

Eritrea Indicator Surveys Data Set

1. Introduction

- 1.1. This document provides additional information on the data collected in Eritrea from 4th July and 10th August 2009 as part of the Enterprise Survey,¹ an initiative of the World Bank.
- 1.2. 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.
- 1.3. The report describes the sampling design of the survey, the structure of the dataset and additional information that may be useful when using the data, including information on non-response rates, the calculation of sample weights and country-specific factors that may have affected survey implementation.

2. Survey Target Population

- 2.1. The whole population, or the universe, covered in the Enterprise Surveys is the non-agricultural economy. It comprises: all manufacturing sectors according to the ISIC Revision 3.1 group classification (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this population 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.
- 2.2. The Indicator Survey (IS) for Eritrea targeted 150 registered establishments with at least five employees. In the Enterprise Surveys, the requirements for registration are defined on a country-by-country basis using the information collected by Doing Business and information from the in-country contractors. In Eritrea, registered firms were defined as firms listed in the National Statistics Bureau and National Chamber of Commerce and Industry register.

¹ The Enterprise Survey initiative includes two types of surveys: Enterprise Surveys and Indicator Surveys. The former is a survey of firms that is meant to be representative of a country's non-agricultural economy. Interviews are conducted using the Core questionnaire, Core + Manufacturing module, or the Core + Retail module. An Indicator Survey is similar to an Enterprise Survey and it is implemented for smaller economies where the sampling strategies inherent in an Enterprise Survey are often not applicable due to the limited universe of firms. The survey instrument is limited to just those questions which form the basis of the global indicators as presented on the Enterprise Surveys website. For a complete list and more detailed information on each type of surveys see <http://www.enterprisesurveys.org/Methodology/CurrentProjects.aspx>

3. Sampling for Registered Establishments

- 3.1. The sample for registered establishments in Eritrea was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*.² Three levels of stratification were used in the Eritrea sample: firm sector, firm size, and geographic region. The original sample design, with specific targets for these strata, is included in Appendix A.
- 3.2. Industry stratification was designed as follows: the universe was stratified into manufacturing and services industries. The initial sample design had a target of 75 interviews in manufacturing and 75 interviews in services.
- 3.3. Size stratification was defined following the standardized definition used for the Enterprise Surveys: 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. The micro sample consists of firms with 1 to 4 employees.
- 3.4. Regional stratification was defined in terms of the geographic regions with the largest commercial presence in the country: Debub, Maekel and Northern Red Sea.

4. Sampling implementation

- 4.1. 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 for the Enterprise Surveys.
- 4.2. The sample frame used for the survey was a combination of two databases. For manufacturing sectors, the official list from the National Statistics Bureau and National Chamber of Commerce and Industry were used. The 2007 year was used. For services sectors, the Services establishments database from the National Chamber of Commerce and Industry National Statistics was used. The 2007 year was used. Both database contained the following information:
 - Name of the firm
 - Contact details
 - ISIC code
 - Number of employees.
- 4.3. Counts from both sample frames shown below.

Universe Figures for Eritrea

² The complete text of the *Sampling Manual* can be found at http://www.enterprisesurveys.org/documents/Sampling_Note.pdf

Region	Size	Manufacturing	Services	Grand Total
Debub	5 to 19	7		7
	20 to 99	25		25
	100+	1		1
Debub Total		33		33
Maekel	5 to 19	17	297	314
	20 to 99	146	52	198
	100+	12	35	47
Maekel Total		175	384	559
Northern Red Sea	5 to 19	2	2	4
	20 to 99	5	5	10
	100+	1	1	2
Northern Red Sea Total		8	8	16
Grand Total		216	392	608

Source: National Statistics Bureau & National Chamber of Commerce and Industry 2007, and National Chamber of Commerce and Industry 2007

- 4.4. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 150 establishments with five or more employees.
- 4.5. The quality of the frame was assessed at the onset of the project through calls 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.
- 4.6. 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 6.9% (20 out of 290 establishments for the IS samples).³ Breaking down by industry, the following numbers of establishments were surveyed:

Manufacturing	Services
94	85

5. Database Structure

- 5.1. The database compiles two different versions of the Indicator Survey questionnaire for all registered establishments. The Services Questionnaire is administered to the establishments in the services sector. The Manufacturing Questionnaire is built upon the Services Questionnaire and adds specific questions relevant to manufacturing. Each version of the questionnaire is identified by the index variable, *a0*.

