

The Kazakhstan 2013 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Kazakhstan between December 2012 and August 2013 as part of the fifth round of the Business Environment and Enterprise Performance Survey (BEEPS V), 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 four times at an interval of three years. This has added an important element of dynamics in the study of business environment in transition countries.

The Enterprise Surveys, through interviews with firms in the manufacturing and services sectors, capture business perceptions on the biggest obstacles to enterprise growth, the relative importance of various constraints to increasing employment and productivity, and the effects of a country’s business environment on its international competitiveness. They are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

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.

II. Sampling Structure

2. The sample for Kazakhstan was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*¹. Stratified random sampling² was preferred over simple random sampling for several reasons³:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following sectors: financial intermediation (group J), real estate and renting activities (group K, 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.

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

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 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 one manufacturing industry, and two service industries (retail, and other services).

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). 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 5 regions (city and the surrounding business area) throughout Kazakhstan.

III. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and region) are required to draw the sample. 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. IPSOS was hired to implement the Kazakhstan 2013 enterprise survey. There were local subcontractors in each of the 5 regions surveyed.

9. The sample frame used for the survey in Kazakhstan was from: Kazakhstan Statistical Committee. The database contained the following information

- Coverage;
- Up to datedness;- Availability of detailed stratification variables;
- Contact name(s).

Counts from the sample frame are shown below.

Sample Frame

Source: Kazakhstan Statistical Committee, 2011.

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	571	226	2227	3024
	20-99	314	59	660	1033
	100+	102	14	146	262
	Total	987	299	3033	4319
East	5-19	565	292	2321	3178
	20-99	285	75	737	1097
	100+	109	13	150	272
	Total	959	380	3208	4547
North	5-19	1363	782	8053	10198
	20-99	770	310	3193	4273
	100+	241	43	493	777
	Total	2374	1135	11739	15248
South	5-19	2513	1306	13045	16864
	20-99	1186	380	4397	5963
	100+	382	95	963	1440
	Total	4081	1781	18405	24267
West	5-19	595	346	4435	5376
	20-99	293	100	1606	1999
	100+	121	20	375	516
	Total	1009	466	6416	7891
Grand Total		9410	4061	42801	56272

10. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 600 establishments with five or more employees.

11. The quality of the frame was assessed at the onset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc.

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 14.9% (566 out of 3789 establishments)⁴. Breaking down by stratified industries, the following sample targets were achieved (using a4a and a6a):

⁴ Based on out of target contacts and impossible to contact establishments

Achieved sample

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	13	14	13	40
	20-99	10	5	3	18
	100+	7	2	1	10
	Total	30	21	17	68
East	5-19	13	16	13	42
	20-99	10	6	3	19
	100+	8	1	1	10
	Total	31	23	17	71
North	5-19	21	33	40	94
	20-99	15	19	21	55
	100+	9	4	2	15
	Total	45	56	63	164
South	5-19	33	40	40	113
	20-99	19	22	26	67
	100+	11	9	4	24
	Total	63	71	70	204
West	5-19	13	19	26	58
	20-99	10	8	5	23
	100+	8	2	2	12
	Total	31	29	33	93
Grand Total		200	200	200	600

IV. Data Base Structure:

13. 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. The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to manufacturing sectors. The third expanded variation, the Retail Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to retail firms. Each variation of the questionnaire is identified by the index variable, *a0*.

14. 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 specific to the Eastern Europe and Central Asia region, 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.

15. 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.

16. 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.

17. All of the following variables contain information from the sampling frame. 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

-*a6a*: coded using the same standard for small, medium, and large establishments as defined above. The code -9 was used to indicate units for which size was undetermined in the sample frame.

-*a4a*: coded using ISIC Rev 3.1 codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), retail (52), and (45, 50, 51, 55, 60-64, 72) for other services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes 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.

19. Note that there are additional variables for location (*a3x*) and size (*11*, *16* and *18*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

20. 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.

