Macedonia Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Macedonia 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 Macedonia 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 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 120 interviews. For the manufacturing industries sample sizes were deflated by about 5% 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 20% to account for under sampling in small firms in manufacturing and 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 Eastern, North-West & West, Skopje, and South.

7. The Macedonia sample contains panel data. The wave 1 panel "Investment Climate Private Enterprise Survey implemented in Macedonia" consisted of 200 establishments interviewed in 2005. A total of 87 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 the Central Registry of Macedonia.

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 15.98% (90 out of 563 establishments).

Sample Frame:

Characteristic of sample frame used:	The list used contained all relevant stratification variables: name of the establishment, contact details, region, city and municipality, ISIC codes, employee number.
Source:	Central Registry of Macedonia
Year of publication:	2008.
Comments on the quality of sample frame:	One of the biggest difficulties was the fact that the establishments did not inform the CR of their new addresses and the information was not up-to-date with the current addresses – the city, municipality and region of the establishment was not always correct. Other information that was not always correct were the number of employees and the activity of the establishment. Also because this data base did not contain phone numbers, the recruitment was hardened.
Year and organization who conducted the last economic census	In Macedonia there never was an economic census

Sample Frame Macedonia

Source: Central Registry of Macedonia

Region	Employees	Sector			
		Manufacturing	**52**	**Residual**	**Grand Total**
Eastern	5-19	422	466	1,051	1,939
	20-99	136	54	203	393
	100+	47	10	42	99
Eastern Total		605	530	1,296	2,431
North-West & West	5-19	468	368	417	1,253
	20-99	283	23	81	387
	100+	89	0	8	97
North-West & West Total		840	391	506	1,737
Skopje	5-19	195	287	305	787
	20-99	96	23	49	168
	100+	23	1	6	30
Skopje Total		314	311	360	985
South	5-19	205	246	299	750
	20-99	103	24	53	180
	100+	63	2	11	76
South Total		371	272	363	1,006
Grand Total		2,130	1,504	2,525	6,159

Sectors included in the Sample:

Original Sectors	Manufactures: *(15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 31, 33, 34, 35, 36, 37)* Services: *(52)* Residual: *(45, 50, 51, 55, 60, 62, 63, 64, 72)*
Added (top up) Sectors	*None*

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 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.

14. 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 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).

-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 Macedonia 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:

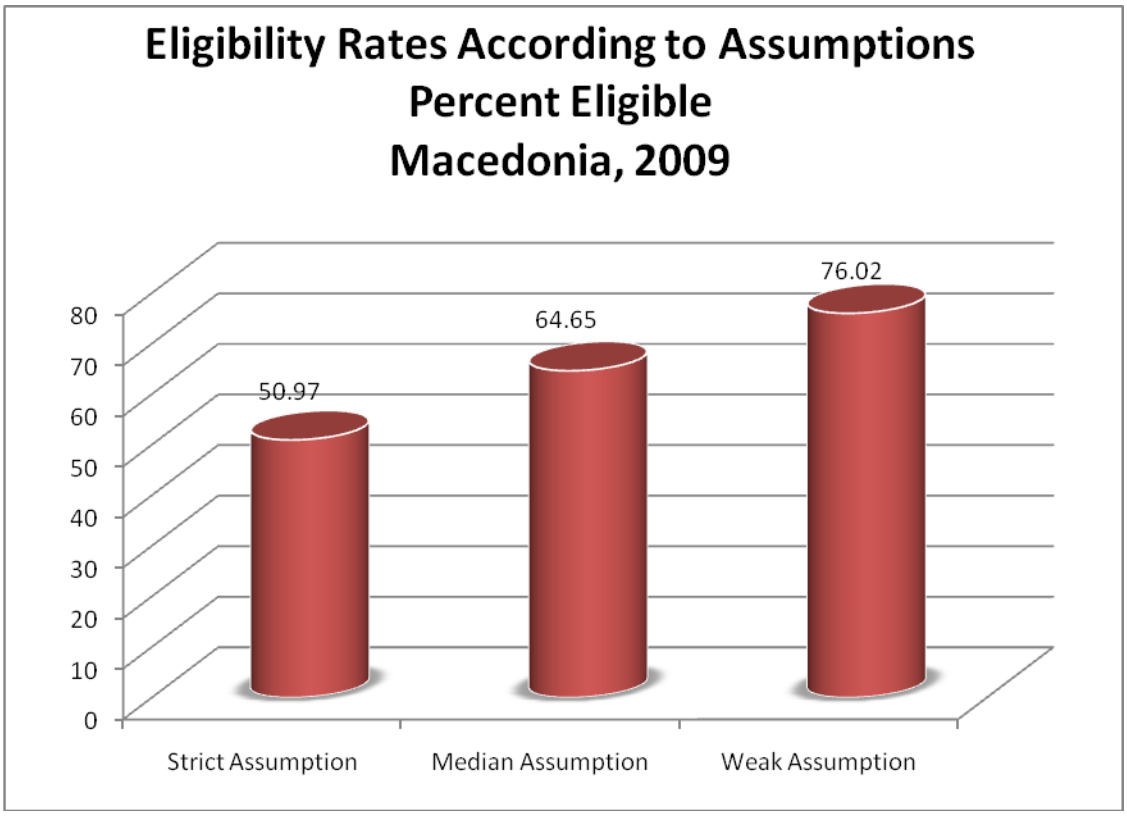
a- Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *w_strict*.