³ Appendix B shows the tabulations for the Eritrea sample of registered firms of response codes that are classified as eligible and non-eligible.

- 5.2. All variables in the database 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 “AF” indicate questions specific to the Africa region; these questions may not have been asked in Enterprise Surveys conducted in countries in other regions. All other variables are global and are present in all country surveys conducted throughout the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” indicates that the variable is alpha-numeric.
- 5.3. The variable *idstd* uniquely identifies each establishment at the global level.
- 5.4. 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. These variables generate the strata cells for each industry/region/size combination. 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 the sampling regions
 - a6a*: coded using the definition for micro, small, medium, and large establishments as discussed 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 industries that comprise the manufacturing, services, and residual categories used in the stratification. These codes include most manufacturing industries (15 to 37), and retail, and IT for services (52, and 72 respectively). All establishments within the ‘other manufacturing’ stratum were coded with *a4a*=2.
 - id*: for panel firms the variable contains the same id used in 2005
 - strata*: identifies sampling strata in the database
- 5.5. Note that these variables may not coincide with reality for some establishments as sample frames may contain information that is later found to be inaccurate.
- 5.6. The surveys were implemented following a two stage procedure. In the first stage a screener questionnaire was administered over the phone to determine sampled establishment’s eligibility for the survey 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 that was collected in the screening phase.
- 5.7. The main questionnaire contains variables for location (*a3x*), industry (*d1a2*), and number of employees (*l1*, *l6* and *l8*) that more accurately reflect describe the characteristics of establishments than the information provided on these variables in the sample frame or the screener.

- 5.8. 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 industry-strata based on information available from the sample frame, whereas variable *d1a2* indicates the actual ISIC code of the main output of the establishment as answered by the interviewee. This variable is probably the most accurate variable with which to classify establishments by activity.
- 5.9. 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 on the sample frame but the actual physical location is in another place.
- 5.10. Variables *l1*, *l6* and *l8* provide a more accurate measure of employment that accounts for both permanent (*l1*) and temporary employment (*l6*, adjusted by duration of employment *l8*). Special efforts were made to make sure that this information was not missing for most establishments. The indicators on the website use the constructed variable *size*, which is computed as follows:

$$\begin{aligned}
 \text{size} &= l1 \text{ if no temporary employment} \\
 &= l1 + l6*(l8/12) \text{ if temporary employment with known duration} \\
 &= l1 + l6*(1/3) \text{ if temporary employment with unknown duration}
 \end{aligned}$$

6. Sample Eligibility and Universe Estimates

- 6.1. Special care is given to the correct computation of universe estimates and weights in the Enterprise Surveys. Considering the varying quality of sample frames across countries, it is important to accurately adjust the universe totals within each region/industry/size stratum to account for the presence of ineligible units in the sampling frame. Information on ineligible firms in the sample frame is used to scale down the universe estimate for each cell by the observed proportion of ineligible units within the cell.
- 6.2. Information on the eligibility of firms in the sample frame is collected during the screening process. Each firm contacted during the screening process, whether interviewed or not, is assigned one of the status codes detailed in the table below.