21. Variables *11*, *16* and *18* 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.

22. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

V. Universe Estimates

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

24. Appendix B shows the overall estimates of the numbers of establishments in Kazakhstan based on the sample frame.

25. For some establishments where contact was not successfully completed during the screening process (because the firm has moved and it is not possible to locate the new location, for example), it is not possible to directly determine eligibility. Thus, different assumptions about the eligibility of establishments result in different adjustments to the universe cells and thus different sampling weights.

26. Three sets of assumptions on establishment eligibility are used to construct sample adjustments using the status code information.

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

Strict eligibility = (Sum of the firms with codes 1,2,3,4,&16) / Total

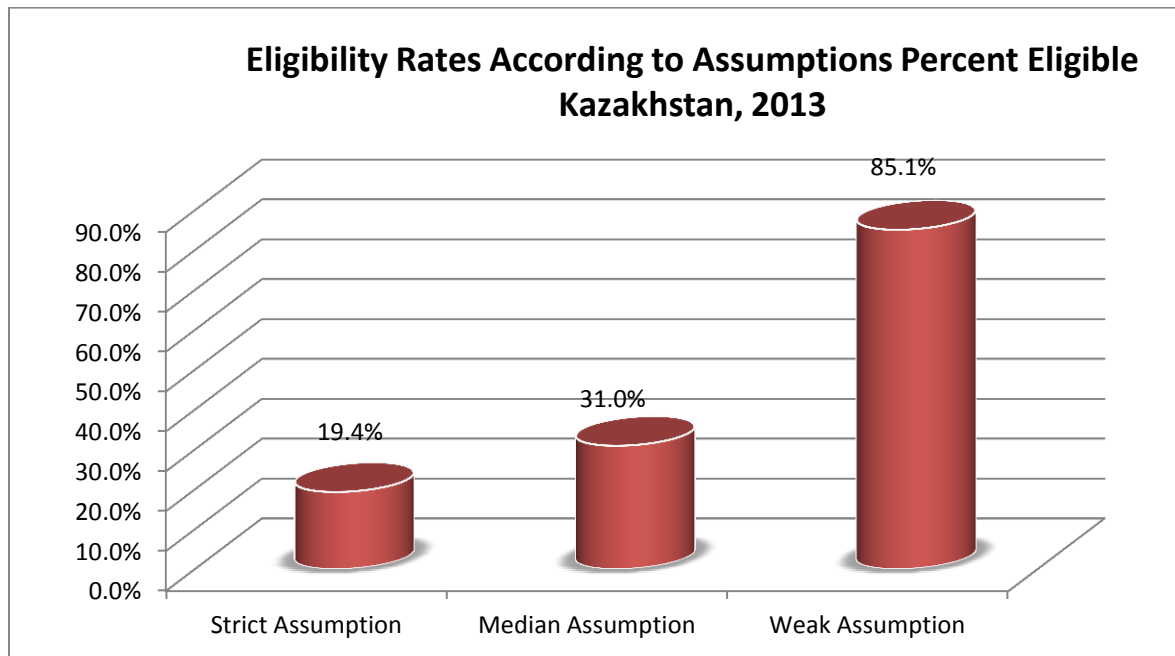
28. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

29. Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible. This definition includes as eligible 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. Under the weak assumption only observed non-eligible units are excluded from universe projections. The resulting weights are included in the variable *wweak*.

Weak eligibility = (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

30. The indicators computed for the Enterprise Survey website use the median weights. The following graph shows the different eligibility rates calculated for firms in the sample frame under each set of assumptions.



31. Universe estimates for the number of establishments in each industry-region-size cell in Kazakhstan were produced for the strict, weak and median eligibility definitions. Appendix D shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

32. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

VI. Weights

33. 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

34. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each region/industry/size stratum to

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

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, no tone on the phone line, answering machine, or fax line⁶, 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.

35. Appendix C shows the cell weights for registered establishments in Kazakhstan.

VII. Appropriate use of the weights

36. 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.

37. 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 a 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.)⁷

38. 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.⁸ 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.

