6	12	25
North-western	< 5	20	4	9	33
	5-19	14	2	6	22
	20-99	12		4	16
	100+	4		2	6
North-western Total		50	6	21	77
South-western	< 5	3	2		5
	5-19	3			3
	20-99	3		2	5
	100+				
South-western Total		9	2	2	13
Northern	< 5		4	3	7
	5-19	1	1	2	4
	20-99	1		4	5
	100+	1		1	2
Northern Total		3	5	10	18
Grand Total		189	46	98	333

Sectors included in the Sample:

Original Sectors	Manufactures: 15-37 Services: 52 Residual: 45, 50, 51, 55, 60-64, 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 3 establishment identifiers, *idstd*, and *id*. The first is a global unique identifier. The second one 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 interviewed in 2005

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

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 characteristics 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 were produced for each of 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:

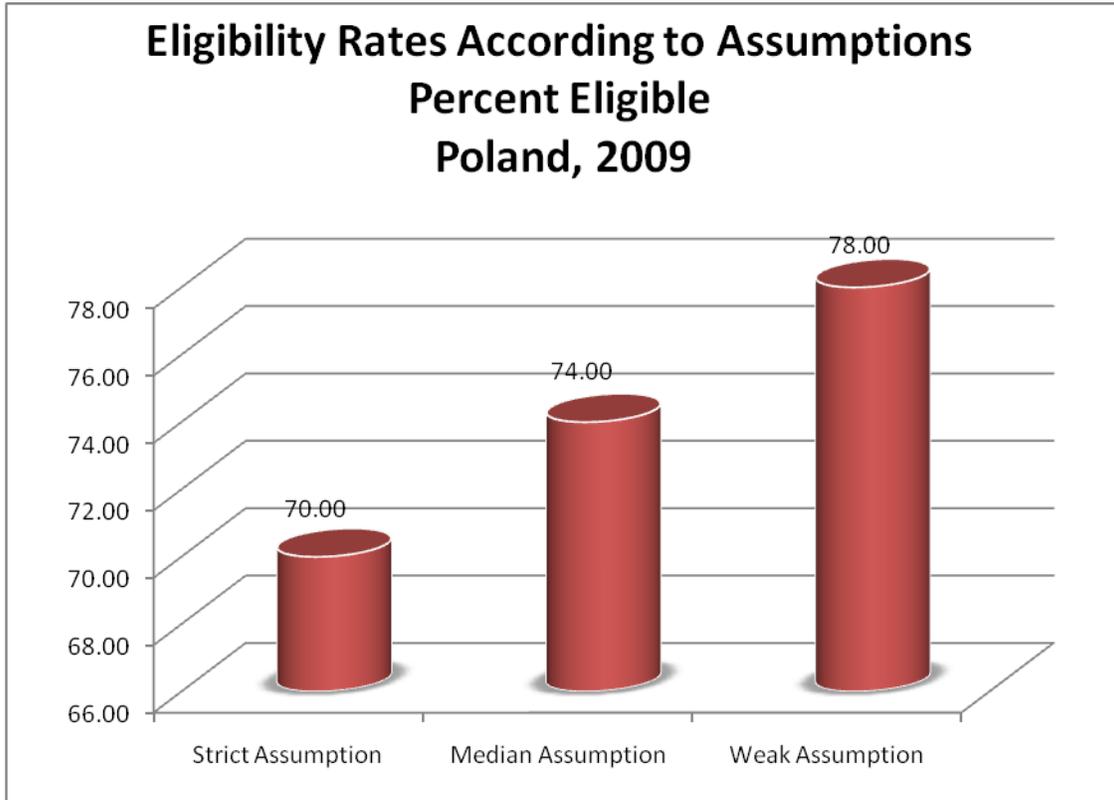
a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *w_strict*.

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 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.

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

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.