Status Code	Eligibility Criteria		
	Strict	Weak	Median
1. Eligible establishment (Correct name and address)	1	1	1
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1	1	1
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	1	1	1
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	1	1	1
16. Panel firm - now less than five employees	1	1	1

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

6.3. The local implementing agency commented that many firms in the sample frame were missing phone numbers, so enumerators often physically traced the firms and administered the screener questionnaire face to face. However, the addresses listed in the sample frame were also sometimes inaccurate, so it was difficult to locate some firms. The tables in Appendix B show the numbers of firms in each status code at the conclusion of the survey.

6.4. 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.

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

6.5.1. 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} .

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

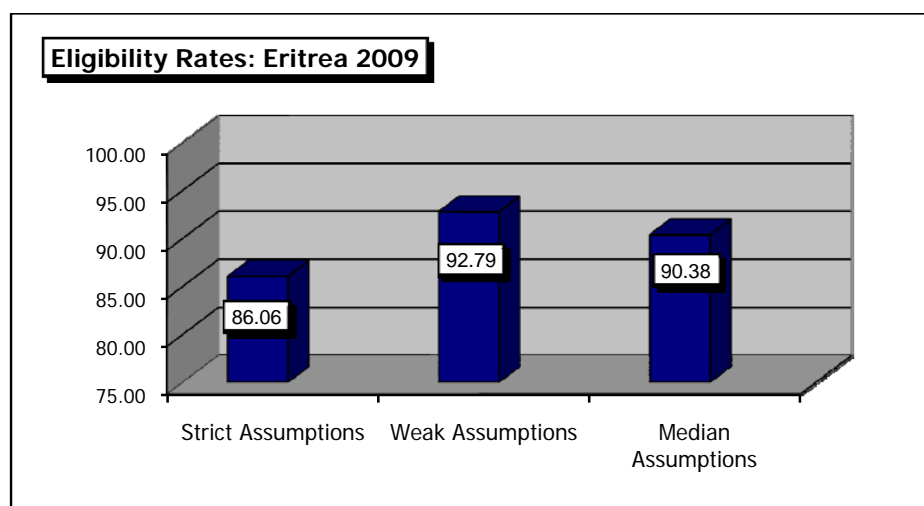
6.5.2. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable w_{median} .

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

6.5.3. 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. The resulting weights are included in the variable w_{weak} . Under the weak assumption only observed non-eligible units are excluded from universe projections.

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

6.6. 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.



6.7. Universe estimates for the number of establishments in each industry-region-size cell in Eritrea were produced for the strict, weak and median eligibility definitions. Appendix C shows the universe estimates of the numbers of registered establishments based on the strict, weak and median relative estimates.

6.8. 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.

7. Weights

7.1. Since the sampling design was stratified and employed differential sampling of the strata, 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.)⁴

7.2. Several sets of weights for each cell were computed using the strict, weak, and median assumptions on establishment eligibility. The first set of estimates calculated proportions using the raw sample count for each cell. However, for many cells the sample numbers of interviewed establishments are small, and eligibility rates and adjusted universe cells projections for those cells are subject to relatively large sampling variations. A set of more robust estimates that use the multiples of the relative eligibility rates for each industry, size, and region was also produced. In addition, in cases where the cell sizes are small, collapsed weights are produced based on combined cells with larger samples than the individual cells, which produce values with smaller sampling variations. The data sets include only the robust weights.⁵

7.3. Appendix D shows the cell weights for registered establishments in Eritrea.

8. Non-response

8.1. The Enterprise Surveys, along with all other surveys, suffer from both survey non-response and 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. Different strategies were used to address these issues.

