Czech Republic Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Czech 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 Czech Republic was selected using stratified random sampling, following the methodology explained in the Sampling Manual[1]. Stratified random sampling[2] was preferred over simple random sampling for several reasons[3]:

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 eight regions. These regions are Praha, Stredni Cechy, Jihozapad, Severozapad, Severovychod, Jihovychod, Stredni Morava, and Moravskoslezsko.

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 frame for the Czech Republic was an official database known as Albertina data [Creditinfo Czech Republic], which is obtained from the complete Business Register [RES] of the Czech Statistical Office. An extract from that frame was sent to the TNS statistical team in London to select the establishments for interview.

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 28% (572 out of 2041 establishments).

Sample Frame Czech Republic – First Frame (Panel)

Source: BEEPS 2005

Region	Employees	Sector			
		Manufacturing	52	Residual	Grand Total
Praha	<5		4	2	6
	5-19	3	1	12	16
	20-99	3	1	5	9
	100+	2		4	6
Praha Total		8	6	23	37
Stredni Cechy	<5		1	1	2
	5-19	1			1
	20-99				
	100+	1	2		3
Stredni Cechy Total		2	3	1	6
Jihozapad	<5		1	2	3
	5-19		1		1
	20-99				
	100+				
Jihozapad Total			2	2	4
Severozapad	<5		6		6
	5-19		5	2	7
	20-99	2			2
	100+	3	1		4
Severozapad Total		5	12	2	19
Severovychod	<5			1	1
	5-19				
	20-99	1	1		2
	100+	1			1
Severovychod Total		2	1	1	4
Jihovychod	<5		1	3	4
	5-19	1	1	1	3
	20-99	3		1	4
	100+			1	1
Jihovychod Total		4	2	6	12
Stredni Morava	<5	2	1	2	5
	5-19	1			1
	20-99	1			1
	100+	3	1		4
Stredni Morava Total		7	2	2	11
Moravskoslezsko	<5	3	4	2	9
	5-19	3	1	4	8
	20-99	3	1		4
	100+	2		5	7
Moravskoslezsko Total		11	6	11	28
Grand Total		39	34	48	121

Sample Frame Czech Republic – Second Frame (Fresh)

Source: Albertina data (Creditinfo Czech Republic) 2007

Region	Employees	Sector			
		Manufacturing	52	Residual	Grand Total
Praha	5-19	1,011	1,070	5,195	7,276
	20-99	643	283	1,187	2,113
	100+	210	83	323	616
Praha Total		1,864	1,436	6,705	10,005
Stredni Cechy	5-19	941	661	2,441	4,043
	20-99	655	109	738	1,502
	100+	243	27	143	413
Stredni Cechy Total		1,839	797	3,322	5,958
Jihozapad	5-19	1,082	701	2,213	3,996
	20-99	841	99	742	1,682
	100+	329	18	97	444
Jihozapad Total		2,252	818	3,052	6,122
Severozapad	5-19	778	627	2,023	3,428
	20-99	593	111	418	1,122
	100+	260	16	70	346
Severozapad Total		1,631	754	2,511	4,896
Severovychod	5-19	1,493	829	3,001	5,323
	20-99	1,072	115	913	2,100
	100+	424	19	107	550
Severovychod Total		2,989	963	4,021	7,973
Jihovychod	5-19	1,675	1,119	3,471	6,265
	20-99	1,140	204	1,134	2,478
	100+	427	33	152	612
Jihovychod Total		3,242	1,356	4,757	9,355
Stredni Morava	5-19	1,279	835	2,326	4,440
	20-99	908	112	757	1,777
	100+	347	17	90	454
Stredni Morava Total		2,534	964	3,173	6,671
Moravskoslezsko	5-19	885	895	2,037	3,817
	20-99	603	133	689	1,425
	100+	210	17	89	316
Moravskoslezsko Total		1,698	1,045	2,815	5,558
Grand Total		18,049	8,133	30,356	56,538

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 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 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:

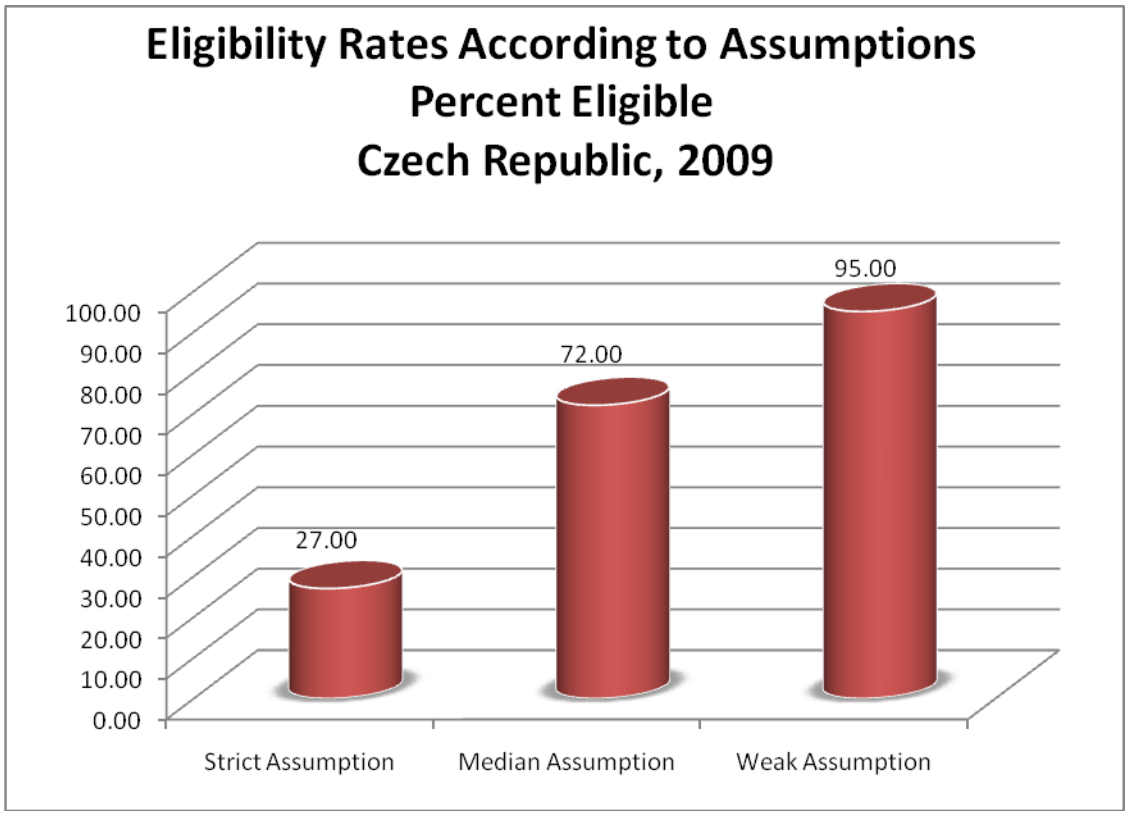
a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *w_strict*.

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.

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

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



26. TNS discussed 'cell collapse' with the World Bank/EBRD and agreed that some cells might reasonably be combined. In those cases revised weights were calculated and also added to the datasets. Hence, analyses are possible using either the original weights or the revised values after cell collapse. Inspection of the weights for the Czech Republic showed a wide range of weights but that little would be gained from any cell collapse. Hence that was not used for the country.

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 observed[7]. 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.

