

Romania Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Romania 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 Romania 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 group of sectors 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 eight regions. These regions are Nord-Est, Sud-Est, Sud, Vest, Nord-Vest, Bucuresti, Sud-Vest, and Centru.

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 establishments responding to the BEEPS 2005 survey where they were within the

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

selected geographical regions and met eligibility criteria. That sample is referred to as the Panel. The second frame used in Romania was the Trade Register of Romania. The full frame was not made available. Instead an extract was selected in Romania according to instructions from the TNS statistical team in London.

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 37% (414 out of 1,115 establishments).

Sample Frame Romania

Source: Trade Register of Romania

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Nord-Est	5-19	1,935	3,309	2,762	8,006
	20-99	1,009	944	296	2,249
	100+	349	148	24	521
Nord-Est Total		3,293	4,401	3,082	10,776
Sud-Est	5-19	1,566	3,662	2,545	7,773
	20-99	873	1,051	253	2,177
	100+	282	210	15	507
Sud-Est Total		2,721	4,923	2,813	10,457
Sud	5-19	1,534	2,994	2,448	6,976
	20-99	944	927	236	2,107
	100+	339	163	18	520
Sud Total		2,817	4,084	2,702	9,603
Vest	5-19	1,572	3,370	1,919	6,861
	20-99	940	866	205	2,011
	100+	330	131	15	476
Vest Total		2,842	4,367	2,139	9,348
Nord-Vest	5-19	2,570	3,809	2,223	8,602
	20-99	1,305	1,048	258	2,611
	100+	426	158	19	603
Nord-Vest Total		4,301	5,015	2,500	11,816
Bucuresti	5-19	2,513	7,303	2,586	12,402
	20-99	1,221	2,235	352	3,808
	100+	446	496	82	1,024
Bucuresti Total		4,180	10,034	3,020	17,234
Sud Vest	5-19	1,021	2,200	1,607	4,828
	20-99	487	581	190	1,258
	100+	178	96	13	287
Sud Vest Total		1,686	2,877	1,810	6,373
Centru	5-19	2,438	4,364	2,383	9,185
	20-99	1,274	1,143	303	2,720
	100+	460	164	18	642
Centru Total		4,172	5,671	2,704	12,547
Grand Total		26,012	41,372	20,770	88,154

Sample Frame Romania

Source: BEEPS 2005

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Nord-Est	< 5	9	3		12
	5-19	24	3	5	32
	20-99	32	3	8	43
	100+	17		5	22
Nord-Est Total		82	9	18	109
Sud-Est	< 5	3		3	6
	5-19	13	3	7	23
	20-99	20	2	7	29
	100+	7	1	1	9
Sud-Est Total		43	6	18	67
Sud	< 5	4	2	1	7
	5-19	13	3	6	22
	20-99	11	2	7	20
	100+	20	1	3	24
Sud Total		48	8	17	73
Vest	< 5	3	1	3	7
	5-19	18	1	6	25
	20-99	9	2	3	14
	100+	6		3	9
Vest Total		36	4	15	55
Nord-Vest	< 5	15	2	2	19
	5-19	22	2	4	28
	20-99	24	3	5	32
	100+	13	1	2	16
Nord-Vest Total		74	8	13	95
Bucuresti	< 5	4	3	1	8
	5-19	11		5	16
	20-99	12	6	7	25
	100+	13	1	2	16
Bucuresti Total		40	10	15	65
Sud Vest	5-19				
	20-99				
	100+				
Sud Vest Total					
Centru	5-19				
	20-99				
	100+				
Centru Total					
Grand Total		323	45	96	464

Sectors included in the Sample:

Original Sectors	Manufactures: 15, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 34, 35, 36 Services: 52 Residual: 45, 50, 51, 55, 60, 61, 62, 63, 72
Added Sectors:	None

4. Data Base Structure:

10. The structure of the data base reflects the fact that three 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 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 (*four digit code*) in the sample frame.