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 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. 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 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.

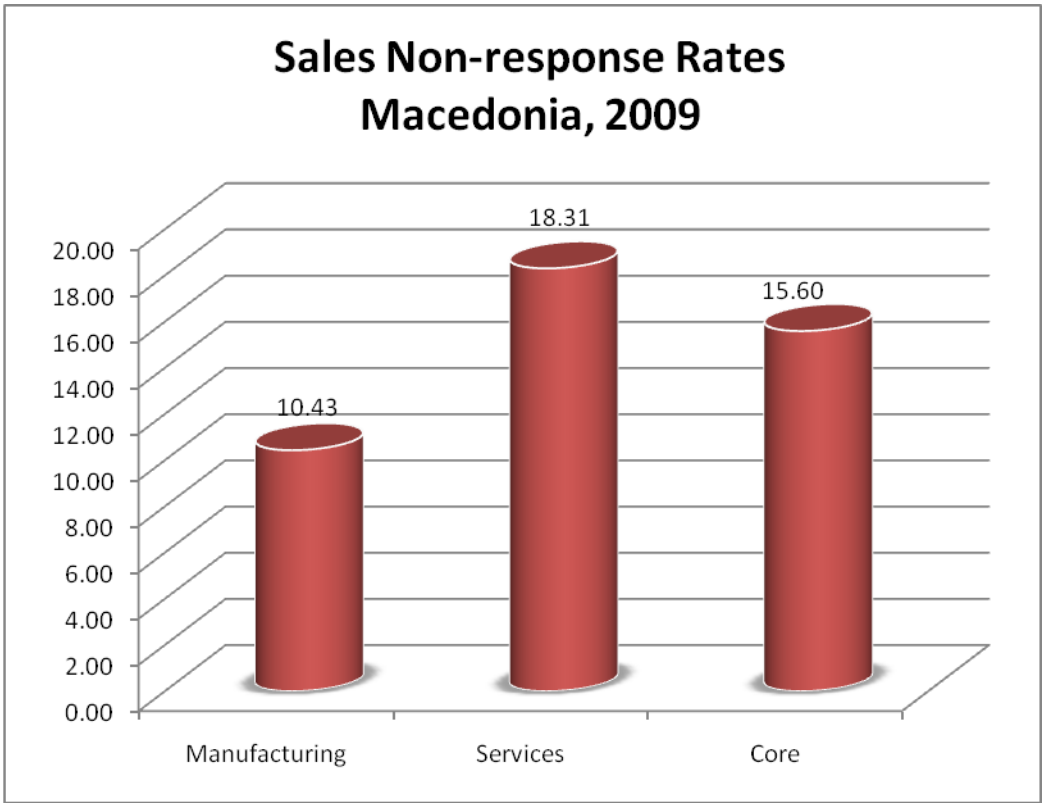
31. Item non-response was addressed by two strategies:

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).