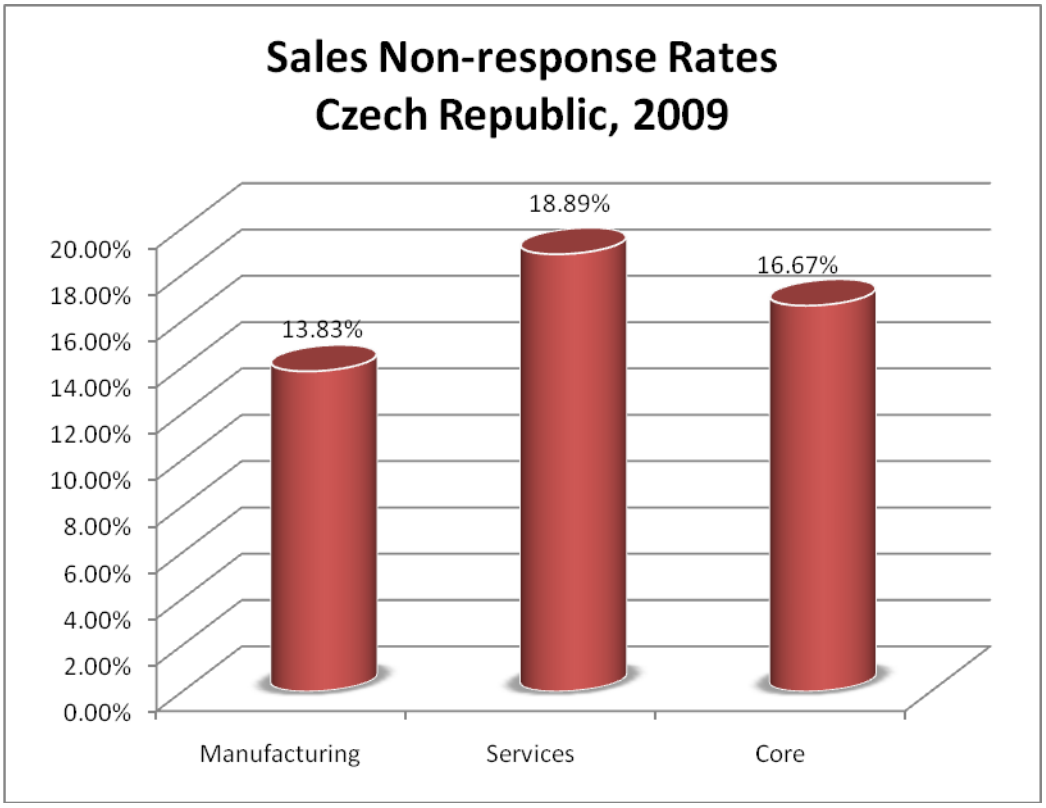
31. Item non-response was addressed by two strategies:

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).

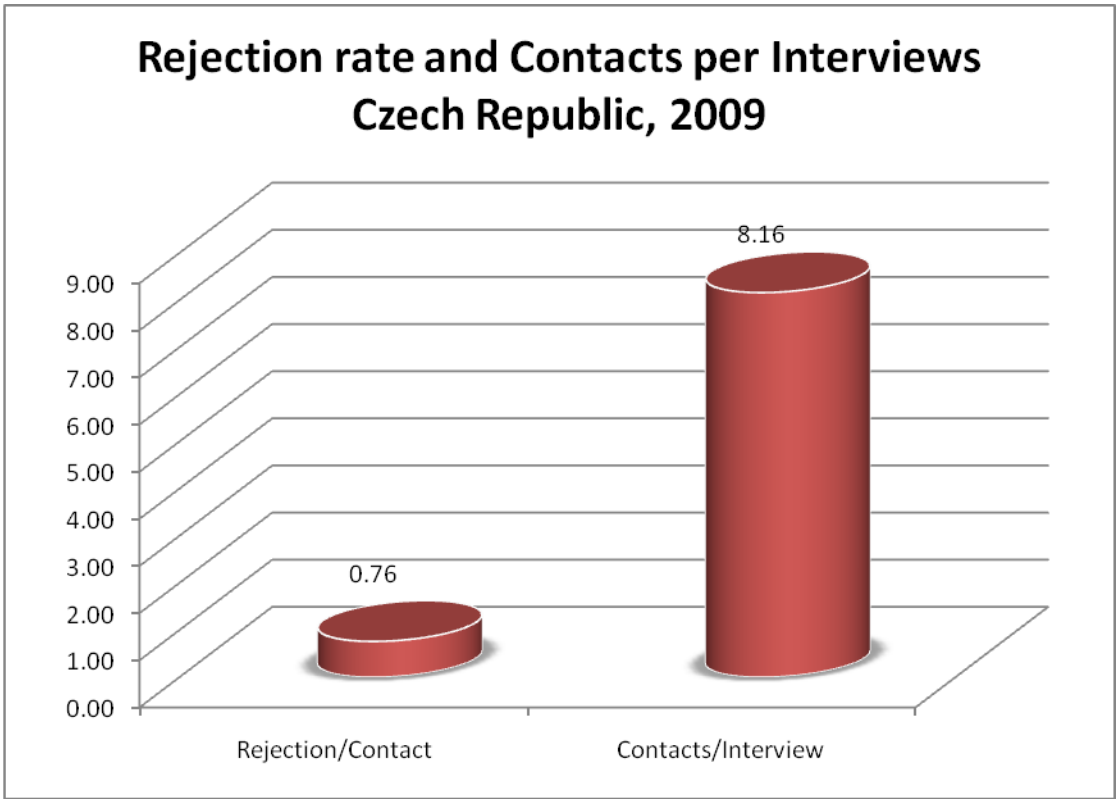
[6] Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

