

Turkey Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Turkey during calendar year 2008 as part of the fourth round of the Business Environment and Enterprise Performance Survey, a joint initiative of the European Bank for Reconstruction and Development and the World Bank, in Turkey.

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 Turkey survey 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.

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.

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

3. Three levels of stratification were used in this country: type of industry, firm size, and geographic region. The original sample design, with specific targets for these strata, is described in Appendix C.
4. Industry stratification was designed in the way that follows: the universe was stratified into 5 manufacturing industries, 1 services industry -retail -, and two residual sectors as defined in the sampling manual. Each manufacturing industry had a target of 160 interviews. The services industry and the two residual sectors had a target of 120 interviews. For the manufacturing industries sample sizes were inflated by about 33% 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.
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 5 regions. These regions are Marmara, Aegean, South, Central Anatolia and Black Sea-Eastern.
7. The Turkey sample contains panel data. The wave 1 panel “Investment Climate Private Enterprise Survey implemented in Turkey” consisted of 1325 establishments interviewed in 2005. A total of 425 establishments have been re-interviewed in the 2008 Business Environment and Enterprise Performance Survey.

3. Sampling implementation

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 (Appendix D) was twofold. Universe estimates were taken from the TOBB database which contains a full list of establishments in manufacturing sectors. TOBB refers to the Union of Chambers and Commodity Exchanges of Turkey. Universe estimates for service sectors were taken from the Statistical Institute of Statistics (SIS) with additional information based on SIC code from the Turkish Studies Institute (TSI). Comparisons were made between estimates in TOBB and SIS to establish that the two sources are comparable and hence can be used side by side.
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

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

confirmed non-eligible units as a proportion of the total number of contacts to complete the survey was 43% (2811 out of 6458 establishments).

Local Agencies team involved in the study:⁵

Local Agency	Name: TNS Piar Country: Turkey Member of the Gallup International Association Alliance with TNS Worldwide Activities since: 1975
Enumerators involved:	Enumerators: 40 Recruiters: 17
Other staff involved:	Fieldwork Coordinators: 17 people Editing: 3 people Data Entry: 5 people Data Processing: 2 people

Local Agency	Name: Ipsos KMG Country: Turkey Member of the ESOMAR Activities since: 2001
Enumerators involved:	Enumerators: 40 Recruiters: 20
Other staff involved:	Fieldwork Coordinators: 3 people Editing: 3 people Data Entry: 5 people Data Processing: 2 people

Sample Frame:

Characteristic of sample frame used:	Sample frame is based on official data from Union of Chambers and Commodity Exchanges of Turkey and Statistical Institute of Statistics (SIS) of Turkey with additional information based on SIC code from the Turkish Studies Institute.
Year:	Data from TOBB database is from 2007. Data from the Statistical Institute of Statistics (SIS) is from 2006.

Sectors included in the Sample:

Original Sectors	Manufactures: 15, 17, 18, 24, , 26, Retail: 52 Residual: 51, 72, 55, 50, 45, 60-64, 25, 27, 28, 29, 31
Added Sectors	

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: - On regions: -
---	--------------------------------

⁵ The survey data collection in Turkey was implemented by two contractors.

Comments on the response rate:	The response rate in this survey it was good. Usually, according to the implementing contractors experience, the target group for ES, top-managers, is very difficult to reach and convince to participate in the survey. The timing of the fieldwork, i.e. during the holiday season in Turkey made it extremely difficult to reach the target. The panel list did not have any contact information for the firms. The contact information was retrieved through the Internet by the field teams.
Comments on the sample design:	

Fieldwork:

Date of Fieldwork	14th April 2008 – 15th January 2009
Country	Turkey
Interview number	Manufactures: 860 Services: 165 Core: 127
Problems found during fieldwork:	- It was hard to convince the top-managers to partake in the survey; - Most of the respondents hesitated to give the financial information of their establishments; - The field team had a difficult time convincing especially the small-scale establishments to participate in the survey;
Other observations:	

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*). 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.

-*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 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). All establishments within the residual stratum were coded with *a4a*=2.

- *interview_no*: the variable contains the 2005 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. Advanced 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. Weights

21. 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 un-weighted 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.)⁶

22. 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,⁷ 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.

23. 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 (eligibility rules are summarized in Appendix A):

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*.