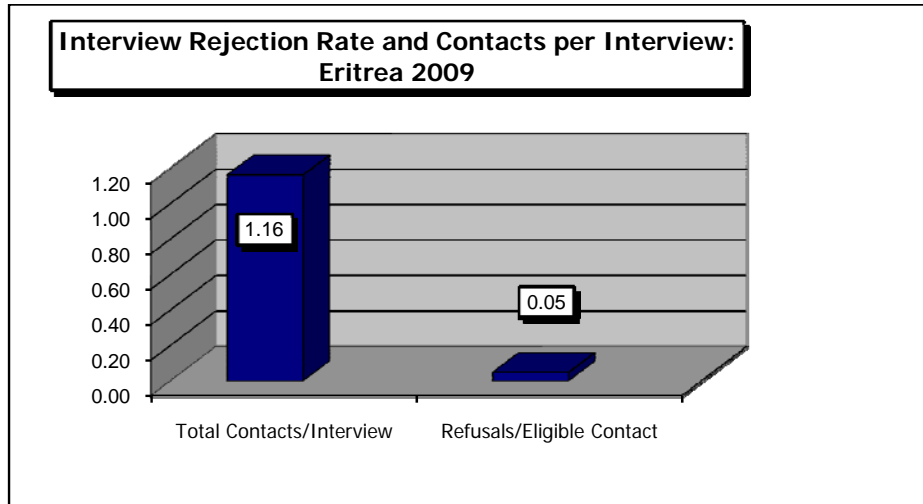
8.2. Survey non-response was addressed by maximizing efforts to contact establishments that were initially sampled. When the survey frame was extracted from the sampling frame, several establishments with the same strata characteristics were randomly selected for each interview and each establishment was assigned a preference number.⁶ Substitutions of replacement establishments were made in order to help achieve targets on the number of interviews for each stratum. Extensive efforts were made to complete interviews with each first preference establishment before contact with a replacement establishment was allowed. At least four attempts were made to contact each sampled establishment for an interview at different times/days of the week before a replacement establishment was allowed to be contacted for an interview.

⁴ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

⁵ See *A Note on Weights* for more details on the weight calculations and using weights with the Enterprise Surveys.

⁶ In cases where the number of contacts initially drawn from the sample frame are insufficient to obtain an interview with the targeted number of establishments in a given strata, additional contacts for that strata may be drawn from the sampling frame. If all establishments in that strata have already been contacted and the sample target has not been reached, the sample design may be adjusted to allow additional interviews in other strata. In Eritrea, the sample design was readjusted from the original design to reflect the small numbers of manufacturing establishments in the country. Additionally, more contacts were issued near the end of fieldwork to compensate for lower than expected response rates in several strata.

8.3. As the following graph shows, the number of contacted establishments per realized interview was 1.16 for the Indicator Survey sample. 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 (e.g., establishments that closed or were in ineligible sectors). Refusal rates are also shown in the graph below. For each establishment eligible for an interview, 0.05 in the Indicator Survey sample refused to participate.