[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 8.16. 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 Czech 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. Samping: Design and Techniques, 1999.

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

Appendix A

Cell Weights – Czech Republic (Strict)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Praha	5-19	80	23	208
	20-99	26	15	70
	100+	30	5	13
Stredni Cechy	5-19	215	28	158
	20-99	59	7	56
	100+	22	7	11
Jihozapad	5-19	141	57	245
	20-99	130	6	65
	100+	26	3	13
Severozapad	5-19	186	35	205
	20-99	84	15	100
	100+	37		17
Severovychod	5-19	151	47	385
	20-99	128	19	92
	100+	13		33
Jihovychod	5-19	48	59	334
	20-99	51	13	64
	100+	16	3	35
Stredni Morava	5-19	67	122	208
	20-99	42	19	48
	100+	44	6	4
Moravskoslezsko	5-19	135	32	176
	20-99	109	6	70
	100+	38	3	14

Cell Weights – Czech Republic (Weak)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Praha	5-19	324	101	1,010
	20-99	88	53	286
	100+	99	19	51
Stredni Cechy	5-19	918	127	801
	20-99	211	26	240
	100+	77	25	46
Jihozapad	5-19	516	219	1,066
	20-99	398	18	236
	100+	77	8	46
Severozapad	5-19	760	151	998
	20-99	287	53	409
	100+	124		68
Severovychod	5-19	481	158	1,464
	20-99	342	54	294
	100+	33		102
Jihovychod	5-19	202	266	1,691
	20-99	182	48	274
	100+	58	10	145
Stredni Morava	5-19	208	402	766
	20-99	110	53	148
	100+	111	16	12
Moravskoslezsko	5-19	425	106	659
	20-99	287	18	221
	100+	99	8	42

Cell Weights – Czech Republic (Median)

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Praha	5-19	225	69	712
	20-99	66	39	220
	100+	82	16	43
Stredni Cechy	5-19	619	84	550
	20-99	156	19	180
	100+	62	20	38
Jihozapad	5-19	378	157	794
	20-99	318	14	192
	100+	67	7	41
Severozapad	5-19	525	102	702
	20-99	217	39	314
	100+	103		57
Severovychod	5-19	363	116	1,127
	20-99	283	44	248
	100+	30		94
Jihovychod	5-19	141	181	1,203
	20-99	139	36	213
	100+	48	8	123
Stredni Morava	5-19	139	262	521
	20-99	80	38	110
	100+	88	13	10
Moravskoslezsko	5-19	286	70	451
	20-99	211	13	165
	100+	79	6	35

Appendix B

Status Codes - Total

	ELIGIBLES	
Eligible	1.Eligible establishment *(Correct name and address)*	543
	2. Eligible establishment *(Different name but same address - the new firm/establishment bought the original firm/establishment)*	4
	3. Eligible establishment *(Different name but same address - the firm/establishment changed its name)*	1
	4. Eligible establishment *(Wrong address - the firm/establishment has changed address and the address could be found)*	2
	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	16
	7. Not a business: private household	32
	8. Ineligible activity: education, agriculture, finances, governments…	39
Unobtainable	91. No reply *(after having called in different days of the week and in different business hours)*	171
	92. Line out of order	16
	93. No tone	16
	10. Answering machine	5
	11. Fax line - data line	8
	12. Wrong address/ moved away and could not get the new references	266
	13. Refuses to answer the screener	919
	14. In process *(the establishment is being called/ is being contacted - previous to ask the screener)*	**375**
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	3
	Total	**2,416**