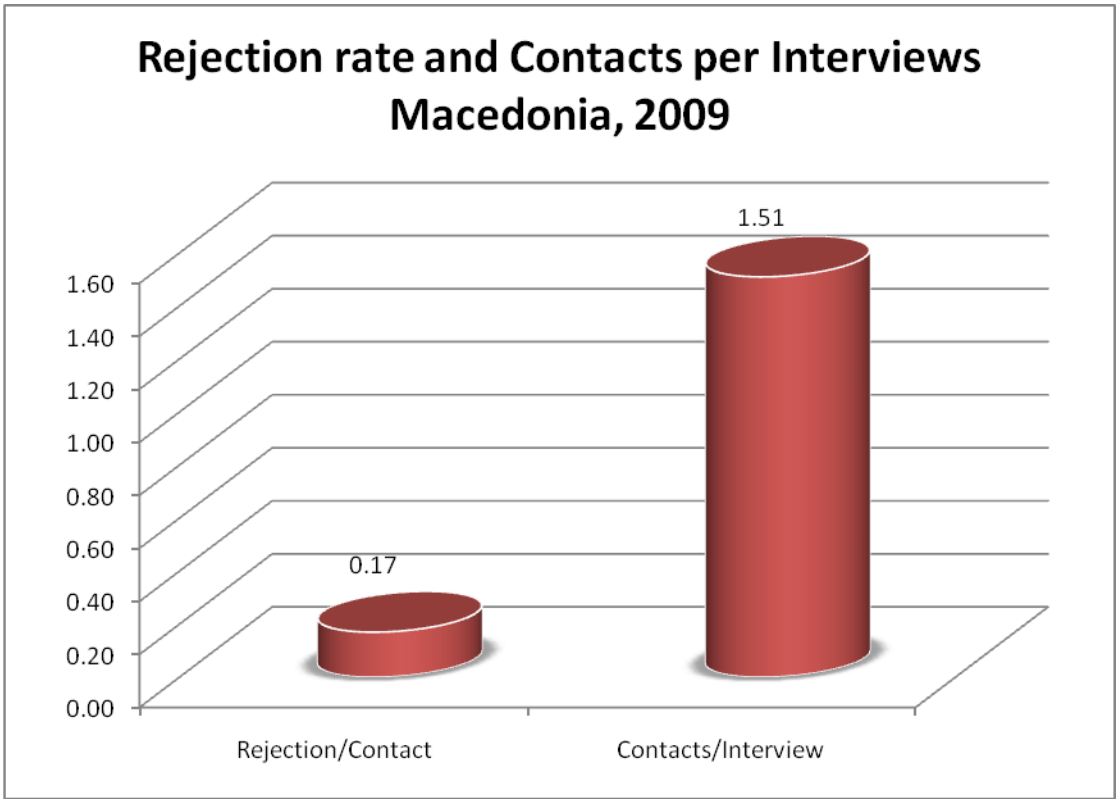
[6] Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

[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 1.51. 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 (1.51) suggests that the main source of error in estimates in the Macedonia 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 Macedonia. 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

Macedonia Strict

Region	Employees	Sector		
		Manufacturing	52	Residual
Eastern	5-19	38	27	44
	20-99	14	4	24
	100+	2		5
North-West & West	5-19	38	24	44
	20-99	14	2	11
	100+	18		2
Skopje	5-19	10	10	6
	20-99	8	1	1
	100+	2	1	1
South	5-19	38	23	44
	20-99	14	2	5
	100+	4	1	1

Macedonia Weak

Region	Employees	Sector		
		Manufacturing	52	Residual
Eastern	5-19	47	38	54
	20-99	17	5	28
	100+	3		6
North-West & West	5-19	47	33	54
	20-99	17	3	14
	100+	22		3
Skopje	5-19	15	16	8
	20-99	12	1	2
	100+	3	1	1
South	5-19	47	31	54
	20-99	17	2	6
	100+	5	2	2

Macedonia Median

Region	Employees	Sector		
		Manufacturing	**52**	**Residual**
Eastern	5-19	38	30	47
	20-99	15	4	26
	100+	3		6
North-West & West	5-19	38	27	47
	20-99	15	2	13
	100+	21		3
Skopje	5-19	12	13	7
	20-99	10	1	2
	100+	2	1	1
South	5-19	38	25	47
	20-99	15	2	6
	100+	5	2	2

