

The Tanzania 2013 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Tanzania between January 2013 and February 2014 as part of the Africa Enterprise Survey 2013, an initiative of the World Bank.

As part of its strategic goal of building a climate for investment, job creation, and sustainable growth, the World Bank has promoted improving business environments 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 Enterprise Surveys currently cover over 130,000 firms in 135 countries, of which 121 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 Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling design of the data, the data set structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

II. Sampling Structure

2. The sample for Tanzania 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.

¹ The complete text can be found at:

http://www.enterprisesurveys.org/~media/FPDKM/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

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/regions.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and region. The original sample design with specific information of the industries and regions chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into three manufacturing industry (food, textiles and garments, furniture, and other manufacturing) and two service sectors (retail and other services).

5. For the Tanzania ES, 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).

6. Regional stratification for the Tanzania ES was defined in five regions (city and the surrounding business area): Arusha, Dar Es Salaam, Mbeya, Mwanza, and Zanzibar.

III. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and region) are required to draw the sample. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. TNS Opinion was hired to implement the Africa ES 2013 roll out, including the Tanzania 2013 ES. In Tanzania the local subcontractor was TNS RMS East Africa Limited.

9. For the Tanzania ES, three sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Tanzania 2006. The World Bank required that attempts should be made to re-interview establishments responding to the Tanzania 2006 survey where they were within the selected geographical regions and met eligibility criteria. Due to the fact that the previous round of surveys seemed to have utilized different stratification criteria (or no stratification at all) and due to the prevalence of small firms and firms located in the capital city in the 2006 sample the

following convention was used. The presence of panel firms was limited to a maximum of 50% of the achieved interviews in each cell. That sample is referred to as the Panel. The second frame was the 2011 Central Registry of Establishment (CRE) of the National Bureau of Statistics (NBS). The sample frame was supplemented by an additional list of establishments from the Tanzania Chamber of Commerce, Industry & Agriculture (TCCIA) out of which 9 interviews were completed. The third frame was 2008 Central Registry of Establishment (CRE) of the Office of Chief Government Statistician (OCGS). The sample frame was used for the establishments in Zanzibar. All databases contained the following information:

- a) Detailed stratification variables;
- b) Location identifiers- address, phone number, email; and
- c) Contact name(s).

Counts from sample frames are shown below.

Source: 2011, National Bureau of Statistics (NBS) and 2008 Office of Chief Government Statistician (OCGS)

Region name	Sampling Employee size	Food	Textile and Apparel	Furniture	Other Manufacturing	Retail	Other Services	TOTAL
Arusha	Small 5-19	17	35	47	38	231	579	947
	Med 20-99	16	0	2	9	9	80	116
	Large 100+	1	4	0	6	4	7	22
		34	39	49	53	244	666	1085
Dar es Salaam	Small 5-19	184	417	560	528	1873	4863	8425
	Med 20-99	41	33	34	194	193	621	1116
	Large 100+	24	12	5	57	19	66	183
		249	462	599	779	2085	5550	9724
Mbeya	Small 5-19	39	194	81	41	92	744	1191
	Med 20-99	7	4	3	1	5	31	51
	Large 100+	1	1	0	1	0	1	4
		47	199	84	43	97	776	1246
Mwanza	Small 5-19	29	66	71	18	172	674	1030
	Med 20-99	11	0	5	4	14	61	95
	Large 100+	6	1	1	0	0	1	9
		46	67	77	22	186	736	1134
Zanzibar	Small 5-19	62	34	146	137	246	353	978
	Med 20-99	5	7	13	8	51	77	161
	Large 100+	0	0	0	1	8	17	26
		67	41	159	146	305	447	1165
Grand Total		443	808	968	1043	2917	8175	14354

10. The enumerated establishments with 5 employees or more were then used as the sample frame for the Tanzania Enterprise Survey with the aim of obtaining interviews of 720 establishments.

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

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 3.3% (100 out of 3029 establishments)⁴.

IV. Data Base Structure:

13. The structure of the data base reflects the fact that 3 different versions of the survey instrument were used for all registered establishments. Questionnaires have common questions (*core* module) and respectfully additional manufacturing and retail 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 ***Retail*** questionnaire (includes the *Core* module plus retail specific questions) and the residual eligible services have been covered using the core module only (***Other Services*** questionnaire). Each variation of the questionnaire is identified by the index variable, *a0*.

14. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1 (some exceptions apply due to comparability reasons). Variable names preceded by a prefix “KEN” and “A2F” indicate questions specific to some countries in Africa, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size and region. Different combinations of these variables generate the strata cells for each industry/region/size

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

combination. A distinction should be made between the variable *a4a* and *d1a2* (industry expressed as ISIC rev. 3.1 code). The former gives the establishment's classification into one of the chosen industry-strata, whereas the latter gives the actual establishment's industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

- a2* is the variable describing sampling regions

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

- a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include manufacturing industries (15 to 37), retail (52), and codes 45, 50, 51, 55, 60, 63, 72 for other services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. In some cases, when the phone numbers were unavailable in the sample frame, the enumerators applied the screeners in person. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a7* to *a11* contain additional information and were also collected in the screening phase.

19. 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*, *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.

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

21. Note that the fiscal years vary by firm as there is no standard for all firms in Tanzania. The start and end dates for the fiscal year for each firm can be found in the *a20* variables in the dataset

V. Universe Estimates

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

23. Appendix B shows the overall estimates of the numbers of establishments in Tanzania based on the sample frame.

24. 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.

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

26. 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*.

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

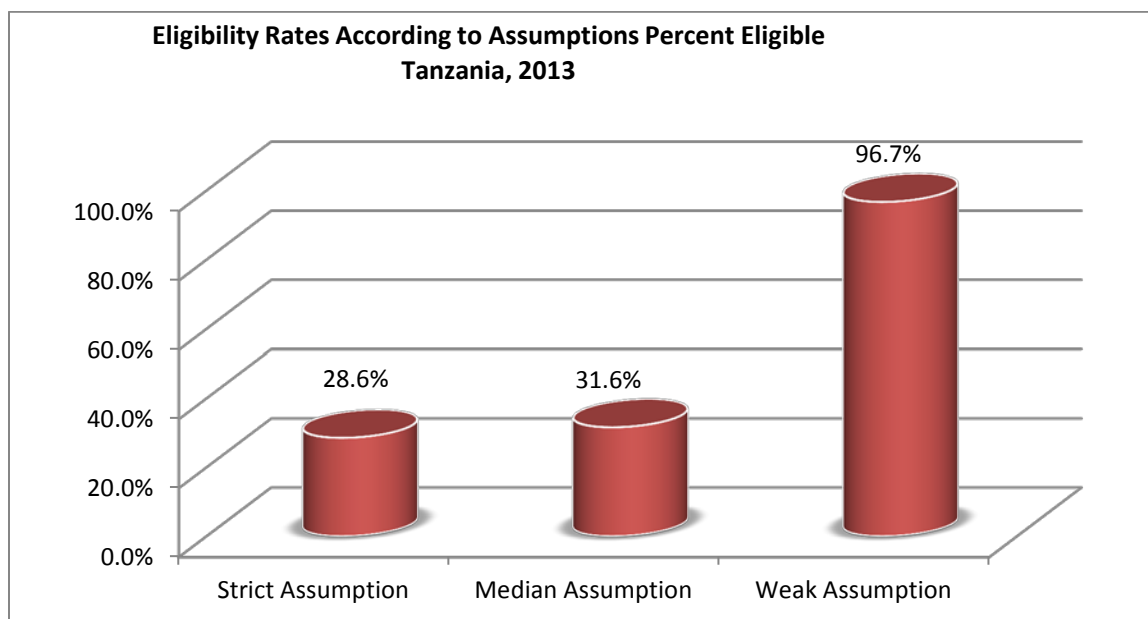
27. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

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

28. 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*.

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

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



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

31. 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

32. 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

33. 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, establishments with less than 5 employees, no reply after having called in different days of the week and in different

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

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 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.

34. Appendix C shows the cell weights for registered establishments in Tanzania.

VII. Appropriate use of the weights

35. 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.

36. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not a strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)⁷

37. 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

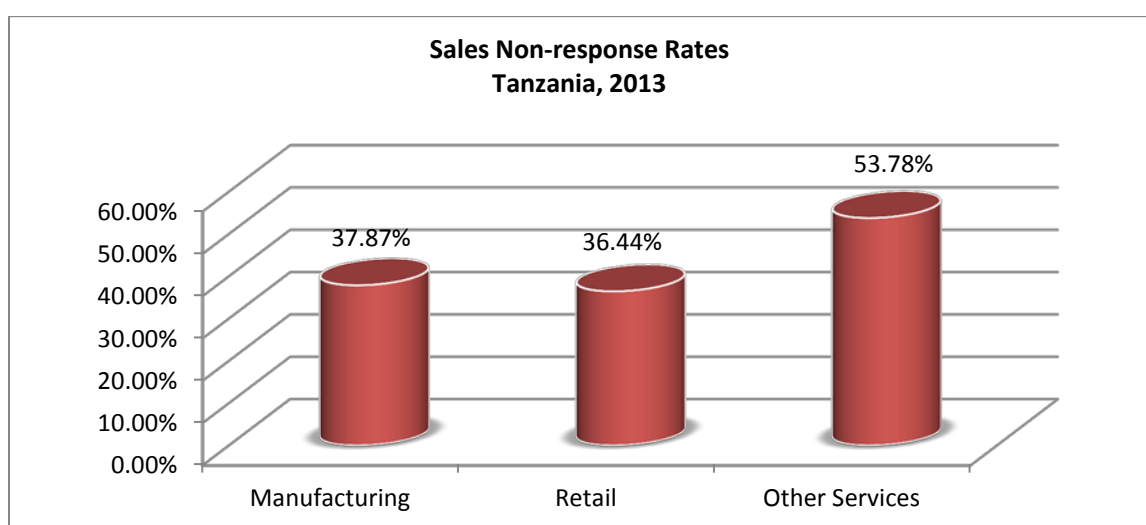
38. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

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

⁷ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

⁸ The use 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.

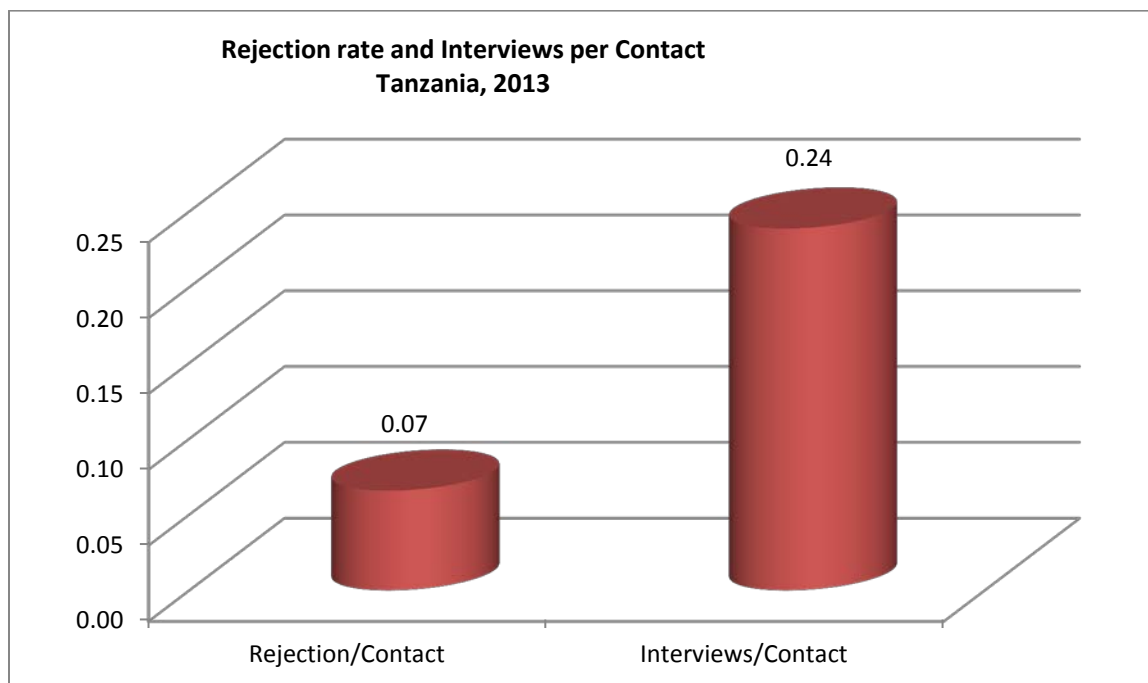
39. Item non-response was addressed by two strategies:
- a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-7).
 - b- Establishments 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 sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the charts below for both enterprise surveys (ES) and micro firms reflect both categories (DKs and NAs).



40. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

41. As the following graphs show, the number of realized interviews per contact contacted establishments was 0.24.⁹ 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 number of rejections per contact was 0.07.

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



42. 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 Tanzania. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

References:

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

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

Appendix A

Status Codes Total:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	814
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	32
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	13
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	6
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	40
	6. The firm discontinued businesses	51
	7. Not a business: private household	1
	8. Ineligible activity: education, agriculture, finances, governments...	3
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	419
	92. Line out of order	151
	93. No tone	166
	94. Phone number does not exist	761
	10. Answering machine	10
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	474
	13. Refuses to answer the screener	83
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	5
	152. Out of target - firm moved abroad	
	153. Impossible to find	
Total		3029

Response Outcomes Total:

Complete interviews (<i>Total</i>)	723
Incomplete interviews	3
Eligible in process	23
Refusals	116
Out of target	95
Impossible to contact	1981
Ineligible - coop.	5
Refusal to the Screener	83
Total	3029

Status Codes Fresh:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	777
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	32
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	13
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	6
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	40
	6. The firm discontinued businesses	48
	7. Not a business: private household	1
	8. Ineligible activity: education, agriculture, finances, governments...	3
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	380
	92. Line out of order	139
	93. No tone	164
	94. Phone number does not exist	522
	10. Answering machine	9
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	463
	13. Refuses to answer the screener	80
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)		0
151. Out of target - outside the covered regions, firm moved abroad		5
152. Out of target - firm moved abroad		
153. Impossible to find		
Total		2682

Response Outcomes Fresh:

Complete interviews (<i>Total</i>)	698
Incomplete interviews	3
Eligible in process	21
Refusals	106
Out of target	92
Impossible to contact	1677
Ineligible - coop.	5
Refusal to the Screener	80
Total	2682

Status Codes Panel:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	37
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	0
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	0
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	3
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	0
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	39
	92. Line out of order	12
	93. No tone	2
	94. Phone number does not exist	239
	10. Answering machine	1
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	11
	13. Refuses to answer the screener	3
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)		0
151. Out of target - outside the covered regions, firm moved abroad		0
152. Out of target - firm moved abroad		
153. Impossible to find		
Total		347

Response Outcomes Panel:

Complete interviews (<i>Total</i>)	25
Incomplete interviews	0
Eligible in process	2
Refusals	10
Out of target	3
Impossible to contact	304
Ineligible - coop.	0
Refusal to the Screener	3
Total	347

Appendix B

Universe Estimates, Tanzania:

Enterprise Survey

Source: 2012, NBS (National Bureau of Statistics) and 2008 Office of Chief Government Statistician (OCGS)

Region name	Sampling Employee size	Food	Textile and Apparel	Furniture	Other Manufacturing	Retail	Other Services	
Arusha	Small 5-19	17	35	47	38	231	579	947
	Med 20-99	16	0	2	9	9	80	116
	Large 100+	1	4	0	6	4	7	22
		34	39	49	53	244	666	1085
Dar es Salaam	Small 5-19	184	417	560	528	1873	4863	8425
	Med 20-99	41	33	34	194	193	621	1116
	Large 100+	24	12	5	57	19	66	183
		249	462	599	779	2085	5550	9724
Mbeya	Small 5-19	39	194	81	41	92	744	1191
	Med 20-99	7	4	3	1	5	31	51
	Large 100+	1	1	0	1	0	1	4
		47	199	84	43	97	776	1246
Mwanza	Small 5-19	29	66	71	18	172	674	1030
	Med 20-99	11	0	5	4	14	61	95
	Large 100+	6	1	1	0	0	1	9
		46	67	77	22	186	736	1134
Zanzibar	Small 5-19	62	34	146	137	246	353	978
	Med 20-99	5	7	13	8	51	77	161
	Large 100+	0	0	0	1	8	17	26
		67	41	159	146	305	447	1165
Grand Total		443	808	968	1043	2917	8175	14354

Appendix C

Strict Cell Weights Fresh Tanzania:

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	1.01	0.95	0.85	1.09	6.16	8.50	3.09
Small (5 to 19)	2.03	1.84	1.55	1.12	16.02	21.38	7.32
Medium (20 to 99)	1.00	1.00	0.00	1.14	1.00	3.13	1.21
Large (100+)	0.00	0.00	1.00	1.00	1.46	1.00	0.74
Dar Es Salaam	3.00	4.57	7.16	5.21	8.65	16.59	7.53
Small (5 to 19)	5.66	11.15	19.15	12.58	9.18	12.48	11.70
Medium (20 to 99)	2.04	1.49	1.02	1.75	15.10	35.10	9.41
Large (100+)	1.31	1.08	1.30	1.30	1.66	2.20	1.48
Mbeya	1.65	0.47	1.04	0.38	1.09	9.04	2.28
Small (5 to 19)	2.10	1.42	2.12	1.13	2.28	24.94	5.66
Medium (20 to 99)	1.85	0.00	1.00	0.00	1.00	1.18	0.84
Large (100+)	1.00	0.00	0.00	0.00	0.00	1.00	0.33
Mwanza	1.01	0.79	0.36	0.85	3.54	7.40	2.33
Small (5 to 19)	1.00	1.38	1.08	1.54	9.63	19.76	5.73
Medium (20 to 99)	2.04	1.00	0.00	1.00	1.00	2.45	1.25
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zanzibar	0.54	0.76	0.91	0.96	5.47	5.44	2.35
Small (5 to 19)	1.61	1.29	1.00	1.87	13.65	11.22	5.11
Medium (20 to 99)	0.00	1.00	1.73	1.00	1.00	5.10	1.64
Large (100+)	0.00	0.00	0.00	0.00	1.76	0.00	0.29
Grand Total	1.44	1.51	2.06	1.70	4.98	9.40	3.52

Median Cell Weights Fresh Tanzania:

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	1.14	0.96	0.86	1.13	6.37	8.76	3.20
Small (5 to 19)	2.42	1.87	1.57	1.15	16.52	21.86	7.56
Medium (20 to 99)	1.00	1.00	0.00	1.23	1.00	3.41	1.27
Large (100+)	0.00	0.00	1.00	1.00	1.59	1.01	0.77
Dar Es Salaam	3.67	4.71	7.31	5.39	9.30	17.79	8.03
Small (5 to 19)	6.76	11.35	19.43	12.86	9.49	12.78	12.11
Medium (20 to 99)	2.59	1.61	1.10	1.90	16.59	38.22	10.33
Large (100+)	1.66	1.17	1.40	1.41	1.82	2.39	1.64
Mbeya	2.23	0.56	1.17	0.45	1.25	10.81	2.75
Small (5 to 19)	2.94	1.69	2.52	1.35	2.76	29.92	6.86
Medium (20 to 99)	2.75	0.00	1.00	0.00	1.00	1.50	1.04
Large (100+)	1.00	0.00	0.00	0.00	0.00	1.00	0.33
Mwanza	1.50	0.90	0.44	0.99	4.43	9.26	2.92
Small (5 to 19)	1.34	1.70	1.33	1.91	12.08	24.54	7.15
Medium (20 to 99)	3.14	1.00	0.00	1.05	1.21	3.24	1.61
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zanzibar	0.66	0.78	0.97	0.99	5.81	5.82	2.51
Small (5 to 19)	1.98	1.35	1.00	1.96	14.46	11.78	5.42
Medium (20 to 99)	0.00	1.00	1.92	1.02	1.00	5.69	1.77
Large (100+)	0.00	0.00	0.00	0.00	1.98	0.00	0.33
Grand Total	1.84	1.58	2.15	1.79	5.43	10.49	3.88

Median Cell Weights Panel Tanzania:

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	0.00	0.00	0.00	0.00	0.33	0.67	0.17
Small (5 to 19)	0.00	0.00	0.00	0.00	1.00	1.00	0.33
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	1.00	0.17
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Dar Es Salaam	1.70	0.00	0.36	1.94	0.39	1.15	0.92
Small (5 to 19)	1.22	0.00	1.08	1.80	1.17	1.45	1.12
Medium (20 to 99)	3.86	0.00	0.00	3.02	0.00	2.01	1.48
Large (100+)	0.00	0.00	0.00	1.00	0.00	0.00	0.17
Mbeya	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mwanza	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zanzibar	0.00	0.00	0.00	0.33	0.00	0.00	0.06
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	1.00	0.00	0.00	0.17
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grand Total	0.34	0.00	0.07	0.46	0.14	0.36	0.23

Weak Cell Weights Fresh Tanzania:

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	2.02	1.76	1.74	2.02	20.47	16.98	7.50
Small (5 to 19)	4.85	3.83	4.21	2.49	56.19	43.96	19.26
Medium (20 to 99)	1.23	1.46	0.00	2.16	1.32	5.51	1.94
Large (100+)	0.00	0.00	1.00	1.41	3.90	1.46	1.30
Dar Es Salaam	9.19	12.63	26.19	15.30	37.51	41.57	23.73
Small (5 to 19)	18.57	31.89	71.61	38.33	44.23	35.24	39.98
Medium (20 to 99)	5.72	3.63	3.26	4.56	62.18	84.73	27.35
Large (100+)	3.28	2.36	3.72	3.02	6.11	4.74	3.87
Mbeya	5.48	1.73	3.80	1.46	5.24	31.62	8.22
Small (5 to 19)	8.82	5.19	10.14	4.39	14.08	90.22	22.14
Medium (20 to 99)	6.63	0.00	1.25	0.00	1.63	3.64	2.19
Large (100+)	1.00	0.00	0.00	0.00	0.00	1.00	0.33
Mwanza	2.57	1.67	1.18	1.98	14.69	18.07	6.69
Small (5 to 19)	2.67	3.47	3.54	4.13	40.78	49.02	17.27
Medium (20 to 99)	5.03	1.54	0.00	1.82	3.29	5.20	2.81
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zanzibar	2.08	2.28	3.28	3.18	29.78	17.31	9.65
Small (5 to 19)	6.25	4.37	3.30	6.73	77.66	37.41	22.62
Medium (20 to 99)	0.00	2.47	6.55	2.82	4.05	14.53	5.07
Large (100+)	0.00	0.00	0.00	0.00	7.64	0.00	1.27
Grand Total	4.27	4.01	7.24	4.79	21.54	25.11	11.16

Weak Cell Weights Panel Tanzania:

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	0.00	0.00	0.00	0.00	1.20	0.81	0.33
Small (5 to 19)	0.00	0.00	0.00	0.00	3.59	1.42	0.83
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	1.00	0.17
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Dar Es Salaam	8.83	0.00	2.54	24.00	2.40	4.49	7.04
Small (5 to 19)	9.02	0.00	7.61	31.93	7.20	7.29	10.51
Medium (20 to 99)	17.46	0.00	0.00	32.87	0.00	6.19	9.42
Large (100+)	0.00	0.00	0.00	7.19	0.00	0.00	1.20
Mbeya	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mwanza	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Zanzibar	0.00	0.00	0.00	1.02	0.00	0.00	0.17
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	3.06	0.00	0.00	0.51
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grand Total	1.77	0.00	0.51	5.00	0.72	1.06	1.51

Appendix D

Strict Universe Estimates

Fresh

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	17	24	16	22	72	326	477
Small (5 to 19)	6	22	12	13	64	278	396
Medium (20 to 99)	11	2	0	5	6	44	67
Large (100+)	0	0	4	4	1	4	13
Dar Es Salaam	73	205	109	227	427	1959	3000
Small (5 to 19)	51	190	96	151	376	1685	2549
Medium (20 to 99)	14	13	9	54	45	246	382
Large (100+)	8	2	4	22	5	29	70
Mbeya	11	21	41	9	17	210	309
Small (5 to 19)	8	21	38	9	14	200	290
Medium (20 to 99)	2	0	3	0	3	9	17
Large (100+)	1	0	0	0	0	1	2
Mwanza	14	29	18	8	43	284	396
Small (5 to 19)	10	26	18	6	39	257	356
Medium (20 to 99)	4	3	0	2	4	27	40
Large (100+)	0	0	0	0	0	0	0
Zanzibar	11	46	10	34	55	126	282
Small (5 to 19)	11	41	8	32	41	101	235
Medium (20 to 99)	0	5	2	2	12	25	46
Large (100+)	0	0	0	0	2	0	2
Grand Total	127	326	194	300	612	2905	4465

Panel

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	0	0	0	0	1	5	6
Small (5 to 19)	0	0	0	0	1	3	4
Medium (20 to 99)	0	0	0	0	0	2	2
Large (100+)	0	0	0	0	0	0	0
Dar Es Salaam	5	0	4	6	4	9	28
Small (5 to 19)	1	0	4	2	4	5	16
Medium (20 to 99)	4	0	0	3	0	4	11
Large (100+)	0	0	0	1	0	0	1
Mbeya	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Mwanza	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Zanzibar	0	0	0	1	0	0	1
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	1	0	0	1
Large (100+)	0	0	0	0	0	0	0
Grand Total	5	0	4	7	5	14	35

