

Slovak Republic Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Slovak Republic 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 Slovak Republic 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 sectors: financial intermediation (group J), real estate and renting activities (group K,

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

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 designs with specific information of the industries and regions chosen are included in the attached Excel file (Sampling Report.xls.)

4. Industry stratification was designed in the way that follows: the universe was stratified into manufacturing industries, services industries, and one residual (core) sector as defined in the sampling manual. Each industry had a target of 90 interviews. For the manufacturing industries sample sizes were inflated by about 10% to account for potential non-response cases when requesting sensitive financial data and also because of likely attrition in future surveys that would affect the construction of a panel. For the other industries (Residuals) sample sizes were inflated by about 4% to account for under sampling in firms in service industries.

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 4 regions. These regions are Bratislava, Západné Slovensko, Stredné Slovensko, and Východné Slovensko.

7. The Slovak Republic sample contains panel data. The wave 1 panel “Investment Climate Private Enterprise Survey implemented in Slovak Republic” consisted of 220 establishments interviewed in 2005. A total of 33 establishments have been re-interviewed in the 2008 Business Environment and Enterprise Performance Survey.

3. Sampling implementation

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

8. 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.)

9. The source of the sample frame was Statistical Office of the Slovak Republic - 2007- Organization database.

10. 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 33.5% (344 out of 1027 establishments).

Sample Frame Slovak Republic

Source: Statistical Office of the Slovak Republic - 2007- Organization database.

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Bratislava	5-19	4,333	5,393	16,074	25,800
	20-99	1,005	885	3,534	5,424
	100+	430	173	1,022	1,625
Bratislava Total		5,768	6,451	20,630	32,849
Západné Slovensko	5-19	10,887	11,969	30,340	53,196
	20-99	3,493	2,116	7,495	13,104
	100+	1,389	520	2,202	4,111
Západné Slovensko Total		15,769	14,605	40,037	70,411
Stredné Slovensko	5-19	6,155	6,678	17,819	30,652
	20-99	1,854	1,126	4,247	7,227
	100+	639	226	1,065	1,930
Stredné Slovensko Total		8,648	8,030	23,131	39,809
Východné Slovensko	5-19	11,744	10,743	27,605	50,092
	20-99	2,965	1,851	6,487	11,303
	100+	1,222	494	1,925	3,641
Východné Slovensko Total		15,931	13,088	36,017	65,036
Grand Total		46,116	42,174	119,815	208,105

Sectors included in the Sample:

Original Sectors	Manufactures: 15,17,18,19,20,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36 Services: 52 Residual: 45,50,51,55,60,61,63,72
Added (top up) Sectors	<i>None</i>

4. Data Base Structure:

11. 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*.

12. 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.

13. There are two 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.

14. As noted above, there are three 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 in the sample frame.

15. 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). All establishments within the residual stratum were coded with *a4a=2*.

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

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

5. Universe Estimates

21. Universe estimates for the number of establishments in each cell in Slovak Republic 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 *pa* 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*.

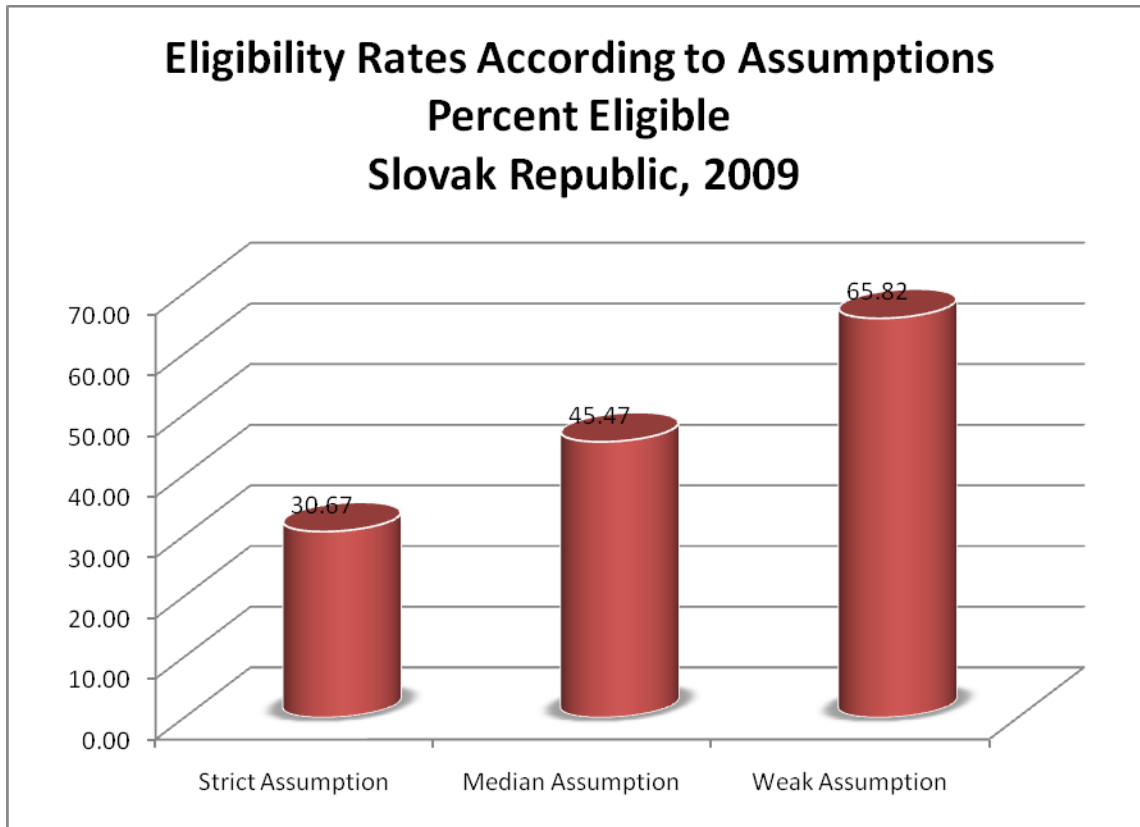
b- Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median*.