Response Outcomes - Total

Complete interviews *(Total)*	250
Incomplete interviews	0
Eligible in process	109
Refusals	191
Out of target	87
Impossible to contact	482
Ineligible - coop.	3
Refusal to the Screener	919
Total	**2,041**

PANEL

Complete interviews *(Total)*	17
Incomplete interviews	0
Eligible in process	5
Refusals	8
Out of target	6
Impossible to contact	34
Ineligible - coop.	0
Refusal to the Screener	39
Total	**109**

	ELIGIBLES	
Eligible	1.Eligible establishment *(Correct name and address)*	29
	2. Eligible establishment *(Different name but same address - the new firm/establishment bought the original firm/establishment)*	0
	3. Eligible establishment *(Different name but same address - the firm/establishment changed its name)*	0
	4. Eligible establishment *(Wrong address - the firm/establishment has changed address and the address could be found)*	1
	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	0
	7. Not a business: private household	3
	8. Ineligible activity: education, agriculture, finances, governments…	3
Unobtainable	91. No reply *(after having called in different days of the week and in different business hours)*	6
	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	28
	13. Refuses to answer the screener	39
	14. In process *(the establishment is being called/ is being contacted - previous to ask the screener)*	**12**
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	Total	**121**

FRESH

Complete interviews *(Total)*	233
Incomplete interviews	0
Eligible in process	104
Refusals	183
Out of target	81
Impossible to contact	448
Ineligible - coop.	3
Refusal to the Screener	880
Total	**1,932**

	ELIGIBLES	
Eligible	1.Eligible establishment *(Correct name and address)*	514
	2. Eligible establishment *(Different name but same address - the new firm/establishment bought the original firm/establishment)*	4
	3. Eligible establishment *(Different name but same address - the firm/establishment changed its name)*	1
	4. Eligible establishment *(Wrong address - the firm/establishment has changed address and the address could be found)*	1
	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	16
	7. Not a business: private household	29
	8. Ineligible activity: education, agriculture, finances, governments…	36
Unobtainable	91. No reply *(after having called in different days of the week and in different business hours)*	165
	92. Line out of order	16
	93. No tone	16
	10. Answering machine	5
	11. Fax line - data line	8
	12. Wrong address/ moved away and could not get the new references	238
	13. Refuses to answer the screener	880
	14. In process *(the establishment is being called/ is being contacted - previous to ask the screener)*	**363**
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	3
	Total	**2,295**

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

Czech Republic Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	14,281	54,599	39,833

Appendix D

Questionnaires:

Problems for the understanding of questions (write question number)	Factum found that respondents on occasion did not know answers during the first interviewer's visit which involved several contacts.
Problems found in the navigability of – questionnaires (for example, skip patterns).	No special problems encountered
Comments on questionnaires length:	Average length in the Czech language is 89.6 minutes. Although the questionnaire is unusually long, the most demanding point was in persuading respondents to partake in the survey. The length of the questionnaire was then accepted in most cases.
Suggestions or other comments on the questionnaire:	N/A

Database

Comments on the data entry program	Data entry program chosen: Confirmit (Factum) Data inserted by the interviewer into programmed script directly during interviewing. NIPO ODIN scripting software used (TNS AISA).
Comments on the data cleaning	In the case of Confirmit, it would have been very helpful to have gotten an export in EXCEL or other such format when doing the cleaning.