Weak Universe Estimates

Fresh

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	28	49	38	44	237	654	1050
Small (5 to 19)	15	46	34	30	225	572	920
Medium (20 to 99)	13	3	0	9	8	77	110
Large (100+)	0	0	4	6	4	6	19
Dar Es Salaam	227	579	398	653	2018	5413	9288
Small (5 to 19)	167	542	358	460	1813	4758	8098
Medium (20 to 99)	40	33	29	141	187	593	1023
Large (100+)	20	5	11	51	18	62	167
Mbeya	43	78	186	35	89	752	1183
Small (5 to 19)	35	78	183	35	84	722	1137
Medium (20 to 99)	7	0	4	0	5	29	44
Large (100+)	1	0	0	0	0	1	2
Mwanza	37	70	60	20	176	694	1058
Small (5 to 19)	27	66	60	17	163	637	970
Medium (20 to 99)	10	5	0	4	13	57	89
Large (100+)	0	0	0	0	0	0	0
Zanzibar	44	152	33	120	289	409	1047
Small (5 to 19)	44	140	26	114	233	337	894
Medium (20 to 99)	0	12	7	6	49	73	146
Large (100+)	0	0	0	0	8	0	8
Grand Total	378	929	716	872	2810	7923	13627

Panel

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	0	0	0	0	4	6	10
Small (5 to 19)	0	0	0	0	4	4	8
Medium (20 to 99)	0	0	0	0	0	2	2
Large (100+)	0	0	0	0	0	0	0
Dar Es Salaam	26	0	30	72	22	42	192
Small (5 to 19)	9	0	30	32	22	29	122
Medium (20 to 99)	17	0	0	33	0	12	63
Large (100+)	0	0	0	7	0	0	7
Mbeya	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Mwanza	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Zanzibar	0	0	0	3	0	0	3
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	3	0	0	3
Large (100+)	0	0	0	0	0	0	0
Grand Total	26	0	30	75	25	48	205

Median Universe Estimates

Fresh

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	18	24	17	23	74	336	491
Small (5 to 19)	7	22	13	14	66	284	406
Medium (20 to 99)	11	2	0	5	6	48	72
Large (100+)	0	0	4	4	2	4	14
Dar Es Salaam	89	210	111	237	444	2024	3115
Small (5 to 19)	61	193	97	154	389	1725	2619
Medium (20 to 99)	18	14	10	59	50	268	419
Large (100+)	10	2	4	24	5	31	77
Mbeya	16	25	48	11	20	252	372
Small (5 to 19)	12	25	45	11	17	239	349
Medium (20 to 99)	3	0	3	0	3	12	21
Large (100+)	1	0	0	0	0	1	2
Mwanza	20	35	23	10	53	355	495
Small (5 to 19)	13	32	23	8	48	319	443
Medium (20 to 99)	6	3	0	2	5	36	52
Large (100+)	0	0	0	0	0	0	0
Zanzibar	14	48	10	35	57	134	299
Small (5 to 19)	14	43	8	33	43	106	248
Medium (20 to 99)	0	5	2	2	12	28	49
Large (100+)	0	0	0	0	2	0	2
Grand Total	156	343	209	316	648	3101	4773

Panel

Row Labels	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	0	0	0	0	1	5	6
Small (5 to 19)	0	0	0	0	1	3	4
Medium (20 to 99)	0	0	0	0	0	2	2
Large (100+)	0	0	0	0	0	0	0
Dar Es Salaam	5	0	4	6	4	10	29
Small (5 to 19)	1	0	4	2	4	6	17
Medium (20 to 99)	4	0	0	3	0	4	11
Large (100+)	0	0	0	1	0	0	1
Mbeya	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Mwanza	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Zanzibar	0	0	0	1	0	0	1
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	1	0	0	1
Large (100+)	0	0	0	0	0	0	0
Grand Total	5	0	4	7	5	15	36

Appendix E
Original Sample Design, Tanzania:

SAMPLE DESIGN (n)