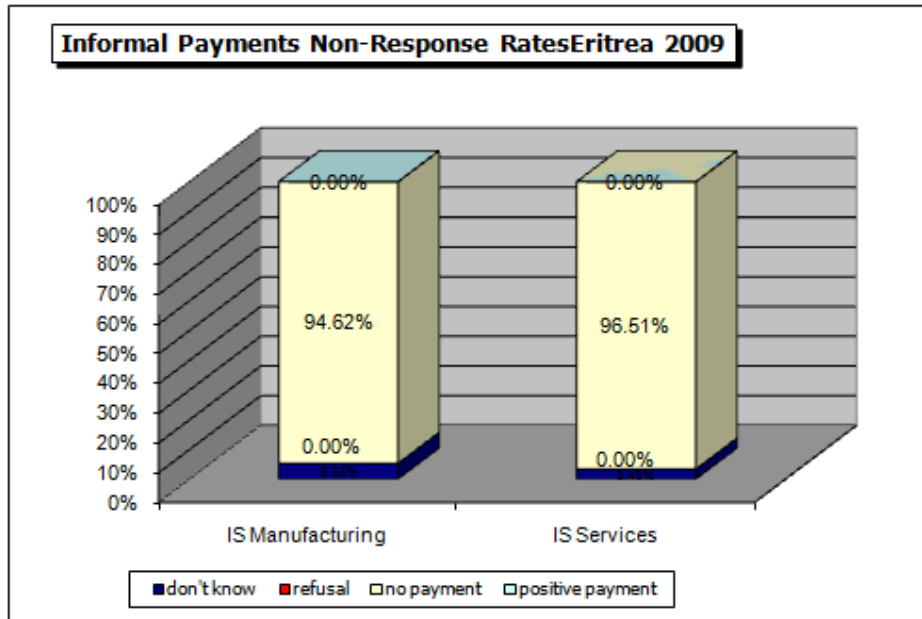


8.4. In completed surveys, item non-response was addressed by two strategies:

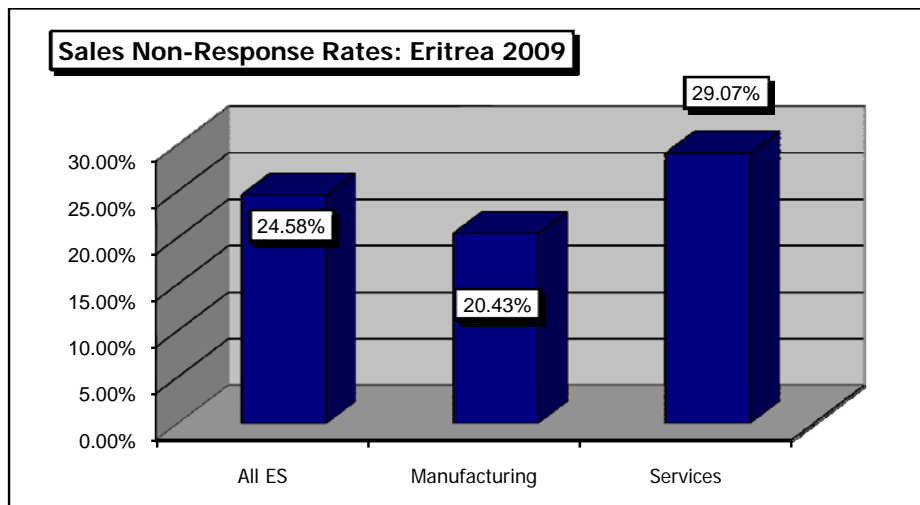
8.4.1. For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond (-7) as a different option from don't know (-9).

8.4.2. Establishments with incomplete information on critical productivity variables including total sales, cost figures and employment levels were re-contacted in order to complete this information and minimize item non-response. However, re-contacts did not fully eliminate low response rates for some items.

8.5. The following graph shows the breakdown of answers about the total amount of informal payments made annually (variables *j7a* and *j7b*) by questionnaire type.



8.6. The following graph shows non-response rates for the sales variable, *d2*.⁷



8.7. This report summarizes statistics on rejection rates, eligibility rates, and item non-response to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and imperfect sampling frames are not unique to Eritrea or the Enterprise Surveys. All surveys suffer from these issues although they may not be made explicit.

9. Country specific comments

9.1. In Eritrea, fieldwork ran from 15th September 2008 to 13th February 2009. The local agency that implemented the Enterprise Survey in Eritrea noted that the political trouble in the country considerably affected the fieldwork.

⁷ Please note that the question on total sales does not have a “refuse to answer” option, thus the non-response rates in the graph above reflect DKs and NAs as well as any missing values.

Managers seemed to be afraid of sharing information that could be perceived as negative by the regime. This made the interviewing process a bit complicated as the respondents were not feeling “free” to discuss issues such as restrictions to trade for fear of victimization.

- 9.2. The local agency employed a total of 22 enumerators and recruiters as well as 1 fieldwork supervisor. Data entry was completed by 3 staff. Few problems were encountered with the understanding of the questionnaire. The local agency stated that both interviewers and respondents found the questionnaire very clear.