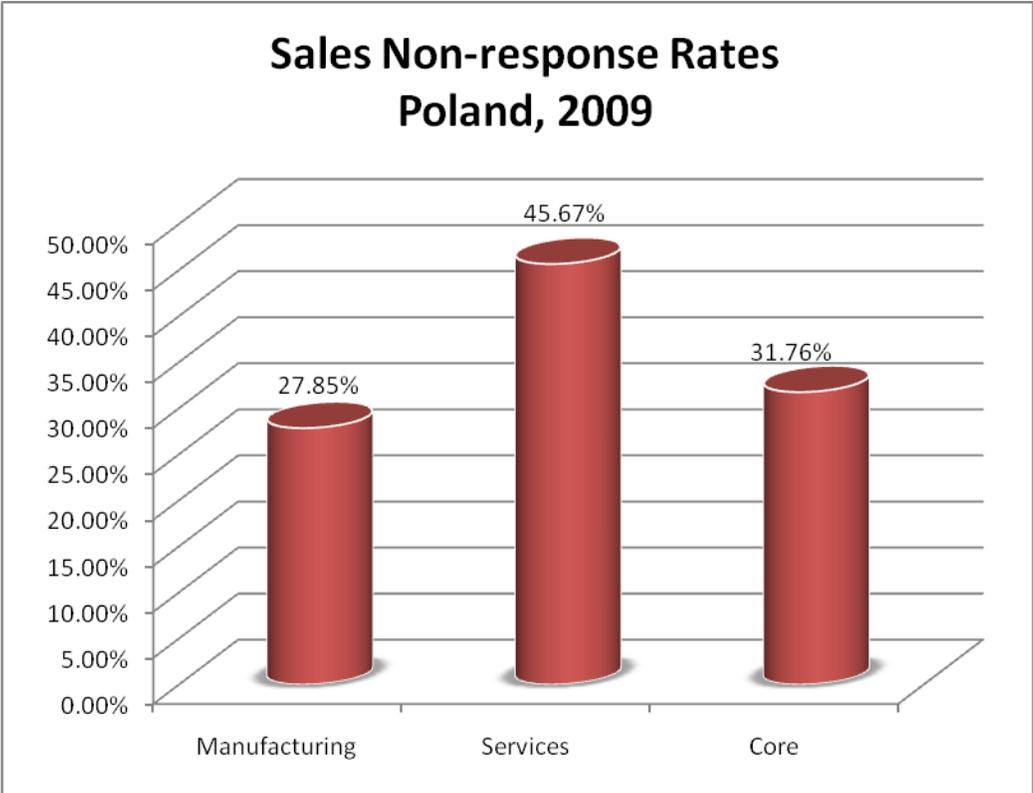
31. Item non-response was addressed by two strategies:

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).

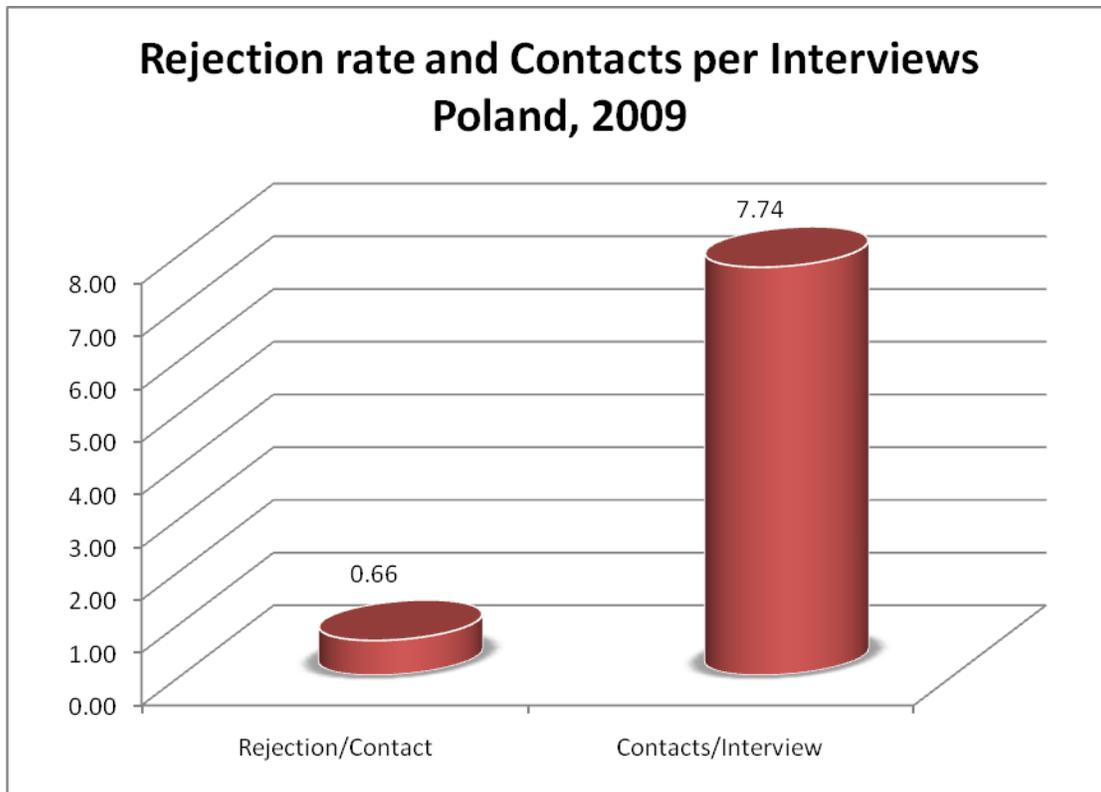
⁶ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