Appendix B

Eligibility Status – Summary Fieldwork Reports TOTAL

Status Codes

	ELIGIBLES	
Eligible	1.Eligible establishment *(Correct name and address)*	386
	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	2
	6. The firm discontinued businesses	16
	7. Not a business: private household	2
	8. Ineligible activity: education, agriculture, finances, governments…	6
Unobtainable	91. No reply *(after having called in different days of the week and in different business hours)*	7
	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	57
	13. Refuses to answer the screener	77
	14. In process *(the establishment is being called/ is being contacted - previous to ask the screener)*	**9**
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	Total	**563**

Response Outcomes

Complete interviews *(Total)*	366
Incomplete interviews	3
Eligible in process	1
Refusals	17
Out of target	26
Impossible to contact	64
Ineligible - coop.	0
Refusal to the Screener	77
Total	**554**

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

Macedonian Establishment Estimates

Cells	Strict	Weak	Median
Un-collapsed Cells	4,640	5,969	5,107
Collapsed Cells	4,612	5,961	5,069

Appendix D

Questionnaires:

Problems for the understanding of questions (write question number)	ECAg12: respondents gave answers for all the licences as a sum, and not an average value. a18 and p4: there was a bit of confusion, when there were other people present during the interview – the answer on a18 would be 1 and on p4 would be 2,3 or four, because of other people present during the interview, despite the respondent.
Problems found in the Navigability of – questionnaires (for example, skip patterns).	No problems here
Comments on questionnaires length:	The interviews lasted an hour on average or more and of course respondents were not very comfortable with the length of questionnaire.
Suggestions or other Comments on the questionnaire:	None

Database

Comments on the data entry program	Data entry program chosen: PERTS Comments: None
Comments on the data cleaning	Data cleaning was done following the validation reports sent by TNS-

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	In the past 15 years, Macedonia is going through a transition period during which unemployment, poverty and overall difficult economic situation are the biggest problems the country is facing today. On the other hand in Macedonia the response rate is usually high, since the culture is hospitable and welcoming.
Relevant country events occurred during fieldwork:	No relevant events during the fieldwork that would affect the results of the survey or the fieldwork.
Other aspects:	None

Appendix E
Original Sample Design

Region	Employees	Sector			
		Manufacturing	52	Residual	Grand Total
Eastern	5-19	13	20	22	55
	20-99	9	25	23	57
	100+	9	5	21	35
Eastern Total		31	50	66	147
North-West & West	5-19	15	15	9	39
	20-99	18	11	9	38
	100+	16		4	20
North-West & West Total		49	26	22	97
Skopje	5-19	6	12	7	25
	20-99	6	10	5	21
	100+	4		3	7
Skopje Total		16	22	15	53
South	5-19	6	10	6	22
	20-99	7	11	6	24
	100+	11	1	5	17
South Total		24	22	17	63
Grand Total		120	120	120	360

TARGET	Sector			
Employees	Manufacturing	52	Residual	Grand Total
5-19	40	57	44	141
20-99	40	57	43	140
100+	40	6	33	79
Grand Total	120	120	120	360

Appendix F

Local Agency team involved in the study:

Local Agency	Name: Brima Country: Macedonia Membership of international organization Gallup International, TNS
Enumerators involved:	Enumerators: 1 Recruiters: 2 * The recruitment was done mostly by the same interviewers that conducted the interviews and additionally by our office – by our field force manager and for companies the project manager on this project
Other staff involved:	Fieldwork Coordinators: 5 Data Entry: 1 Data Processing: 1

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: None On regions: Since some of the addresses of the establishment were wrong (as given by the Central Registry), the results was that our first preference establishments were sometimes done in a different municipality and region than planed, but still within and eligible region.
Comments on the response rate:	We did not have many problems with the refusals; the refusal rate was as expected from our previous experiences. We managed to convince some of the respondents that refused to participate, but still some refusals were definite. The refusals were not related to any specific region, sector of activity and interviewer.
Comments on the sample design:	None
Other comments:	None

Fieldwork:

Date of Fieldwork	5th of September
Country	Macedonia
Interview number	Manufactures: 115 Services: 142 Core: 109

Problems found during fieldwork:	The issues we had to deal with during fieldwork were expected and since we have experience in the B2B surveys and skilled interviews on the filed, we managed to solve the problems efficiently. Those issues referred to the following: - recruitment process, - ensuring responses to more delicate issues - locating the establishment in cases where the addresses were wrong and there was no phone number
Other observations:	None

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.