VIII. Non-response

39. 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

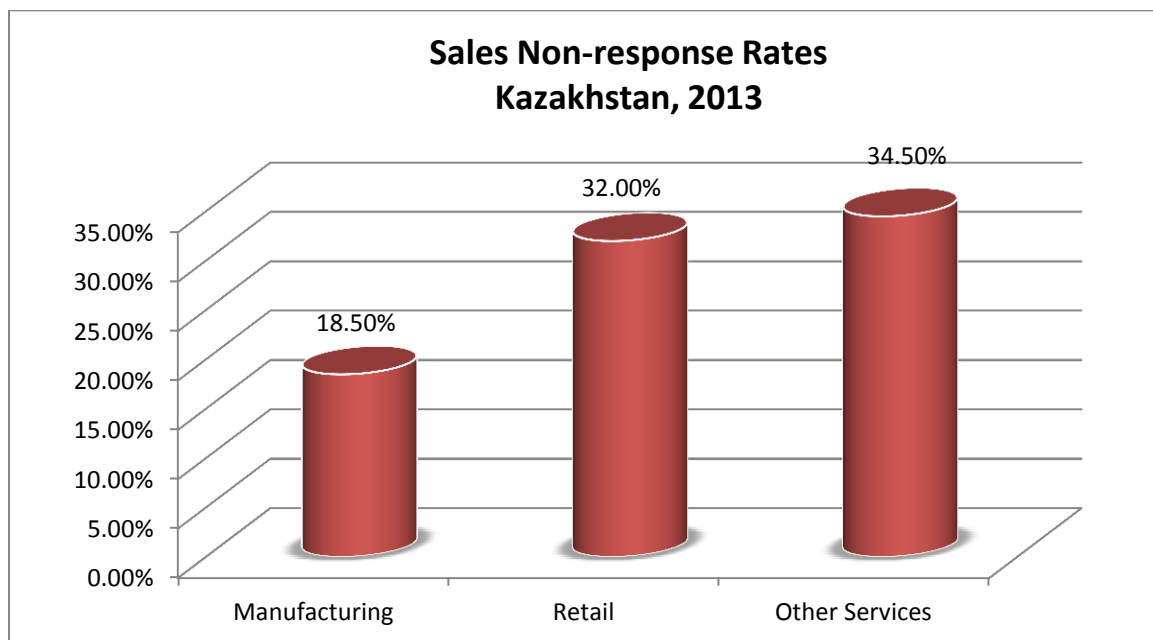
6 For the surveys that implemented a screener over the phone.

7 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.

8 The use of 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.

refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

40. 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 a different option from don't know (-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 sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the chart below reflects both categories (DKs and NAs).



41. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. 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.

42. As the following graph shows, the number of realized interviews per contacted establishment was 0.16⁹. 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

⁹ The estimate is based on the total number of firms contacted including ineligible establishments.

the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The number of rejections per contact was 0.14.



43. Details on the rejection rate, eligibility rate, and item non-response are available at the strata level. 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 Kazakhstan. 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

Status Codes Total:

ELIGIBLES	
1. Eligible establishment (Correct name and address)	728
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)	8
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	7
16. Panel firm - now less than five employees	1
5. The establishment has less than 5 permanent full time employees	18
6. The firm discontinued businesses	436
7. Not a business: private household	46
8. Ineligible activity: education, agriculture, finances, governments...	37
151. Out of target - outside the covered regions, firm moved abroad	29
152. Out of target - firm moved abroad	0
153. Impossible to find	3
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	507
92. Line out of order	141
93. No tone	59
94. Phone number does not exist	298
10. Answering machine	34
11. Fax line - data line	2
12. Wrong address/ moved away and could not get the new references	1074
13. Refuses to answer the screener	409
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	1
Total	3838

Response Outcomes Total:

Complete interviews (<i>Total</i>)	600
Incomplete interviews	0
Eligible in process	0
Refusals	144
Out of target	29
Impossible to contact	2115
Ineligible - coop.	29
Refusal to the Screener	409
Total	3835

Status Codes Fresh:

ELIGIBLES	
1. Eligible establishment (<i>Correct name and address</i>)	629
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>)	7
4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	5
16. Panel firm - now less than five employees	1
5. The establishment has less than 5 permanent full time employees	17
6. The firm discontinued businesses	401
7. Not a business: private household	44
8. Ineligible activity: education, agriculture, finances, governments...	36
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	467
92. Line out of order	134
93. No tone	57
94. Phone number does not exist	287
10. Answering machine	33
11. Fax line - data line	2
12. Wrong address/ moved away and could not get the new references	1039
13. Refuses to answer the screener	342
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	1
151. Out of target - outside the covered regions, firm moved abroad	28
152. Out of target - firm moved abroad	0
153. Impossible to find	3
Total	3537

Response Outcomes Fresh:

Complete interviews (<i>Total</i>)	517
Incomplete interviews	0
Eligible in process	0
Refusals	125
Out of target	31
Impossible to contact	2019
Ineligible - coop.	28
Refusal to the Screener	346
Total	3537

Status Codes Panel:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	99
	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>)	2
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	1
	6. The firm discontinued businesses	35
	7. Not a business: private household	2
	8. Ineligible activity: education, agriculture, finances, governments...	1
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	40
	92. Line out of order	7
	93. No tone	2
	94. Phone number does not exist	11
	10. Answering machine	1
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	35
	13. Refuses to answer the screener	63
	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	1
	152. Out of target - firm moved abroad	0
	153. Impossible to find	0
	Total	301

Response Outcomes Panel:

Complete interviews (<i>Total</i>)	83
Incomplete interviews	0
Eligible in process	0
Refusals	19
Out of target	1
Impossible to contact	96
Ineligible - coop.	1
Refusal to the Screener	63
Total	301

Appendix B

Sampling Frame, Kazakhstan:

Source: Kazakhstan Statistical Committee, 2011

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	571	226	2227	3024
	20-99	314	59	660	1033
	100+	102	14	146	262
	Total	987	299	3033	4319
East	5-19	565	292	2321	3178
	20-99	285	75	737	1097
	100+	109	13	150	272
	Total	959	380	3208	4547
North	5-19	1363	782	8053	10198
	20-99	770	310	3193	4273
	100+	241	43	493	777
	Total	2374	1135	11739	15248
South	5-19	2513	1306	13045	16864
	20-99	1186	380	4397	5963
	100+	382	95	963	1440
	Total	4081	1781	18405	24267
West	5-19	595	346	4435	5376
	20-99	293	100	1606	1999
	100+	121	20	375	516
	Total	1009	466	6416	7891
Grand Total		9410	4061	42801	56272

Appendix C

Kazakhstan, administrative divisions

Source: http://en.wikipedia.org/wiki/Provinces_of_Kazakhstan



Regions	Grouping used for stratification purposes in BEEPS V
Centre	Centre (Karagandinskaya)
East	East (Vostochno-Kazakhstanskaya)
North	North (Astana, Akmolinskaya, Severo-Kazakhstanskaya, Kostanayskaya, Pavlodarskaya)
South	South (Almaty, Kyzylordinskaya, Zhambylskaya, Yuzhno-Kazakhstanskaya, Almatinskaya)
West	West (Mangistauskaya, Atyrauskaya, Aktyubinskaya, Zapadno-Kazakhstanskaya)

Appendix D

Strict Cell Weights Kazakhstan – Panel

Region	Employees	Manufacturing	Retail	Other Services
Center	5-19	3.3		
	20-99	7.7		
	100+	5.4	2.6	6.8
East	5-19	1.0	1.0	
	20-99		1.6	1.9
	100+	4.4	1.0	
North	5-19	1.1	6.9	1.1
	20-99	2.6	6.6	1.6
	100+	7.0	4.5	1.8
South	5-19	3.5	1.0	
	20-99	2.7	1.6	1.9
	100+	5.1	1.5	2.4
West	5-19	1.2	3.7	1.8
	20-99	2.4	3.4	4.6
	100+	1.8	1.2	4.4

Strict Cell Weights Kazakhstan – Fresh

ok	Employees	Manufacturing	Retail	Other Services
Center	5-19	18.2	5.4	41.7
	20-99	18.3	4.7	74.1
	100+	9.9	6.1	
East	5-19	11.0	3.0	17.8
	20-99	5.9	5.7	100.0
	100+	3.4		21.8
North	5-19	18.4	4.7	29.2
	20-99	19.1	4.6	35.8
	100+	9.9	3.6	
South	5-19	12.9	6.1	33.9
	20-99	15.8	3.9	26.4
	100+	9.6	3.1	52.9
West	5-19	10.4	3.4	23.5
	20-99	9.6	3.6	89.8
	100+	6.9		71.6