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 7.74. 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 Poland. 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 – Poland (Strict)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	150	216	391
	20-99	61	18	144
	100+	108	11	46
Southern	5-19	100	275	520
	20-99	108	17	129
	100+	105	6	35
Eastern	5-19	96	188	271
	20-99	92	23	90
	100+	67	9	23
North-western	5-19	85	259	264
	20-99	113	25	254
	100+	73	10	56
South-western	5-19	119	386	988
	20-99	82	32	225
	100+	97	11	64
Northern	5-19	211	437	712
	20-99	119	43	166
	100+	117	9	94

Collapsed Cells

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	150	216	417
	20-99	61	18	144
	100+	108	11	46
Southern	5-19	100	275	417
	20-99	108	17	129
	100+	105	6	35
Eastern	5-19	96	188	417
	20-99	92	23	90
	100+	67	9	23
North-western	5-19	85	259	417
	20-99	113	25	254
	100+	73	10	56
South-western	5-19	119	386	417
	20-99	82	32	225
	100+	97	11	64
Northern	5-19	211	437	417
	20-99	119	43	166
	100+	117	9	94

Cell Weights – Poland (Weak)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	192	263	470
	20-99	72	20	161
	100+	126	12	51
Southern	5-19	127	333	624
	20-99	128	19	144
	100+	122	7	38
Eastern	5-19	117	216	309
	20-99	104	25	95
	100+	74	9	24
North-western	5-19	101	293	295
	20-99	125	26	264
	100+	79	10	57
South-western	5-19	143	443	1,120
	20-99	92	34	237
	100+	107	11	66
Northern	5-19	241	477	767
	20-99	126	43	166
	100+	122	9	92

Collapsed Cells

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	192	263	478
	20-99	72	20	161
	100+	126	12	51
Southern	5-19	127	333	478
	20-99	128	19	144
	100+	122	7	38
Eastern	5-19	117	216	478
	20-99	104	25	95
	100+	74	9	24
North-western	5-19	101	293	478
	20-99	125	26	264
	100+	79	10	57
South-western	5-19	143	443	478
	20-99	92	34	237
	100+	107	11	66
Northern	5-19	241	477	478
	20-99	126	43	166
	100+	122	9	92

Cell Weights – Poland (Median)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	172	232	425
	20-99	67	18	150
	100+	117	11	48
Southern	5-19	116	299	573
	20-99	120	18	136
	100+	116	6	37
Eastern	5-19	108	197	289
	20-99	100	23	92
	100+	72	9	23
North-western	5-19	96	275	284
	20-99	123	25	262
	100+	78	10	57
South-western	5-19	135	412	1,067
	20-99	90	32	233
	100+	105	11	65
Northern	5-19	232	450	742
	20-99	126	42	166
	100+	122	9	93

