

The Kazakhstan 2019 Enterprise Surveys Data Set

I. Introduction

This document provides additional information on the data collected in Kazakhstan between January and October of 2019. The survey was part of a joint project of the European Bank for Reconstruction and Development (EBRD), the European Investment Bank (EIB) and the World Bank Group (WBG). The objective of the Enterprise Survey is to gain an understanding of what firms experience in the private sector.

As part of its strategic goal of building a climate for investment, job creation, and sustainable growth, the World Bank has promoted improving the business environment as a key strategy for development, which has led to a systematic effort in collecting enterprise data across countries. The Enterprise Surveys (ES) are an ongoing World Bank project in collecting both objective data based on firms' experiences and enterprises' perception of the environment in which they operate.

The ES currently cover over 185,000 firms in 151 countries, of which 143 have been surveyed following the standard methodology. This allows for better comparisons across countries and across time. Data are used to create statistically significant business environment indicators that are comparable across countries. The ES 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.

This 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

The sample for 2019 Kazakhstan ES was selected using stratified random sampling, following the methodology explained in the *Sampling Note*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-

¹ The complete text can be found at http://www.enterprisesurveys.org/~media/GIAWB/EnterpriseSurveys/Documents/Methodology/Sampling_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

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.

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 C.

Industry stratification was designed in the way that follows: the universe was stratified into six manufacturing industries and two services industries: Food and Beverages (ISIC Rev. 4 codes 10 and 11), Garments (ISIC code 14), Non-Metallic Mineral Products (ISIC code 23), Fabricated Metal Products (ISIC code 25), Machinery and Equipment (ISIC code 28), Other Manufacturing (ISIC codes 12, 13, 15-22, 24, 26, 27, 29, 30-33), Retail (ISIC code 47), and Other Services (ISIC codes 41-43, 45, 46, 49-53, 55, 56, 58, 61, 62, 79, 95).

For the Kazakhstan ES, size stratification was defined as follows: small (5 to 19 employees), medium (20 to 99 employees), and large (100 or more employees).

Regional stratification for the Kazakhstan ES was done across eleven regions: Akmola Region; Aktobe Region; Almaty; Almaty Region; Nur-Sultan; Atyrau Region; Mangystau and West Kazakhstan; East Kazakhstan; Karaganda Region; Kostanay, North Kazakhstan, Pavlodar and Kyzylorda Region, South Kazakhstan, Jambyl.

III. Sampling implementation

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.

Ipsos, the main contractor, in collaboration with Ipsos-Kazakhstan implemented the Kazakhstan 2019 ES.

The sample frame consisted of listings of establishments from two sources: for panel firms, the list of 600 firms from the Kazakhstan 2013 ES was used; and for fresh firms (i.e., firms not covered in 2013), a listing of establishments from the National Committee on Statistics, was used. The establishments in the listing are all registered as businesses with the Public Registration Center.

Table 1: Kazakhstan ES Sample Frame (Fresh and Panel Combined)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	292	43	127	126	43	745	1702	10083	14692
	Medium (20-99)	104	14	22	22	19	101	120	789	
	Large (100 or more)	40	1	17	7	11	52	44	168	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	321	53	227	91	25	775	1885	11176	16278
	Medium (20-99)	88	11	60	19	8	152	129	946	
	Large (100 or more)	26	7	29	5	1	56	36	152	
Akmola Region	Small (5-19)	123	10	70	17	5	193	429	2576	3919
	Medium (20-99)	45	3	20	1	6	32	31	272	
	Large (100 or more)	8	0	4	1	4	19	6	44	
Aktobe Region	Small (5-19)	59	12	61	36	9	235	702	5288	6959
	Medium (20-99)	20	3	20	7	4	32	40	307	
	Large (100 or more)	11	0	13	1	1	17	16	65	
Almaty	Small (5-19)	344	149	265	251	76	1760	4934	25692	37449
	Medium (20-99)	65	31	41	45	9	275	321	2572	
	Large (100 or more)	41	1	18	10	2	56	85	406	
Almaty Region	Small (5-19)	153	22	152	55	13	288	748	3301	5597
	Medium (20-99)	61	6	47	19	4	62	52	455	
	Large (100 or more)	27	4	7	7	1	29	7	77	
Nur-Sultan	Small (5-19)	163	60	254	128	29	781	3457	18650	25902
	Medium (20-99)	25	14	38	29	4	95	158	1722	
	Large (100 or more)	8	0	10	4	1	15	36	221	
Atyrau Region	Small (5-19)	34	10	21	19	7	122	574	3158	4583
	Medium (20-99)	6	3	8	2	3	28	27	438	
	Large (100 or more)	0	0	3	1	1	7	12	99	
Mangystau and West Kazakhstan	Small (5-19)	92	22	88	51	19	346	906	6763	9243
	Medium (20-99)	29	4	24	7	6	59	72	561	
	Large (100 or more)	11	1	13	2	7	22	19	119	
East Kazakhstan	Small (5-19)	139	26	61	54	21	359	950	4437	6998
	Medium (20-99)	54	5	15	15	8	82	64	503	
	Large (100 or more)	23	2	11	8	5	35	16	105	
Karaganda Region	Small (5-19)	182	57	164	140	34	722	900	8663	11889
	Medium (20-99)	38	11	21	27	11	97	62	541	
	Large (100 or more)	17	2	9	4	6	53	23	105	
		2649	587	1940	1211	403	7702	18563	110454	143509