Appendix A

Original Sample Design

Region	Size	Manufacturing	Services	Grand Total
Debub	5 to 19	7		7
	20 to 99	12		12
	100+	1		1
Debub Total		20		20
Maekel	5 to 19	17	22	39
	20 to 99	20	21	41
	100+	12	25	37
Maekel Total		49	68	117
Northern Red Sea	5 to 19	2	1	3
	20 to 99	3	3	6
	100+	1	3	4
Northern Red Sea Total		6	7	13
Grand Total		75	75	150

Appendix B

Enterprise Survey Sample

Status Codes

1. Eligible establishment (Correct name and address)	179
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	0
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	0
16. Panel firm - now less than five employees	0
5. The establishment has less than 5 permanent full time employees	0
6. The firm discontinued businesses	3
7. Not a business: private household	4
8. Ineligible activity: education, agriculture, finances, governments...	7
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	3
92. Line out of order	1
93. No tone	1
10. Answering machine	0
11. Fax line - data line	0
12. Wrong address/ moved away and could not get the new references	0
13. Refuses to answer the screener	9
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	0
152. Out of target - firm moved abroad	1
Total	208

Response Outcomes

Complete interviews (Total)	179
Incomplete interviews	0
Eligible in process	0
Refusals	0
Out of target	14
Impossible to contact	5
Ineligible - coop.	1
Refusal to the Screener	9
Total	208

Appendix C

Eritrea Strict Universe Estimates

Region	Size	Manufacturing	Services	Grand Total
Debub	5 to 19	7		7
	20 to 99	25		25
	100+	1		1
Debub Total		33		33
Maekel	5 to 19	16	251	267
	20 to 99	139	44	183
	100+	10	25	35
Maekel Total		165	320	484
Northern Red Sea	5 to 19	1	1	3
	20 to 99	4	3	7
	100+	1	1	2
Northern Red Sea Total		6	6	12
Grand Total		204	325	529

Eritrea Weak Universe Estimates

Region	Size	Manufacturing	Services	Grand Total
Debub	5 to 19	7		7
	20 to 99	25		25
	100+	1		1
Debub Total		33		33
Maekel	5 to 19	17	277	294
	20 to 99	146	48	194
	100+	11	28	39
Maekel Total		174	354	527
Northern Red Sea	5 to 19	2	1	3
	20 to 99	4	4	8
	100+	1	1	2
Northern Red Sea Total		7	6	13
Grand Total		213	360	573

Eritrea Median Universe Estimates

Region	Size	Manufacturing	Services	Grand Total
Debub	5 to 19	7		7
	20 to 99	25		25
	100+	1		1
Debub Total		33		33
Maekel	5 to 19	16	256	272
	20 to 99	146	47	193
	100+	10	27	37
Maekel Total		173	330	502
Northern Red Sea	5 to 19	2	1	3
	20 to 99	4	4	8
	100+	1	1	2
Northern Red Sea Total		7	6	13
Grand Total		212	336	548

Appendix D

Eritrea Strict Cell Weights

Region	Size	Manufacturing	Services
Dehub	5 to 19	1	
	20 to 99	2	
	100+	1	
Maekel	5 to 19	1	10
	20 to 99	4	1
	100+	1	1
Northern Red Sea	5 to 19	1	1
	20 to 99	1	2
	100+	1	1

Eritrea Weak Cell Weights

Region	Size	Manufacturing	Services
Dehub	5 to 19	1	
	20 to 99	2	
	100+	1	
Maekel	5 to 19	1	11
	20 to 99	4	2
	100+	1	1
Northern Red Sea	5 to 19	2	1
	20 to 99	1	2
	100+	1	1

Eritrea Median Cell Weights

Region	Size	Manufacturing	Services
Dehub	5 to 19	1	
	20 to 99	2	
	100+	1	
Maekel	5 to 19	1	10
	20 to 99	4	1
	100+	1	1
Northern Red Sea	5 to 19	2	1
	20 to 99	1	2
	100+	1	1