Collapsed Cells

Region	Employees	Sector		
		Manufacturing	52	Residual
Central	5-19	172	232	448
	20-99	67	18	150
	100+	117	11	48
Southern	5-19	116	299	448
	20-99	120	18	136
	100+	116	6	37
Eastern	5-19	108	197	448
	20-99	100	23	92
	100+	72	9	23
North-western	5-19	96	275	448
	20-99	123	25	262
	100+	78	10	57
South-western	5-19	135	412	448
	20-99	90	32	233
	100+	105	11	65
Northern	5-19	232	450	448
	20-99	126	42	166
	100+	122	9	93

Appendix B
Eligibility Status – Summary Fieldwork Reports

TOTAL

Status Codes

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	2,822
	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	52
	7. Not a business: private household	11
	8. Ineligible activity: education, agriculture, finances, governments...	172
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	72
	92. Line out of order	0
	93. No tone	0
	10. Answering machine	3
	11. Fax line - data line	11
	12. Wrong address/ moved away and could not get the new references	86
	13. Refuses to answer the screener	142
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	504
	151. Out of target - outside the covered regions, firm moved abroad	152
	152. Out of target - firm moved abroad	0
Total	4,027	

Response Outcomes

Complete interviews (<i>Total</i>)	455
Incomplete interviews	1
Eligible in process	548
Refusals	1,818
Out of target	235
Impossible to contact	172
Ineligible - coop.	152
Refusal to the Screener	142
Total	3,523

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 (<i>the establishment is being called/ is being contacted – previous to ask the screener</i>)	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

Poland Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	66,539	76,014	71,342
Collapsed Cells	66,632	76,081	71,432

Appendix D

Questionnaires:

Problems for the understanding of questions (write question number)	<p>In a4b respondents had lot of problems to correctly identify company sector;</p> <ul style="list-style-type: none"> • In d1a2 respondents had problems with ISIC code. In some cases their activities involves both production of certain goods and selling them. List was not complete enough for needs of some and codes given from outside the list; • In d1a1x it was difficult to indicate main product especially when company produces several products; • In d2 it was not automatically clear for some respondents whether it is net value or gross value figure that was required; • Questions b4 and ECAb7a are sensitive and respondents did not understand their purpose; • b5, b6 – some comprehension problems occurred with differences between starting operations and being registered; • The scale for the obstacles questions was difficult to use for respondents; • f1 - question is not clear. Respondents did not immediately understand the concept of capacity utilization; • In l10 it is difficult to understand the notion of “formal training”; • n2i respondents sometimes mixed this up with the value requested in d2.
Problems found in the navigability of – questionnaires (for example, skip patterns).	In Poland CAPI was used so no such problems occurred.
Comments on questionnaires length:	According to many respondents (and interviewers) the questionnaire was too long. The average length of the interview was about 57 minutes.
Suggestions or other comments on the questionnaire:	<ul style="list-style-type: none"> • There were problems with differences between licences, permits and certificates; • Respondents felt that the questionnaire contains too many questions about facts and figures and not enough about their situation and barriers.

Database

Comments on the data entry program	<p>Data entry program chosen: In Poland CAPI was used so no data entry was needed.</p> <p>SPSS 14.0 PL for Windows was used for cleaning the data.</p>
Comments on the data cleaning	<p>As in Poland CAPI was used there were no problems with skip patterns.</p> <p>Some problems occurred in financial questions with number “0”.</p>

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	During the fieldwork the economic crisis started to be an issue in Poland, but it seems not to have had an influence on the fieldwork.
Relevant country events occurred during fieldwork:	-
Other aspects:	-

Appendix E
Original Sample Design

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Central	5-19	15	13	12	40
	20-99	12	13	13	38
	100+	11	17	16	44
Central Total		38	43	41	122
Southern	5-19	12	14	13	39
	20-99	11	10	13	34
	100+	11	12	12	35
Southern Total		34	36	38	108
Eastern	5-19	8	9	9	26
	20-99	9	12	8	29
	100+	9	9	9	27
Eastern Total		26	30	26	82
North-western	5-19	9	9	9	27
	20-99	11	10	10	31
	100+	12	9	9	30
North-western Total		32	28	28	88
South-western	5-19	4	5	5	14
	20-99	5	5	5	15
	100+	6	6	5	17
South-western Total		15	16	15	46
Northern	5-19	12	10	12	34
	20-99	12	10	11	33
	100+	11	7	9	27
Northern Total		35	27	32	94
Grand Total		180	180	180	540

Appendix F.