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.

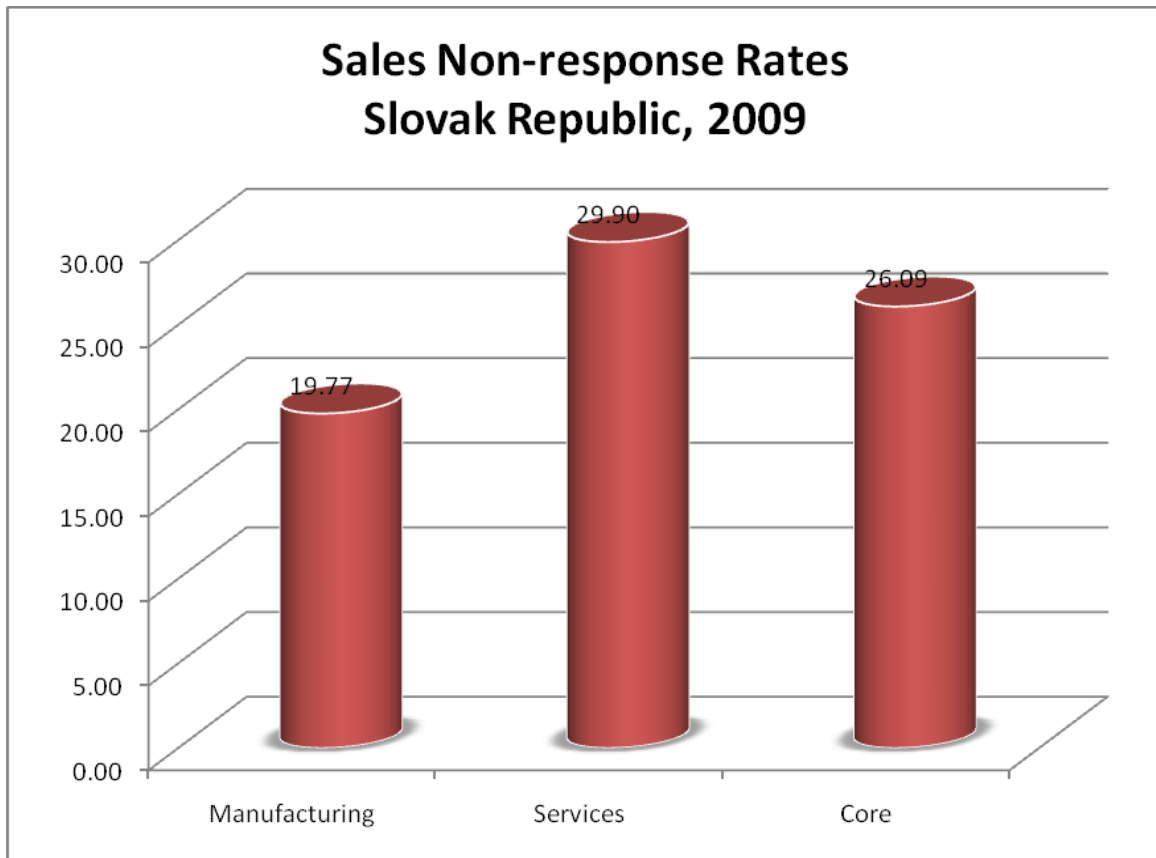
31. Item non-response was addressed by two strategies:

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).