14. All of the following variables contain information from the sampling frame and were defined with the sampling design. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

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

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

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

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

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

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

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

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

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

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

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

5. Universe Estimates

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

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

6. Weights

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

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

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

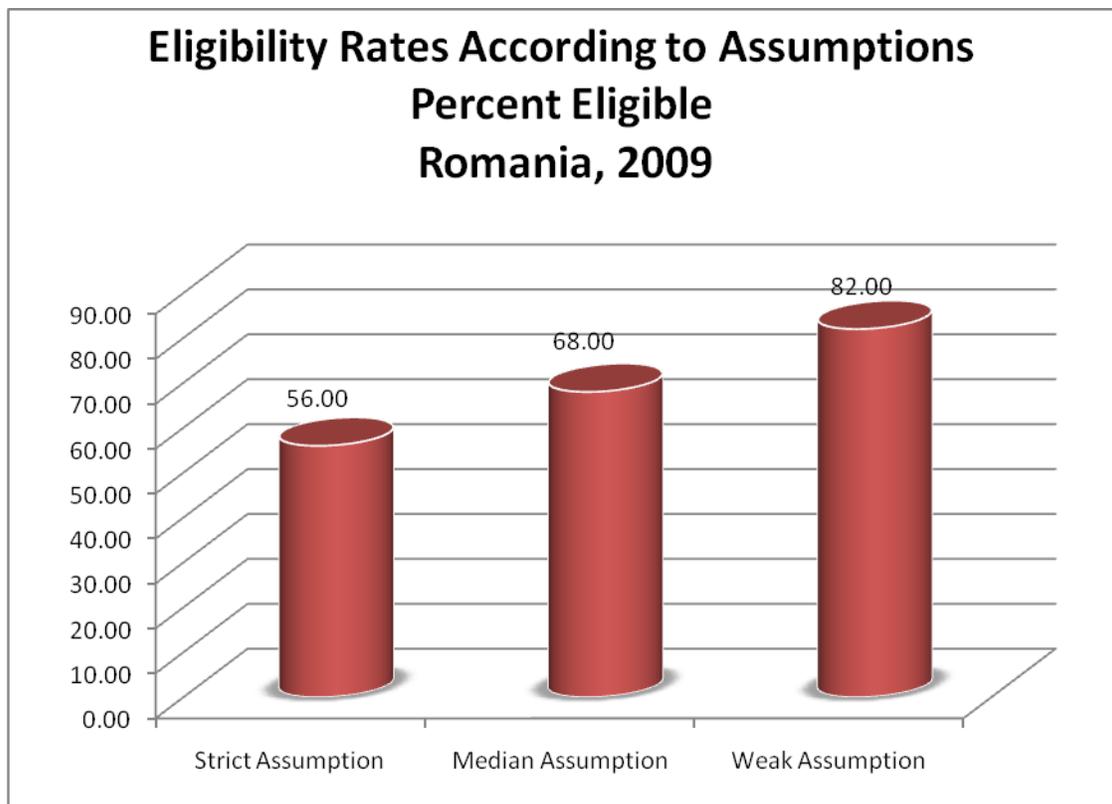
a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable w_strict .

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

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.



26. Within each of these assumptions regarding eligibility, a pair of weight sets was calculated. Proportions were calculated using the raw sample count for each cell.

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

7. Appropriate use of the weights

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

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

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

8. Non-response

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

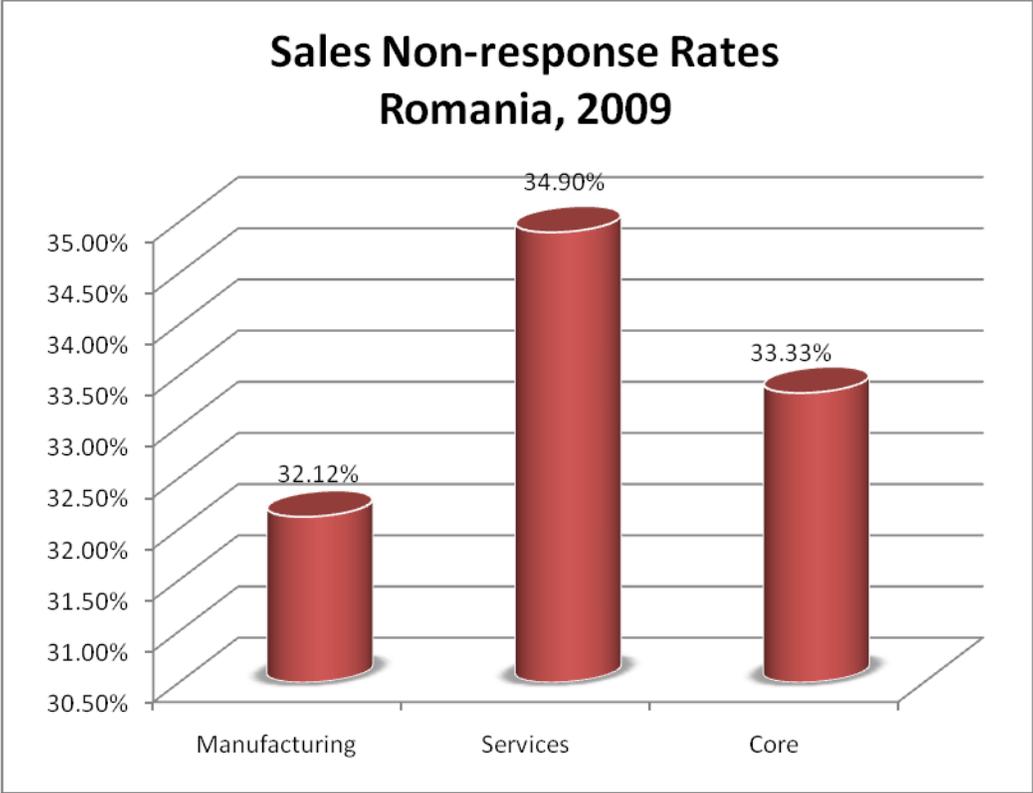
31. Item non-response was addressed by two strategies:

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

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

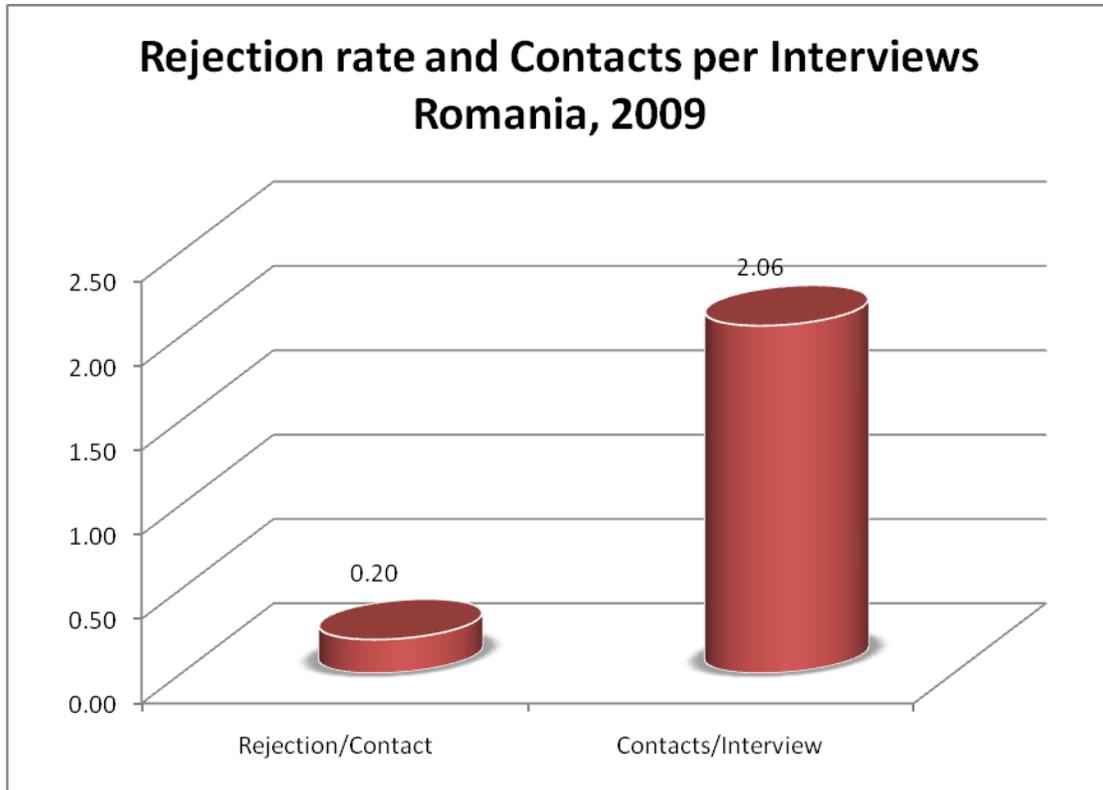
⁶ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

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



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

33. As the following graph shows, the number of contacted establishments per realized interview was 2.06. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units.