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	The global economic crisis influenced the level of respondents' willingness to participate but there shouldn't be any bias in the data validity. February-March was also difficult as it's the tax returns period.
Relevant country events occurred during fieldwork:	None
Other aspects:	N/A

Appendix E

Original Sample Design

Region	Employees	Sector			
		Manufacturing	52	Residual	Grand Total
Praha	5-19	3	5	7	15
	20-99	3	7	6	16
	100+	3	11	9	23
Praha Total		9	23	22	54
Stredni Cechy	5-19	3	3	3	9
	20-99	3	3	3	9
	100+	3	4	4	11
Stredni Cechy Total		9	10	10	29
Jihozapad	5-19	4	3	3	10
	20-99	4	3	3	10
	100+	4	2	3	9
Jihozapad Total		12	8	9	29
Severozapad	5-19	3	3	3	9
	20-99	3	3	2	8
	100+	3	2	2	7
Severozapad Total		9	8	7	24
Severovychod	5-19	5	3	4	12
	20-99	5	3	4	12
	100+	5	3	3	11
Severovychod Total		15	9	11	35
Jihovychod	5-19	5	5	4	14
	20-99	5	5	5	15
	100+	5	4	4	13
Jihovychod Total		15	14	13	42
Stredni Morava	5-19	4	4	3	11
	20-99	4	3	4	11
	100+	4	2	3	9
Stredni Morava Total		12	9	10	31
Moravskoslezsko	5-19	3	4	3	10
	20-99	3	3	3	9
	100+	3	2	2	7
Moravskoslezsko Total		9	9	8	26
Grand Total		90	90	90	270

Appendix F

Local Agencies involved in the study:

Local Agency	Name: TNS AISA, s.r.o. Country: The Czech Republic Membership of international organisation: TNS, WPP, ESOMAR, SIMAR Activities since: 1990
Name of Project Manager	Jana Rajsnerová
Name and position of other key persons of the project:	Petr Šobotník, Project Executive Olga Nebeská, Project Executive Kateřina Sixtová, Executive Assistant F2F Martina Chauturová, Executive Assistant F2F
Enumerators involved:	Enumerators: 111 Recruiters: interviewers also involved in recruitment
Other staff involved:	Fieldwork Coordinators: 2 Editing: 0 Data Entry: 0 Data Processing: 3

Local Agency	Name: Factum Invenio, s.r.o. Country: Czech Republic Membership of international organisation: Factum Group, Esomar, British Chamber of Commerce Chambre de Commerce Franco - Tchèque Activities since: 1991
Name of Project Manager	Jan Nalezený
Name and position of other key persons of the project:	Jan Nalezený – Country Manager Šárka Vidlařová – Field Work Manager
Enumerators involved:	Enumerators: 144 Recruiters: 5 No enumerator worked on recruitment
Other staff involved:	Fieldwork Coordinators: 1 Editing: 4 Data Entry: 2 Data Processing: -

Sample Frame:

Characteristic of sample frame used:	Database ALBERTINA (http://www.albertina.cz/czech/afm/p_poparo.html)
Source:	Albertina - Creditinfo Czech Republic. Drawn from the complete RES (database of the Czech Statistical Office)
Year of publication:	2007
Comments on the quality of sample frame:	TNS AISA statistics: 18.2 % wrong addresses 4.4 % unable to contact by telephone 2.2 % out of target
Year and organism who conducted the last economic census	Czech Statistical Office
Other sources for companies statistics	Meritum (not used for this survey)

Sample:

Comments/ problems on sectors and regions selected in the sample:	Czech Republic was divided in accordance with NUTS2. Given the required coverage, a large field-force was necessary which posed some challenges for the field organization.
Comments on the response rate:	Initially, the response rate was proving problematic. Respondents were often very reluctant to participate a priori; the market is "over-surveyed" and respondents claim to be too busy to find time to participate. Consequently, TNS AISA was taken on board to help achieve the target number of interviews.
Comments on the sample design:	From an organisational point-of-view it was difficult to handle replacements when 2nd and 3rd priorities of a record number were situated in other districts or regions than 1st priority and thus for another team of interviewers.

Fieldwork:

Date of Fieldwork	4th of September 2008 – 11th of March 2009
Country	Czech Republic
Interview number	Manufactures: 94 Services: 85 Core: 71
Problems found during fieldwork:	Refusals caused mainly by respondents' lack of time. The length and depth of the questionnaire and sampling method made the coordination of interviewers' work generally demanding – especially when there was a mistake or data missing in a questionnaire which required being sorted out by call-backs.
Other observations:	Respondents often interested in the survey

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.