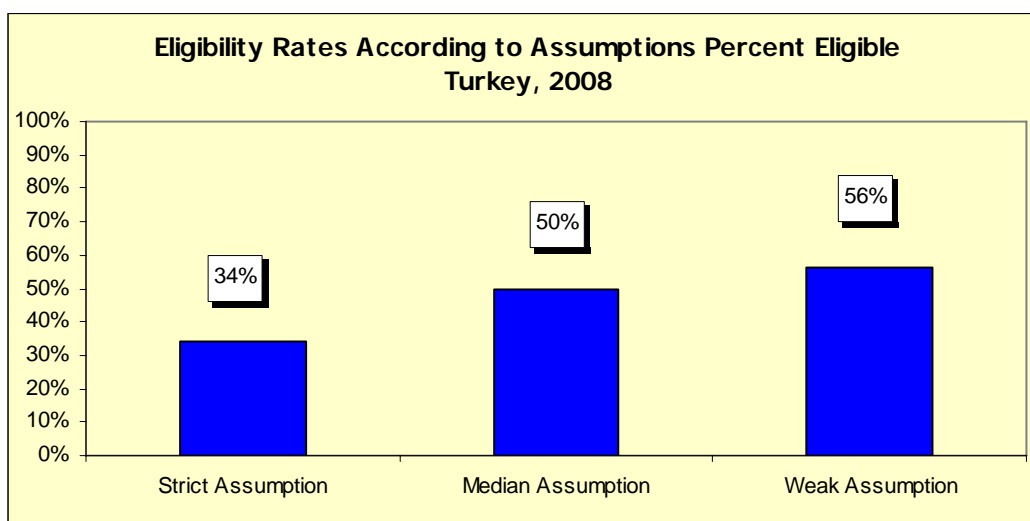
b- Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median*.

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.

⁶ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

⁷ Only for panel sample establishments with less than 5 employees were considered eligible.



24. 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.

For Turkey there was a need to combine the two regions Central Anatolia and Black Sea due to low numbers of interviews in Black Sea region. No other cell collapsing was deemed advantageous.

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

6. Appropriate use of the weights

26. 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.

27. 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.)⁸

28. 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 observed⁹. 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.

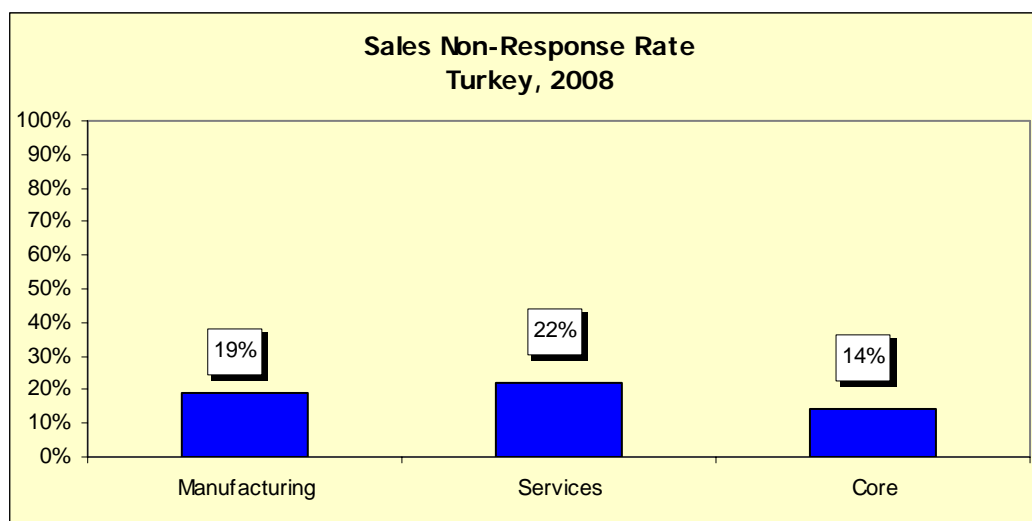
7. Non-response

29. 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.

30. 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 (-7).

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).



31. 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

⁸ 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.

⁹ 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.

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.

32. As the following graph shows, the number of contacted establishments per realized interview was 5.60. Details on rejection rates, eligibility rates, 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 the Republic of Turkey. 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 and Eligibility Status

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 (Moved and traced)	1	1	1
16. Panel firm - now less than five employees	1	1	1
5. The establishment has less than 5 permanent full time employees (no panel)	0	0	0
6. The firm discontinued businesses	0	0	0
7. Not a business: Private household	0	0	0
8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	0	0	0
151. Out of target - outside the covered regions	0	0	0
152. Out of target - moved abroad	0	0	0
91. No reply after having called in different days of the week and in different business hours	0	1	1
92. Line out of order	0	1	1
93. No tone	0	1	1
10. Answering machine	0	1	1
11. Fax line- data line	0	1	0
12. Wrong address/ moved away and could not get the new references	0	1	0
13. Refuses to answer the screener	0	1	1
14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0	0	0

Eligibility Status	TOTAL
1. Eligible establishment (Correct name and address)	2125
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	9
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	15
4. Eligible establishment (Moved and traced)	24
5. The establishment has less than 5 permanent full time employees	0
6. The firm discontinued businesses	86
7. Not a business: Private household	47
8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	2562
10. Answering machine	0
11. Fax line- data line	46
12. Wrong address/ moved away and could not get the new references	166
13. Refuses to answer the screener	820
14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	134
16. Panel firm - now less than five employees	2
91. No reply after having called in different days of the week and in different business hours	190
92. Line out of order	124
93. No tone	62
151. Out of target - outside the covered regions	46
152. Out of target - moved abroad	0
Grand Total	6458