Median Cell Weights Kazakhstan – Panel

Region	Employees	Manufacturing	Retail	Other Services
Center	5-19		4.0	
	20-99		10.2	
	100+		7.2	2.9 9.4
East	5-19		1.0	1.4
	20-99			2.9 4.5
	100+		10.0	1.9
North	5-19		1.6	8.6 1.7
	20-99		4.4	9.2 2.8
	100+		12.0	6.3 3.1
South	5-19		6.9	1.0
	20-99		5.9	2.9 4.2
	100+		11.3	2.8 5.6
West	5-19		1.3	3.4 2.1
	20-99		3.0	3.5 5.9
	100+		2.3	1.3 5.6

Median Cell Weights Kazakhstan – Fresh

Region	Employees	Manufacturing	Retail	Other Services
Center	5-19		20.7	6.5 53.1
	20-99		20.0	5.4 90.6
	100+		11.1	7.2
East	5-19		19.4	5.6 35.1
	20-99		10.0	10.1 189.8
	100+		5.9	42.4
North	5-19		27.2	7.2 48.1
	20-99		27.1	6.8 56.7
	100+		14.4	5.4
South	5-19		20.0	9.9 58.8
	20-99		23.5	6.2 44.0
	100+		14.7	5.0 90.6
West	5-19		18.6	6.3 47.0
	20-99		16.5	6.5 172.7
	100+		12.2	141.1

Weak Cell Weights Kazakhstan - Panel

Region	Employees	Manufacturing	Retail	Other Services	
Center	5-19		4.8		
	20-99		10.0		
	100+		7.0	2.9	11.7
East	5-19		1.2	2.6	
	20-99			4.2	7.9
	100+		13.9	2.7	
North	5-19		2.4	13.3	3.3
	20-99		5.3	11.6	4.3
	100+		14.4	7.9	4.8
South	5-19		12.1	1.8	
	20-99		8.3	4.3	7.6
	100+		15.8	4.1	9.9
West	5-19		2.2	5.9	4.4
	20-99		4.0	4.9	10.1
	100+		3.0	1.8	9.6

Weak Cell Weights Kazakhstan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	
Center	5-19		40.9	13.9	153.5
	20-99		28.6	8.4	190.2
	100+		13.2	9.4	
East	5-19		64.9	20.5	172.0
	20-99		24.3	26.8	674.8
	100+		12.0		125.9
North	5-19		76.5	22.1	198.8
	20-99		55.4	15.1	169.9
	100+		24.5	10.1	
South	5-19		56.7	30.5	244.2
	20-99		48.3	13.8	132.5
	100+		25.2	9.3	227.4
West	5-19		49.1	18.1	181.4
	20-99		31.6	13.6	484.1
	100+		19.5		329.8

Appendix E

Strict Universe Estimates Kazakhstan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	3	0	0	3
	20-99	8	0	0	8
	100+	5	3	7	15
	Total	16	3	7	26
East	5-19	5	3	0	8
	20-99	0	6	4	10
	100+	4	1	0	5
	Total	9	10	4	23
North	5-19	5	7	2	14
	20-99	8	13	6	27
	100+	7	4	4	15
	Total	20	25	12	57
South	5-19	4	9	0	13
	20-99	5	5	4	14
	100+	5	5	2	12
	Total	14	18	6	38
West	5-19	2	7	6	15
	20-99	5	7	9	21
	100+	5	2	4	12
	Total	12	17	19	48
Grand Total		72	72	48	193