34. Details on rejections rates, eligibility rates, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to the Romania. 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 – Romania

Strict Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Nord-Est	5-19	65	244	346
	20-99	22	48	15
	100+	16	8	2
Sud-Est	5-19	109	363	387
	20-99	39	116	24
	100+	18	30	1
Sud	5-19	106	218	319
	20-99	73	88	34
	100+	34	14	3
Vest	5-19	87	228	155
	20-99	45	114	14
	100+	38	37	2
Nord-Vest	5-19	119	341	203
	20-99	67	104	18
	100+	29	20	2
Bucuresti	5-19	257	366	78
	20-99	79	138	13
	100+	24	12	2
Sud Vest	5-19	151	398	248
	20-99	100	116	48
	100+	31	19	3
Centru	5-19	221	484	166
	20-99	205	113	43
	100+	44	27	6

Cell Weights – Romania

Weak Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Nord-Est	5-19	197	636	868
	20-99	61	112	35
	100+	41	18	5
Sud-Est	5-19	202	576	590
	20-99	65	164	33
	100+	29	41	1
Sud	5-19	220	385	541
	20-99	134	138	52
	100+	60	21	4
Vest	5-19	209	468	305
	20-99	96	209	24
	100+	78	65	3
Nord-Vest	5-19	222	542	311
	20-99	112	148	25
	100+	46	26	3
Bucuresti	5-19	579	705	144
	20-99	159	238	22
	100+	47	20	4
Sud Vest	5-19	195	440	263
	20-99	115	115	46
	100+	35	19	3
Centru	5-19	286	536	177
	20-99	237	111	41
	100+	49	27	6

Cell Weights – Romania

Median Weights

Individual Cell Weights

Region	Employees	Sector		
		Manufacturing	52	Residual
Nord-Est	5-19	106	399	528
	20-99	36	76	23
	100+	25	13	3
Sud-Est	5-19	127	421	417
	20-99	44	129	25
	100+	21	34	1
Sud	5-19	129	264	360
	20-99	85	102	37
	100+	40	17	3
Vest	5-19	119	313	197
	20-99	60	151	17
	100+	51	50	2
Nord-Vest	5-19	136	387	215
	20-99	74	114	19
	100+	32	22	3
Bucuresti	5-19	345	490	97
	20-99	103	178	16
	100+	32	16	3
Sud Vest	5-19	168	440	256
	20-99	107	116	48
	100+	34	19	3
Centru	5-19	244	533	170
	20-99	218	114	43
	100+	48	27	6

Appendix B

Status Codes - Total

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	528
	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>)	69
	16. Panel firm - now less than five employees	23
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	18
	7. Not a business: private household	28
	8. Ineligible activity: education, agriculture, finances, governments...	18
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	159
	92. Line out of order	41
	93. No tone	56
	10. Answering machine	12
	11. Fax line - data line	16
	12. Wrong address/ moved away and could not get the new references	42
	13. Refuses to answer the screener	106
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	1
Total		1,119

Response Outcomes - Total

Complete interviews (<i>Total</i>)	541
Incomplete interviews	6
Eligible in process	32
Refusals	43
Out of target	64
Impossible to contact	326
Ineligible - coop.	1
Refusal to the Screener	106
Total	1,119

PANEL

Complete interviews (<i>Total</i>)	92
Incomplete interviews	3
Eligible in process	32
Refusals	24
Out of target	46
Impossible to contact	241
Ineligible - coop.	1
Refusal to the Screener	24
Total	463

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	79
	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>)	1
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	48
Ineligible	16. Panel firm - now less than five employees	23
	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	12
	7. Not a business: private household	21
Unobtainable	8. Ineligible activity: education, agriculture, finances, governments...	13
	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	116
	92. Line out of order	36
	93. No tone	40
	10. Answering machine	8
	11. Fax line - data line	13
	12. Wrong address/ moved away and could not get the new references	28
	13. Refuses to answer the screener	24
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0	
Out of target	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	1
Total		463

FRESH SAMPLE

Complete interviews (<i>Total</i>)	449
Incomplete interviews	3
Eligible in process	0
Refusals	19
Out of target	18
Impossible to contact	85
Ineligible - coop.	0
Refusal to the Screener	82
Total	656