Appendix B

Questionnaires:

Problems for the understanding of questions	No special problems encountered
Problems found in the navigability of – questionnaires (for example, skip patterns).	No special problems encountered
Comments on questionnaires length:	Most of the respondents have found the questionnaire very long and they got tired during the interview. Some of them interrupted the interview because of the length.
Suggestions or other comments on the questionnaire:	

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	Because of the declining economy in Turkey in the last 7 years (since 2001 economic crises), a lot of establishments were bankrupt or closed. This is a factor that could be noted during the analysis of the survey results.
Relevant country events occurred during fieldwork:	None
Other aspects:	None

Appendix C

Sample Design

		Sector								
Region	Employee size	15	17	18	24	26	52	Other Manufacturing	Residual	Grand Total
	5 to 19	15	14	25	31	17	15	5	16	138
	20 to 99	10	29	24	41	30	25	17	16	192
	100 AND OVER	16	31	35	19	8	25	12	17	163
<i>Marmara Total</i>		<i>41</i>	<i>74</i>	<i>84</i>	<i>91</i>	<i>55</i>	<i>65</i>	<i>34</i>	<i>49</i>	<i>493</i>
	5 to 19	7	8	5	10	17	8	5	6	66
	20 to 99	10	10	12	13	26	5	6	5	87
	100 AND OVER	6	12	15	0	0	5	10	5	53
<i>Aegean Total</i>		<i>23</i>	<i>30</i>	<i>32</i>	<i>23</i>	<i>43</i>	<i>18</i>	<i>21</i>	<i>16</i>	<i>206</i>
	5 to 19	12	5	10	17	10	3	15	11	83
	20 to 99	11	17	8	0	5	2	5	9	57
	100 AND OVER	0	14	0	0	0	2	0	5	21
<i>South Total</i>		<i>23</i>	<i>36</i>	<i>18</i>	<i>17</i>	<i>15</i>	<i>7</i>	<i>20</i>	<i>25</i>	<i>161</i>
	5 to 19	23	5	7	17	19	11	15	7	104
	20 to 99	15	5	14	7	18	6	0	6	71
	100 AND OVER	10	5	0	0	0	6	10	5	36
<i>Central Anatolia Total</i>		<i>48</i>	<i>15</i>	<i>21</i>	<i>24</i>	<i>37</i>	<i>23</i>	<i>25</i>	<i>18</i>	<i>211</i>
	5 to 19	15	5	0	5	10	3	5	5	48
	20 to 99	10	0	5	0	0	2	15	6	38
	100 AND OVER	0	0	0	0	0	2	0	1	3
<i>Black Sea - Eastern Total</i>		<i>25</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>10</i>	<i>7</i>	<i>20</i>	<i>12</i>	<i>89</i>
Grand Total		160	160	160	160	160	120	120	120	1160
Total Employee Size	5 to 19	72	37	47	80	73	40	45	45	439
	20 to 99	56	61	63	61	79	40	43	42	445
	100 AND OVER	32	62	50	19	8	40	32	33	276

Appendix D

Sample Frame Turkey

Source: TOBB 2007 and SIS 2006

Region	Size	15	17	18	24	26	52	Other Manuf	Residual	Grand Total
Marmara	5-19	2,060	3,299	3,697	716	363	8,801	8,344	4,244	31,524
	20-99	580	1,579	3,569	320	266	1,650	4,899	882	13,745
	100+	141	389	692	107	64	550	942	276	3,161
Marmara Total		2,781	5,267	7,958	1,143	693	11,001	14,185	5,403	48,430
Aegean	5-19	1,496	642	222	190	163	2,601	2,051	1,018	8,383
	20-99	398	564	475	73	238	488	1,257	224	3,717
	100+	122	204	164	26	48	163	305	75	1,107
Aegean Total		2,016	1,410	861	289	449	3,251	3,613	1,317	13,206
South	5-19	826	718	122	172	82	878	1,439	365	4,603
	20-99	180	426	96	22	50	165	386	77	1,401
	100+	27	155		10	3	55	39	24	312
South Total		1,033	1,299	218	204	135	1,098	1,864	466	6,316
Black Sea - Eastern	5-19	2,567	198	206	331	306	4,532	5,070	1,912	15,122
	20-99	580	146	316	66	234	850	2,127	381	4,700
	100+	111	76	104	16	43	283	351	107	1,092
Black Sea-Eastern Total		3,258	420	626	413	583	5,665	7,548	2,401	20,914
Grand Total		9,088	8,396	9,663	2,049	1,860	21,015	27,210	9,586	88,867

Industry sectors based on ISIC Rev 3.1