Source: World Bank and the National Committee on Statistics

Table 2: Kazakhstan Sample Frame (Panel)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	2	1	0	2	2	4	19	16	96
	Medium (20-99)	4	1	0	0	5	2	13	15	
	Large (100 or more)	3	0	1	0	1	1	3	1	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	6	1	3	0	0	6	30	36	141
	Medium (20-99)	5	1	1	0	1	4	12	19	
	Large (100 or more)	1	0	3	1	0	1	8	2	
Akmola Region	Small (5-19)	0	0	2	0	0	2	4	3	28
	Medium (20-99)	1	1	0	0	1	1	4	7	
	Large (100 or more)	0	0	0	0	1	0	0	1	
Aktobe Region	Small (5-19)	0	0	0	1	0	5	7	10	52
	Medium (20-99)	1	1	1	1	2	1	7	5	
	Large (100 or more)	2	0	3	0	1	1	3	0	
Almaty	Small (5-19)	1	0	1	1	1	6	6	11	54
	Medium (20-99)	0	2	0	1	0	4	8	6	
	Large (100 or more)	0	0	1	1	0	0	3	1	
Almaty Region	Small (5-19)	1	0	0	0	0	1	1	3	9
	Medium (20-99)	0	0	1	1	0	0	1	0	
Nur-Sultan	Small (5-19)	0	0	1	0	0	0	7	9	40
	Medium (20-99)	0	1	0	0	0	3	1	11	
	Large (100 or more)	1	0	0	0	0	1	2	3	
Atyrau Region	Small (5-19)	0	0	1	0	0	2	0	6	13
	Medium (20-99)	0	0	0	0	1	0	0	2	
	Large (100 or more)	0	0	0	0	0	0	0	1	
Mangystau and West Kazakhstan	Small (5-19)	1	0	0	0	0	0	6	7	29
	Medium (20-99)	1	0	1	0	0	2	4	3	
	Large (100 or more)	0	0	1	0	1	0	1	1	
East Kazakhstan	Small (5-19)	0	0	1	0	0	5	12	16	71
	Medium (20-99)	2	0	0	1	3	8	9	6	
	Large (100 or more)	1	0	1	3	0	1	0	2	
Karaganda Region	Small (5-19)	2	1	2	1	1	8	11	8	67
	Medium (20-99)	1	0	3	0	0	3	7	9	
	Large (100 or more)	1	0	2	0	1	3	2	1	
		37	10	30	14	22	75	191	221	600

Necessary measures were taken to ensure the quality of the frame; however, the sample frame was not immune to the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc.

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 40.5% (4674 out of 11530 establishments)⁴.

Breaking down by industry and size, the following sample targets were achieved (based on the sampling information):

Table 3: Achieved Interviews (Fresh and Panel Combined)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar										
	Small (5-19)	11	4	9	8	6	5	4	9	120
	Medium (20-99)	6	0	2	3	1	13	6	5	
	Large (100 or more)	2	0	1	3	1	7	5	6	
	Medium and Large (20+)	0	3	0	0	0	0	0	0	
Kyzylorda Region, South Kazakhstan, Jambyl										
	Small (5-19)	6	10	6	5	10	4	17	8	167
	Medium (20-99)	14	0	15	6	0	3	4	12	
	Large (100 or more)	3	0	9	1	0	19	7	2	
	Medium and Large (20+)	0	4	0	0	2	0	0	0	
Akmola Region										
	Small (5-19)	0	2	5	0	0	24	14	31	117
	Medium (20-99)	0	1	2	0	0	2	0	26	
	Large (100 or more)	0	0	1	0	0	1	0	2	
	Medium and Large (20+)	0	0	0	0	0	0	2	0	
	Small, Medium and Large (5+)	2	0	0	1	1	0	0	0	
Aktobe Region										
	Small (5-19)	5	4	9	13	2	6	7	5	120
	Medium (20-99)	7	1	1	0	2	5	8	6	
	Large (100 or more)	6	0	6	0	1	6	6	11	
	Medium and Large (20+)	0	0	0	3	0	0	0	0	
Almaty										
	Small (5-19)	14	19	11	20	12	10	4	18	168