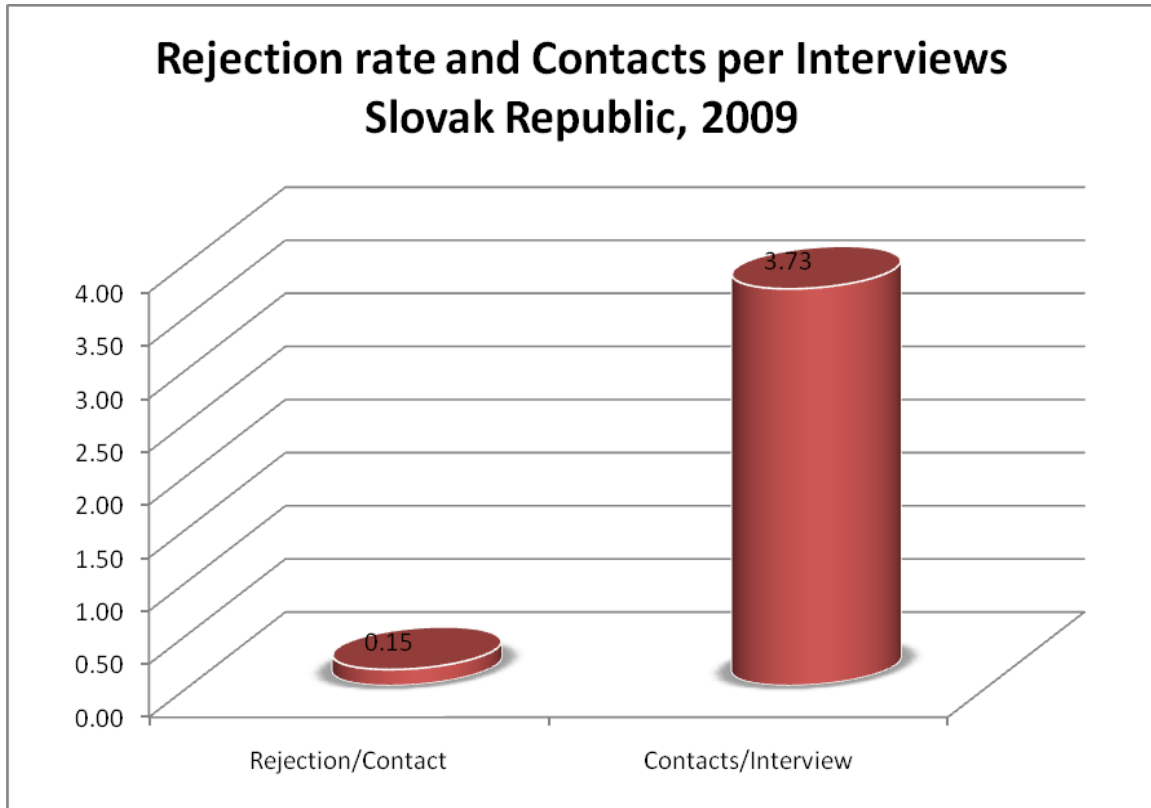
⁶ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

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 3.73. 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. The relatively low ratio of contacted establishments per realized interview (3.73) suggests that the main source of error in estimates in the Slovak Republic may be selection bias and not frame inaccuracy.



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 Slovak Republic. 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

Slovak Republic Strict

Region	Employees	Sector		
		Manufacturing	52	Residual
Bratislava	5-19	315	416	779
	20-99	160	119	192
	100+	69	23	56
Západné Slovensko	5-19	383	626	995
	20-99	161	89	772
	100+	113	29	114
Stredné Slovensko	5-19	313	863	542
	20-99	82	42	225
	100+	37	19	57
Východné Slovensko	5-19	383	327	1,130
	20-99	127	105	406
	100+	72	22	52

Slovak Republic Weak

Region	Employees	Sector		
		Manufacturing	52	Residual
Bratislava	5-19	751	1,055	2,141
	20-99	311	247	432
	100+	130	47	122
Západné Slovensko	5-19	690	1,199	2,069
	20-99	237	140	1,314
	100+	161	44	189
Stredné Slovensko	5-19	911	2,680	1,826
	20-99	196	108	621
	100+	85	48	152
Východné Slovensko	5-19	930	845	3,168
	20-99	251	223	930
	100+	139	45	116