Local Agency team involved in the study:

Local Agency	Name: TNS OBOP Country: Poland Membership of international organisation: ESOMAR Activities since: 1958
Name of Project Manager	Agata Zadrożna
Name and position of other key persons of the project:	Ewelina Gałka – Fieldwork Manager
Enumerators involved:	Enumerators: 103 Recruiters: 107 In Poland enumerators worked as recruiters because in the second part of the fieldwork we changed the recruitment technique from telephone to face-to-face. The number of telephone recruiters was 3.
Other staff involved:	Fieldwork Coordinators: 21 Editing: 1 Data Entry: - Data Processing: 1

Sample Frame:

Characteristic of sample frame used:	Data base of Polskie Firmy contain the most active companies and business addresses in Poland, including commercial (manufacturing, trade, services) and non-commercial (administration, education, societies) activities. It is created by a private company from various sources and is updated regularly.
Source:	Polskie Firmy - Warszawa
Year of publication:	2008
Comments on the quality of sample frame:	Quality of the sample frame is relatively good (worse for small companies). However, in the sample there were 52 inactive firms, 11 households and 86 firms which address cannot be found.
Year and organism who conducted the last economic census	There were no economic censuses in Poland.
Other sources for companies statistics	Central Statistical Office keeps National Official Business Register http://www.stat.gov.pl/bip/regon_ENG_HTML.htm .

Sample:

Comments/ problems on sectors and regions selected in the sample:	<p>On sectors:</p> <ul style="list-style-type: none">• Many companies are officially production but in fact they are services;• State-owned companies and co-operatives were out of target but were not excluded from the sample. During the fieldwork 172 cases of out of target companies were encountered (the most often in Northern and North-western region). <p>On regions:</p> <ul style="list-style-type: none">• We had more problems in big cities like Warsaw (Central region), Poznań (Northern-western region), Cracow (Southern region or Wrocław (South-western region) – appointments were rescheduled most often there.• It was much easier to conduct interviews in smaller towns; in smaller towns and in smaller companies people have more time.
Comments on the response rate:	<ul style="list-style-type: none">• Lots of refusals (2076 out of 454 successful interviews);• Lack of time was the most common reason for refusal. Appointments are re-scheduled many times. It sometimes happened that an appointment was set but when the interviewer came to the firm there was no contact with respondent (respondents even refused to answer phone calls).
Comments on the sample design:	<ul style="list-style-type: none">• In the first phase of the survey (when respondents were recruited by phone) sample design with three preferences was the reason for a very slow progress. It takes lot of time to get in touch with potential respondents with preference one and during that period preferences two and three cannot be contacted.• Placing emphasis on response rate (generally low among companies in Poland) has proven to be ineffective in terms of time.

Fieldwork:

Date of Fieldwork	25.08.2008 – 18.03.2009
Country	Poland
Interview number	Manufactures: 166 Services: 126 Core: 163
Problems found during fieldwork:	<ul style="list-style-type: none"> • In December and January it was very difficult to conduct any interviews (it is year end and start and companies are busy with financial issues). Respondents refused to accept an appointment during this period; • Respondents do not believe in confidentiality and even if they do, they are still afraid of saying too much. If an establishment was part of a bigger firm the interviewer was sent to the central location; • In bigger companies interviewers had problems reaching a potential respondent as they were not let into the buildings. Everything had to be settled through the reception desk. It often happened that it is the receptionist/secretary who refuses to let the interviewer in (calling by phone is not very helpful here as receptionists do not want to put the recruiter through to top manager); • In general the target group for the BEEPS survey was very difficult to reach; • In big companies several respondents had to be involved in the interview. One person cannot answer all questions. It caused problems in cases where the respondent was not willing to consult with other colleagues.
Other observations:	<ul style="list-style-type: none"> • Respondents wanted to know the questions before the interview. Because of the methodology we were not able to tell them before the interview what questions were in the questionnaire. This made getting consent for the interview difficult; • Many respondents proposed to fill in the questionnaire themselves if it could have been sent by e-mail. It is much easier to interview respondents in firms by phone and it is really difficult to get consent for a face-to-face interview.

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.