Region name	Sampling Employee size	Food 15	Textile and Apparel 17-18	Furniture 36	Other Manufacturing	Retail 52	Other Services	TOTAL
Arusha	Small 5-19	5	12	11	10	5	5	48
	Med 20-99	9	0	2	9	7	5	32
	Large 100+	1	4	0	6	4	5	20
		15	16	13	25	16	15	100
Dar es Salaam	Small 5-19	5	6	12	11	49	58	141
	Med 20-99	5	27	26	34	5	5	102
	Large 100+	10	12	5	38	7	5	77
		20	45	43	83	61	68	320
Mbeya	Small 5-19	5	24	17	15	5	5	71
	Med 20-99	7	4	3	1	5	5	25
	Large 100+	1	1	0	1	0	1	4
		13	29	20	17	10	11	100
Mwanza	Small 5-19	5	14	13	10	5	5	52
	Med 20-99	11	0	5	4	14	5	39
	Large 100+	6	1	1	0	0	1	9
		22	15	19	14	19	11	100
Zanzibar	Small 5-19	5	8	13	12	5	5	48
	Med 20-99	5	7	11	8	5	5	41
	Large 100+	0	0	0	1	5	5	11
		10	15	24	21	15	15	100
Grand Total		80	120	119	160	121	120	720

Completed Interviews, Tanzania:

	Food	Furniture	Textile and Apparel	Other Manufacturing	Retail	Other Services	Grand Total
Arusha	14	14	12	20	32	16	108
Small (5 to 19)	3	12	8	12	14	7	56
Medium (20 to 99)	11	2	0	4	14	8	39
Large (100+)	0	0	4	4	4	1	13
Dar Es Salaam	24	28	21	63	158	53	347
Small (5 to 19)	10	17	9	13	138	45	232
Medium (20 to 99)	8	9	9	32	7	5	70
Large (100+)	6	2	3	18	13	3	45
Mbeya	6	15	21	8	17	9	76
Small (5 to 19)	4	15	18	8	8	6	59
Medium (20 to 99)	1	0	3	0	8	3	15
Large (100+)	1	0	0	0	1	0	2
Mwanza	12	22	17	6	24	8	89
Small (5 to 19)	10	19	17	4	13	4	67
Medium (20 to 99)	2	3	0	2	11	4	22
Large (100+)	0	0	0	0	0	0	0
Zanzibar	7	37	9	20	14	16	103
Small (5 to 19)	7	32	8	17	9	3	76
Medium (20 to 99)	0	5	1	3	5	12	26
Large (100+)	0	0	0	0	0	1	1
Grand Total	63	116	80	117	245	102	723

Appendix F

Local Agency team involved in the study:

Local Agencies	Name: TNS RMS East Africa Limited Country: Tanzania Activities since: 1975
Enumerators involved:	Enumerators: 40 Recruiters: 27
Other staff involved:	Fieldwork Coordinators: 1 Data Processing: 3

Sample Frame:

Characteristic of sample frame used:	Census of registered businesses operating in Tanzania Census of registered businesses operating in Zanzibar Panel list from WB (companies interviewed in 2006) List from Tanzanian Chamber of Commerce
Source:	2011, Tanzania National Bureau of Statistics (NBS)) 2011, Tanzania Chamber of Commerce (TCCIA) 2008, Zanzibar Office of Chief Government Statistician (OCGS)
Year:	2011 and 2008

Sectors included in the Sample:

Original Sectors	The manufacturing sector comprises all manufacturing establishments as mentioned in group D The service sector includes Retail (Group H – 52); Group F (construction), Groups G without 52 and H (Hotel and restaurant), Group I (transport, storage, and communications) and subsector 72 from Group K
Added (top up) Sectors	None

Fieldwork and country situation:

Date of Fieldwork	30 th January.2013 to 4 th February 2014
Country	Tanzania
Use of CAPI	<ul style="list-style-type: none">• Computer-assisted personal interviewing (CAPI) was used to collected data. The SODA mobile research app was used
Problems found during fieldwork:	<ul style="list-style-type: none">• There was a high drop-out rate among interviewers which was a significant challenge during the course of the project.• Among the respondents, there was fear that interviewers represented the TRA (Tanzania Revenue Authority). This created some discomfort during interviews and some respondents refused to provide information when it came to financial questions.• Some respondents were unwilling to share some information which they termed as “sensitive” particularly on their financial position• The respondents largely did not perceive the direct impact of these surveys. There were also several projects of the same type being run by other research firms and it happened that this survey coincided. As such, this brought about some challenges in interviewing respondents who had already been interviewed by another firm.
Country specific situation	Fiscal years differ across firms