Slovak Republic Median

Region	Employees	Sector		
		Manufacturing	52	Residual
Bratislava	5-19	420	560	1,080
	20-99	196	148	246
	100+	84	29	72
Západné Slovensko	5-19	545	899	1,475
	20-99	212	118	1,059
	100+	148	38	156
Stredné Slovensko	5-19	605	1,689	1,094
	20-99	147	77	421
	100+	66	35	106
Východné Slovensko	5-19	576	497	1,772
	20-99	176	148	588
	100+	100	31	75

Appendix B Eligibility Status – Summary Fieldwork Reports
TOTAL

Complete interviews

Status Codes

ELIGIBLES		
Eligible	1. Eligible establishment <i>(Correct name and address)</i>	294
	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>	1
	4. Eligible establishment <i>(Wrong address - the firm/establishment has changed address and the address could be found)</i>	19
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	18
	6. The firm discontinued businesses	41
	7. Not a business: private household	15
	8. Ineligible activity: education, agriculture, finances, governments...	39
Unobtainable	91. No reply <i>(after having called in different days of the week and in different business hours)</i>	85
	92. Line out of order	31
	93. No tone	11
	10. Answering machine	11
	11. Fax line - data line	3
	12. Wrong address/ moved away and could not get the new references	82
	13. Refuses to answer the screener	138
	14. In process <i>(the establishment is being called/ is being contacted - previous to ask the screener)</i>	230
	151. Out of target - outside the covered regions, firm moved abroad	6
	152. Out of target - firm moved abroad	2
	Total	1,027

Response Outcomes

Complete interviews <i>(Total)</i>	275
Incomplete interviews	5
Eligible in process	18
Refusals	17
Out of target	113
Impossible to contact	223
Ineligible - coop.	8
Refusal to the Screener	138
Total	797

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
5. The establishment has less than 5 employees	0	0	0
6. The firm discontinued businesses/ unattainable	0	0	0
7. Not a business: Private	0	0	0
8. Not a business: Education or Government	0	0	0
9. No reply (after having called in different days of the week and in different business hours) out of order, no tone	0	0	1
10. Answering machine	0	1	1
11. Fax line	0	1	1
12. Wrong address/ moved away and could not get the new references	0	0	1
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
15. Out of target - cooperative, outside the covered regions	0	0	0

Slovak Republic Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	75,236	173,809	113,749
Collapsed Cells			

Appendix D

Questionnaires:

Problems for the understanding of questions (write question number)	None
Problems found in the Navigability of – questionnaires (for example, skip patterns).	No problems here
Comments on questionnaires length:	Length was adequate
Suggestions or other Comments on the questionnaire:	None

Database

Comments on the data entry program	Data entry program chosen: CONFIRMIT Comments: Good feedback generally from team working with CONFIRMIT
Comments on the data cleaning	None

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	None
Relevant country events occurred during fieldwork:	None
Other aspects:	None

Appendix E Original Sample Design

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Bratislava	5-19	4	5	5	14
	20-99	3	4	5	12
	100+	4	4	5	13
Bratislava Total		11	13	15	39
Západné Slovensko	5-19	10	10	10	30
	20-99	11	11	10	32
	100+	11	11	11	33
Západné Slovensko Total		32	32	31	95
Stredné Slovensko	5-19	5	6	6	17
	20-99	6	6	6	18
	100+	5	5	5	15
Stredné Slovensko Total		16	17	17	50
Východné Slovensko	5-19	11	9	9	29
	20-99	10	9	9	28
	100+	10	10	9	29
Východné Slovensko Total		31	28	27	86
Grand Total		90	90	90	270

Appendix F

Local Agency team involved in the study:

Local Agency	Name: ACRC s.r.o. Country: Slovak Republic Membership of international organisation: none Activities since: various
Enumerators involved:	Recruiters: 2 Enumerators: 12 (there were no enumerators doing recruitment)
Other staff involved:	Fieldwork Coordinators: 3 Editing: 2 Data Entry: 5 Data Processing: 2

Sample Frame:

Characteristic of sample frame used:	All requirements were met
Source:	Statistical Office of the Slovak Republic - Organization database (National Register of Businesses)
Year of publication:	2007
Comments on the quality of sample frame:	None
Year and organization who conducted the last economic census	Slovak Statistical Institute

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: None. On regions: None.
Comments on the response rate:	None
Comments on the sample design:	None
Other comments:	None.

Fieldwork:

Date of Fieldwork	3.9.2008 – 10.3.2009
Country	Slovak Republic
Interview number	Manufactures:86 Services: 97 Core: 92
Problems found during fieldwork:	None.
Other observations:	None