⁴ Based on out of target and ineligible contacts

	Medium (20-99)	0	0	7	7	3	6	3	2	
	Large (100 or more)	0	0	1	1	1	3	2	2	
	Medium and Large (20+)	20	2	0	0	0	0	0	0	
Almaty Region	Small (5-19)	18	0	11	4	0	14	3	2	118
	Medium (20-99)	6	0	7	6	0	19	2	2	
	Large (100 or more)	3	0	1	2	0	6	2	5	
	Small, Medium and Large (5+)	0	3	0	0	2	0	0	0	
Nur-Sultan	Small (5-19)	11	4	14	19	0	12	3	34	159
	Medium (20-99)	0	3	6	0	0	15	2	4	
	Large (100 or more)	0	0	1	0	0	3	5	10	
	Medium and Large (20+)	6	0	0	6	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	1	0	0	0	
Atyrau Region	Small (5-19)	3	0	4	3	0	8	9	33	118
	Medium (20-99)	1	0	2	0	0	3	3	29	
	Large (100 or more)	0	0	1	0	0	1	4	11	
	Medium and Large (20+)	0	0	0	1	0	0	0	0	
	Small, Medium and Large (5+)	0	1	0	0	1	0	0	0	
Mangystau and West Kazakhstan	Small (5-19)	7	3	9	14	4	9	8	6	119
	Medium (20-99)	0	0	1	1	0	11	4	7	
	Large (100 or more)	0	0	1	1	0	1	4	15	
	Medium and Large (20+)	11	1	0	0	1	0	0	0	
East Kazakhstan	Small (5-19)	10	0	10	13	3	3	6	6	119
	Medium (20-99)	13	0	4	4	2	8	7	4	
	Large (100 or more)	4	0	2	4	1	2	2	4	
	Small, Medium and Large (5+)	0	7	0	0	0	0	0	0	
Karaganda Region	Small (5-19)	11	2	16	16	0	4	6	5	121
	Medium (20-99)	17	0	0	0	0	10	5	2	
	Large (100 or more)	3	0	0	0	0	1	5	5	
	Medium and Large (20+)	0	2	3	3	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	5	0	0	0	
		220	76	178	168	62	244	169	329	1446

Table 4: Achieved Interviews (Panel)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North										
Kazakhstan, Pavlodar	Small (5-19)	1	0	0	0	0	0	2	4	15
	Medium (20-99)	0	0	0	0	1	1	2	1	
	Large (100 or more)	0	0	0	0	0	1	2	0	
Kyzylorda Region, South										
Kazakhstan, Jambyl	Small (5-19)	4	0	3	0	0	2	8	4	47
	Medium (20-99)	2	0	1	0	0	1	2	9	
	Large (100 or more)	1	0	2	1	0	1	5	0	
	Medium and Large (20+)	0	1	0	0	0	0	0	0	
Akmola Region										
	Small (5-19)	0	0	0	0	0	0	0	1	3
	Medium (20-99)	0	0	0	0	0	1	0	0	
	Medium and Large (20+)	0	0	0	0	0	0	1	0	
Aktobe Region										
	Small (5-19)	0	0	0	1	0	1	3	3	20
	Medium (20-99)	0	0	0	0	0	0	4	0	
	Large (100 or more)	1	0	2	0	1	0	3	0	
	Medium and Large (20+)	0	0	0	1	0	0	0	0	
Almaty										
	Small (5-19)	0	0	0	0	1	0	2	0	4
	Medium (20-99)	0	0	0	0	0	0	1	0	
Almaty Region										
	Medium (20-99)	0	0	0	1	0	0	0	0	1
Nur-Sultan										
	Small (5-19)	0	0	0	0	0	0	1	4	9
	Medium (20-99)	0	1	0	0	0	0	0	2	
	Large (100 or more)	0	0	0	0	0	0	1	0	
Atyrau Region										
	Small (5-19)	0	0	1	0	0	1	0	6	10
	Medium (20-99)	0	0	0	0	0	0	0	1	
	Small, Medium and Large (5+)	0	0	0	0	1	0	0	0	
Mangystau and West										
Kazakhstan	Small (5-19)	0	0	0	0	0	0	4	3	11
	Medium (20-99)	0	0	0	0	0	1	1	0	
	Large (100 or more)	0	0	0	0	0	0	1	1	
East Kazakhstan										
	Small (5-19)	0	0	0	0	0	0	3	3	15
	Medium (20-99)	0	0	0	0	0	3	2	1	
	Large (100 or more)	0	0	0	2	0	1	0	0	
Karaganda Region										
	Small (5-19)	0	0	0	0	0	1	2	0	5
	Medium (20-99)	0	0	0	0	0	0	1	0	
	Medium and Large (20+)	0	0	1	0	0	0	0	0	
		9	2	10	6	4	15	51	43	140

IV. Data Base Structure:

The structure of the data base reflects the fact that 2 different versions of the survey instrument were used for all registered establishments. Questionnaires have common questions (*core* module) and respectfully additional manufacturing- and services-specific questions. The eligible manufacturing industries have been surveyed using the *Manufacturing* questionnaire (includes the *core* module, plus manufacturing specific questions). Retail firms have been interviewed using the *Services* questionnaire (includes the *core* module plus retail specific questions) and the residual eligible services have been covered using the *Services* questionnaire (includes the *core* module). Each variation of the questionnaire is identified by the index variable, *a0*.

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 (some exceptions apply due to comparability reasons). Variable names preceded by the prefix prefix “BM” or “BMG” indicate questions specific to Kazakhstan and other countries in Europe and Central Asia 2018/2019 and Middle East and North Africa 2019, 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.