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	449
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	1
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	0
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	21
Ineligible	16. Panel firm - now less than five employees	0
	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	6
	7. Not a business: private household	7
Unobtainable	8. Ineligible activity: education, agriculture, finances, governments...	5
	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	43
	92. Line out of order	5
	93. No tone	16
	10. Answering machine	4
	11. Fax line - data line	3
	12. Wrong address/ moved away and could not get the new references	14
	13. Refuses to answer the screener	82
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0	
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
Total		656

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

$$= (\text{Sum of the numbers with codes 1,2,3,4,\&16}) / \text{Total}$$

Weak eligibility

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

Median eligibility

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

Romania Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	51,438	83,519	61,381

Appendix D
Questionnaires:

Problems for the understanding of questions (write question number)	Some of the questions were quite complicated and were not very well understood by the respondents (see repeated DK/NA answers).
Problems found in the navigability of – questionnaires (for example, skip patterns).	None.
Comments on questionnaires length:	The questionnaire was considered to be too long by us, our coordinators, our interviewers and the respondents. One of our regional coordinators even refused to work with such a long questionnaire so we had to find another one for that specific region.
Suggestions or other comments on the questionnaire:	None.

Database

Comments on the data entry program	Data entry program chosen: Conformat Comments: The interface was not very user-friendly and not very fast for data entry. For optimum results, our data-entry operators had to use both the keyboard and the mouse and this increased the overall time spent on punching in the data. Everything would have worked much faster and smoother if we would have used our own data entry software and deliver the database in SPSS/Excel/whatever.
Comments on the data cleaning	We had no direct access to the database so we weren't able to run any filters or cleaning programs on it. The cleaning process took longer because all corrections needed to be implemented in excel files (data validation reports) provided by TNS Opinion.

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	The answers for questions regarding informal payments were hardly obtained and sometimes the interviewers signaled that the respondents didn't seem to be very sincere when answering such questions. From our previous experience, due to various reasons – fiscal, political, grey-market economy - we can say that large multinational and national companies in Romania have quite strict rules regarding answering such surveys and some of them definitely refuse to participate due to internal regulations.
Relevant country events occurred during fieldwork:	Christmas and New Year Holidays; parliamentary election on November 28
Other aspects:	None

Appendix E
Original Sample Design

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Nord-Est	5-19	8	6	9	23
	20-99	8	7	9	24
	100+	7	6	7	20
Nord-Est Total		23	19	25	67
Sud-Est	5-19	6	7	8	21
	20-99	6	7	7	20
	100+	6	8	5	19
Sud-Est Total		18	22	20	60
Sud	5-19	6	6	8	20
	20-99	7	6	7	20
	100+	7	6	5	18
Sud Total		20	18	20	58
Vest	5-19	6	7	6	19
	20-99	7	6	6	19
	100+	7	5	4	16
Vest Total		20	18	16	54
Nord-Vest	5-19	10	7	7	24
	20-99	10	7	7	24
	100+	9	6	6	21
Nord-Vest Total		29	20	20	69
Bucuresti	5-19	10	14	9	33
	20-99	9	15	10	34
	100+	10	19	24	53
Bucuresti Total		29	48	43	120
Sud Vest	5-19	4	4	5	13
	20-99	4	4	5	13
	100+	4	4	4	12
Sud Vest Total		12	12	14	38
Centru	5-19	10	9	8	27
	20-99	9	8	9	26
	100+	10	6	5	21
Centru Total		29	23	22	74
Grand Total		180	180	180	540

Achieved Sample

Region	Employees	Sector			Grand Total
		Manufacturing	52	Residual	
Nord-Est	5-19	9	5	3	17
	20-99	15	8	8	31
	100+	8	8	5	21
Nord-Est Total		32	21	16	69
Sud-Est	5-19	7	6	4	17
	20-99	12	6	7	25
	100+	9	5	11	25
Sud-Est Total		28	17	22	67
Sud	5-19	6	7	4	17
	20-99	6	6	4	16
	100+	5	7	4	16
Sud Total		17	20	12	49
Vest	5-19	7	7	6	20
	20-99	9	4	8	21
	100+	4	2	5	11
Vest Total		20	13	19	52
Nord-Vest	5-19	11	7	7	25
	20-99	11	7	10	28
	100+	9	6	6	21
Nord-Vest Total		31	20	23	74
Bucuresti	5-19	4	10	17	31
	20-99	7	9	15	31
	100+	9	24	22	55
Bucuresti Total		20	43	54	117
Sud Vest	5-19	5	5	6	16
	20-99	4	5	4	13
	100+	5	5	5	15
Sud Vest Total		14	15	15	44
Centru	5-19	8	8	13	29
	20-99	5	10	7	22
	100+	9	6	3	18
Centru Total		22	24	23	69
Grand Total		184	173	184	541