Strict Universe Estimates Kazakhstan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	218	76	542	836
	20-99	164	24	222	410
	100+	59	6	0	65
	Total	442	106	764	1312
East	5-19	88	40	231	359
	20-99	59	11	100	171
	100+	24	0	22	46
	Total	171	51	353	575
North	5-19	295	149	1108	1552
	20-99	229	78	608	915
	100+	79	11	0	90
	Total	603	238	1716	2556
South	5-19	412	189	1355	1956
	20-99	268	75	633	975
	100+	96	19	159	273
	Total	776	282	2146	3204
West	5-19	114	57	540	711
	20-99	77	22	269	368
	100+	35	0	72	106
	Total	226	79	881	1186
Grand Total		2218	756	5860	8833

Median Universe Estimates Kazakhstan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	4	0	0	4
	20-99	10	0	0	10
	100+	7	3	9	19
	Total	21	3	9	34
East	5-19	5	4	0	9
	20-99	0	12	9	21
	100+	10	2	0	12
	Total	15	18	9	42
North	5-19	8	9	3	20
	20-99	13	18	11	43
	100+	12	6	6	24
	Total	33	33	21	87
South	5-19	7	9	0	16
	20-99	12	9	8	29
	100+	11	8	6	25
	Total	30	26	14	70
West	5-19	3	7	6	16
	20-99	6	7	12	25
	100+	7	3	6	15
	Total	15	16	24	55
Grand Total		115	96	77	289

Median Universe Estimates Kazakhstan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	248	90	690	1029
	20-99	180	27	272	479
	100+	66	7	0	74
	Total	494	125	962	1581
East	5-19	155	73	456	685
	20-99	100	20	190	310
	100+	42	0	42	84
	Total	297	93	689	1079
North	5-19	434	230	1830	2494
	20-99	325	115	964	1404
	100+	115	16	0	131
	Total	874	361	2794	4029
South	5-19	640	306	2354	3300
	20-99	399	117	1056	1573
	100+	147	30	272	449
	Total	1186	453	3682	5321
West	5-19	205	107	1080	1392
	20-99	132	39	518	689
	100+	61	0	141	202
	Total	398	146	1739	2283
Grand Total		3250	1178	9865	14293

Weak Universe Estimates Kazakhstan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	5	0	0	5
	20-99	10	0	0	10
	100+	7	3	12	22
	Total	22	3	12	36
East	5-19	6	8	0	14
	20-99	0	17	16	33
	100+	14	3	0	17
	Total	20	27	16	63
North	5-19	12	13	7	32
	20-99	16	23	17	56
	100+	14	8	10	32
	Total	42	44	33	120
South	5-19	12	16	0	28
	20-99	17	13	15	45
	100+	16	12	10	38
	Total	45	41	25	111
West	5-19	4	12	13	29
	20-99	8	10	20	38
	100+	9	4	10	22
	Total	21	25	43	89
Grand Total		150	141	129	420

Weak Universe Estimates Kazakhstan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	490	195	1996	2681
	20-99	257	42	571	870
	100+	79	9	0	89
	Total	827	246	2566	3639
East	5-19	519	266	2236	3022
	20-99	243	54	675	971
	100+	84	0	126	210
	Total	846	320	3037	4203
North	5-19	1225	706	7555	9485
	20-99	664	257	2889	3810
	100+	196	30	0	226
	Total	2085	993	10444	13521
South	5-19	1814	945	9770	12528
	20-99	821	262	3181	4263
	100+	252	56	682	990
	Total	2886	1263	13633	17782
West	5-19	540	307	4172	5019
	20-99	253	81	1452	1787
	100+	97	0	330	427
	Total	890	389	5954	7233
Grand Total		7534	3210	35634	46379

Appendix F

Original Sample Design, Kazakhstan:

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Center	5-19	13	14	13	40
	20-99	10	5	3	18
	100+	7	2	1	10
	Total	30	21	17	68
East	5-19	13	16	13	42
	20-99	10	6	3	19
	100+	8	1	1	10
	Total	31	23	17	71
North	5-19	21	33	40	94
	20-99	15	19	21	55
	100+	9	4	2	15
	Total	45	56	63	164
South	5-19	33	40	40	113
	20-99	19	22	26	67
	100+	11	9	4	24
	Total	63	71	70	204
West	5-19	13	19	26	58
	20-99	10	8	5	23
	100+	8	2	2	12
	Total	31	29	33	93
Grand Total		200	200	200	600