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.

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 based on the sample frame, whereas the latter gives the establishment’s actual industry classification (four-digit code) based on the main activity at the time of the survey.

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 or outdated 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.

-*a4a*: coded following the stratification by sector as defined above.

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. However, sometimes the phone numbers were unavailable in the sample frame, and thus the

enumerators applied the screeners in person. Interviews were conducted using Computer-assisted personal interviewing (CAPI) in Kazakhstan. The variables *a4b* and *a6c* contain the industry and size of the establishment from the screener questionnaire.

Note that there are variables for size (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes. Variables *l1* (number of permanent full-time workers at the end of the last complete fiscal year), *l6* (number of full-time seasonal workers employed during last complete fiscal year) and *l8* (average length of employment of full-time temporary employees during last complete fiscal year) 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.

The firms interviewed had several fiscal years. Most firms had January to December 2018 as their last complete fiscal year. Variables *a20m* (starting month of last complete fiscal year) and *a20y* (last complete fiscal year) can be used to obtain the last complete fiscal year for each firm.

For questions pertaining to monetary amounts, the unit is the Kazakhstani Tenge (KZT).

V. Universe Estimates

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

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.

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

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*.

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

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*.

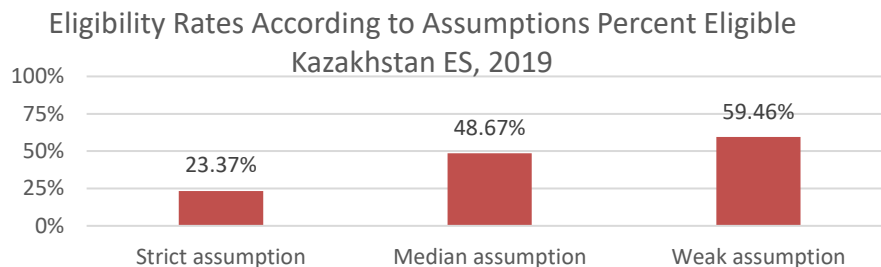
$$\text{Median eligibility} = (\text{Sum of the firms with codes } 1,2,3,4,16,10,11, \& 13) / \text{Total}$$

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*.

$$\text{Weak eligibility} = (\text{Sum of the firms with codes, } 1,2,3,4,16,10,11,13,91,92,93,94,12) / \text{Total}$$

The indicators computed for the ES 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.



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 B shows the universe estimates of the numbers of registered establishments that fit the criteria of the ES.

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

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

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 account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, 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

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

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

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.

Due to non-response rates, some stratification cells were collapsed for the purposes of weighting. The following cells have been transformed as follows:

- (i) Medium and large firms are treated as one cell in:
 - Akmola for Retail
 - Aktobe for Fabricated Metal Products
 - Almaty for Food
 - Almaty for Garments
 - Nur-Sultan for Food
 - Nur-Sultan for Fabricated Metal Products
 - Atyrau for Fabricated Metal Products
 - Mangystau for Food
 - Mangystau for Garments
 - Mangystau for Machinery & Equipment
 - Karaganda for Non-Metallic Mineral Products
 - Karaganda for Fabricated Metal Products
 - Kostanay, North Kazakhstan, Pavlodar for Garments
 - Kyzylorda Region, South Kazakhstan, Jambyl for Large Machinery & Equipment, Pavlodar
 - Kyzylorda Region, South Kazakhstan, Jambyl for Garments
 - Karaganda Region for Garments.
- (ii) Small, medium, and large firms are treated as one cell in:
 - Akmola for Food
 - Akmola for Machinery & Equipment
 - Akmola for Fabricated Metal Products
 - Almaty for Garments
 - Almaty for Machinery & Equipment
 - Nur-Sultan for Machinery & Equipment
 - Atyrau Region for Machinery & Equipment
 - East Kazakhstan for Garments
 - Atyrau Region for Garments
 - Karaganda Region for Machinery & Equipment.

VII. Appropriate use of the weights

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.

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 have the advantage of providing an estimate that is independent of the sample design. This latter

point may be quite relevant for the ES 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

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.

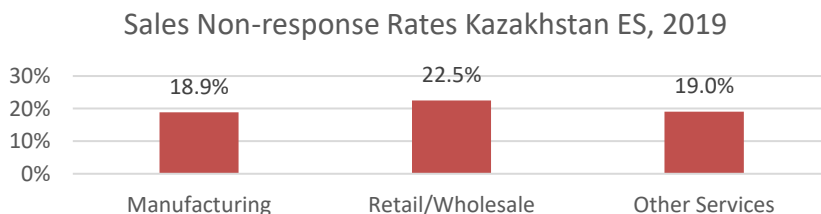
VIII. Non-response

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.

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 (-8) as a different option from don't know (-9).

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 for this specific question, refusals were not separately identified from "Don't know" responses.