Appendix F

Local Agency team involved in the study:

Local Agency	Name: Center for Urban and Regional Sociology - CURS Country: Romania Membership of international organisation: - Activities since: 1990
Name of Project Manager	Catalin Augustin Stoica
Name and position of other key persons of the project:	Ionela Şufaru Augustin Abraham
Enumerators involved:	Enumerators: 55 Recruiters: 56 (most of the interviewers did the recruitment themselves. For the panel sample, the recruitment was conducted from our central headquarters). 2 interviewers did not carry out screeners 3 recruiters didn't carry out complete interviews
Other staff involved:	Fieldwork Coordinators: 51 – 8 at the regional level and 42 at county levels Editing: 4 Data Entry: 5 Data Processing: 2

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.

Sample Frame:

Characteristic of sample frame used:	Panel sample provided by the World Bank and EBRD – 464 establishments; Fresh sample of 3570 establishment bought from The National Trade Register Office and selected by The National Trade Register Office specialists` based on the instructions provided by TNS Opinion
Source:	The National Trade Register Office - official source
Year:	2007
Comments on the quality of sample frame:	<p>Poor quality of the panel sample:</p> <ul style="list-style-type: none"> - many of the firms didn't correspond to the regions they were assigned to in the panel sample (for example, firms from Constanta which is in the far East of Romania were assigned to the North-West region); - many of the firms didn't correspond to the sectors they were assigned to in the panel sample and they never had the given ISIC codes; - many firms were outside the sampling zone and they never had ISIC codes that belonged to this range; - many firms didn't exist at all or their contact details were not accurate – for these last ones we tried to find other sources for their contacts but we weren't able to solve all of them; - many managers from these firms did not remember to have participated in the BEEPS survey in 2005 or to have ever been contacted for such a survey; - in the end, we were able to use only about 25% of the firms from the panel sample; <p>The fresh sample was much better and had an acceptable response rate. However, we encountered problems because a limited number of replacement contacts was provided, and sample top ups had to be approved by the World Bank and EBRD task managers every time there was a request. This delayed our time spent on the fieldwork considerably.</p>
Year and organization who conducted the last economic census:	National Institute of Statistics – 2007
Other sources for companies' statistics:	National Institute of Statistics - Romanian Statistical Yearbook 2007. Data sources: Statistical Business Register managed by the National Institute of Statistics that is a statistical instrument for keeping identification data for all legal units carrying out an economic or social activity. The register is updated based on the following sources: Fiscal Register, Trade Register, Balance sheet of economic operators. Romanian Statistical Yearbook presents in table no. 15.20 Active local units from industry, construction, trade and other services, by development region, activity of national economy at level of NEAC Classification Rev.1 and by size class (in accordance with Eurostat criteria the following interval are used: 0-9 employees, 10-49 employees, 50-249 employees, 250 employees and over).

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: None On regions: None
Comments on the response rate:	Quite low rate because of the length and structure of the questionnaire
Comments on the sample design:	The ample design was quite complicated, strict and didn't allow us much flexibility, which in Romania is very much appreciated due to the economic environment. Many firms appear and disappear from one year to the other, the refusal rate for such surveys is generally quite high, and many of the selected firms were from rural far areas that were not easy at all to access.
Other comments:	None

Fieldwork:

Date of Fieldwork	September 2008 – December 2008
Country	Romania
Interview number	Manufactures: 193; Services: 192; Core: 156
Problems found during fieldwork:	<p>Many respondents initially agreed to complete the interview when they were screened but later, when they saw how long the questionnaire was, refused to do it.</p> <p>Some of the respondents became bored during the interview and refused to continue. Some of them rescheduled the interview and then were unreachable (didn't answer the phone, didn't respect the meeting times and dates). However, most of them categorically refused to carry on with the interview.</p> <p>Due to the low flexibility level of this survey and the limited sample given by TNS London, some of our interviewers had to carry repeated visits of over 200 km's (back and forth) to far rural villages for one single questionnaire, fact which was pretty annoying and increased our costs considerably.</p> <p>Difficulties signaled in talking to several different managers (HR, Finance, etc.) for getting the required answers for one single questionnaire.</p>
Other observations:	None