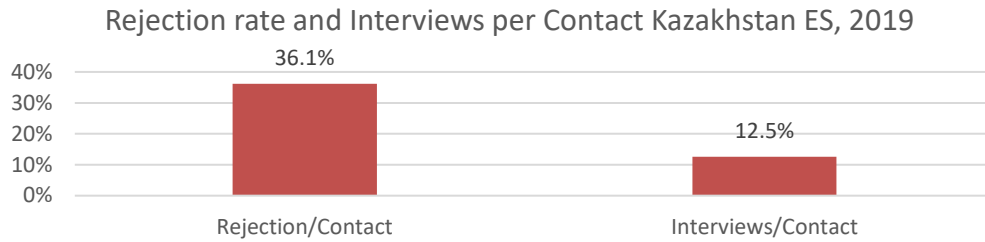


As the following graph shows, the number of interviews per contacted establishments was 0.13.⁹ This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The share of rejections per contact was 0.36.

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 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.

⁹ The estimate is based on the total no. of firms contacted including ineligible establishments.



Details on the rejection rate, eligibility rate, 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 Kazakhstan. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

References:

Cochran, William G., *Sampling Techniques*, New York, New York: John Wiley & Sons, 1977.

Deaton, Angus, *The Analysis of Household Surveys*, Baltimore, Maryland: Johns Hopkins University Press, 1998.

Levy, Paul S. and Stanley Lemeshow, *Sampling of Populations: Methods and Applications*, New York, New York: John Wiley & Sons, 1999.

Lohr, Sharon L. *Sampling: Design and Techniques*, Boston, Massachusetts: Brookes/Cole, 1999.

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

Appendix A

Status Codes Enterprise Survey (ES) :

0	Screening in process	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
2695	Eligible	1. Eligible establishment (Correct name and address) 2461 2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment) 39 3. Eligible establishment (Different name but same address - the firm/establishment changed its name) 40 4. Eligible establishment (Moved and traced) 147 16. Eligible establishment (Panel Firm - now less than five employees; this code applies only to panel firms.) 8	
2917	Screener refusal	13. Refuses to answer the screener	2917
3810	Ineligible	5. The establishment has less than 5 permanent full time employees 718 616. The firm discontinued businesses - (Establishment went bankrupt) 131 618. The firm discontinued businesses - (Original establishment disappeared and is now a different firm) 128 619. The firm discontinued businesses - (Establishment was bought out by another firm) 95 620. The firm discontinued businesses - (It was impossible to determine for what reason) 1468 621. The firm discontinued businesses - (Other) 929 71. Ineligible legal status: not a business, but private household 78 72. Ineligible legal status: cooperatives, non-profit organizations, etc. 49 8. Ineligible activity: Education, Agriculture, Finances, Government, etc. 214	
864	Out of Target	151. Out of target - outside the covered regions 127 152. Out of target - moved abroad 0 153. Out of target - Not registered with Statistical Authority 11 154. Out of target - establishment is HQ without production or sales of goods or services 6 155. Out of target - establishment was not in operation for the entirety of last fiscal year 636 156. Duplicated firm within the sample 7 157. Out of target - location that is not HQ and does not have financial statements prepared separately 77	
1244	Unobtainable	91. No reply after having called in different days of the week and in different business hours 184 92. Line out of order 2 93. No tone 0 94. Phone number does not exist 0 10. Answering machine 0 11. Fax line- data line 0 12. Wrong address/ moved away and could not get the new references 1058	
11530	Total contacted		

Response Outcomes : Kazakhstan ES 2019 :

Target and totals	Sample target	1440
	Sample target completion rate	100.4%
	Total contacts available in frame	143509
	Total contacts issued	11779
	Total contacts contacted	11530

Screening phase	Screening in process	0
	Eligibles	2695
	Screener refusal	2917
	Ineligible + out of target	4674
	Unobtainable	1244
Interview phase (only if eligible)	Complete interviews without extra module	35
	Complete interviews with extra module	1411
	Eligible in process + incomplete interviews	0
	Interview refusal	1249

Percent breakdown (relative to total contacted)	Screening in process rate	0.0%
	Screener refusal rate	25.3%
	Ineligible + out of target rate	40.5%
	Unobtainable rate	10.8%
	Interview conversion rate	12.5%
	Eligible in process + incomplete interviews rate	0.0%
	Interview refusal rate	10.8%

Appendix B: Universe Estimate Based on Sampling Weights

Strict Universe Estimates – Fresh:

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	50	7	22	24	8	133	298	1092	1945
	Medium (20-99)	27	0	6	6	3	27	31	128	
	Large (100 or more)	11	0	5	3	3	15	12	29	
	Medium and Large (20+)	0	5	0	0	0	0	0	0	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	102	16	72	32	10	257	613	2248	3998
	Medium (20-99)	42	0	29	10	0	75	63	285	
	Large (100 or more)	13	0	15	1	0	30	20	49	
	Medium and Large (20+)	0	10	0	0	6	0	0	0	
Akmola Region	Small (5-19)	0	2	10	0	0	29	63	236	461
	Medium (20-99)	0	1	4	0	0	7	0	37	
	Large (100 or more)	0	0	1	0	0	5	0	7	
	Medium and Large (20+)	0	0	0	0	0	0	10	0	
	Small, Medium and Large (5+)	40	0	0	5	4	0	0	0	
Aktobe Region	Small (5-19)	25	5	26	17	4	104	306	1425	2190
	Medium (20-99)	13	2	13	0	3	21	26	124	
	Large (100 or more)	8	0	9	0	1	12	11	28	
	Medium and Large (20+)	0	0	0	7	0	0	0	0	
Almaty	Small (5-19)	61	26	47	50	15	328	900	2901	5124
	Medium (20-99)	0	0	11	13	3	77	88	434	
	Large (100 or more)	0	0	5	3	1	17	25	74	
	Medium and Large (20+)	35	10	0	0	0	0	0	0	
Almaty Region	Small (5-19)	25	0	25	10	0	49	124	339	753
	Medium (20-99)	15	0	11	6	0	19	13	70	
	Large (100 or more)	7	0	2	2	0	8	2	13	
	Small, Medium and Large (5+)	0	8	0	0	5	0	0	0	
Nur-Sultan	Small (5-19)	36	13	56	31	0	179	775	2586	4247
	Medium (20-99)	0	4	12	0	0	32	53	357	
	Large (100 or more)	0	0	4	0	0	6	13	50	
	Medium and Large (20+)	13	0	0	15	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	13	0	0	0	
Atyrau Region	Small (5-19)	5	0	4	3	0	19	87	295	518
	Medium (20-99)	1	0	2	0	0	6	6	61	
	Large (100 or more)	0	0	1	0	0	2	4	15	
	Medium and Large (20+)	0	0	0	1	0	0	0	0	
	Small, Medium and Large (5+)	0	3	0	0	2	0	0	0	

Mangystau and West Kazakhstan	Small (5-19)	16	4	15	14	4	62	160	738	1204
	Medium (20-99)	0	0	6	2	0	16	19	92	
	Large (100 or more)	0	0	4	1	0	6	5	21	
	Medium and Large (20+)	13	2	0	0	5	0	0	0	
East Kazakhstan	Small (5-19)	42	0	19	18	7	114	296	856	1707
	Medium (20-99)	25	0	7	8	4	39	30	145	
	Large (100 or more)	11	0	5	4	3	18	8	33	
	Small, Medium and Large (5+)	0	15	0	0	0	0	0	0	
Karaganda Region	Small (5-19)	54	16	49	46	0	224	273	1628	2679
	Medium (20-99)	17	0	0	0	0	45	28	152	
	Large (100 or more)	8	0	0	0	0	27	11	32	
	Medium and Large (20+)	0	7	16	19	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	26	0	0	0	
		713	156	512	353	129	2008	4376	16580	24827

Median Universe Estimates – Fresh:

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	98	16	42	51	14	280	559	2751	4479
	Medium (20-99)	49	0	10	13	8	54	56	304	
	Large (100 or more)	21	0	9	4	6	31	23	73	
	Medium and Large (20+)	0	9	0	0	0	0	0	0	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	168	30	119	58	13	455	968	4766	7748
	Medium (20-99)	65	0	44	17	0	126	93	569	
	Large (100 or more)	21	0	24	2	0	52	29	103	
	Medium and Large (20+)	0	18	0	0	8	0	0	0	
Akmola Region	Small (5-19)	0	4	29	0	0	88	172	857	1496
	Medium (20-99)	0	2	11	0	0	21	0	128	
	Large (100 or more)	0	0	3	0	0	14	0	23	
	Medium and Large (20+)	0	0	0	0	0	0	26	0	
	Small, Medium and Large (5+)	98	0	0	13	8	0	0	0	
Aktobe Region	Small (5-19)	39	9	40	29	6	175	457	2858	4090
	Medium (20-99)	19	3	19	0	4	34	37	234	
	Large (100 or more)	12	0	14	0	1	20	16	56	
	Medium and Large (20+)	0	0	0	11	0	0	0	0	
Almaty	Small (5-19)	170	81	131	150	37	979	2397	10371	16808
	Medium (20-99)	0	0	29	38	6	216	220	1464	
	Large (100 or more)	0	0	14	9	2	49	65	260	
	Medium and Large (20+)	90	30	0	0	0	0	0	0	

Almaty Region	Small (5-19)	79	0	78	34	0	166	378	1385	2726
	Medium (20-99)	44	0	34	17	0	51	37	269	
	Large (100 or more)	22	0	6	7	0	27	6	51	
	Small, Medium and Large (5+)	0	25	0	0	13	0	0	0	
Nur-Sultan	Small (5-19)	56	23	87	53	0	302	1170	5242	7961
	Medium (20-99)	0	7	18	0	0	52	75	683	
	Large (100 or more)	0	0	5	0	0	9	19	98	
	Medium and Large (20+)	20	0	0	24	0	0	0	0	
Atyrau Region	Small, Medium and Large (5+)	0	0	0	0	16	0	0	0	953
	Small (5-19)	8	0	5	5	0	31	129	589	
	Medium (20-99)	2	0	3	0	0	10	9	115	
	Large (100 or more)	0	0	1	0	0	3	4	29	
	Medium and Large (20+)	0	0	0	1	0	0	0	0	
Mangystau and West Kazakhstan	Small, Medium and Large (5+)	0	4	0	0	4	0	0	0	3467
	Small (5-19)	38	10	37	26	8	162	371	2302	
	Medium (20-99)	0	0	14	5	0	39	42	269	
	Large (100 or more)	0	0	9	2	0	16	12	64	
East Kazakhstan	Medium and Large (20+)	29	4	0	0	9	0	0	0	2738
	Small (5-19)	58	0	26	27	9	169	392	1519	
	Medium (20-99)	32	0	9	11	5	55	37	243	
	Large (100 or more)	15	0	7	6	3	26	10	57	
Karaganda Region	Small, Medium and Large (5+)	0	21	0	0	0	0	0	0	5748
	Small (5-19)	97	33	88	90	0	434	472	3776	
	Medium (20-99)	29	0	0	0	0	82	46	333	
	Large (100 or more)	14	0	0	0	0	50	19	72	
	Medium and Large (20+)	0	13	28	35	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	37	0	0	0	
		1392	343	991	738	216	4278	8345	41911	58216

Weak Universe Estimates – Fresh:

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Kostanay, North Kazakhstan, Pavlodar										
	Small (5-19)	154	27	65	77	23	409	872	4772	7435
	Medium (20-99)	74	0	15	18	10	75	83	506	
	Large (100 or more)	30	0	12	6	8	40	32	111	
	Medium and Large (20+)	0	14	0	0	0	0	0	0	
Kyzylorda Region, South Kazakhstan, Jambyl										
	Small (5-19)	196	39	133	64	16	493	1119	6123	9533
	Medium (20-99)	73	0	48	18	0	131	104	703	
	Large (100 or more)	22	0	24	3	0	50	31	117	

Akmola Region	Medium and Large (20+)	0	20	0	0	9	0	0	0	1763
	Small (5-19)	0	6	31	0	0	93	193	1067	
	Medium (20-99)	0	2	12	0	0	21	0	153	
	Large (100 or more)	0	0	2	0	0	13	0	26	
Aktobe Region	Medium and Large (20+)	0	0	0	0	0	0	25	0	5047
	Small, Medium and Large (5+)	99	0	0	12	9	0	0	0	
	Small (5-19)	45	11	45	32	7	188	524	3647	
	Medium (20-99)	21	4	20	0	4	35	41	287	
Almaty	Large (100 or more)	12	0	13	0	1	19	17	63	21321
	Medium and Large (20+)	0	0	0	11	0	0	0	0	
	Small (5-19)	204	106	152	172	47	1090	2850	13701	
	Medium (20-99)	0	0	32	42	7	231	251	1860	
Almaty Region	Large (100 or more)	0	0	14	10	2	49	69	303	3156
	Medium and Large (20+)	95	34	0	0	0	0	0	0	
	Small (5-19)	88	0	84	36	0	172	416	1696	
	Medium (20-99)	47	0	35	17	0	50	39	317	
Nur-Sultan	Large (100 or more)	22	0	5	6	0	24	5	55	9791
	Small, Medium and Large (5+)	0	27	0	0	13	0	0	0	
	Small (5-19)	65	29	98	59	0	324	1340	6672	
	Medium (20-99)	0	9	20	0	0	54	83	836	
Atyrau Region	Large (100 or more)	0	0	5	0	0	9	20	111	2042
	Medium and Large (20+)	20	0	0	23	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	17	0	0	0	
	Small (5-19)	16	0	9	10	0	59	258	1311	
Mangystau and West Kazakhstan	Medium (20-99)	4	0	5	0	0	18	16	247	4196
	Large (100 or more)	0	0	2	0	0	5	8	58	
	Medium and Large (20+)	0	0	0	2	0	0	0	0	
	Small, Medium and Large (5+)	0	9	0	0	6	0	0	0	
East Kazakhstan	Small (5-19)	44	13	40	28	9	172	419	2889	3364
	Medium (20-99)	0	0	15	5	0	40	45	325	
	Large (100 or more)	0	0	8	2	0	15	12	71	
	Medium and Large (20+)	29	4	0	0	10	0	0	0	
Karaganda Region	Small (5-19)	69	0	29	31	11	185	455	1964	6813
	Medium (20-99)	36	0	10	12	6	57	42	302	
	Large (100 or more)	16	0	7	6	4	25	11	65	
	Small, Medium and Large (5+)	0	24	0	0	0	0	0	0	
Karaganda Region	Small (5-19)	109	41	95	97	0	452	525	4667	74460
	Medium (20-99)	31	0	0	0	0	82	49	395	
	Large (100 or more)	14	0	0	0	0	46	19	79	
	Medium and Large (20+)	0	14	26	32	0	0	0	0	
	Small, Medium and Large (5+)	0	0	0	0	38	0	0	0	
		1633	433	1113	830	255	4725	9972	55499	74460

Appendix C: Original Sample Design

Original Sample Design (Fresh)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Akmola Region	Small (5-19)	10	3	14	5	2	2	2	2	97
	Medium (20-99)	13	1	6	0	2	6	2	2	
	Large (100 or more)	2	0	1	0	1	6	2	13	
Aktobe Region	Small (5-19)	3	4	9	11	3	2	2	2	84
	Medium (20-99)	6	1	6	2	1	2	2	2	
	Large (100 or more)	3	0	3	0	0	5	4	11	
Almaty	Small (5-19)	2	21	2	10	10	2	2	5	131
	Medium (20-99)	2	9	8	13	3	2	2	2	
	Large (100 or more)	12	0	5	3	1	11	2	2	
Almaty Region	Small (5-19)	2	7	2	17	4	2	2	2	112
	Medium (20-99)	18	2	14	5	1	2	2	2	
	Large (100 or more)	8	1	2	2	0	9	2	4	
Nur-Sultan	Small (5-19)	2	18	2	29	9	2	2	4	129
	Medium (20-99)	8	4	11	9	1	2	2	2	
	Large (100 or more)	2	0	3	1	0	4	10	2	
Atyrau Region	Small (5-19)	10	3	6	6	2	29	2	2	148
	Medium (20-99)	2	1	2	1	1	8	8	29	
	Large (100 or more)	0	0	1	0	0	2	4	29	
Mangystau and West Kazakhstan	Small (5-19)	2	7	6	15	6	2	2	2	98
	Medium (20-99)	8	1	7	2	2	2	2	2	
	Large (100 or more)	3	0	4	1	2	7	5	8	
East Kazakhstan	Small (5-19)	2	8	2	12	6	2	2	2	88
	Medium (20-99)	4	2	5	4	2	2	2	2	
	Large (100 or more)	7	1	3	2	2	10	2	2	
Karaganda Region	Small (5-19)	2	17	2	6	7	2	2	2	89
	Medium (20-99)	2	3	4	8	3	2	2	2	
	Large (100 or more)	5	1	2	1	2	8	2	2	
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	2	13	2	3	3	2	2	2	85
	Medium (20-99)	2	4	3	7	4	2	2	2	
	Large (100 or more)	11	0	5	2	3	5	2	2	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	2	13	2	3	5	2	2	2	79
	Medium (20-99)	2	3	2	6	2	2	2	2	
	Large (100 or more)	8	2	8	1	0	4	2	2	
		167	150	154	187	90	152	87	153	1140

Original Sample Design (Panel)

		Food	Garments	Non Metallic Mineral Products	Fabricated Metal Products	Machinery and Equipment	Other Manufacturing	Retail	Other Services	Grand Total
Akmola Region	Small (5-19)	0	0	2	0	0	2	3	2	23
	Medium (20-99)	1	1	0	0	1	1	3	5	
	Large (100 or more)	0	0	0	0	1	0	0	1	
Aktobe Region	Small (5-19)	0	0	0	1	0	4	5	2	36
	Medium (20-99)	1	1	1	1	2	1	5	4	
	Large (100 or more)	2	0	2	0	1	1	2	0	
Almaty	Small (5-19)	1	0	1	1	1	2	2	5	29
	Medium (20-99)	0	2	0	1	0	3	3	2	
	Large (100 or more)	0	0	1	1	0	0	2	1	
Almaty Region	Small (5-19)	1	0	0	0	0	1	1	2	8
	Medium (20-99)	0	0	1	1	0	0	1	0	
	Large (100 or more)	0	0	0	0	0	0	0	0	
Nur-Sultan	Small (5-19)	0	0	1	0	0	0	5	7	31
	Medium (20-99)	0	1	0	0	0	2	1	8	
	Large (100 or more)	1	0	0	0	0	1	2	2	
Atyrau Region	Small (5-19)	0	0	1	0	0	2	0	5	12
	Medium (20-99)	0	0	0	0	1	0	0	2	
	Large (100 or more)	0	0	0	0	0	0	0	1	
Mangystau and West Kazakhstan	Small (5-19)	1	0	0	0	0	0	5	3	22
	Medium (20-99)	1	0	1	0	0	2	3	2	
	Large (100 or more)	0	0	1	0	1	0	1	1	
East Kazakhstan	Small (5-19)	0	0	1	0	0	2	2	2	32
	Medium (20-99)	2	0	0	1	2	6	5	2	
	Large (100 or more)	1	0	1	2	0	1	0	2	
Karaganda Region	Small (5-19)	2	1	2	1	1	2	2	2	31
	Medium (20-99)	1	0	2	0	0	2	2	2	
	Large (100 or more)	1	0	2	0	1	2	2	1	
Kostanay, North Kazakhstan, Pavlodar	Small (5-19)	2	1	0	2	2	2	2	2	35
	Medium (20-99)	3	1	0	0	4	2	2	2	
	Large (100 or more)	2	0	1	0	1	1	2	1	
Kyzylorda Region, South Kazakhstan, Jambyl	Small (5-19)	5	1	2	0	0	2	2	2	41
	Medium (20-99)	4	1	1	0	1	3	2	2	
	Large (100 or more)	1	0	2	1	0	1	6	2	
		33	10	26	13	20	48	73	77	300